

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ
Издательство "Новые технологии"

СОДЕРЖАНИЕ

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

- Казиков П. В. Оценка эффективности генетических алгоритмов многокритериальной оптимизации. Часть 1. 2
Елтаренко Е. А. Конъюнктивные операторы агрегирования в многокритериальных задачах 7
Зак Ю. А. Распределение множества заданий и определение оптимальных очередностей их выполнения на параллельных машинах методами динамического программирования 14
Трудоношин В. А., Федорук В. Г. Сравнение математических моделей шарнирного соединения 20

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

- Норенков И. П., Уваров М. Ю., Грушин М. А., Овешников А. А. Классификация документов на основе кластеризации онтологий 24
Андрейчикова О. Н., Андрейчиков А. В. Что делать с авторским правом в сети Интернет? Поиск решения методом аналитических сетей 28
Баранов Д. В. Логические уравнения в эквивалентных LP-структурах. 35

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И СЕТИ

- Барский А. Б. Двух- и трехкомпьютерная кластеризация *Grid*-ресурсов для оптимизированной сборки результатов распределенных вычислений по *SPMD*-технологии 43

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И УПРАВЛЕНИИ

- Сологуб Р. А. Порождение регрессионных моделей поверхности волатильности биржевых опционов 47
Галиев Д. Р., Исавнин А. Г. Модели выбора оптимального инвестиционного портфеля с нечетко-множественными доходностями 52

КОДИРОВАНИЕ И ОБРАБОТКА СИГНАЛОВ

- Агиевич С. Н. Теоретические основы сплайн-алгебраического гармонического анализа сигналов систем радиосвязи 58
Захаров А. В., Богданов М. Р., Мухамедьянов А. Р. Аналитический метод решения задачи позиционирования 63

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

- Гриняк В. М., Девятисильный А. С. Нечеткая система экспертных оценок характера движения судна 66
Contents 71

- Приложение. Лейпяля Т., Шилов В. В. Из истории вычислительной техники в России XIX — начала XX веков

Главный редактор
НОРЕНКОВ И. П.

Зам. гл. редактора
ФИЛИМОНОВ Н. Б.

Редакционная
коллегия:

- АВДОШИН С. М.
АНТОНОВ Б. И.
БАРСКИЙ А. Б.
БОЖКО А. Н.
ВАСЕНИН В. А.
ГАЛУШКИН А. И.
ГЛОРИОЗОВ Е. Л.
ДОМРАЧЕВ В. Г.
ЗАГИДУЛЛИН Р. Ш.
ЗАРУБИН В. С.
ИВАННИКОВ А. Д.
ИСАЕНКО Р. О.
КОЛИН К. К.
КУЛАГИН В. П.
КУРЕЙЧИК В. М.
ЛЬВОВИЧ Я. Е.
МАЛЬЦЕВ П. П.
МЕДВЕДЕВ Н. В.
МИХАЙЛОВ Б. М.
НЕЧАЕВ В. В.
ПАВЛОВ В. В.
ПУЗАНКОВ Д. В.
РЯБОВ Г. Г.
СОКОЛОВ Б. В.
СТЕМПКОВСКИЙ А. Л.
УСКОВ В. Л.
ФОМИЧЕВ В. А.
ЧЕРМОШЕНЦЕВ С. Ф.
ШИЛОВ В. В.

Редакция:

- БЕЗМЕНОВА М. Ю.
ГРИГОРИН-РЯБОВА Е. В.
ЛЫСЕНКО А. В.
ЧУГУНОВА А. В.

Информация о журнале доступна по сети Internet по адресу <http://novtex.ru/IT>.
Журнал включен в систему Российского индекса научного цитирования.
Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

УДК 004.89+004.021

П. В. Казаков, канд. техн. наук, доц.,
Брянский государственный
технический университет,
e-mail: pvk_mail@list.ru

Оценка эффективности генетических алгоритмов многокритериальной оптимизации. Часть 1

Рассматриваются способы оценки качества решения задач многокритериальной оптимизации с применением генетических алгоритмов. Отмечаются различные индикаторы, а также их наборы для количественной оценки эффективности этих эволюционных методов.

Ключевые слова: многокритериальная оптимизация, принципы Парето, граница Парето, индикаторы эффективности, многокритериальные генетические алгоритмы

Введение

Задача многокритериальной оптимизации (МО) [1, 2] часто возникает в различных исследованиях, носящих как теоретический, так и практический характер. Известно, что особенности МО и, прежде всего, необходимость одновременной оптимизации по множеству несоизмеримых и конфликтующих критериев требуют принципиально других математических и алгоритмических средств для поиска решений, чем при однокритериальной оптимизации. Также результатом МО является целое множество оптимальных решений, точное определение которых в настоящее время остается нерешенной *NP*-проблемой. Поэтому актуальным является разработка новых эффективных численных методов автоматизации решения задачи МО. Одними из них являются генетические алгоритмы [3], а именно их специальные версии для многокритериальной оптимизации [4, 5]. Принципы поиска решений, используемые в таких генетических алгоритмах, различаются между собой, что по-разному влияет на качество получаемых результатов. Их количественная оценка во многом определяет эффективность выбранного алгоритма. В то же время опыт анализа получаемых результатов многокритериальной оптимизации показал, что вследствие ее

специфики процедура такой оценки является не тривиальной, а представляет собой отдельную задачу.

Как известно, эффективность любого оптимизационного алгоритма зависит от качества найденных решений и времени, затраченного на их поиск. Генетические алгоритмы многокритериальной оптимизации (МГА) не являются здесь исключением. Однако, если для однокритериальных и многокритериальных задач оптимизации правила определения времени поиска решений во многом совпадают, то оценка их качества принципиально различается. В однокритериальных генетических алгоритмах качество решений может быть определено простым сравнением значений целевой функции — единственного критерия. Как известно, в процессе многокритериальной оптимизации решения оцениваются одновременно по всем критериям на основе специальных принципов Парето [1, 2]. Согласно им оптимальным считается множество несравнимых между собой решений, образующих в пространстве критериев различные топологии (границы Парето). В связи с этим процедура оценки качества решений, найденных МГА, усложняется и требует привлечения нескольких критериев эффективности.

1. Постановка задачи многокритериальной оптимизации

В общем случае задача многокритериальной оптимизации [1] включает n оптимизируемых переменных, m критериев оптимальности и k ограничений. Математически постановка задачи МО может быть представлена следующим образом:

$$Y = F(x) = (f_1(x), f_2(x), \dots, f_m(x)) \rightarrow \min;$$

$$\varphi(x) = (\varphi_1(x), \varphi_2(x), \dots, \varphi_k(x)); x = (x_1, x_2, \dots, x_n) \in X;$$

$$y = (y_1, y_2, \dots, y_m) \in Y; (f_1, f_2, \dots, f_m) \in F.$$

Множество X здесь означает пространство решений, имеющее произвольную форму — непрерывную, дискретную, смешанную, Y — пространство критериев, F — обобщенный векторный критерий.

При выполнении соответствующих ограничений данные пространства сужаются, образуя множества $X_\varphi = \{x \in X | \varphi(x) \leq 0\}$ и $Y_\varphi = F(X_\varphi) = \bigcup_{x \in X_\varphi} \{f(x)\}$ (рис. 1).

Цель многокритериальной оптимизации — сужение множества X_φ до так называемого множества недоминируемых решений. Для определения таких решений используются принципы Парето-доминирования и Парето-оптимальности.

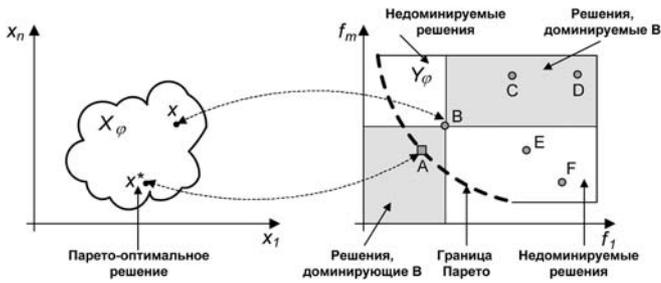


Рис. 1. Отображение пространств решений и критериев

Принцип Парето-доминирования, $\forall x^{(1)}, x^{(2)} \in X_\varphi$:

- 1) $x^{(1)} > x^{(2)}$ ($x^{(1)}$ строго доминирует $x^{(2)}$), если $f_i(x^{(1)}) < f_i(x^{(2)})$, $i = 1, \dots, m$;
- 2) $x^{(1)} \geq x^{(2)}$ ($x^{(1)}$ доминирует $x^{(2)}$), если $f_i(x^{(1)}) \leq f_i(x^{(2)}) \wedge \exists i: f_i(x^{(1)}) < f_i(x^{(2)})$, $i = 1, \dots, m$;
- 3) $x^{(1)} \not> x^{(2)}$ ($x^{(1)}$ не доминирует $x^{(2)}$), если $(f_i(x^{(1)}) \not\leq f_i(x^{(2)}) \wedge f_i(x^{(2)}) \not\leq f_i(x^{(1)}))$, $i = 1, \dots, m$.

На рис. 1 решение B строго доминирует решение C, D, но само доминируется решением A. По отношению к другим изображенным точкам B является недоминируемым, несравнимым.

Принцип Парето-оптимальности состоит в том, что решение задачи МО есть множество недоминируемых по отношению ко всем остальным решений. Точка x^* называется оптимальной по Парето (эффективной, не улучшаемой), если $\nexists x \in X_\varphi: x \geq x^*$. На рис. 1 такие точки изображены квадратами. Множество эффективных точек образует множество Парето $X_P = \{x^* \in X_\varphi | \nexists x \in X_\varphi: x \geq x^*\}$, которое в общем случае может включать любое число решений. Соответственно, в пространстве критериев множество решений из X_P образует так называемую компромиссную кривую, или границу Парето $Y_P = f(X_P) = \bigcup_{x \in X_P} \{f(x)\}$.

2. Особенности оценки результатов многокритериальной оптимизации

В общем случае критерии эффективности многокритериальных генетических алгоритмов могут быть сформулированы следующим образом [6]:

- минимальное отличие результирующего множества Парето-оптимальных решений от глобального множества Парето (если оно известно);
- максимальная равномерность распределения решений вдоль найденной МГА границы Парето;
- максимальная мощность найденного множества Парето;
- минимальное время поиска решений.

Анализ этих критериев позволяет сделать вывод, что, с одной стороны, сама задача оценки эффективности МГА является многокритериальной, с другой стороны, количественное значение ни од-

ного из критериев не позволяет однозначно судить о качестве работы генетического алгоритма. Так, для подавляющего большинства практических задач нет информации об оптимальных решениях или математическая модель задачи имеет ограничения, что приводит к изначальной фрагментации границы Парето. В итоге первые два критерия при оценке качества решений могут приобрести условный характер. Что касается мощности результирующего множества Парето, то большое число найденных решений становится при оценке действительно важным только в сочетании с высокими показателями первых двух критериев. Известно, что при работе МГА основной объем процессорного времени приходится именно на вычисления значений критериев. Поэтому при большом пространстве критериев и решений сам их поиск МГА, его настройка и оценка эффективности становятся вычислительно высокотрудоемким процессом. Это приводит к необходимости ограничения времени поиска (например, для систем реального времени), что автоматически оказывает влияние на остальные критерии эффективности.

Таким образом, отмеченные критерии следует воспринимать как стратегические направления для совершенствования МГА. В то время как на практике для более гибкой и объективной оценки эффективности МГА используются специальные индикаторы (метрики) [6, 7], уточняющие соответствующие критерии эффективности. Как правило, такой индикатор представляет собой функциональную зависимость с аргументами в виде одного или нескольких множеств (границ) Парето. В зависимости от потребностей в исследовании МГА могут быть использованы различные наборы этих индикаторов — бенчмарки (*benchmarks*), одни для сравнения результатов разных МГА между собой при решении прикладных задач; другие при исследовании нового МГА на тестовых задачах с известными правильными решениями; третьи при настройке МГА на высокую скорость поиска и т. д.

Анализ существующих индикаторов позволяет разделить их на две группы — по числу оцениваемых множеств Парето, участвующих в вычислении индикаторов. Первая группа — унарные индикаторы — предназначена для абсолютной количественной оценки отдельного множества Парето. Цель второй группы — бинарных индикаторов — заключается в определении преимуществ двух множеств Парето относительно друг друга. На рис. 2 показан графический смысл некоторых из унарных и бинарных индикаторов; запись $A \underset{I}{>} B$, $A \underset{I}{<} B$ означает соответственно превосходство множества Парето A над B или их равнозначность по индикатору I.

Анализ опыта использования разных индикаторов [6] позволяет сделать вывод, что не все из них могут быть использованы на практике, в частности,

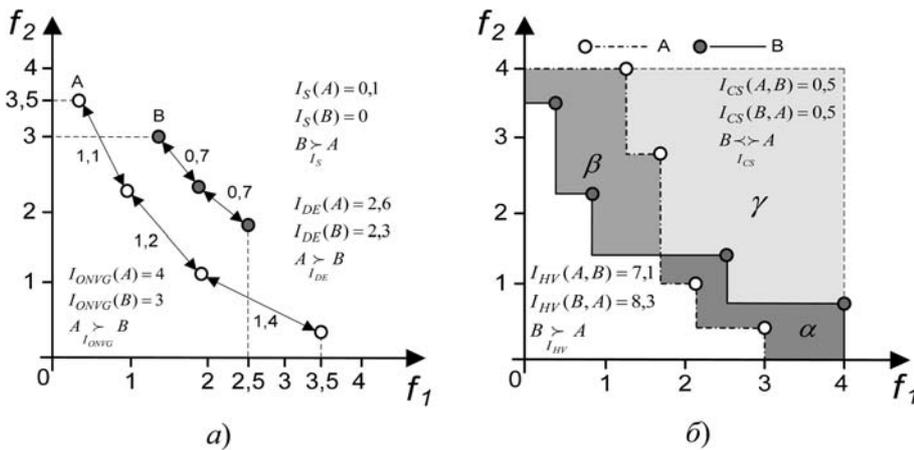


Рис. 2. Графический смысл некоторых индикаторов эффективности: а — унарные индикаторы; б — бинарные индикаторы

вследствие сложности их вычисления, а также не-универсальности. Например, индикатор I_{PFV} (Pareto Front Visualization) не относится ни к одной из двух групп, так как основывается на непосредственном визуальном анализе границы Парето [8]. Это позволяет внешне оценить ее свойства относительно других границ. Очевидно, что применение этого индикатора имеет смысл для числа критериев $m \leq 3$. Также необходимость выполнения анализа человеком может стать источником ошибок ввиду сложности топологий границ Парето. Далее рассматривается практический набор сначала унарных, затем бинарных индикаторов, вычисляемых в пространстве как критериев, так и решений и позволяющих всесторонне оценить результаты работы МГА.

3. Унарные индикаторы

Индикатор I_{ONVG} (Overall Nondominated Vector Generation) определяет мощность найденного множества Парето: $I_{ONVG}(A) = |A|$, $A \subseteq X_P$. Большее значение индикатора соответствует лучшему результату (рис. 2, а).

Индикатор I_S (Spacing) может быть использован для оценки равномерности распределения решений вдоль границы Парето и вычисляется следующим образом:

$$I_S(A) = \frac{\sum_{i=1}^{|A|-1} |d_i - \bar{d}|}{(|A|-1) \cdot \bar{d}},$$

где $A \subseteq X_P$; d_i — расстояние между каждой парой соседних решений, предварительно упорядоченных по некоторому критерию; \bar{d} — среднее всех значений d_i .

Данный индикатор принимает значения $I_S \geq 0$. Значение I_S , близкое к нулю, свидетельствует о достаточно равномерном распределении решений вдоль границы Парето (рис. 2, а). Однако она может

быть изначально фрагментирована вследствие особенностей решаемой задачи, что может привести к снижению объективности этого индикатора. Поэтому для повышения качества оценки равномерности заполнения границы Парето дополнительно могут быть использованы следующие два индикатора [9].

Индикатор $I_{SD(\epsilon)}$ (spacing distribution in ϵ -area) позволяет оценить число фрагментов границы Парето, где решения плотно сгруппированы в некоторой окрестности $\epsilon > 0$ относительно $A \subseteq Y_P$:

$$I_{SD(\epsilon)}(A) = \frac{1}{|A|-1} \sum_{a \in A} |\{b \in A \mid \|a - b\| > \epsilon\}|.$$

$I_{SD(\epsilon)}$ принимает значения из интервала $[0, |A|]$, они пропорциональны степени равномерности заполнения границы Парето. Значение $I_{SD(\epsilon)} = |A|$ означает, что для каждого решения в пространстве критериев в окрестности ϵ нет других решений.

Индикатор I_{DE} (dimensions extent) позволяет оценить максимальную протяженность границы Парето по каждой из размерностей (рис. 2, а). Чем больше его значение, тем шире диапазон изменения каждого критерия границы Парето:

$$I_{DE}(A) = \sqrt{\sum_{i=1}^m \max\{|f_i(a) - f_i(b)| \mid a, b \in A, A \subseteq X_P\}}.$$

Индикатор I_{OT} (overall time computing) ориентирован исключительно на оценку вычислительной эффективности МГА при определении некоторого множества Парето $A \subseteq X_P$. Как известно, у МГА все процессорное время тратится на вычисление значений критериев и выполнение генетических операторов. При рациональной программной реализации именно последнее будет определять сравниваемую временную сложность МГА, решающих одинаковую задачу. Поэтому в ряде задач многокритериальной оптимизации, например, связанных с системами реального времени, при одинаковых или достаточно близких значениях других индикаторов показатель I_{OT} может оказаться решающим. Меньшее значение этого индикатора соответствует лучшему показателю. Это свидетельствует о минимальном количестве и интенсивности применения генетических операторов для получения нового поколения. В итоге увеличение числа критериев в задаче оптимизации не приведет к резкому росту ее вычислительной сложности.

Унарные индикаторы являются необходимыми, но, как правило, недостаточными показателями эффективности МГА. Так, большое значение ин-

дикатора I_{ONVG} совершенно не позволяет судить о качестве найденных решений, а также эффективности используемого алгоритма. В то же время, группа унарных индикаторов имеет относительно невысокую вычислительную сложность. Это делает возможным их самостоятельное использование при оценке найденных решений непосредственно в процессе работы МГА. Качество же результирующего множества Парето должно оцениваться в сочетании с бинарными индикаторами.

4. Бинарные индикаторы

Индикатор I_{HV} (*hyper-volume*) вычисляется как гиперобъем m -мерной области, заключенной внутри гиперкуба между границей Парето и его вершиной с координатами $\{f_i^{\max}\}$, $i = 1, \dots, m$. Для случая $m = 2$ I_{HV} может быть определен как сумма площадей прямоугольников с координатами левого нижнего, правого верхнего углов соответственно $(y_1^{(j)}, y_2^{(j)})$, (f_1^{\max}, f_2^{\max}) , где $j = 1, \dots, |A|$, — число точек на границе Парето.

Пусть $A, B \subseteq X_P$ — два множества Парето, найденных МГА, тогда через I_{HV} можно также определить объем области пространства критериев, ограниченной A, B , но в которой, например, только решения из A являются недоминируемыми:

$$I_{HV}(A, B) = I_{HV}(A + B) - I_{HV}(B).$$

Как следует из примера на рис. 2, б, область размера α лимитируется только границей A , а область β , соответственно, только границей B , при этом γ — обеими границами. Принимая во внимание, что $I_{HV}(A, B) = \alpha$ и $I_{HV}(B, A) = \beta$, получим

$$\alpha + \beta + \gamma = I_{HV}(A + B);$$

$$\alpha + \gamma = I_{HV}(A);$$

$$\beta + \gamma = I_{HV}(B).$$

Большее значение I_{HV} соответствует лучшей оценке. В этом примере $I_{HV}(B, A) > I_{HV}(A, B)$, что показывает превосходство границы B над A . В случае $I_{HV}(A, B) = 0$ и $I_{HV}(B, A) > 0$ можно судить о том, что все решения из A доминируются соответствующими решениями из B .

Индикатор I_{CS} (*two set coverage*) позволяет определить, в каком отношении два множества Парето доминируют друг друга. I_{CS} принимает значения из интервала $[0, 1]$ и определяется следующим образом:

$$I_{CS}(A, B) = \frac{|\{b \in B | \exists a \in A: a \geq b\}|}{|B|},$$

где $A, B \subseteq X_P$, $|A| = |B|$.

Так, значение $I_{CS}(A, B) = 1$ означает, что все решения из B доминируются соответствующими решениями из A . Обратная ситуация $I_{CS}(A, B) = 0$ свидетельствует о том, что множество B в сравнении

с A целиком является недоминируемым. Для этого индикатора не выполняется условие $I_{CS}(A, B) = 1 - I_{CS}(B, A)$, также в случае трех оцениваемых множеств Парето ($A, B, C \subseteq X_P$) I_{CS} не транзитивна [7]. При вычислении I_{CS} не учитывается топология расположения решений на границах Парето, что может снизить точность полученной оценки. Например, на рис. 2, б $I_{CS}(A, B) = I_{CS}(B, A) = 0,5$, однако видно, что граница B в целом ближе к глобальной границе Парето, чем A . Необходимо заметить, что эта особенность учитывается в индикаторе I_{HV} , поэтому значение I_{CS} следует анализировать совместно с I_{HV} .

Индикатор I_{GD} (*generational distance*) позволяет оценить степень близости некоторой границы Парето $A \subseteq Y_P$ и заданной эталонной границы $B \subseteq Y_P$; $B = Y_P$, если известна глобальная граница Парето:

$$I_{GD}(A, B) = \frac{1}{|A|} \sum_{a \in A} \{\|a - b\| | b \in B\}.$$

Таким образом, I_{GD} вычисляется как сумма расстояний в пространстве критериев между каждым вектором $a \in A$ и ближайшим к нему вектором $b \in B$. Меньшее значение I_{GD} соответствует лучшей оценке.

Индикатор I_{UF} (*measure over a set of utility functions*) предназначен для сравнительной оценки двух множеств Парето $A, B \subseteq X_P$ относительно заданного множества U функций полезности. Он основывается на вычислении вероятности, что A лучше B на этом множестве:

$$I_{UF}(A, B, U, p) = \int_{u \in U} E(A, B, u) p(u) du;$$

$$E(A, B, u) = \begin{cases} 1, & \text{если } u^*(A) < u^*(B), \\ 1/2, & \text{если } u^*(A) = u^*(B), \\ 0, & \text{если } u^*(A) > u^*(B); \end{cases}$$

$$\forall x^{(1)}, x^{(2)} : x^{(1)} \succ x^{(2)} \mid u(x^{(1)}) < u(x^{(2)});$$

$$u^*(A) = \min_{x \in A} \{u(x)\}, \quad u^*(B) = \min_{x \in B} \{u(x)\};$$

$$U(A \succ B) = \{u \in U \mid u^*(A) < u^*(B)\},$$

где $u : R^m \rightarrow R$ ставит в соответствие каждой точке пространства критериев некоторое значение полезности; $p(u)$ — выражает плотность вероятности функции полезности $u \in U$.

Этот индикатор может оказаться определяющим при решении задачи многокритериальной оптимизации с последующим выбором окончательного решения с применением заданной функции полезности.

5. Тестовые наборы индикаторов

Представленные индикаторы позволяют полноценно и объективно оценить эффективность МГА в разных ситуациях его применения. На практике число используемых индикаторов, а также их сочетания в наборе могут быть различными. Это позво-

Наборы индикаторов для разных ситуаций оценки эффективности МГА

Случай оценки эффективности МГА	Набор индикаторов
Сравнение нового (существующего) МГА на тестовых задачах МО с известным глобальным множеством Парето	$\{I_{ONVG}, I_S, I_{DE}, I_{GD}, I_{OT}\}$
Сравнение результатов нескольких МГА (одного МГА с разной настройкой) при решении некоторой прикладной задачи МО	$\{I_{ONVG}, I_S, I_{SD(\epsilon)}, I_{DE}, I_{HV}, I_{CS}, I_{UF}, I_{OT}\}$
Оценка эффективности МГА непосредственно в процессе его работы	$\{I_{ONVG}, I_S, I_{SD(\epsilon)}, I_{DE}\}$

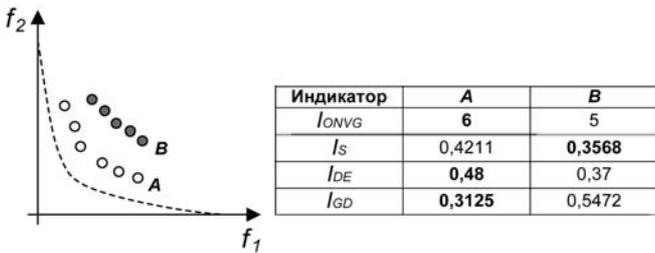


Рис. 3. Количественная оценка двух

ляет количественно оценить именно те показатели качества работы МГА, которые наиболее важны в данной ситуации их применения. В частности, для основных случаев использования МГА предлагаются следующие наборы индикаторов (см. таблицу).

На рис. 3 показана графическая иллюстрация для случая, когда нужно сравнить две границы Парето с известной глобально оптимальной третьей.

Приведенные на рис. 3 две границы Парето *A*, *B* оценены по четырем индикаторам. Их значения (лучшие выделены полужирным шрифтом) показывают, что граница Парето *B* превосходит *A* только по равномерности. По трем остальным индикаторам, а также по доступному и наиболее объективному в данном случае визуальному анализу, в качестве итоговой следует выбрать множество Парето, связанное с границей *A*.

Возможны ситуации, когда анализ значений индикаторов не позволяет сделать выбор среди найденных множеств Парето. Например, по одним показателям некоторое множество Парето лучше, по другим — хуже. В таком случае следует прежде всего учитывать те индикаторы, которые позволяют оценить данное множество Парето как наиболее близкое к оптимальному (желаемому для исследователя). Среди представленных такими индикаторами могут быть I_{HV} , I_{CS} , I_{GD} , I_{UF}

Заключение

Таким образом, использование различных комбинаций индикаторов позволяет объективно оце-

нить качество решений, полученных МГА. Кроме того, анализ значений индикаторов дает некоторое представление об особенностях многомерного пространства поиска, топологии полученной границы Парето. В частности, это облегчает ЛПР-формирование системы предпочтений относительно полученных Парето-оптимальных решений, например, более важными могут оказаться обособленно расположенные на границе Парето решения, или, наоборот, компактно расположенные друг с другом. Также особо следует подчеркнуть возможность использования систем индикаторов для оценки масштабируемости МГА, т. е. сохранения им эффективности при усложнении решаемых задач. Очевидно, что высокие показатели отдельных индикаторов при решении некоторой задачи МО не гарантируют подобный результат в других ситуациях. Исследование динамики изменения значений разных индикаторов при усложнении решаемых МГА задач многокритериальной оптимизации позволяет как сузить, так и расширить потенциал его возможностей, обосновать необходимость использования другого МГА либо создания нового. В следующей части статьи подобным образом исследуется масштабируемость наиболее применяемых сейчас МГА второго поколения — SPEA2 [10], NSGA-II [11].

Список литературы

1. **Соболь И. М., Статников Р. Б.** Выбор оптимальных параметров в задачах со многими критериями. 2-е изд., перераб. и доп. М.: Дрофа, 2006. 175 с.
2. **Ногин В. Д.** Принятие решений в многокритериальной среде: количественный подход. 2-е изд., испр. и доп. М.: ФИЗМАТЛИТ, 2004. 176 с.
3. **Гладков Л. А., Курейчик В. В., Курейчик В. М.** Генетические алгоритмы. М.: ФИЗМАТЛИТ, 2006. 320 с.
4. **Deb K.** Multi-Objective Optimization Using Evolutionary Algorithms. Wiley, 2009. 536 p.
5. **Казаков П. В.** Генетические алгоритмы многокритериальной оптимизации. Обзор. // Информационные технологии. 2011. № 9. С. 2—8.
6. **Zitzler E., Thiele L., Laumanns M., Fonseca C., Grunert da Fonseca V.** Performance Assessment of Multiobjective Optimizers: An Analysis and Review // IEEE Transactions on Evolutionary Computation. 2003. N 7 (2). P. 117—132.
7. **Knowles J., Corne D.** On Metrics for Comparing Nondominated Sets. In Congress on Evolutionary Computation (CEC'2002). 2002. Vol. 1. P. 711—716.
8. **Лотов А. В.** Компьютерная визуализация оболочки Эджворта—Парето и ее применение в интеллектуальных системах поддержки принятия решений. // Информационные технологии и вычислительные системы. 2002. № 1. С. 83—100.
9. **Zitzler E.** Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications. Ph. D thesis, Swiss Federal Institute of Technology. Zurich, 1999.
10. **Zitzler E., Laumanns M., Thiele L.** SPEA2: Improving the Strength Pareto Evolutionary Algorithm // Proc. of the EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems. 2002. P. 95—100.
11. **Deb K., Pratap A., Agarwal S., Meyarivan T.** A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II // IEEE Transactions on Evolutionary Computation. 2002. N 6 (2). P. 182—197.

Е. А. Елтаренко, канд. техн. наук, доц.,
Национальный исследовательский
ядерный университет "МИФИ"
E-mail: EAEltarenko@mephi.ru

Конъюнктивные операторы агрегирования в многокритериальных задачах

Рассматривается класс конъюнктивных предпочтений. Для него сформулированы аксиомы предпочтений. Анализируются вопросы проверки аксиом в предпочтениях лица, принимающего решения. Излагаются вопросы генерации операторов для описания конъюнктивных предпочтений. Приведены примеры конъюнктивных операторов.

Ключевые слова: принятие решений, многокритериальные задачи, агрегирование критериев, описание предпочтений, конъюнктивные операторы

Введение

Статья является развитием работ [1, 2], в которых излагается подход к агрегированию критериев в многокритериальных задачах (МКЗ), основанный на описании предпочтений лица, принимающего решения (ЛПР), операторами агрегирования. В них сформулированы аксиомы предпочтений, при выполнении которых подход может использоваться. В указанных работах приведены операторы, позволяющие описать предпочтения с определенными свойствами. Каждый оператор строится с использованием двух генерирующих функций $\varphi(x)$ и $\psi(x)$, которые определяют его свойства. Применение различных классов генерирующих функций приводит к операторам с разными свойствами. Поэтому можно ставить задачу формулирования свойств предпочтений, а затем определять класс генерирующих функций, которые позволят построить операторы, описывающие эти предпочтения. В настоящей статье демонстрируется данный подход. Сначала сформулируем свойство предпочтений, довольно часто встречающееся на практике.

При выборе ответственных решений, например, связанных с проектированием сложных объектов, таких как атомные станции, транспортные узлы и т. п., во множество критериев включают безопасность. При этом, как правило, к критерию безопасности предъявляется особое требование — значение, близкое к нулю, означает низкую оценку варианта решения в целом. Если критерии переведены в относительные единицы u_j ($j = 1, \dots, m$), то низким значениям критерия соответствуют значения, близкие к нулю, а высокие — единице. Проводя аналогию с операциями с булевыми переменными, можно

сказать, что имеет место операция конъюнкции. В конъюнктивных операторах малые значения одних критериев не компенсируются другими критериями.

Часто в таких случаях используется свертка

$$P(u_1, \dots, u_m) = \prod_{j=1}^m u_j^{V_j}, \text{ где } V_j \text{ — веса критериев.}$$

Вместе с тем, для описания данного класса предпочтений можно построить множество операторов агрегирования. Выбор конкретного оператора определяется предпочтениями ЛПР в решаемой задаче.

Наряду с конъюнктивными предпочтениями можно сформулировать и дизъюнктивные, хотя последние на практике встречаются реже. В дизъюнктивном операторе значение хотя бы одного критерия, равного единице, приводит к единичному значению всего оператора.

Для выделенных классов конъюнктивных и дизъюнктивных предпочтений необходимо:

- сформулировать аксиомы предпочтений;
- определить процедуры проверки аксиом;
- разработать процедуры генерации операторов агрегирования;
- определить требования к генерирующим функциям для операторов, описывающих данные предпочтения.

Материал данной статьи основывается на работах [1, 2]. Поэтому рекомендуется ознакомиться с указанными работами.

1. Аксиомы конъюнктивных и дизъюнктивных предпочтений

Имеется множество критериев u_1, u_2, \dots, u_m , измеряемых в относительных единицах в интервале $[0; 1]$. Используем следующие обозначения: \bar{u}_j — дополнение критерия u_j , т. е. множество критериев, за исключением u_j ; в дополнении $\bar{u}_{j,s}$ исключаются u_j и u_s . Обозначения предпочтений ЛПР $P(u_1, u_2, \dots, u_m) = P(u_j, \bar{u}_{j,s}) = P(u_j, u_s, \bar{u}_{j,s})$ эквивалентны.

Аксиома 1 монотонности предпочтений приведена в работе [2].

Аксиома 2 важности критериев для конъюнктивных предпочтений. Для любых двух критериев с весами $V_j > V_s$ и любым фиксированным дополнением $\bar{u}'_{j,s}$ выполняется $P(u_j, u_s^+, \bar{u}'_{j,s}) < P(u_j^+, u_s, \bar{u}'_{j,s})$, где $u_j^+ = u_s^+ = 1$ и любые значения $u_j > u_s > 0$. Из $V_j > V_s$ следует $P(u_j^+, u_s, \bar{u}'_{j,s}) = P(u_j, u_s^+, \bar{u}'_{j,s})$.

Отметим, что аксиома важности не совпадает с определением качественной важности критериев в работе [3, стр. 26–28]. В указанной работе считается, что критерий j важнее s , если $P(u_j^-, u_s^+, \bar{u}'_{j,s}) < P(u_j^+, u_s^-, \bar{u}'_{j,s})$, где уровни u^- и u^+ — произволь-

ные значения. В аксиоме 2 верхний уровень u^+ всегда должен быть равен единице. Это означает, что понятие качественной важности — частное по отношению к важности критериев в аксиоме 2.

Аксиома 3 конъюнктивности предпочтений включает три условия:

а) множественности объектов с нулевым предпочтением: для любого дополнения \bar{u}_j и $u_j^- = 0$ предпочтение $P(u_j^-, \bar{u}_j) = 0$;

б) единственности объекта с единичным предпочтением: $P(u_j^+, \bar{u}_j^+) = 1$ только при равенстве всех критериев единице;

в) для любых значений критериев u_j, \dots, u_m предпочтение $P(u_j, \dots, u_m) < \sum_{j=1}^m V_j u_j$.

Аксиома 4 дизъюнктивности предпочтений также включает три условия:

а) единственности объекта с нулевым значением: $P(u_j^-, \bar{u}_j^-) = 0$ только при равенстве всех критериев нулю;

б) множественности объектов с единичным предпочтением: для любого дополнения \bar{u}_j и $u_j^+ = 1$ предпочтение $P(u_j^+, \bar{u}_j) = 1$;

в) для любых значений критериев u_j, \dots, u_m предпочтение $P(u_j, \dots, u_m) > \sum_{j=1}^m V_j u_j$.

Понятно, что аксиомы (свойства) 3 и 4 взаимоисключающие.

Рассмотрим вопросы проверки вышеприведенных аксиом в предпочтениях ЛПР.

Сначала ЛПР должен задать веса критериев V_j ($j = 1, 2, \dots, m$), которые следует пронормировать, чтобы $\sum_{j=1}^m V_j = 1$.

Затем проверяем аксиому 2 важности критериев для конъюнктивных предпочтений. Для этого выбираем два уровня значений критериев: $u^+ = 1$ и $u^{-(1)} > 0$ и предлагаем ЛПР дать оценки следующих объектов: $W_j^{-(1)} = P(u_j^+, \bar{u}_j^{-(1)})$, $W_j^{+(1)} = P(u_j^{-(1)}, \bar{u}_j^+)$ ($j = 1, 2, \dots, m$). Отметим, что $P(u_j^+, \bar{u}_j^+) = 1$, а $P(u_j^{-(1)}, \bar{u}_j^{-(1)})$ есть оценка предпочтений при равных значениях критериев $u_j = u^{-(1)}$.

Если порядок $W_j^{-(1)}$ совпадает с порядком весов V_j , а порядок $W_j^{+(1)}$ — обратный порядку V_j , то для заданных уровней u^+ и $u^{-(1)}$ выполняются требования аксиомы. Изменив нижний уровень на $u^{-(2)}$ и вновь проведя опрос для оценок $W_j^{-(2)}$ и $W_j^{+(2)}$, проверим выполнение аксиомы и т. д.

Проверка аксиомы важности для дизъюнктивных предпочтений проводится по той же схеме, но в ка-

честве уровней значений критериев используются: нижний уровень $u^- = 0$ и верхний $u^{+(1)} > 0$. Тогда ЛПР дает оценки следующих объектов: $W_j^{-(1)} = P(u_j^{+(1)}, \bar{u}_j^-)$, $W_j^{+(1)} = P(u_j^-, \bar{u}_j^{+(1)})$. Условие выполнения аксиомы важности критериев такое же, как и для конъюнктивных предпочтений.

Чтобы проверить условие (а) аксиомы 3 конъюнктивности предпочтений, предъявляем ЛПР для оценки объекты $P(u_j^-, \bar{u}_j^-)$, где $u_j^- = 0$, а \bar{u}_j^- — любое дополнение критерия. Если все оценки равны нулю, то условие выполняется.

Проверка условия (в) (сравнение предпочтений с аддитивными) затруднительна для ЛПР, так как для этого потребовалось бы предъявить ему для оценки очень большое число объектов. Эти вопросы будут рассматриваться далее.

2. Генерация конъюнктивных операторов

Аксиома монотонности и аксиома важности критериев справедливы для функционала следующего вида:

$$P(u_j, \dots, u_m) = \psi \left(\sum_{j=1}^m \varphi_j^{-1}(V_j, u_j) \right), \quad (1)$$

где $\varphi_j^{-1}(V_j, u_j)$ — обратная функция от $\varphi_j(V_j, x)$ (вес V_j — параметр).

Функции $\psi(x)$ и $\varphi_j(V_j, x)$ — монотонно убывающие и для каждой пары критериев, в которой $V_j > V_s$, между функциями $\varphi_j(V_j, x)$, $\varphi_s(V_s, x)$ должны выполняться неравенства $\varphi_j(V_j, u) > \varphi_s(V_s, u)$ для конъюнктивных предпочтений. При $V_j = V_s$ $\varphi_j(V_j, u) = \varphi_s(V_s, u)$.

Для построения $\varphi_j(V_j, x)$ ($j = 1, 2, \dots, m$) будем использовать одну функцию $\varphi(x)$ с параметром V_j , который позволит формировать разные $\varphi_j(V_j, x) = \varphi(V_j, x)$. Такое ограничение на $\varphi_j(V_j, x)$ связано с выполнением условия (в) аксиомы 3 конъюнктивных предпочтений. Эти вопросы обсуждаются в пп. 3 и 6.

Как и в работах [1, 2], будем формировать оператор агрегирования на основе двух генерирующих функций $\psi(x)$, $\varphi(x)$.

Для выполнения условия (а) аксиомы 3 генерирующие функции $\varphi(x)$ и $\psi(x)$ должны иметь

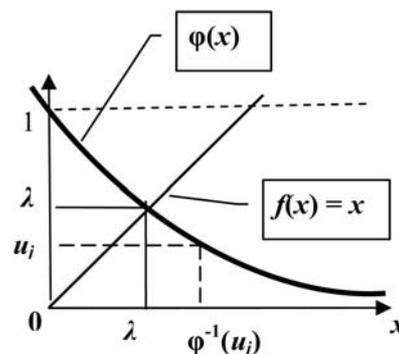


Рис. 1. Генерирующая функция конъюнктивного оператора

$\lim_{x \rightarrow \infty} (\varphi(x)) = 0$ (рис. 1), тогда предел обратной функции $\lim_{u \rightarrow 0} (\varphi^{-1}(u)) = \infty$. Чтобы выполнялось условие

(б) аксиомы 3, обе генерирующие функции при нуле должны быть равны единице: $\varphi(0) = \psi(0) = 1$. В п. 5 приведены примеры видов генерирующих функций: степенная, показательная, гиперболическая, тангенциальная.

На основе генерирующих функций $\varphi(x)$ и $\psi(x)$ сформируем следующий оператор:

$$h(u_1, \dots, u_m) = \psi \left(\sum_{j=1}^m V_j \varphi^{-1}(u_j) \right). \quad (2)$$

Этот оператор в работе [1] определен как оператор типа h . Убывающие функции $\varphi_j(V_j, x) = \varphi(x/V_j)$ упорядочиваются в соответствии со значениями весов, т. е. при $V_j > V_s$ функция $\varphi_j(V_j, x) > \varphi_s(V_s, x)$, значит, аксиома 2 важности критериев выполняется.

В операторах типа g , который в работе [1] определяется выражением $g(u_1, \dots, u_m) = \psi \left(\sum_{j=1}^m \varphi^{-1}(V_j u_j) \right)$,

вес критерия учитывается в качестве множителя u_j в обратной функции $\varphi_j^{-1}(V_j u_j)$. В этом случае порядок

$\varphi_{jg}(x) = \frac{1}{V_j} \varphi(x)$ ($j = 1, \dots, m$) — обратный порядок весов, что означает $\varphi_{jg}(V_j, x) < \varphi_{sg}(V_s, x)$ при

$V_j > V_s$. Для выполнения аксиомы 2 в качестве $\varphi_{jg}(x)$ следует взять $V_j \varphi(x)$, тогда тип g для конъюнктивных операторов будет иметь вид

$$g(u_1, \dots, u_m) = \psi \left(\sum_{j=1}^m \varphi^{-1}(u_j/V_j) \right). \quad (3)$$

Использование V_j в качестве делителя u_j эквивалентно введению в качестве весовой характеристики в формуле (3) $\tilde{V}_j = \frac{1}{V_j}$. Новые значения весов

$\tilde{V}_j = \frac{1}{V_j}$ следует пронормировать $\tilde{V}_j = \frac{1}{V_j} / \sum_{s=1}^m \frac{1}{V_s}$,

чтобы выполнялось условие $\sum_{j=1}^m \tilde{V}_j = 1$. Отношение между новыми весами — обратное к исходным:

$\frac{\tilde{V}_j}{\tilde{V}_s} = \frac{V_s}{V_j}$. Так как веса измеряются в шкале отношений,

то интерпретация V_j и \tilde{V}_j как важности критериев одинакова.

Обозначим выражение, являющееся в операторе (3) аргументом функции $\psi(x)$, через $Fg = \sum_{j=1}^m \varphi^{-1}(u_j/V_j)$,

с учетом введенных весов \tilde{V}_j $Fg = \sum_{j=1}^m \varphi^{-1}(\tilde{V}_j u_j)$.

Чтобы выполнялось условие (б) аксиомы 3 (единственности объекта с единичным предпочтением), необходимо Fg отмасштабировать — при всех критериях, равных единице, Fg должна равняться нулю. Это можно сделать так же, как для операторов мягкой логики в работе [2]:

а) ввести масштабный коэффициент k в качестве множителя, который будет вычисляться из уравнения

$$Fg_a = \sum_{j=1}^m \varphi^{-1}(k \tilde{V}_j) = 0, \text{ тогда оператор примет вид}$$

$$g_a(u_1, \dots, u_m) = \psi \left(\sum_{j=1}^m \varphi^{-1}(k \tilde{V}_j u_j) \right); \quad (4)$$

б) ввести масштабный коэффициент d в качестве слагаемого $Fg_b = \sum_{j=1}^m \varphi^{-1}(\tilde{V}_j u_j) - d$, где $d = \sum_{j=1}^m \varphi^{-1}(\tilde{V}_j) = \sum_{j=1}^m d_j$, в результате получим оператор типа g_b :

$$g_b(u_1, \dots, u_m) = \psi \left(\sum_{j=1}^m \varphi^{-1}(\tilde{V}_j u_j) - d \right). \quad (5)$$

Приведенные выше типы операторов в общем виде можно записать как $P(u_1, \dots, u_m) = \psi \left(\sum_{j=1}^m \varphi_j^{-1}(V_j, u_j) \right)$,

где $\varphi_j(V_j, x)$ — генерирующая функция для каждого критерия u_j , полученная на основе одной функции $\varphi(x)$ и параметра V_j . Для оператора (2) $\varphi_{jh}(x) =$

$= \varphi(x/V_j)$, для оператора (4) $\varphi_{jga}(x) = \frac{1}{k \tilde{V}_j} \varphi(x)$, а для (5)

$\varphi_{jgb}(x) = \frac{1}{\tilde{V}_j} (\varphi(x - d_j))$.

В работе [1] рассмотрены вопросы влияния генерирующих функций на значения оператора. Функция $\varphi(x)$ определяет выпуклость линий безразличия оператора. Функция $\psi(x)$ выполняет роль учета совместного влияния на значение оператора всех критериев. Оператор агрегирования

$$P(u_1, \dots, u_m) = \psi(F(u_1, \dots, u_m)),$$

где $F(u_1, \dots, u_m) = \sum_{j=1}^m \varphi_j^{-1}(V_j, u_j)$.

Выпуклость функций будем характеризовать параметром λ , который будет определять параметр функции из уравнения $\varphi(\lambda) = \lambda$ (см. рис. 1). Для

рассматриваемых генерирующих функций параметр выпуклости λ определен в интервале $(0; 1)$.

Таким образом, на основе пары генерирующих функций получили три типа операторов (2), (4), (5), различающихся способом учета весов и принципами масштабирования.

На рис. 2 приведен пример конъюнктивного оператора для двух критериев. Так как функция $\varphi(x)$ выпукла вниз, то поверхности безразличия в конъюнктивных операторах выпуклы вниз. Параметр выпуклости λ_φ функции $\varphi(x)$ вместе с весами определяет выпуклость поверхностей (линий) безразличия.

Анализируя оператор g_a , введем в рассмотрение $\varphi_{g_a}(x) = \frac{1}{k} \varphi(x)$, где $k > 1$ — масштабный коэффициент. Тогда оператор (4) запишется в виде

$$g_a = \psi \left(\sum_{j=1}^m \varphi_{g_a}^{-1}(\tilde{V}_j u_j) \right).$$

Оператор

$$g_b = \psi \left(\sum_{j=1}^m \varphi_{g_b}^{-1}(\tilde{V}_j u_j) - d \right),$$

где $\varphi_{g_b}(x) = \varphi(x)$. Используя одну генерирующую функцию $\varphi(x)$, для оператора типа g получили две функции, выпуклые вниз, $\varphi_{g_a}(x)$ и $\varphi_{g_b}(x)$, которые

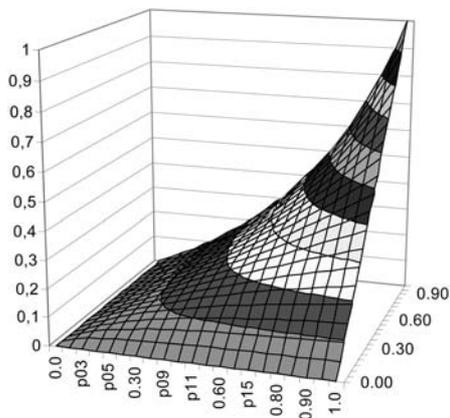


Рис. 2. Конъюнктивный оператор для двух критериев

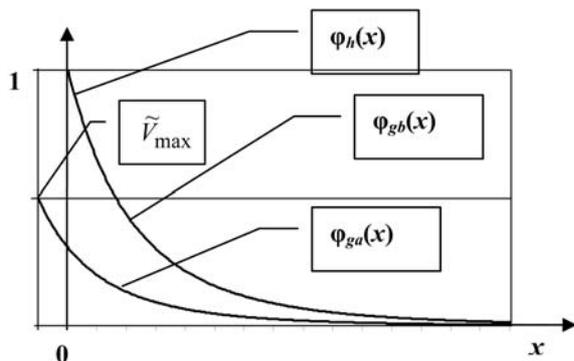


Рис. 3. Генерирующие функции $\varphi(x)$ конъюнктивных операторов разных типов

различаются условиями нормировки $Fg(u_1, \dots, u_m)$. В операторах типа g_b используется только часть генерирующей функции $\varphi(x)$ на интервале $[0; \tilde{V}_{\max}]$ в положительной области аргумента (рис. 3). В операторах $\varphi_h(x) = \varphi(x)$ $x \in [0; \infty)$, а в $\varphi_{g_a}(x)$ используются и отрицательные значения аргумента.

3. Требования к генерирующим функциям для выполнения аксиомы конъюнктивности

Аксиома 3 конъюнктивности отличается от аксиом мягкой конъюнкции, приведенных в работе [2], тем, что для конъюнктивности выполняется свойство множественности объектов с нулевым предпочтением, а для мягкой конъюнкции существует единственный объект с минимальным нулевым предпочтением: $P(u_j^-, \bar{u}_j^-) = 0$ при $u^- = 0$.

Выполнение свойства множественности объектов с нулевым предпочтением для конъюнктивных операторов обеспечивается классом монотонно убывающих генерирующих функций $\varphi(x)$ с $\lim_{x \rightarrow \infty} (\varphi(x)) = 0$

и $\psi(x)$ с $\lim_{x \rightarrow \infty} (\psi(x)) = 0$. Требования к генерирующим

функциям $\varphi(x)$ и $\psi(x)$ для выполнения условия (в) аксиомы конъюнктивности $P(u_j, \dots, u_m) <$

$< \sum_{j=1}^m V_j u_j$ будут такими же, как для мягкой конъюнкции, рассмотренные в работе [2]. Поэтому

приведем их без доказательства.

В [2] введено понятие линии экстремальных значений оператора (ЛЭЗО). Это линия в пространстве критериев, описываемая параметрически через параметр $t \in [0; 1]$, на которой $F(u_1, \dots, u_m) =$

$= \sum_{j=1}^m \varphi_j^{-1}(V_j, u_j)$ принимает экстремальные значения.

Для конъюнктивных операторов это линия максимальных значений $F_{\max}(t)$. В [2] приведены

правила построения ЛЭЗО для выпуклых функций, которые справедливы и для $\varphi(x)$ конъюнктивных операторов. Поэтому генерирующие функции

$\varphi(x)$ и $\psi(x)$ должны быть выпуклы.

Теорема 1. Чтобы оператор был конъюнктивным, необходимо и достаточно, чтобы функции $\varphi(x)$ и $\psi(x)$ были выпуклы, и на линии максимальных значений операторов функция $F_{\max}(t)$ была больше $\psi^{-1}(t)$.

Для операторов типа h ЛЭЗО представляет собой линию при равных значениях переменных, а сам оператор образует одномерную функцию, получаемую из оператора (2) при равных значениях $u_j = u$:

$$h(u) = \psi \left(\sum_{j=1}^m V_j \varphi^{-1}(u) \right) = \psi(\varphi^{-1}(u)). \quad (6)$$

В [1] функция (6) определена как *функция равных значений переменных* (ФРЗП).

Условия теоремы 1 для операторов типа h можно сформулировать более наглядно.

Теорема 2 о существовании конъюнктивных операторов типа h . Если генерирующие функции $\varphi(x)$ и $\psi(x)$ выпуклы вниз и $\varphi(x) > \psi(x)$, то $h(u_1, \dots, u_m) <$

$$< \sum_{j=1}^m V_j u_j, \text{ т. е. оператор } h \text{ конъюнктивный.}$$

Используя понятие функции равных значений переменных, на основе теоремы 2 сформулируем следующее очевидное утверждение.

Утверждение. Для всех типов конъюнктивных операторов функция равных значений переменных меньше линейной. Для операторов типа h условие $h(u) < u$ является необходимым и достаточным, а для типов g — только необходимым.

4. Генерация дизъюнктивных операторов

Для генерации данного класса операторов требуются выпуклые вверх генерирующие функции $\varphi_d(x)$ и $\psi_d(x)$, имеющие $\lim_{x \rightarrow \infty} (\varphi_d(x)) = 1$. В качестве таких функций можно использовать $\varphi_d(x) = 1 - \varphi(x)$, $\psi_d(x) = 1 - \psi(x)$, где $\varphi(x)$ и $\psi(x)$ — рассмотренные ранее генерирующие функции для конъюнктивных операторов. В результате получим для операторов

$$\text{типа } h_d(u_1, \dots, u_m) = 1 - \psi \left(\sum_{j=1}^m V_j \varphi^{-1}(1 - u_j) \right), \text{ что}$$

соответствует закону двойственности (де Моргана) [3, с. 385] $h_d(u_1, \dots, u_m) = 1 - h_c(1 - u_1, \dots, 1 - u_m)$.

Дизъюнктивные операторы типа g_a и g_b также формируются из конъюнктивных на основе закона двойственности.

5. Примеры видов генерирующих функций и конъюнктивных операторов

Различных видов генерирующих функций можно сформировать довольно много. В качестве примеров в табл. 1 и 2 приведены: степенная, показательная, гиперболическая и тангенциальная функции и образованные на их основе конъюнктивные операторы с $\varphi(x)$ и $\psi(x)$ одинакового вида. Как указывалось ранее, выпуклость функций определяет параметр $\lambda \in (0; 1)$. Параметры функций рассчитываются через λ в результате решения уравнения $\varphi(\lambda) = \lambda$.

Следует сказать, что в одном операторе могут использоваться в качестве $\psi(x)$ и $\varphi(x)$ разные виды функций. Чтобы оператор был конъюнктивным, требуется выполнение условий теоремы 1 или 2. Для примера построения такого оператора типа h используем две генерирующие функции: $\varphi(x) = \frac{1}{ax + 1}$ (гиперболическая) и $\psi(x) = (x + 1)^{-p}$ (степенная).

$$\text{Сначала запишем выражение } Fh = \frac{1}{a} \left(\sum_{j=1}^m \frac{V_j}{u_j} - 1 \right),$$

$$\text{итоговый вид оператора: } h(u_1, \dots, u_m) = \left(\frac{1}{a} \left(\sum_{j=1}^m \frac{V_j}{u_j} - 1 \right) + 1 \right)^{-p}. \text{ Остается определить параметры } a \text{ и } p,$$

Таблица 1

Примеры генерирующих функций и конъюнктивных операторов

Название оператора	Генерирующая функция	Параметр	Обратная функция $\varphi^{-1}(u)$	Конъюнктивный оператор $h_c(u_1, \dots, u_m)$	ФРЗП $h_c(u)$
Степенной	$(x + 1)^{-p}$	$p = -\frac{\ln(\lambda)}{\ln(1 + \lambda)}$	$u^{-1/p_\varphi} - 1$	$\left[\sum_{j=1}^m V_j u_j^{-1/p_\varphi} \right]^{-p_\psi}$	$[u^{-1/p_\varphi}]^{-p_\psi}$
Показательный	e^{-ax}	$a = -\frac{\ln(\lambda)}{\lambda}$	$-\frac{1}{a_\varphi} \ln(u)$	$\left(\prod_{j=1}^m u_j \right)^{\frac{a_\psi}{a_\varphi}}$	$(u)^{\frac{a_\psi}{a_\varphi}}$
Гиперболический	$\frac{1}{ax + 1}$	$a = \frac{1 - \lambda}{\lambda^2}$	$\frac{1}{a_\varphi} \left(\frac{1}{u} - 1 \right)$	$\frac{1}{\frac{a_\psi}{a_\varphi} \left(\sum_{j=1}^m V_j \frac{1}{u_j} - 1 \right) + 1}$	$\frac{u}{\frac{a_\psi}{a_\varphi} (1 - u) + u}$
Тангенциальный	$1 - \frac{2}{\pi} \operatorname{arctg}(ax)$	$a = \frac{1}{\lambda} \operatorname{tg} \left(\frac{\pi}{2} (1 - \lambda) \right)$	$\frac{1}{a_\varphi} \operatorname{tg} \left(\frac{\pi}{2} (1 - u) \right)$	$1 - \frac{2}{\pi} \operatorname{arctg} \left(\frac{a_\psi}{a_\varphi} \sum_{j=1}^m V_j \operatorname{tg} \left(\frac{\pi}{2} (1 - u_j) \right) \right)$	$1 - \frac{2}{\pi} \operatorname{arctg} \left(\frac{a_\psi}{a_\varphi} \operatorname{tg} \left(\frac{\pi}{2} (1 - u) \right) \right)$

Примечание: $\lambda \in (0; 1)$ — параметр выпуклости генерирующей функции.

Примеры конъюнктивных операторов типа g

Название оператора	Масштабный коэффициент k	Конъюнктивный оператор типа $g_a(u_1, \dots, u_m) = \psi\left(\sum_{j=1}^m \varphi^{-1}(k \tilde{V}_j u_j)\right)$	Масштабный коэффициент d	Конъюнктивный оператор типа $g_b(u_1, \dots, u_m) = \psi\left(\sum_{j=1}^m \varphi^{-1}(\tilde{V}_j u_j) - d\right)$
Степенной	$\left(\frac{1}{m} \sum_{j=1}^m \tilde{V}_j^{-1/p_\varphi}\right)^{p_\varphi}$	$\left[\sum_{j=1}^m (k \tilde{V}_j u_j)^{-1/p_\varphi} - m + 1\right]^{-p_\psi}$	$\sum_{j=1}^m \tilde{V}_j^{-1/p_\varphi} - m$	$\left[\sum_{j=1}^m (\tilde{V}_j)^{-1/p_\varphi} (u_j^{-1/p_\varphi} - 1) + 1\right]^{-p_\psi}$
Показательный	$\left(\prod_{j=1}^m \tilde{V}_j\right)^{-1/m}$	$\prod_{j=1}^m (k \tilde{V}_j u_j)^{\frac{a_\psi}{a_\varphi}}$	$\ln\left(\prod_{j=1}^m \tilde{V}_j\right)^{-1/a_\varphi}$	$\prod_{j=1}^m u_j^{\frac{a_\psi}{a_\varphi}}$ (не отвечает требованиям к операторам, так как не содержит весов)
Гиперболический	$\frac{1}{m} \sum_{j=1}^m \frac{1}{\tilde{V}_j}$	$\frac{1}{\frac{a_\psi}{a_\varphi} \left(\sum_{j=1}^m \frac{1}{k \tilde{V}_j u_j} - m\right) + 1}$	$\frac{1}{a_\varphi} \left(\sum_{j=1}^m \frac{1}{\tilde{V}_j} - m\right)$	$\frac{1}{\frac{a_\psi}{a_\varphi} \left(\sum_{j=1}^m \frac{1}{\tilde{V}_j} \left(\frac{1}{u_j} - 1\right)\right) + 1}$
Тангенциальный	Из уравнения: $\sum_{j=1}^m \operatorname{tg}\left(\frac{\pi}{2}(1 - k \tilde{V}_j)\right) = 0$	$1 - \frac{2}{\pi} \operatorname{arctg}\left(\frac{a_\psi}{a_\varphi} \sum_{j=1}^m \operatorname{tg}\left(\frac{\pi}{2}(1 - k \tilde{V}_j u_j)\right)\right)$	$\frac{1}{a_\varphi} \sum_{j=1}^m \operatorname{tg}\left(\frac{\pi}{2}(1 - \tilde{V}_j)\right)$	$1 - \frac{2}{\pi} \operatorname{arctg}\left(\frac{a_\psi}{a_\varphi} \sum_{j=1}^m \operatorname{tg}\left(\frac{\pi}{2}(1 - \tilde{V}_j u_j)\right) - a_\varphi d\right)$

при которых этот оператор будет конъюнктивным, т. е. надо выполнить условие теоремы 2: $\varphi(x) > \psi(x)$. Производные генерирующих функций: $\varphi'(x) = -a(ax + 1)^{-2} < 0$; $\psi'(x) = -p(x + 1)^{-(p+1)} < 0$. При a и p больше 1 $\varphi'(x) > \psi'(x)$, $x \in [0; \infty)$. С учетом, что $\psi(0) = \varphi(0) = 1$, получаем $\varphi(x) > \psi(x)$. Поэтому оператор конъюнктивный при $p \geq a > 1$.

Генерирующие функции для конъюнктивных операторов можно сформировать из генерирующих функций операторов мягкой конъюнкции, приведенных в работе [2]. Если генерирующая функция $\varphi(x) \in [0; 1]$ для квазиконъюнктивного оператора, то $\varphi_c(x) = \varphi\left(\frac{1}{x+1}\right)$ отвечает требованиям генерирующей функции конъюнктивного оператора. Действительно, $\varphi_c(0) = \varphi(1) = 1$, значение $\varphi_c(x) = 0$ достигается при $x \rightarrow \infty$, если $\varphi(x)$ выпукла вниз, то и $\varphi_c(x)$ выпукла вниз.

Рассмотрим примеры построения генерирующих функций $\varphi_c(x)$ из $\varphi(x)$ операторов мягкой конъюнкции.

Степенная генерирующая функция для квазиконъюнктивного оператора $\varphi(x) = x^p$ для конъюнктивного будет равна $\varphi_k(x) = (1 + x)^{-p}$. В табл. 1 эта генерирующая функция и используется.

Показательная генерирующая функция, выпуклая вниз, квазиконъюнктивного оператора $\varphi(x) = \frac{e^{ax} - 1}{e^a - 1}$ (см. [2]) для конъюнктивного будет:

$$\varphi_c(x) = \frac{e^{a/(x+1)} - 1}{e^a - 1}. \text{ Параметр } a \text{ определяется из уравнения } \lambda = \frac{e^{a/(\lambda+1)} - 1}{e^a - 1}.$$

Интересно отметить, что приведенные в табл. 1 конъюнктивные операторы типа h совпадают с производственными функциями, рассматриваемыми в экономике [5]. Функция Кобба—Дугласа имеет вид:

$$\prod_{j=1}^m x_j^{a_j}, \text{ где } x_j \text{ — переменные, } a_j = \text{const, причем } \sum_{j=1}^m a_j = 1. \text{ Она совпадает с показательным оператором при равенстве генерирующих функций } a_\varphi \text{ и } a_\psi \text{ (см. табл. 1).}$$

Функция Солоу имеет вид $\left(\sum_{j=1}^m a_j x_j^{-p}\right)^{-v/p}$, где p и v — параметры. Приняв $p = 1/p_\varphi$ и $v/p = v p_\varphi = p_\psi$, получим, что производственная функция Солоу совпадает со степенным конъюнктивным оператором типа h .

Такое совпадение объясняется сходством требований к производственным функциям и конъюнктивным операторам: монотонности по всем переменным, равенства нулю функции при одной или более переменных, равных нулю.

6. Идентификация конъюнктивных операторов

Алгоритм идентификации конъюнктивных операторов такой же, как для операторов мягкой конъюнкции, изложенный в работе [2]. В качестве оце-

нок ЛПР используются данные, полученные при проверке аксиомы важности критериев (см. п. 1, аксиома 2).

На практике могут встретиться случаи, когда в результате проверки свойства множественности объектов с нулевым предпочтением $P(u_j^-, \bar{u}_j^+) = 0$, где $u^- = 0$, а \bar{u}_j^+ — любое дополнение критерия, оказалось, что не для всех критериев оно выполняется. Рассмотрим такую ситуацию на примере.

Пусть из четырех критериев оно выполняется только для первого, т. е. $P(0, u_2', u_3', u_4') = 0$, а для остальных

$$\begin{aligned} P(u_1', 0, u_3', u_4') > 0, P(u_1', u_2', 0, u_4') > 0, \\ P(u_1', u_2', u_3', 0) > 0. \end{aligned} \quad (7)$$

Это означает, что генерирующая функция $\varphi_1(x)$ для первого критерия должна отличаться от $q(x)$ для 2-, 3- и 4-го критериев, как показано на рис. 4. Функция $\psi(x)$ должна соответствовать конъюнктивному оператору, т. е. с $\lim_{x \rightarrow \infty} (\psi(x)) = 0$. Вместе с

тем, с двумя или несколькими генерирующими функциями $\varphi_1(x)$ и $q(x)$ разного типа не обеспечивается условие (в) аксиомы конъюнктивности:

$$P(u_j, \dots, u_m) < \sum_{j=1}^m V_j u_j \quad (\text{см. Леммы 1, 2 и теорему 1}$$

в работе [2], в которых рассматривается только одна выпуклая генерирующая функция $\varphi(x)$).

Неравенства (7) указывают на то, что для агрегирования трех последних критериев подходит оператор мягкой конъюнкции. Следует сказать, что убывающая функция $q(x)$, представленная на рис. 4, получается из возрастающей функции операторов мягкой конъюнкции: $q(x) = \varphi(1 - x)$. Поэтому, чтобы обеспечить выполнение условия (в) аксиомы 3 для рассматриваемого примера, следует из критериев 2, 3, 4 сформировать комплексный критерий $U_2(u_{2,1}, u_{2,2}, u_{2,3})$ с $u_{2,1} = u_2, u_{2,2} = u_3, u_{2,3} = u_3$. Для агрегирования этих критериев используем оператор мягкой конъюнкции. Соответственно, веса критериев $V_{2,1}, V_{2,2}, V_{2,3}$ нормируются, чтобы их сумма равнялась единице.

Тогда для критериев u_1 и U_2 выполняется $P(0, U_2) = 0$, а оценка $P(u_1', 0) = P(u_1', 0, 0, 0)$ будет

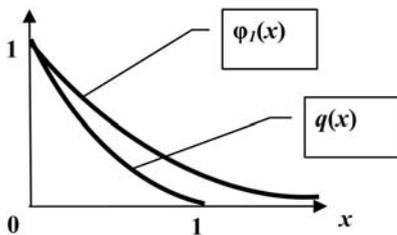


Рис. 4. Генерирующие функции для примера с $P(u_1', 0, u_3', u_4') > 0$

много меньше оценок (7), т. е. $P(u_1', 0)$ будет близка к нулю. Взяв в качестве веса для комплексного критерия U_2 сумму $\widehat{V}_2 = V_2 + V_3 + V_3$, можно определить генерирующие функции конъюнктивного оператора.

Использование двух уровней для агрегирования всех четырех критериев позволяет выполнить свойство (в) $P(u_j, \dots, u_m) < \sum_{j=1}^m V_j u_j$. Действительно, если для u_1 и U_2 выполняется $P(u_1, U_2) < V_1 u_1 + \widehat{V}_2 U_2(u_{2,1}, u_{2,2}, u_{2,3})$, где $\widehat{V}_2 = V_2 + V_3 + V_3$, и для $U_2(u_{2,1}, u_{2,2}, u_{2,3}) < V_{2,1}' u_{2,1} + V_{2,2}' u_{2,2} + V_{2,3}' u_{2,3}$, где $V_{2,j}' = \frac{V_j}{V_2 + V_3 + V_3} = \frac{V_j}{\widehat{V}_2}$, то и общий результат агрегирования будет $P(u_1, U_2) < V_1 u_1 + V_2 u_2 + V_3 u_3 + V_4 u_4$.

Заключение

Рассмотрены специальные классы предпочтений ЛПР в многокритериальных задачах, названных конъюнктивными и дизъюнктивными предпочтениями. Для них сформулированы аксиомы предпочтений. Приводится процедура опроса лица, принимающего решения, для проверки аксиом. Определены свойства генерирующих функций и типы операторов агрегирования, которые описывают такие предпочтения.

Приведены примеры видов конъюнктивных операторов: степенной, показательный, гиперболический, тангенциальный. Число разных операторов определяется используемыми генерирующими функциями. Приведенные в статье требования к генерирующим функциям позволяют строить их.

Следует отметить, что некоторые конъюнктивные операторы типа h совпадают с производственными функциями, рассматриваемыми в экономике, что объясняется сходством требований к производственным функциям и конъюнктивным операторам.

Список литературы

1. Елтаренко Е. А. Аппроксимация предпочтений в многокритериальных задачах // Информационные технологии. 2011. № 6. С. 22–32.
2. Елтаренко Е. А. Операторы мягкой логики в многокритериальных задачах // Информационные технологии. 2011. № 10. С. 8–18.
3. Подиновский В. В. Введение в теорию важности критериев в многокритериальных задачах принятия решений. М.: Физматлит, 2007. 65 с.
4. Корн Г., Корн Т. Справочник по математике (для научных работников и инженеров). М.: Наука, 1973. 832 с.
5. Клейнер Г. Б. Производственные функции: Теория, методы, применение. М.: Финансы и статистика, 1986. 239 с.

Ю. А. Зак, д-р техн. наук, независимый исследователь, консультант, Германия,
e-mail: yuriy_zack@hotmail.com

Распределение множества заданий и определение оптимальных очередностей их выполнения на параллельных машинах методами динамического программирования

Предлагаются математические модели и изучаются свойства задач построения выполнения расписаний на параллельных и различных по техническим характеристикам машинах в условиях учета потерь на переналадку машин, ограничений на времена выполнения заданий и допустимые сроки работы машин. Конструируются операторы исключения из рассмотрения подмножеств расписаний, не содержащих допустимых и неоптимальных решений. На основе установленных свойств рассмотрены алгоритмы решения рассматриваемых задач методами динамического программирования.

Ключевые слова: расписания на параллельных машинах, потери на переналадку, ограничения на сроки выполнения заданий и времена работы машин, динамическое программирование

Введение

Задачи распределения множества подлежащих выполнению заданий по рабочим станциям (машинам) и определения оптимальных последовательностей их выполнения на каждой машине в условиях потерь времени на переналадку имеют большое количество приложений в календарном планировании производства, маршрутизации перевозок, организации обслуживания и выполнения ремонтных работ, организации вычислительного процесса и т. п. (см. [1, 4, 17]).

Методам решения данного класса задач уделялось значительное место в монографиях и периодической литературе [1–8]. Известны подходы, связанные с построением линейных и нелинейных целочисленных моделей и с использованием методов математического программирования. Алгоритмы решения задач данного класса достаточно большой размерности, представляющих практический интерес, требуют больших объемов вычислений [5, 7].

Наиболее широкое распространение получили приближенные методы решения задач данного класса с использованием генераторов случайных расписаний [4, 6, 8, 10], использующих различные правила предпочтения [4], эвристические подходы [7, 12], а также генетические алгоритмы и эволюционные стратегии [8, 11, 13]. Алгоритмы решения данной задачи без учета ограничений на директивные

сроки выполнения заданий методами построения кратчайших допустимых путей на графах приведены в работе автора [14–17]. Указанные подходы позволили в ряде случаев находить эффективные расписания выполнения заданий для многих практических приложений. Однако наличие жестких ограничений на директивные сроки выполнения заданий в ряде случаев затрудняет процесс генерирования допустимых расписаний и существенно увеличивает затраты на поиск. Кроме того, отсутствие нижних оценок значения критерия оптимальности для построенного расписания не позволяет объективно оценить эффективность полученного решения. В монографии автора [17] исследованы свойства допустимых и оптимальных планов сформулированной задачи и предложены алгоритмы ее решения методами ветвей и границ.

В данной работе изучаются свойства задач такого класса, на основе которых конструируются операторы исключения из рассмотрения подмножеств расписаний, не содержащих допустимых и неоптимальных решений. На основе установленных свойств предлагаются алгоритмы решения рассматриваемых задач методами динамического программирования алгоритмами построения допустимых и экстремальных путей на графе.

1. Постановка и математическая модель задачи

На K , $k = 1, \dots, K$, рабочих станциях (машинах) должны быть выполнены N различных работ (заданий), $i, j = 1, \dots, N$. Каждое из заданий должно выполняться только на одной машине и без разрывов времени в процессе его выполнения. Пусть заданы:

- директивные сроки допустимого времени наиболее раннего срока начала b_i и наиболее позднего срока завершения каждого из заданий B_i , $i = 1, \dots, N$;
- матрица времен выполнения каждого из заданий на всех машинах $\bar{t}^k = (t_1^k, t_2^k, \dots, t_i^k, \dots, t_N^k)$, $k = 1, \dots, K$;
- h^k и H^k — наиболее ранние допустимые сроки начала и завершения выполнения работ на каждой k -й машине;
- $\bar{\tau}^k = |\tau_{ij}^k|$, $i, j = 0, 1, \dots, N$, $k = 1, \dots, K$, — матрицы времен потерь времени на переналадку при переходе k -й машины от выполнения одного задания к другому. На пересечении i -й строки и j -го столбца этих матриц стоят потери времени на переналадку k -й машины при переходе после выполнения i -го задания к j -му. В 0-й строке каждой матрицы $\bar{\tau}^k$ заданы времена настройки машины при переходе ее из состояния, в котором она находится в момент начала выполнения расписания работ, в режим выполнения i -го задания. Элементы 0-го столбца определяют затраты времени на переход k -й машины после завершения выполнения j -го задания в режим простоя (или возвращения машины из j -го пункта на базу).

Необходимо найти распределение всего множества заданий по машинам, а также определить последовательности всех назначенных на каждой машине заданий, обеспечивающие выполнение всех ограничений на установленные сроки их выполнения T_i , и минимизировать время завершения всего комплекса работ (критерий оптимальности F_1). В качестве других критериев оптимальности могут быть выбраны F_2 — минимальное средневзвешенное время работы машин, необходимое для выполнения всего комплекса работ, а также F_3 — наиболее равномерная загрузка всех работающих машин.

Частным случаем рассматриваемых задач являются задачи построения последовательностей выполнения N заданий на K параллельных и различных по производительности машинах в условиях, когда потерями времени на переналадку машин можно пренебречь, т. е. $\tau_{ij}^k = 0, i, j = 0, 1, \dots, N; k = 1, \dots, K$, и случаи, когда времена выполнения каждого из заданий и потеря на переналадку для всех машин одинаковы, т. е.

$$t_j^k = t_j \text{ и } \tau_{ij}^k = a_{ij}, i, j = 0, 1, \dots, N; k = 1, \dots, K.$$

Построим матрицы суммарных затрат времени на выполнения j -го задания и переналадок для каждой k -й машины при переходе от выполнения i -го к j -му заданию, $A^k = \|\tau_{ij}^k + t_j^k\|, i, j = 0, 1, \dots, N, k = 1, \dots, K$, где $i = 0$ — режим начального состояния (остановка машины),

$$a_{ij}^k = \begin{cases} \tau_{ij}^k + t_j^k, & \text{если } i = 1, \dots, N, \\ \tau_{ij}^k, & \text{если } i = 0. \end{cases} \quad (1)$$

Необходимо разбить все множество выполняемых заданий \tilde{I} на K непересекающихся подмножеств \tilde{I}^k , т. е. $\tilde{I}^k \cap \tilde{I}^l = \emptyset, k, l = 1, \dots, K, k \neq l; \bigcup_{k=1}^K \tilde{I}^k = \tilde{I}$, определить последовательности \bar{U}^k выполнения подмножеств этих заданий на каждой машине, каждая из которых должна начинаться и завершаться в начальной вершине $i^k = 0$.

Пусть последовательность \bar{U}^k имеет вид $\bar{U}^k = \{0, i_1^k, \dots, i_p^k, \dots, i_n^k, 0\}$. Времена начала и завершения стоящих в последовательности \bar{U}^k заданий определяются согласно выражениям

$$\begin{aligned} \theta_{i_1}^k &= \max(h^k + \tau_{0, i_1}^k, b_{i_1}), \quad T_{i_1}^k = \theta_{i_1}^k + t_{i_1}^k; \\ \theta_{i_2}^k &= \max(T_{i_1}^k + 1 + \tau_{i_1, i_2}^k, b_{i_2}), \quad T_{i_2}^k = \theta_{i_2}^k + t_{i_2}^k, \dots, \\ \theta_{i_n}^k &= \max(T_{i_{n-1}}^k + 1 + \tau_{i_{n-1}, i_n}^k, b_{i_n}), \\ T_{i_n}^k &= \theta_{i_n}^k + t_{i_n}^k. \end{aligned} \quad (2)$$

Обозначим $R^k = T_{i_n}^k + a_{i_n, 0}^k, k = 1, \dots, K$, — время завершения выполнения расписания работ на k -й машине. Здесь $T_{i_n}^k$ и R^k — соответственно время завер-

шения стоящего последним в последовательности \bar{U}^k задания и время завершения выполнения расписания.

На строящиеся последовательности выполнения заданий накладываются ограничения:

- время работы каждой машины не должно превышать выделенный для ее работы ресурс времени

$$R^k \leq H^k, k = 1, \dots, K, \quad (3)$$

- время завершения выполнения каждого из заданий не должно превышать заданные граничные значения B_i

$$T_{i_p}^k \leq B_{i_p}, p = 1, \dots, n, k = 1, \dots, K. \quad (4)$$

В качестве критериев оптимальности могут рассматриваться:

- завершение выполнения расписания работ в кратчайшие сроки

$$F_1 = \min_{1 \leq k \leq K} \max_{1 \leq k \leq K} R^k; \quad (5)$$

- минимизация средневзвешенного суммарного времени работы машин

$$F_2 = \min \sum_{k=1}^K w^k R^k, \text{ где } 0 \leq w^k \leq 1, k = 1, \dots, K; \sum_{k=1}^K w^k = 1. \quad (6)$$

В качестве частного случая может быть принято $w^k = 1, k = 1, \dots, K$.

Критерий наиболее равномерной загрузки машин имеет вид:

$$F_3 = \min_{1 \leq k \leq K} \max_{1 \leq s \leq K} \max_{k > s} |R^k - R^s|. \quad (7)$$

Здесь $|R^k - R^s|$ — абсолютное значение разности двух величин.

Ниже рассмотрены алгоритмы решения сформулированной задачи методами динамического программирования [2, 4, 5, 17]. В процессе решения строятся различные допустимые частичные планы (последовательности) выполнения работ на каждой машине. В процессе этих построений проводится отсев недопустимых и не содержащих оптимальных решений последовательностей.

На втором этапе решения проводится проверка, возможно ли из имеющихся последовательностей выполнения заданий на каждой машине сконструировать расписание выполнения всех заданий, удовлетворяющее всем ограничениям задачи и являющееся оптимальным согласно выбранному критерию оптимальности. В процессе решения задачи на первом и втором этапах используются сформулированные ниже правила сравнения и отсева неперспективных вариантов.

2. Свойства допустимых и оптимальных расписаний

Правило отсева 1. Если при построении последовательности выполнения заданий на некоторой

k -й машине при включении i_l^k -го задания на последнее место в строящейся подпоследовательности $\bar{U}^k = \{0, i_1^k, \dots, i_{l-1}^k, i_l^k\}$ выполняется одно из следующих неравенств

$$T_{i_l}^k = \theta_{i_l}^k + t_{i_l}^k > B_{i_l}, \quad T_{i_l}^k = T_{i_l}^k + a_{i_l,0}^k > H^k, \quad (8)$$

где значения $\theta_{i_l}^k$ определяются по формулам (2), то это задание не может быть включено на l -е место в последовательности выполнения заданий на k -й машине.

Правило отсева 2. Если для двух последовательностей $\bar{U}^{1k} = \{0, i_{i_1}^{1k}, \dots, i_{i_p}^{1k}\}$ и $\bar{U}^{2k} = \{0, i_{i_1}^{2k}, \dots, i_{i_r}^{2k}\}$ выполнения заданий на одной и той же k -й машине справедливы соотношения

$$i_{i_p}^{1k} = i_{i_r}^{2k}, \quad \tilde{I}^{2k} \subseteq \tilde{I}^{1k}, \quad T_{i_p}^{1k} \leq T_{i_r}^{2k}, \quad (9)$$

то последовательность \bar{U}^{2k} может быть отброшена как неперспективная.

Следствия правила отсева 2.

При построении подпоследовательности $\bar{U}^k = \{0, i_1^k, \dots, i_{l-1}^k, i_l^k\}$ выполнение заданий на k -й машине осуществляется на основании использования правила отсева 2 по результатам проверки соотношений (9), исключается из рассмотрения как неперспективная либо вновь построенная подпоследовательность, либо все ранее построенные неперспективные подпоследовательности.

Для продолженных во все допустимые вершины графа последовательностей выполнения заданий на одной и той же машине может быть использовано следующее правило предпочтения.

Правило отсева 3. Если для двух построенных и недопустимых для дальнейшего продолжения последовательностей $\bar{U}^{1k} = \{0, i_{i_1}^{1k}, \dots, i_{i_p}^{1k}\}$ и $\bar{U}^{2k} = \{0, i_{i_1}^{2k}, \dots, i_{i_r}^{2k}\}$ выполнения заданий на одной и той же k -й машине, для которых индексы стоящих последними в последовательностях заданий $i_{i_p}^{1k}$ и $i_{i_r}^{2k}$ могут быть различными, справедливы соотношения

$$\tilde{I}^{2k} \subseteq \tilde{I}^{1k}, \quad R^{1k} \leq R^{2k}, \quad (10)$$

то последовательность \bar{U}^{2k} может быть отброшена как неперспективная.

Эти правила отсева используются на 1-м этапе решения задачи.

На втором этапе решения в качестве допустимых последовательностей выполнения заданий \bar{U}^q на первых Q машинах, $q = 1, \dots, Q$, где $Q = 2, \dots, K$, рассматриваются только такие, у которых подмножества включенных в них заданий имеют следующие свойства:

$$\tilde{J}^Q = \bigcup_{q=1}^Q \tilde{J}^q, \quad \tilde{J}^q \cap \tilde{J}^p = \emptyset, \quad q \neq p, \quad q, p = 1, \dots, Q. \quad (11)$$

Для следующей и еще нерассмотренной k -й машины \bar{U}^k с подмножеством заданий \tilde{I}^k рассматриваются только такие метки, для которых справедливо соотношение $\tilde{J}^Q \cap \tilde{I}^k = \emptyset$. Обозначим такое объединение последовательностей Ω^Q .

Правило отсева 4. Если для двух построенных объединений последовательностей $\Omega^{1,q}$ и $\Omega^{2,q}$, $q = 1, \dots, Q$, подмножеств $\tilde{J}^{1,Q}$ и $\tilde{J}^{2,Q}$ при решении задачи по критерию F_1 справедливы соотношения

$$\tilde{J}^{2,Q} \subseteq \tilde{J}^{1,Q}, \quad \max_{1 \leq q \leq Q} R^{1,q} \leq \max_{1 \leq q \leq Q} R^{2,q} \quad (12)$$

либо при решении задачи по критерию F_2 справедливы соотношения

$$\tilde{J}^{2,Q} \subseteq \tilde{J}^{1,Q}, \quad \sum_{q=1}^Q w^q R^{1,q} \leq \sum_{q=1}^Q w^q R^{2,q}, \quad (13)$$

то объединение частичных планов $\Omega^{2,q}$ может быть отброшено как неперспективное.

Сформулированные правила отсева неперспективных продолжений используются в работе описанного выше алгоритма.

3. Вычислительная схема алгоритма

При построении различных вариантов последовательностей выполнения заданий на каждой машине метка s -й строящейся последовательности содержит следующие признаки:

v_1^s — номер машины; v_2^s — номер метки для этой машины; v_3^s — номер задания, включенного на данное место в последовательность; v_4^s — номер предыдущей метки (продолжаемой на этом шаге последовательности); v_5^s — возможность и перспективность продолжения данной последовательности. При этом $v_5^s = 0$, если данная последовательность может быть продолжена, и $v_5^s = 1$ — в противном случае, $v_5^s = 2$, если эта метка исключена из рассмотрения как неперспективная и не должна рассматриваться на 2-м этапе работы алгоритма;

$\bar{U}^s = \{i_0^k, i_1^k, i_2^k, \dots, i_n^k\}$ — построенная последовательность выполнения заданий;

$T^s = T_{i_n}^k$ — время завершения стоящего в данной последовательности задания;

R^s — время работы машины (с учетом возвращения в исходное состояние i_0^k), необходимое для выполнения данной последовательности работ.

На втором этапе алгоритма осуществляется объединение в общее расписание построенных последовательностей. При этом метка каждого рассматриваемого g -го варианта состоит из следующих признаков: g — номер варианта; $\bar{U}^{g,k} = \{i_1^{g,k}, i_2^{g,k}, \dots, i_n^{g,k}\}$, $k = 1, \dots, K$, — последовательности выполнения заданий на каждой машине; $R^{g,k}$, $k = 1, \dots, K$, — вектор суммарных времен работы каждой машины.

3.1. Алгоритм 1-го этапа

В начале работы 1-го этапа алгоритма сформируем начальные ($s = 0$) метки для каждой k -й машины:

$$v_1^s = k, v_2^s = 0, v_3^s = i_0, v_4^s = \emptyset, v_5^s = 0;$$

$$\bar{U}^s = \{i_0\}, T^s = R^s = 0.$$

Для каждого значения $k = 1, \dots, K$ выполняем следующие построения.

Шаг 1. Среди меток k -й машины находим s -ю метку, для которой $v_5^s = 0$. Если такую метку удастся найти, то переходим к шагу 2. Если таких меток не существует, то построены все допустимые последовательности выполнения заданий для данной машины и переходим к следующей машине. Если таких меток не существует ни для одной из машин, то построены все допустимые последовательности выполнения заданий на данной машине и переходим к шагу 4.

Шаг 2. Рассматриваем s -ю метку. Подмножество заданий, для которых выполняются все условия

$$D_j^{s,k} = \max(b_j, T^{s,k} + \tau_{i_n, j}^{s,k}) + t_j^k \leq B_j,$$

$$D_j^{s,k} + a_{j,0}^k \leq H^k, j \in \tilde{I} / \tilde{J}^{s,k}, \quad (14)$$

обозначим $\hat{J}^{s,k} = \{j_1^{s,k}, \dots, j_\lambda^{s,k}, \dots, j_M^{s,k}\}$. В выражении (14) $\tilde{J}^{s,k}$ — подмножество заданий, входящих в последовательность \bar{U}^s .

Пусть $M^{s,k}$ — число заданий в подмножестве $\hat{J}^{s,k}$. Если $\hat{J}^{s,k} = \emptyset$, то полагаем для s -й метки $v_5^s = 1$ и переходим к шагу 1.

Если $\hat{J}^{s,k} \neq \emptyset$, то образуем $M^{s,k}$ новых меток, продолжив и поставив на последнее место последовательности \bar{U}^s каждое из заданий подмножества $\tilde{J}^{s,k}$.

Признаки каждой из этих меток имеют вид

$$v_1^{s+\lambda} = k, v_2^{s+\lambda} = s + \lambda, v_3^{s+\lambda} = j_\lambda^{s+\lambda}, v_4^{s+\lambda} = s,$$

$$v_5^{s+\lambda} = 0; \bar{U}^{s+\lambda} = \{i_0^{s,k}, \dots, i_n^{s,k}, j_\lambda^{s,k}\};$$

$$T^{s+\lambda} = \max(T_{i_n}^{s,k} + \tau_{i_n, j_\lambda}^k, b_{j_\lambda}) + t_{j_\lambda}^k,$$

$$R^{s+\lambda} = T^{s+\lambda} + a_{j_\lambda, i_0}^k; \lambda = 1, 2, \dots, M^{s,k}. \quad (15)$$

Полагаем для s -й метки признак $v_5^s = 1$ и переходим к шагу 3.

Шаг 3. Проводим попарно сравнение каждой из вновь образованных последовательностей $\bar{U}^{s+\lambda}$ со всеми хранящимися в банке данных последовательностями. Если для одной из рассматриваемых пар выполняются условия правила отсева 2, то соответствующая неперспективная метка с индексом γ исключается из рассмотрения, т. е. полагаем для

нее значение $v_5^\gamma = 2$. Выполняем переиндексацию всех рассматриваемых на данной машине вариантов. Исключив варианты, для которых $v_5^\gamma = 2$, переходим к шагу 1.

Шаг 4. Проводим попарно сравнение всех меток, для которых $v_5^s = 1$. Если в результате такого сравнения для какой-то из пар меток выполняется правило отсева 3, то для неперспективной и подлежащей отсеву метки полагаем значение $v_5^s = 2$. После выполнения этой процедуры и переиндексации всех перспективных меток, для которых $v_5^s = 1$, алгоритм 1-го этапа для k -й машины завершает свою работу.

Выполнив описанный выше алгоритм для каждой из рассматриваемых машин, переходим к этапу 2 алгоритма. Отметим, что если все параллельные машины имеют одинаковые технические характеристики и одни и те же допустимые времена начала и завершения выполнения работ, то на 1-м этапе построение альтернативных последовательностей выполнения заданий строится только для одной машины.

3.2. Алгоритм второго этапа

Метка рассматриваемых вариантов на 2-м этапе решения задачи имеет следующие признаки:

u^s — номер составной метки; $Z^s = |z^{sk}|$ — номера включаемых меток каждой машины; $\tilde{\Pi}^s$ — множество заданий, включенных в расписание выполнения работ на всех машинах; $F_1^s = R_{\min}^s$ — минимальное время расписания выполнения работ на всех машинах; F_2^s — средневзвешенное суммарное время работы машин.

Рассматриваются только допустимые и перспективные метки последовательностей выполнения заданий на каждой машине, для которых значение $v_5^{pk} = 1$. Каждая из этих меток имеет свой индекс $p^k = 1, \dots, P^k$. Обозначим через $F_1^{\min}(s)$, $F_2^{\min}(s)$ наилучшие значения соответствующих показателей на данном s -м шаге алгоритма. Введем вектор номеров рассмотренных меток каждой машины $L = |\lambda^k|$, $k = 1, \dots, K$, где $\lambda^k = 1, \dots, m^k$, а также счетчик номера анализируемой машины δ^k , $k = 1, \dots, K$. В начале процесса полагаем $s = 0$; $\lambda^k = 0$, $\delta^k = 0$, $k = 1, \dots, K$; $F_1^{\min}(s) = F_2^{\min}(s) = \infty$.

Алгоритм представляет собой последовательное выполнение следующих шагов.

Шаг 1. Полагаем $s := (s + 1)$. Формируем начальные признаки новой метки. Для $k = K, (K - 1), \dots, 1$, выполняем следующие действия.

1.1. Если $\lambda^k \leq (m^k + 1)$, то полагаем $z^{sk} = (\lambda^k + 1)$, $\lambda^k := (\lambda^k + 1)$; $z^{gk} = \lambda^g$, $g = (k - 1), \dots, 1$.

1.2. Если $\lambda^k = (m^k + 1)$, то полагаем $z^{sk} = \lambda^k = 1$, $z^{s, k-1} = (\lambda^{k-1} + 1)$, $\lambda^{k-1} := (\lambda^{k-1} + 1)$. При этом если $\lambda^{k+1} = (m^{k+1} + 1)$, то, положив $k := (k - 1)$, переходим к пункту 1.1. В противном случае полагаем $z^{gk} = \lambda^g$, $g := (k - 2)$, ..., 1.

Если $\lambda^k = m^k + 1$, $k = 1, \dots, K$, то переходим к шагу 4. В противном случае переходим к шагу 2.

Шаг 2. Полагаем $\tilde{\Pi}^s = \emptyset$, $F_1^s = F_2^s = 0$. Последовательно для индексов машин $k = 1, \dots, K$ выбираем метку, номер которой равен z^{sk} , подмножество выполняемых заданий для которой $\tilde{J}(z^{sk})$. Если $\tilde{J}(z^{sk}) \cap \tilde{\Pi}^s = \emptyset$, то определяем $\tilde{\Pi}^s = \tilde{\Pi}^s \cup \tilde{J}(z^{sk})$ и вычисляем в зависимости от выбранного критерия оптимальности значение:

$$F_1^s = \max[F_1^s, R^k(z^{sk})] \text{ или } F_2^s := F_2^s + w^k R^k(z^{sk}),$$

$$\text{или } F_3^s = \max_{1 \leq k \leq K} \max_{\substack{1 \leq s \leq K \\ k > s}} [R^k(z^{sk}) - R^s]. \quad (16)$$

Если $F_1^s \geq F_1^{\min}(s)$ или $F_2^s \geq F_2^{\min}(s)$, или $F_3^s \geq F_3^{\min}(s)$, либо если $\tilde{J}(z^{sk}) \cap \tilde{\Pi}^s \neq \emptyset$, то полагаем $\lambda^k := (\lambda^k + 1)$ и переходим к шагу 1. В противном случае переходим к шагу 3.

Шаг 3. Если $\tilde{\Pi}^s \neq \tilde{I}$, то переходим к шагу 1. Если $\tilde{\Pi}^s = \tilde{I}$, то, если $F_1^s < F_1^{\min}(s)$, полагаем $F_1^s = F_1^{\min}(s)$.

Если $F_2^s < F_2^{\min}(s)$, то полагаем $F_2^s = F_2^{\min}(s)$. Запоминаем номер метки s^* , соответствующей полученному наилучшему расписанию на данном шаге. Переходим к шагу 1.

Шаг 4. Получено оптимальное решение задачи, которое содержится в метке с номером s^* . Находим номера меток с номерами z^{s^*k} , соответствующие последовательностям выполнения заданий на каждой k -й машине $\bar{U}^k(z^{s^*k})$, а также времена работы машин $R^k(z^{s^*k})$, $k = 1, 2, \dots, K$. На основе этой информации могут быть определены времена начала и завершения выполнения каждого из заданий на данной машине. На этом алгоритм решения задачи завершает свою работу.

Следует отметить, что в практических расчетах решения задачи по критериям (5) или (7) минимизации времени выполнения расписания работ, в целях сокращения числа вариантов на 1-м этапе решения задачи, могут быть заданы вспомогательные дополнительные ограничения на время работы машин $\hat{H}^k < H^k$, $k = 1, \dots, K$. Это сократит также и объем вычислений на 2-м этапе алгоритма.

4. Иллюстративные примеры

4.1. Пример 1

На двух машинах необходимо выполнить шесть заданий. Суммарные времена выполнения заданий и потерь на переналадки при переходе от выполнения i -го заданий к j -му для каждой из машин сведены в табл. 1, 2.

Граничные сроки завершения каждого из заданий приведены в табл. 3.

Выполнение производственной программы должно начинаться и завершаться в состоянии $i = 0$. Необходимо определить подмножества заданий, выполняемых на каждой из машин, и последовательности их выполнения, которые обеспечили бы выполнение всех ограничений на сроки завершения заданий и минимизировали бы значение следующих критериев оптимальности:

1) выполнение производственной программы в кратчайшие сроки, $F_1 = \min \max(T_1, T_2)$;

2) минимизация суммарного времени работы машин $F_2 = \min(T_1 + T_2)$.

Таблица 1
Суммарные времена выполнения заданий и потерь на переналадки для 1-й машины

№ заданий	Суммарные времена выполнения заданий и переналадок						
	0	1	2	3	4	5	6
0	∞	7	5	6	4	6	8
1	13	∞	14	15	12	16	17
2	20	18	∞	20	21	20	17
3	24	24	25	∞	26	23	27
4	18	14	16	15	∞	18	16
5	13	13	16	16	14	∞	15
6	23	21	22	20	24	23	∞

Таблица 2
Суммарные времена выполнения заданий и потерь на переналадки для 2-й машины

№ заданий	Суммарные времена выполнения заданий и переналадок						
	0	1	2	3	4	5	6
0	∞	8	6	5	4	7	9
1	16	∞	16	17	15	18	14
2	21	20	∞	22	21	23	21
3	28	27	26	∞	28	29	26
4	22	18	19	20	∞	21	17
5	16	17	16	18	19	∞	15
6	15	24	23	26	25	27	∞

Таблица 3
Граничные сроки завершения заданий

Граничные сроки завершения выполнения заданий						
1	2	3	4	5	6	
51	22	65	50	37	75	

А. Строим допустимые последовательности выполнения заданий на каждой машине, осуществляя отсев неперспективных вариантов. Для сокращения объема изложения признаки меток v_1^s, v_5^s в описании примера опущены.

1-я машина

- 1) 1, 0, $\bar{U}_1^1 = \{1\}, 13, 20$;
- 2) 2, 1, $\bar{U}_2^1 = \{1, 3\}, 37, 43$;
- 3) 3, 1, $\bar{U}_3^1 = \{1, 4\}, 27, 31$;
- 4) 4, 1, $\bar{U}_4^1 = \{1, 5\}, 26, 31$;
- 5) 5, 1, $\bar{U}_5^1 = \{1, 6\}, 34, 42$;
- 6) 6, 2, $\bar{U}_6^1 = \{1, 3, 6\}, 57, 65$;
- 7) 7, 3, $\bar{U}_7^1 = \{1, 4, 3\}, 53, 59$;
- 8) 8, 3, $\bar{U}_8^1 = \{1, 4, 6\}, 51, 59$;
- 9) 9, 4, $\bar{U}_9^1 = \{1, 5, 3\}, 49, 55$;
- 10) 10, 4, $\bar{U}_{10}^1 = \{1, 5, 4\}, 44, 48$;
- 11) 11, 4, $\bar{U}_{11}^1 = \{1, 5, 6\}, 49, 57$;
- 12) 12, 5, $\bar{U}_{12}^1 = \{1, 6, 4\}, 50, 54$;
- 13) 13, 7, $\bar{U}_{13}^1 = \{1, 4, 3, 6\}, 73, 81$;
- 14) 14, 9, $\bar{U}_{14}^1 = \{1, 5, 3, 6\}, 69, 77$;
- 15) 15, 10, $\bar{U}_{15}^1 = \{1, 5, 4, 6\}, 68, 76$.

2-я машина

- 1) 0, $\bar{U}_1^2 = \{2\}, 21, 27$; 2) 0, $\bar{U}_2^2 = \{3\}, 28, 33$;
- 3) 0, $\bar{U}_3^2 = \{4\}, 22, 26$; 4) 0, $\bar{U}_4^2 = \{5\}, 16, 23$;
- 5) 0, $\bar{U}_5^2 = \{6\}, 26, 35$; 6) 1, $\bar{U}_6^2 = \{2, 3\}, 48, 53$;
- 7) 1, $\bar{U}_7^2 = \{2, 4\}, 40, 46$; 8) 1, $\bar{U}_8^2 = \{2, 6\}, 46, 55$;
- 9) 2, $\bar{U}_9^2 = \{3, 4\}, 48, 52$; 10) 2, $\bar{U}_{10}^2 = \{3, 5\}, 46, 53$;
- 11) 2, $\bar{U}_{11}^2 = \{3, 6\}, 54, 63$; 12) 3, $\bar{U}_{12}^2 = \{4, 3\}, 50, 54$;
- 13) 3, $\bar{U}_{13}^2 = \{4, 5\}, 41, 48$; 14) 3, $\bar{U}_{14}^2 = \{4, 6\}, 47, 56$;
- 15) 4, $\bar{U}_{15}^2 = \{5, 3\}, 45, 52$; 16) 4, $\bar{U}_{16}^2 = \{5, 4\}, 41, 51$;
- 17) 4, $\bar{U}_{17}^2 = \{5, 6\}, 43, 52$; 18) 5, $\bar{U}_{18}^2 = \{6, 3\}, 52, 57$;
- 19) 5, $\bar{U}_{19}^2 = \{5, 6\}, 47, 51$;
- 20) 6, $\bar{U}_{20}^2 = \{2, 3, 6\}, 67, 73$;
- 21) 7, $\bar{U}_{21}^2 = \{2, 4, 6\}, 65, 74$;

- 22) 9, $\bar{U}_{22}^2 = \{3, 4, 6\}, 73, 82$;
- 23) 10, $\bar{U}_{23}^2 = \{3, 5, 6\}, 63, 72$;
- 24) 12, $\bar{U}_{24}^2 = \{4, 3, 6\}, 76, 85$;
- 25) 13, $\bar{U}_{25}^2 = \{4, 5, 6\}, 68, 75$.

Следует отметить, что метка 24 рассматриваемых вариантов для 2-й машины, т. е. последовательность $\bar{U}_{24}^2 = \{4, 3, 6\}$, исключается как неперспективная, так как метка 22 с последовательностью выполнения заданий $\bar{U}_{22}^2 = \{3, 4, 6\}$ является более перспективной.

В. На втором этапе решения строим допустимые и наиболее эффективные объединения подмножеств выполняемых заданий на двух машинах.

Допустимыми являются только четыре объединения меток:

- 1) $\bar{U}_9^1 = \{1, 5, 3\}, T_9^1 = 55$, и $\bar{U}_{21}^2 = \{2, 4, 6\}, T_{21}^2 = 74$;
- 2) $\bar{U}_{10}^1 = \{1, 5, 4\}, T_{10}^1 = 54$, и $\bar{U}_{20}^2 = \{2, 3, 6\}, T_{20}^2 = 73$;
- 3) $\bar{U}_{14}^1 = \{1, 5, 3, 6\}, T_{14}^1 = 77$, и $\bar{U}_7^2 = \{2, 4\}, T_7^2 = 46$;
- 4) $\bar{U}_{15}^1 = \{1, 5, 4, 6\}, T_{15}^1 = 76$, и $\bar{U}_6^2 = \{2, 3\}, T_6^2 = 53$.

По критерию (5) минимизации времени завершения выполнения всех заданий оптимальным является расписание $\bar{U}_{10}^1 = \{1, 5, 4\}$ & $\bar{U}_{20}^2 = \{2, 3, 6\}$ со значением критерия оптимальности, равным $F_1^b = \max(54, 73) = 73$.

По критерию (6) минимизации суммарного времени работы машин оптимальным является расписание $\bar{U}_{14}^1 = \{1, 5, 3, 6\}$ & $\bar{U}_7^2 = \{2, 4\}$ со значением критерия оптимальности, равным $F_1^c = (77 + 46) = 123$.

4.2. Пример 2 (одинаковые по техническим характеристикам машины)

Суммарные времена переналадок и выполнения заданий для каждой машины сведены в табл. 4, а граничные сроки завершения выполнения у каж-

Таблица 4

Суммарные времена переналадок и выполнения заданий для каждой машины

№	0	1	2	3	4	5
0	∞	5	7	8	4	6
1	15	∞	15	17	20	23
2	16	16	∞	21	19	16
3	17	16	19	∞	22	18
4	15	21	17	20	∞	19
5	16	22	14	16	20	∞

Таблица 5

Граничные сроки завершения выполнения заданий

1	2	3	4	5
30	25	48	52	43

дого из заданий — в табл. 5. Время завершения выполнения всех работ каждой машиной не должно превышать значения 60.

Процесс решения иллюстративного примера 2.

А. Построения допустимых и перспективных расписаний работы каждой машины.

- 1) 0; (0, 1); 15, 20. 2) 0; (0, 2); 16, 23.
- 3) 0; (0, 3); 17, 25. 4) 0; (0, 4); 15, 19.
- 5) 0; (0, 5); 16, 22. 6) 1; (0, 1, 3); 31, 39.
- 7) 1; (0, 1, 4); 36, 40. 8) 1; (0, 1, 5); 37, 43.
- 9) 2; (0, 2, 3); 37, 45. 10) 2; (0, 2, 4); 33, 40.
- 11) 2; (0, 2, 5); 30, 36. 12) 3; (0, 3, 4); 35, 39.
- 13) 3; (0, 3, 5); 33, 39. 14) 4; (0, 4, 3); 37, 45.
- 15) 4; (0, 4, 5); 35, 41. 16) 5; (0, 5, 3); 34, 42.
- 17) 5; (0, 5, 4); 35, 39. 18) 6; (0, 1, 3, 4); 51, 55.
- 19) 11; (0, 2, 5, 3); 46, 54. 20) 11; (0, 2, 5, 4); 49, 53.
- 21) 13; (0, 3, 5, 4); 52, 56.

Б. Построение допустимых перспективных планов работы двух машин (допустимые объединения подмножеств):

- а) 6) 1; (0, 1, 3); 31, 39 и 20) 11; (0, 2, 5, 4); 49, 53.
- б) 7) 1; (0, 1, 4); 36, 40 и 19) 11; (0, 2, 5, 3); 46, 54.

По всем рассматриваемым критериям оптимальности наиболее предпочтительным является вариант а). В соответствии с этим решением 1-я машина выполняет задание 1 со сроком завершения, равным 15, затем задание 3 со сроком завершения, равным 31, и возвращается на базу в момент времени 39. 2-я машина выполняет задание 2 со сроком завершения, равным 16, затем задание 5 со сроком завершения, равным 30, после чего задание 3 со сроком завершения, равным 49, и возвращается на базу в момент времени 53.

Список литературы

1. Конвей Р. В., Максвелл В. Л., Миллер Л. В. Теория расписаний. М.: Физматгиз. Наука, 1975. 359 с.
2. Танаев В. С., Шкурба В. В. Введение в теорию расписаний. М.: Физматгиз. Наука, 1975. 256 с.
3. Танаев В. С., Ковалев М. Я., Шафранский Я. М. Теория расписаний. Групповые технологии. Минск: Изд. Института технической кибернетики НАН Беларуси, 1998. 289 с.
4. Domschke W., Scholl A., Vob S. Produktionsplanung. Ablauforganisations- und torische Aspekte. — Berlin, Heidelberg: Springer Verlag, 2005. 456 s.
5. Brucker P. Scheduling Algorithms. Berlin, Heidelberg und New York Springer-Verlag, 1998.
6. Blazewicz J., Ecker K. H., Pesch E., Schmidt G., Weglarz J. Scheduling Computer and Manufacturing Processes. Berlin-Heidelberg: Springer Verlag, 2001.
7. Herrmann J. Supply Chain Scheduling. Transaktionskostentheorie; Parallele Maschinen; Heuristik; Optimierungsmodelle. Berlin-Heidelberg: Gabler Verlag, 2010. 162 s.
8. Pinedo M. Scheduling: Theory, Algorithms and Systems. Berlin-Heidelberg: Springer Verlag, 1999.
9. Blazewicz J., Domschke W., Pesch E. The job shop scheduling problem: Conventional and new solution techniques // Eur. J. Oper. Res. 1996. P. 93, 1—33.
10. Herrmann J. Supply Chain Scheduling. Transaktionskostentheorie; Parallele Maschinen; Heuristik; Optimierungsmodelle. Berlin-Heidelberg: Gabler Verlag, 2010. 162 s.
11. Szelke E., Kerr R. M. Artificial Intelligence in Reactive Scheduling. London: Chapman & Hall, 1995.
12. Батишев Д. И., Гудман Э. Д., Норенков И. П., Прилуцкий М. Х. Метод комбинирования эвристик для решения комбинаторных задач упорядочения и распределения ресурсов // Информационные технологии. 1997. № 2. С. 29—32.
13. Курейчик В. М. Генетические алгоритмы. Таганрог: Изд. ТРТУ, 1998. 242 с.
14. Зак Ю. А. Алгоритмы решения задач "N коммивояжеров" // Кибернетика. 1972. № 2. С. 99—106.
15. Зак Ю. А. Об одном классе задач разбиения на допустимые подмножества и упорядочения // Техническая кибернетика. 1970. № 15. С. 27—36.
16. Зак Ю. А. Определение порядка выполнения независимых операций на параллельных машинах // Изв. АН СССР. Техническая кибернетика. 1969. № 2. С. 15—20.
17. Зак Ю. А. Прикладные задачи теории расписаний и маршрутизации перевозок. М.: URSS, 2011. 394 с.

УДК 004.925.8

В. А. Трудоношин, канд. техн. наук, доц.,

В. Г. Федорук, канд. техн. наук, доц.,

МГТУ им. Н. Э. Баумана

E-mail: trudonoshin@mail.ru

Сравнение математических моделей шарнирного соединения

Рассмотрены три варианта построения математической модели шарнира при моделировании плоских механических систем. Выявлены особенности этих моделей, выполнены вычислительные эксперименты, даны рекомендации по применению.

Ключевые слова: математическое моделирование, базис, модель, шарнир, механизм, динамика

Для моделирования динамики и кинематики плоских механизмов достаточно давно используют программные комплексы типа ПА9 [1], "Универсальный механизм" [2], Pradis [3], AMESim [4]. В составе библиотек математических моделей элементов этих комплексов присутствует модель шарнирного соединения, которое является одним из самых распространенных типов соединения в машинах и механизмах. Поэтому при моделировании динамики механизмов математической модели шарнира следует уделить особое внимание. Зачастую информация о реализации математической модели скрыта от пользователя. В данной статье будут рассмотрены три варианта построения модели шарнира для плоского движения (двумерная механическая система) и отмечены их особенности.

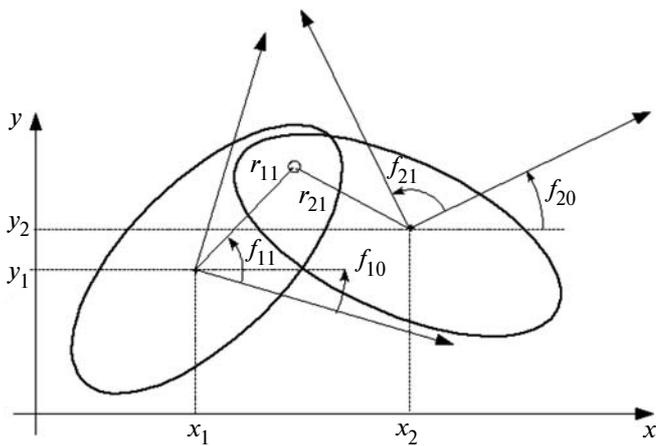


Рис. 1

На рис. 1 представлено шарнирное соединение тел с необходимыми для математических зависимостей обозначениями. С центром масс каждого тела связана локальная система координат, в которой задается положение шарнира через радиус-вектор и угол его поворота.

Уравнения, связывающие положения центров масс, будут выглядеть следующим образом:

$$\begin{aligned} x_1 - x_2 &= r_{11} \cos(f_{11} + f_{10}) - r_{21} \cos(f_{21} + f_{20}); \\ y_1 - y_2 &= r_{11} \sin(f_{11} + f_{10}) - r_{21} \sin(f_{21} + f_{20}). \end{aligned} \quad (1)$$

Модель 1. Поскольку фазовыми переменными механической системы являются силы (переменная типа потока) и скорости (переменная типа потенциала) [5], то, дифференцируя по времени уравнения (1), получим

$$\begin{aligned} v_x &= -w_1 r_{11} \sin(f_{11} + f_{10}) + w_2 r_{21} \sin(f_{21} + f_{20}); \\ v_y &= w_1 r_{11} \cos(f_{11} + f_{10}) - w_2 r_{21} \cos(f_{21} + f_{20}), \end{aligned} \quad (2)$$

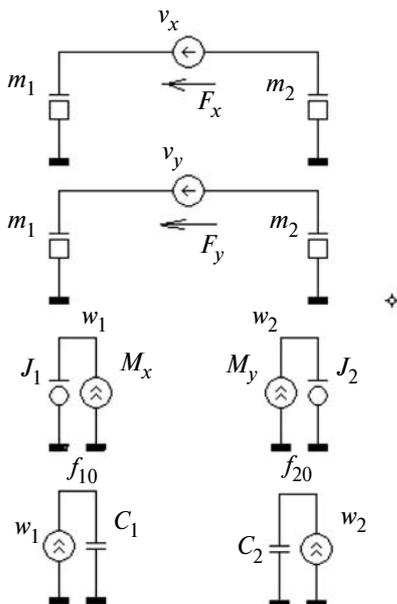


Рис. 2

где v_x — взаимная скорость перемещения центров масс тел вдоль оси X ; v_y — взаимная скорость перемещения центров масс тел вдоль оси Y ; w_1 и w_2 — угловые скорости тел относительно центра масс.

Полученные зависимости на эквивалентной схеме системы можно интерпретировать как источники скорости, включенные между массами. Силы, действующие в этих источниках, есть реакции в шарнире, которые создают вращательный момент относительно центра масс. Уравнения моментов имеют следующий вид:

$$\begin{aligned} M_1 &= F_x r_{11} \cos(f_{11} + f_{10}) + F_y r_{11} \sin(f_{11} + f_{10}); \\ M_2 &= -F_x r_{21} \cos(f_{21} + f_{20}) - F_y r_{21} \sin(f_{21} + f_{20}). \end{aligned} \quad (3)$$

Здесь M_1 — вращательный момент для тела 1; M_2 — вращательный момент для тела 2; F_x, F_y — реакции в шарнире по осям соответственно X и Y .

Эквивалентная схема шарнирного соединения представлена на рис. 2.

В эквивалентную схему включены два интегратора, которые позволяют через угловые скорости w_1 и w_2 получать углы f_{10} и f_{20} поворота локальных систем координат, присутствующие в зависимостях (2) и (3).

Математическая модель этого объекта в расширенном узловом базисе [6] будет выглядеть следующим образом:

$$\begin{aligned} \frac{dV_{x1}}{dt} - (V_{x1} - V_{x1}^{n-1})/h &= 0; \\ \frac{dV_{x2}}{dt} - (V_{x2} - V_{x2}^{n-1})/h &= 0; \\ \frac{dV_{y1}}{dt} - (V_{y1} - V_{y1}^{n-1})/h &= 0; \\ \frac{dV_{y2}}{dt} - (V_{y2} - V_{y2}^{n-1})/h &= 0; \\ \frac{dw_1}{dt} - (w_1 - w_1^{n-1})/h &= 0; \\ \frac{dw_2}{dt} - (w_2 - w_2^{n-1})/h &= 0; \\ \frac{dU_{c1}}{dt} - (U_{c1} - U_{c1}^{n-1})/h &= 0; \\ \frac{dU_{c2}}{dt} - (U_{c2} - U_{c2}^{n-1})/h &= 0. \end{aligned} \quad (4)$$

$$\begin{aligned} m_1 \frac{dV_{x1}}{dt} - F_x &= 0; \quad m_2 \frac{dV_{x2}}{dt} + F_x = 0; \\ m_1 \frac{dV_{y1}}{dt} - F_y &= 0; \quad m_2 \frac{dV_{y2}}{dt} + F_y = 0; \\ J_1 \frac{dw_1}{dt} + F_x r_{11} \cos(f_{10} + f_{11}) + F_y r_{11} \sin(f_{10} + f_{11}) &= 0; \\ J_2 \frac{dw_2}{dt} + F_x r_{21} \cos(f_{20} + f_{21}) + F_y r_{21} \sin(f_{20} + f_{21}) &= 0; \\ \frac{dU_{c1}}{dt} - w_1 &= 0; \quad \frac{dU_{c2}}{dt} - w_2 = 0; \end{aligned} \quad (5)$$

$$\begin{aligned}
& v_{x1} - v_{x2} + w_1 r_{11} \sin(f_{11} + f_{10}) - \\
& - w_2 r_{21} \sin(f_{21} + f_{20}) = 0; \\
& v_{y1} - v_{y2} - w_1 r_{11} \cos(f_{11} + f_{10}) + \\
& + w_2 r_{21} \cos(f_{21} + f_{20}) = 0.
\end{aligned} \quad (6)$$

Здесь m_1, m_2 — массы первого и второго тел; J_1, J_2 — центральные моменты инерции первого и второго тел; n — номер шага интегрирования; U_{c1}, U_{c2} — емкостные напряжения.

Уравнения (4) представляют собой формулы численного интегрирования, в данном случае неявного метода Эйлера, (5) — уравнения равновесия для узлов эквивалентной схемы и (6) — уравнения идеальных источников скорости.

Вектор неизвестных (базис) составляют следующие переменные:

$$\frac{dV_{x1}}{dt}, \frac{dV_{x2}}{dt}, \frac{dV_{y1}}{dt}, \frac{dV_{y2}}{dt}, \frac{dw_1}{dt}, \frac{dw_2}{dt}, \frac{dU_{c1}}{dt}, \frac{dU_{c2}}{dt}, \\
V_{x1}, V_{x2}, V_{y1}, V_{y2}, w_1, w_2, f_{10}, f_{20}, F_x, F_y,$$

где V_{x1}, V_{x2} — скорости центров масс тел 1 и 2 соответственно вдоль оси x ; V_{y1}, V_{y2} — скорости центров масс тел 1 и 2 соответственно вдоль оси y .

Решить систему уравнений можно, например, методом Ньютона, после чего выполняется переход к следующему шагу интегрирования.

Модель 2. Модель шарнира можно получить из модели двумерной пружины, для которой задана нулевая начальная длина. Модель двумерной пружины можно получить из следующих соотношений (рис. 3):

$$F_{\text{пр}} = k(L - L_0);$$

$$\begin{aligned}
L = & (((x_2 + r_{21} \cos(f_{21} + f_{20}) - \\
& - (x_1 + r_{11} \cos(f_{11} + f_{10}))^2 + ((y_2 + \\
& + r_{21} \sin(f_{21} + f_{20}) - (y_1 + r_{11} \sin(f_{11} + f_{10}))^2)^{1/2}; \\
F_x = & F_{\text{пр}} \cos \alpha; \quad F_y = F_{\text{пр}} \sin \alpha; \quad (7)
\end{aligned}$$

$$\sin \alpha = ((y_2 + r_{21} \sin(f_{21} + f_{20}) - (y_1 + r_{11} \sin(f_{11} + f_{10}))) / L;$$

$$\cos \alpha = ((x_2 + r_{21} \cos(f_{21} + f_{20}) - (x_1 + r_{11} \cos(f_{11} + f_{10}))).$$

Здесь $F_{\text{пр}}$ — сила в пружине; k — коэффициент жесткости; L_0 — начальная длина пружины.

Эквивалентная схема выглядит аналогично схеме модели 1, только вместо источников скорости стоят источники силы с компонентными уравнениями (7). Математическую модель объекта составят уравнения (4) и (5), где значения источников F_x и F_y вычисляются по формулам (7), а из вектора неизвестных будут исключены переменные F_x и F_y .

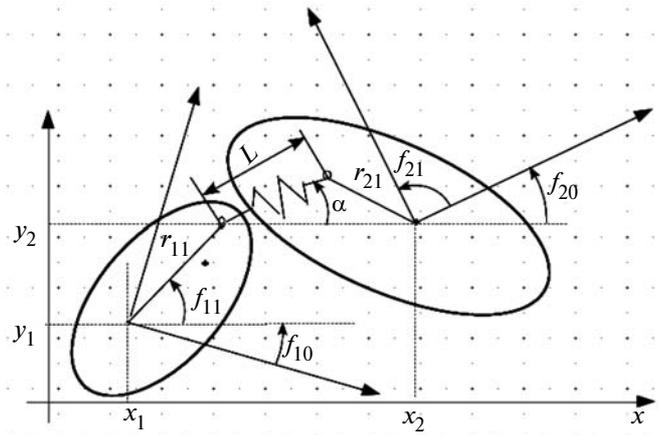


Рис. 3

Модель 3. Уравнения источников скорости (6) для модели 1 можно заменить на уравнения источников перемещений

$$x_1 - x_2 - r_{11} \cos(f_{11} + f_{10}) + r_{21} \cos(f_{21} + f_{20}) = 0;$$

$$y_1 - y_2 - r_{11} \sin(f_{11} + f_{10}) + r_{21} \sin(f_{21} + f_{20}) = 0$$

при условии, что в модель добавляются интеграторы для определения координат центров масс.

Численные эксперименты с приведенными выше моделями показывают их следующие особенности: модель 1 и модель 3 чувствительны к кинематическим заклиниваниям, т. е. в определенных ситуациях возможно вырождение матрицы Якоби; модель 1 вследствие погрешностей методов интегрирования может приводить к тому, что абсолютные координаты шарнира для тела 1 и тела 2 станут различаться; модель 2 может приводить к значительным вычислительным затратам при высоком коэффициенте жесткости.

Все три варианта моделей реализованы в универсальных комплексах моделирования сложных технических объектов с подсистемами различной физической природы ПА8 и ПА9, разработанных на кафедре САПР МГТУ им. Н. Э. Баумана. Для сопоставления рассмотренных математических моделей двумерного шарнира были выполнены численные эксперименты с использованием комплекса ПА8. В нем реализован расширенный узловый метод формирования математической модели объекта исследования. Численное интегрирование системы обыкновенных дифференциальных уравнений реализовывалось неявным методом Эйлера первого порядка точности с применением метода Ньютона на каждом шаге интегрирования.

В первом численном эксперименте моделировался процесс свободного колебания физического маятника (рис. 4) массой 1 кг и моментом инерции $1 \text{ кг} \cdot \text{м}^2$ на интервале модельного времени $0 \dots 100 \text{ с}$. К центру массы маятника прилагалась сила со значением 10 Н, направленная в отрицательном направлении оси y .

Таблица 1

Мо- дель	Общее число шагов	Число неудачных шагов	Общее число итераций	Число потерянных итераций	Число периодов	Амплитуда последнего полупериода, рад.
1	34718	215	71109	466	23,5	1,1
2	513832	254104	1066465	542762	20,8	1,87
3	620215	308107	1471862	829417	20,5	1,91

Таблица 2

Мо- дель	Мини- мальный шаг, с	Допусти- мая погр- ешность	Общее число шагов	Число неудачных шагов	Общее число итераций	Число потерянных итераций
1	10^{-12}	$3 \cdot 10^{-6}$	735983	122	1171577	216
2	10^{-12}	10^{-5}	773835	338526	1339316	677052
3	10^{-5}	$5 \cdot 10^{-5}$	745823	372729	1586080	848907

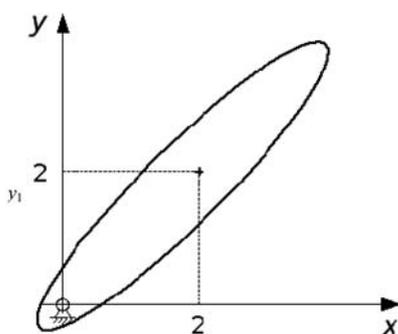


Рис. 4

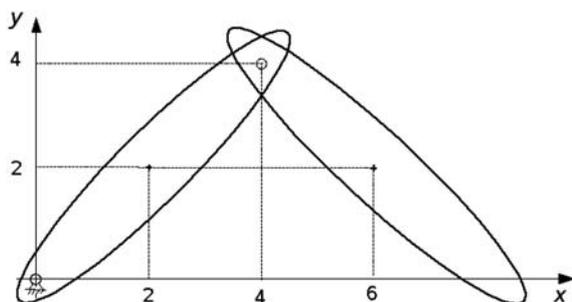


Рис. 5

Все расчеты проводили с точностью 10^{-4} (относительно текущего максимального по абсолютной величине значения каждой переменной, составляющей базис) при минимально допустимом значении шага интегрирования 10^{-5} с. Значение жесткости для модели 2 равнялось $2 \cdot 10^{-7}$ Н/м. Результаты эксперимента представлены в табл. 1.

Все неудачные шаги признаны таковыми по соображениям точности. Для всех моделей характерно затухание колебательного переходного процесса, что обусловлено особенностями неявного метода Эйлера. Однако для расчета с использованием

модели 1, кроме того, наблюдается нарушение "целостности" объекта — к окончанию моделирования шарнир на теле отклоняется от точки 0,0 на расстояние 0,83 м в положительном направлении оси Y. Данная модель показала себя наиболее экономичной, но и наименее адекватной, что объясняется отсутствием в ней непосредственного "контроля" положения шарнира на неподвижном основании. Точность расчетов с использованием модели 1 можно улучшить, повышая точность численного интегрирования (выбирая метод более высокого порядка, понижая погрешность интегрирования и минимальный допустимый шаг).

Экономичность и адекватность моделей 2 и 3 приблизительно одинаковы.

В эксперименте 2 исследовался объект, состоящий из двух тел, связанных шарнирами взаимно и с неподвижным основанием (рис. 5).

На центр масс каждого тела действует сила с значением 10 Н, направленная вниз. Численное интегрирование выполнялось на интервале 0...20 с модельного времени. Параметры метода Эйлера (допустимая погрешность интегрирования и минимально возможный шаг) подбирали так, чтобы расчеты для разных моделей требовали приблизительно одинакового числа шагов, это обеспечивало примерно одинаковый вид колебаний на всем интервале интегрирования. Результаты эксперимента приведены в табл. 2.

Выводы. В случаях, когда не требуется моделирование многоциклического процесса, можно рекомендовать модель 1 как самую экономичную, в противном случае — модель 2 или 3, причем при возможном заклинивании механизма следует использовать модель 2.

Список литературы

1. **Применение** комплекса ПА9 для проектирования объектов машиностроения. URL: <http://wwwcdl.bmstu.ru/Press/Press.html>
2. **Универсальный** механизм Express. URL: <http://machinery.ascon.ru/software/tasks/items/?prcid=6&prpid=1116>
3. **PRADIS** — анализ динамики систем различной физической природы. URL: <http://www.laduga.ru/pradis/pradis.shtml>
4. **1D Multi-Domain** System Simulation. URL: <http://www.lmsintl.com/imagine-amesim-1-d-multi-domain-system-simulation>
5. **Трудоношин В. А., Пивоварова Н. В.** САПР. Математические модели технических объектов. — М.: Высш. шк. 1986.
6. **Норенков И. П., Трудоношин В. А., Федорук В. Г.** Метод формирования математических моделей для адаптируемых программных комплексов анализа радиоэлектронных схем // Радиотехника. 1986. № 9.

УДК 004.4(075.8)

И. П. Норенков, д-р техн. наук, проф., зав. каф.,
М. Ю. Уваров, инженер, **М. А. Грушин**, студент,
А. А. Овешников, студент,
МГТУ им. Н. Э. Баумана
E-mail: norenkov@rk6.bmstu.ru

Классификация документов на основе кластеризации онтологий*

Для применения метода кластеризации онтологий к решению задач классификации документов необходимо располагать частными онтологиями, соответствующими рубрикам, задаваемым для классификации. Статья посвящена автоматизации создания частных онтологий на основе имеющейся общей онтологии определенной отрасли знаний. Приведены экспериментальные данные по определению эффективности решения задачи классификации и формированию частной онтологии.

Ключевые слова: онтология, классификация документов, кластеризация онтологий

Введение

В общей проблеме управления знаниями ведущее место занимают задачи поиска и обработки знаний, извлекаемых из текстовых документов (задачи TextMining). К задачам TextMining относятся кластеризация и классификация документов, их аннотирование, прогнозирование событий, поддержка принятия решений и др. Большие объемы корпоративных документальных баз знаний и продолжающийся процесс их пополнения ужесточают требования к показателям эффективности интеллектуальных систем управления знаниями. Поэтому развитие методов TextMining происходит в направлении повышения степени автоматизации решения задач с одновременным ростом показателей полноты и точности получаемых результатов.

Одним из методов, относящихся к этому направлению, является метод кластеризации онтологий (КО) [1]. Он может быть отнесен к группе методов, основанных на знаниях, по классификации

методов, представленной в работе [2]. Метод КО базируется на использовании паттернов, представляющих собой словосочетания с определенными ролями составляющих слов (слотов), и выявлении экземпляров паттернов заданных типов в текстах документов. В соответствии каждому слоту ставится ролевой кластер онтологии, концепты которого выполняют в паттернах определенную роль.

В альтернативной группе методов для классификации используют обучающие выборки (классификация с учителем), являющиеся задаваемыми пользователем подмножествами документов, отвечающих запросам пользователя и задающих ориентиры для оперирования документами в расширенном информационном пространстве. Например, задачей без учителя является задача кластеризации, в которой выявляются обучающие подмножества в виде кластеров документов по критериям их сходства. Эти кластеры далее применяют для решения задачи классификации с учителем, т. е. для распределения документов основной коллекции по установленным рубрикам.

Обзор методов классификации документов имеется в ряде источников, например [3, 4]. Для кластеризации документов обычно применяют векторные методы, в которых критерием близости документов является оценка расстояния между векторами, характеризующими словарный состав, с одной стороны, документа, с другой — рубрики. Семантический образ рубрики может представлять собой взвешенный список поисковых запросов [5]. Вопросы качества работы классификаторов рассматриваются в [6].

Особенностью метода КО является первоначальное распределение по кластерам (рубрикам) не документов, а концептов онтологии. Это обусловлено тем, что в этом методе для решения задач TextMining вместо обучающих выборок документов используются частные онтологии, поэтому необходимо их создание. Частные онтологии описывают семантику сравнительно узких предметных областей. Например, в образовательной сфере частные онтологии соответствуют отдельным учебным дисциплинам или крупным разделам таких дисциплин, в отличие от более общих онтологий, характеризующих, в частности, понятийный состав направлений подготовки специалистов в вузах или группы научных специальностей.

Кластеризация концептов имеет следующие отличия от кластеризации документов. Во-первых,

* Работа выполнена при финансовой поддержке РФФИ (проект 12-07-00222а).

необязательно должны быть распределены по кластерам все имеющиеся концепты. Среди них могут быть малоинформативные концепты, которые приблизительно в равной мере относятся к разным предметным областям. Во-вторых, имеется предварительная информация об искомым кластерах в виде одного или нескольких ключевых концептов, характерных для приложения.

Получение частных онтологий может быть автоматизировано. Данная статья посвящена решению задач классификации документов и автоматизации процедуры создания частных онтологий, используемых в методе КО. При этом предполагается, что общая онтология направления, в рамках которого функционирует корпоративная база знаний и требуется решать задачи TextMining, создана и поддерживается.

Метод кластеризация онтологий

Кластеризация онтологий может быть тематической и ролевой.

Тематические кластеры представляют собой подмножества тематически родственных концептов. При ролевой кластеризации объединение концептов в кластеры происходит по идентичности ролей, выполняемых концептами в словосочетаниях.

Метод КО основан на использовании для решения задач TextMining ролевых кластеров онтологий и формируемых на их основе семантических паттернов, являющихся сложными концептами. Сложный концепт $R(a, b, c...)$ — отношение (словосочетание) R между простыми концептами $a, b, c...$. При этом простой концепт — это концепт, выражаемый однословным термином (или словосочетанием, в котором отдельные слова не являются терминами концептов в данной онтологии). Каждый простой концепт выполняет в сложном концепте определенную роль. Концепты распределяются по кластерам в соответствии с ролью, которую они выполняют в сложных концептах.

Во многих случаях целесообразно выделять такие роли, как "объект", "свойство", "действие", "средство". Специфика предметной области отражается, как правило, концептами кластера "объект", в то время как кластеры "свойство", "действие", "средство" в значительной части инвариантны к приложениям. Следует также отметить, что некоторые концепты могут в разных словосочетаниях выполнять разные роли и потому могут быть включены в более чем один ролевой кластер.

Примеры решения задач поиска и обработки знаний, извлекаемых из документов, с помощью метода КО приведены в работе [1].

Задача информационного поиска ИП

$$\text{ИП} = \{\text{ОО}, \text{КД}, \text{ПОЗ}, \text{РП}\}$$

заключается в определении подмножества РП документов в коллекции документов КД, релевант-

ных запросу ПОЗ, с помощью общей онтологии ОО, в которой выделены ролевые кластеры. Особенностью метода КО является поиск по словосочетаниям, включающим как задаваемые пользователем ключевые слова из кластера "Объект", так и автоматически добавляемые к ним термины из других ролевых кластеров.

Задача аннотирования документов (АД)

$$\text{АД} = \{\text{ЧО}, \text{КД}, \text{АН}, \text{МА}\}$$

есть задача построения аннотации АН документов из коллекции документов КД на основе совокупности заданных частных онтологий ЧО с помощью метода аннотирования МА. В настоящее время в качестве МА преимущественно применяют экстрактивные методы аннотирования (в отличие от абстрактных или конструктивных), в соответствии с которыми аннотации состояются из характерных фрагментов текста (а не из отдельных слов или синтезируемых фраз). Фрагментами аннотации в случае применения метода КО являются предложения, включающие сложные концепты. Вместо предварительного обучения системы используется предварительная разработка кластеризованных онтологий ЧО.

Задача классификации текстов (документов) (КТ)

$$\text{КТ} = \{\text{ЧО}, \text{КД}, \text{Q}, \text{МК}\}$$

имеет целью определение матрицы $Q = |q_{ki}|$, элементами которой являются q_{ki} — коэффициенты принадлежности k -го документа i -й рубрике (тематическому кластеру). Коэффициенты принадлежности рассчитывают в соответствии с методом МК, например, как число концептов из i -й ЧО, приходящееся на 1000 слов k -го документа.

Эксперименты по оценке качества классификации с помощью метода КО

Для экспериментов была случайным образом сформирована выборка из архива статей научно-технического журнала "Информационные технологии". Размер выборки 200 статей. В этой выборке требовалось выполнить автоматическую классификацию статей по пяти рубрикам и сравнить полученные результаты с данными экспертного распределения, в качестве которого принято распределение статей по рубрикам, реализованное в журнале. К выбранным рубрикам относятся "Программирование" (P1), "Вычислительные системы и сети" (P2), "Интеллектуальные системы" (P3), "Безопасность информации" (P4), "Моделирование и оптимизация" (P5). Применительно к каждой статье рубрику, под которой статья опубликована в журнале, будем называть основной.

Приняты следующие обозначения. Пусть \mathcal{E}_i — множество статей, отнесенных экспертом к i -й рубрике. Множества статей A_i, B_i, C_i образуются при

Таблица 1

Номер статьи	Рубрики				
	P1	P2	P3	P4	P5
4	2,683	0,894	0	0	3,577
20	0,597	—	—	—	0
64	1,999	0,999	—	—	0,999
198	0,998	—	—	1,197	6,788

исполнении классифицирующей системы. $A_i \subset \mathcal{D}_i$ — подмножество статей из i -й рубрики, для которых i -я рубрика является основной; $B_i \subset \mathcal{D}_i$ — подмножество статей из i -й рубрики, отнесенных к одной из неосновных рубрик; C_i — подмножество статей, для которых i -я рубрика не является основной, но отнесенных системой классификации к i -й рубрике. Появление элемента в B_i является ошибкой первого рода, а в C_i — ошибкой второго рода, и эти события фиксируются, если максимальный элемент $q_{k\max}$ в k -й строке матрицы Q находится в столбце неосновной рубрики.

В качестве примера в табл. 1 приведены несколько строк матрицы Q . Для статей с номерами 4, 20 и 64 основной рубрикой является P5. При этом статья 4 попадает в множество A_5 , так как максимальный коэффициент принадлежности этой статьи находится в столбце P5. В следующих строках табл. 1 имеет место несовпадение оценок эксперта и системы классификации, поэтому статьи 20 и 64, во-первых, попадают в множество B_5 , во-вторых, в множество C_1 , поскольку их максимальные коэффициенты относятся к столбцу P1 табл. 1, т. е. к рубрике, не являющейся основной. Статья 198 оказывается во множестве C_5 , поскольку максимальный коэффициент в 198-й строке матрицы Q находится в столбце P5.

Число ошибок первого рода позволяет оценить полноту классификации по i -й рубрике, определяемую по формуле

$$P_i = |A_i| / (|A_i| + |B_i|).$$

Число ошибок второго рода влияет на точность классификации, оцениваемой по формуле

$$T_i = |A_i| / (|A_i| + |C_i|).$$

Результаты экспериментальной оценки полноты и точности классификации с использованием метода кластеризации онтологий приведены в табл. 2.

При анализе полученных результатов следует учитывать следующие факторы.

Во-первых, оценки P_i и T_i выполнены в предположении идеальности экспертного распределения статей по рубрикам, каковым в действительности оно не является хотя бы по причине тематической неоднозначности статей.

Во-вторых, качество результатов с очевидностью зависит от характеристик используемых частных онтологий, задающих семантические образы рубрик. Вопросы автоматизации создания частных онтологий будут рассмотрены в следующем разделе данной статьи.

В-третьих, имеются возможности сравнения качества классификации с помощью излагаемого подхода и альтернативного метода на основе простых запросов, в соответствии с которым семантическими образами рубрик являются наборы характеристик для рубрики ключевых слов.

Метод простых запросов был применен к выборке из 200 статей, использованной в предыдущих экспериментах. В качестве результатов рассматривались мощности L_i множеств статей, отобранных в качестве релевантных запросу по каждой рубрике. Например, для рубрики "Программирование" был применен запрос, состоящий из ключевых слов "Программа" и "Программное обеспечение". Одновременно были оценены мощности M_i множеств статей, имеющих ненулевые значения коэффициентов q_{ki} в матрице Q . Очевидно, что в идеальном случае однозначности тематики каждой статьи и с повышением качества классификации числа M_i и L_i будут стремиться к значению $|\mathcal{D}_i|$. Поскольку для всех рубрик в эксперименте получено $M_i < L_i$, метод кластеризации онтологий заметно более адекватен задачам классификации документов, чем метод простых запросов.

Создание частных онтологий. Задача создания частной онтологии

$$СЧО = \{ЧО, ОО, КД, П\} \quad (1)$$

является задачей выделения тематических кластеров ЧО из известной общей онтологии ОО на базе заданной коллекции КД документов с помощью

Таблица 2

№	Рубрика	$ \mathcal{D}_i $	$ A_i $	$ B_i $	$ C_i $	P_i	T_i	M_i	L_i
1	Программирование	14	8	6	7	0,57	0,53	81	130
2	Вычислительные системы и сети	36	28	8	13	0,78	0,68	115	163
3	Интеллектуальные системы	34	15	19	6	0,44	0,63	73	135
4	Безопасность информации	14	10	4	1	0,71	0,91	28	59
5	Моделирование и оптимизация	42	30	12	21	0,71	0,59	102	174
	Итого:	140	91	49	48	0,65	0,66		

Концепт	m_{ij}/r_i	w_i	Концепт	m_{ij}/r_i	w_i
устройство	0/90	0	время	36/382	1,02
база данных	2/101	0,21	концепт; класс; сущность; понятие; таксон; десигнат	28/278	1,09
сигнал	3/118	0,27	язык	12/115	1,13
уравнение	3/107	0,30	математическая модель	11/96	1,24
программа	6/191	0,34	информационная технология	24/207	1,25
информационная система	4/102	0,42	индекс	12/100	1,30
компьютер; ЭВМ; электронная вычислительная машина	5/120	0,45	нейронная сеть	13/68	2,06
знания	7/128	0,59	память; накопитель; запоминающее поколение	14/44	2,44
программное обеспечение	6/110	0,59	целевая функция; функция полезности	25/64	4,22
атрибут; слот; роль	9/145	0,67	расписание	13/26	5,40
веб; web; WWW	11/142	0,84	ген	13/18	7,8
объект	19/232	0,88	хромосома	15/15	10,8
HTTP	12/142	0,91	популяция	18/18	10,8
интерфейс	8/94	0,92	мутация	17/17	10,8
сеть	17/200	0,92	кроссовер	14/14	10,8
информация	30/346	0,97	ГА	41/41	10,8
данные	39/424	0,99			

метода кластеризации П. Поскольку большинство концептов в кластерах "свойство", "действие", "средство" в значительной мере одинаково для различных частных онтологий, при решении задачи (1) можно ограничиться кластером "объект".

Кластеризацию концептов предлагается выполнять в три этапа. На первом этапе в коллекции КД ищется целевое подмножество документов, соответствующее запросу, состоящему из заданных ключевых терминов. На втором этапе из целевого подмножества извлекаются имеющиеся в документах термины, соответствующие концептам онтологии ОО. На третьем этапе из найденного множества терминов выделяется подмножество терминов информативных концептов, из которых и формируется частная онтология.

Информативность i -го термина характеризуется отношением

$$w_i = h_{ij}/H_i,$$

где $h_{ij} = m_{ij}/N_j$, $H_i = r_i/N$, m_{ij} — число документов, отнесенных к j -й частной онтологии и включающих i -й концепт; N_j — число документов, отнесенных к j -й онтологии, N — число документов в коллекции, r_i — число документов в коллекции, включающих i -й концепт.

В табл. 3 приведены результаты эксперимента по отбору концептов в слот "Объект" частной онтологии "Генетические алгоритмы". Коллекция документов представляла собой набор научно-технических статей по проблемам информационных технологий, $N = 444$. По запросу "Генетические алгоритмы" после первого этапа получено $N_j = 41$.

Курсивом выделены концепты с $w_i < 0,6$ (их можно назвать чужими, поскольку они с высокой вероятностью могут войти в иные частные онтологии). Концепты с $0,6 > w_i > 2,0$ — нейтральные (или малоинформативные), ими целесообразно пренебречь. Полужирным шрифтом показаны концепты с $w_i > 2,0$, это собственные концепты, которые должны быть отобраны в формируемую ЧО "Генетические алгоритмы".

Закключение. Метод кластеризации онтологий разработан в целях повышения эффективности решения задач поиска и обработки знаний, извлекаемых из текстовых документов. В статье выполнена экспериментальная оценка полноты и точности решения конкретной задачи классификации с помощью этого метода и предложен способ расчета частных онтологий, с необходимостью используемых в методе кластеризации онтологий.

Список литературы

1. Норенков И. П., Уваров М. Ю. Задачи обработки знаний на основе ролевой кластеризации онтологий // Информационные технологии. 2012. № 2. С. 19—24.
2. Лукашевич Н. В. Методы автоматической рубрикации и оценка их качества. — <http://like-money.ru/stati/259-metody-avtomaticheskoy-rubrikaczii-i-oczenka-ix-kachestva>
3. Агеев М. С., Добров Б. В., Лукашевич Н. В. Автоматическая рубрикация текстов: методы и проблемы // Учен. зап. Казан. гос. ун-та. Сер. Физ.-матем. науки. 2008. Т. 150, кн. 4. С. 25—40.
4. Ермакова Л. М. Методы классификации текстов и определения качества контента // Вестник Пермского университета. Математика. Механика. Информатика. 2011. Вып. 3 (7). С. 47—53.
5. Шабанов В. И., Андреев А. М. Метод классификации текстовых документов, основанный на полнотекстовом поиске // Труды РОМИП'2003. — СПб: Изд. НИИ химии СПбГУ, 2003. С. 52—71.
6. Агеев М. С. Методы автоматической рубрикации текстов, основанные на машинном обучении и знаниях экспертов: дисс. на соискание ученой степени канд. физ.-мат. наук. МГУ, 2005. 136 с.

О. Н. Андрейчикова¹, д-р техн. наук, проф.,
А. В. Андрейчиков², д-р техн. наук, проф.,

¹Всероссийская государственная
налоговая академия Минфина России

²Московский государственный университет
путей сообщений (МИИТ)

E-mail: alexandrol@mail.ru

Что делать с авторским правом в сети Интернет? Поиск решения методом аналитических сетей

Проведено исследование актуальной проблемы охраны авторских прав в сети Интернет, на основе аналитических сетей выполнен многокритериальный анализ данной проблемы с рассмотрением возможных вариантов решений и связанных с ними выгод, издержек и рисков. В качестве альтернатив рассматривались различные меры регулирования Интернет-контента. При этом учитывались интересы трех категорий акторов — создателей контента, потребителей и провайдеров. Полученные результаты показывают, что наиболее перспективной альтернативой в будущем является свободное использование Интернет-контента. Альтернативы, связанные с ужесточением наказаний за нарушения авторских прав, не приведут к приемлемому для общества решению проблемы.

Ключевые слова: аналитические сети, авторское право, Интернет, контент, приоритет

Введение

Появление Интернета изменило мир и поставило перед человечеством ряд новых проблем. Классические законы экономики описывают рыночный обмен материальными активами. Экономика знаний не подчиняется этим законам, поскольку интеллектуальные продукты имеют нематериальную природу, а механизмы *распространения* знаний отличаются от товарно-денежного обмена. Ключевое отличие интеллектуальных продуктов (знаний) от материальных товаров связано со свойством неограниченной тиражируемости информации. Развитие информационных технологий и, в частности, появление сети Интернет практически свело к нулю затраты на копирование нематериальных (интеллектуальных) продуктов. Неудивительно, что при этом возрос уровень пиратства (нарушения авторского права). Самый простой путь связан с ужесточением наказаний за нарушение авторских прав, но это ведет к нарушению других прав человека, например, прав на получение информации, на образование, на свободу слова и т. д. При отказе от охраны авторских прав пропадают материальные

стимулы для создания интеллектуальных продуктов, правообладатели несут убытки, а создатели произведений лишаются шансов заработать на жизнь. В настоящее время во всем мире ведутся дебаты о том, что делать с авторским правом в сети Интернет [1—4, 6—9].

В настоящей статье предпринята попытка систематизировать и проанализировать известные данные и мнения по рассматриваемому вопросу с применением метода аналитических сетей [5, 12]. Аналитические сети (Analytic Network Process, ANP) — уникальная методика многокритериального анализа решений, которая позволяет исследовать сложные проблемы с взаимно зависимыми элементами (критериями и альтернативами). В отличие от других методик принятия решений, которые, как правило, требуют соблюдения определенных условий независимости критериев и альтернатив, методика ANP ориентирована на исследование и оценку взаимного влияния элементов рассматриваемой проблемы. Соблюдение условий независимости критериев и альтернатив весьма удобно с точки зрения математической обработки, но во многих случаях такое допущение не позволяет адекватно описать реальность. Помимо ANP, взаимное влияние элементов слабо структурированных проблем можно изучать с помощью методик искусственного интеллекта и имитационного моделирования, которые обычно требуют использования количественных данных и специальных языков.

1. Сущность ANP

Аналитические сети (ANP) — обобщение метода анализа иерархий [11]. Этот подход подразумевает представление исследуемой проблемы в виде сети, содержащей взаимовлияющие кластеры, каждый из которых объединяет семантически однородные элементы. Кластеры могут соответствовать уровням иерархии или принадлежать сети с обратными связями. Исследователь строит сеть, которая представляет его знания о проблеме и дает возможность оценить приоритеты кластеров и их элементов относительно так называемых управляющих критериев — экономических, политических, социальных и др. Цель исследования заключается в определении приоритетов альтернатив по каждому управляющему критерию и последующем обобщении этих приоритетов, на основании которого можно упорядочить альтернативы с учетом многих критериев. Подобная процедура может выполняться для оценки выгод, издержек, возможностей и рисков рассматриваемых решений. Если необходимо, можно получить интегральный результат, используя тот или иной принцип обобщения приоритетов альтернатив по всем четырем категориям качества решений. Вычисление приоритетов выполняется

так же, как в методе анализа иерархий, — на основе экспертных суждений и парных сравнений кластеров и их элементов.

Методика ANP содержит следующие основные этапы [5, 12].

1. Формулирование главной цели исследования.
2. Построение сетевой модели, которая может представлять собой единственную сеть или совокупность сетей, описывающих различные аспекты проблемы, например, выгоды, издержки, возможности и риски.
3. Парные сравнения кластеров и элементов, в процессе которых формируются матрицы парных сравнений на основе суждений экспертов, выраженных с помощью 9-балльной шкалы отношений [11]. Этот и последующие этапы выполняются для каждой сети в модели.
4. Вычисление главных собственных векторов матриц парных сравнений, которые интерпретируются как векторы приоритетов сравниваемых элементов относительно других элементов.
5. Формирование суперматрицы из векторов приоритетов элементов кластеров, которые записываются в ее столбцы. Элементы суперматрицы показывают приоритеты влияния элементов, указанных слева, на элементы, указанные сверху.
6. Нормирование столбцов суперматрицы путем умножения ее блоков на приоритеты соответствующих кластеров. Последние получаются из матриц парных сравнений кластеров по управляющим критериям. В результате формируется взвешенная суперматрица, в которой сумма элементов каждого столбца равна единице.
7. Последовательное возведение взвешенной суперматрицы в целочисленные степени до достижения предела, существование которого доказано. Предельные приоритеты могут интерпретироваться как важность, вклад или предпочтительность элементов, подверженных взаимному влиянию.
8. Если модель содержит несколько сетей, необходим синтез обобщенного результата. В таких случаях должен быть хотя бы один кластер, который присутствует во всех сетях. Как правило, такой кластер содержит рассматриваемые альтернативы решения. Интегральные приоритеты альтернатив вычисляются на основе того или иного вида свертки предельных приоритетов, полученных для каждой сети, и весовых коэффициентов важности самих сетей.

2. Постановка задачи исследования

Главная цель исследования — найти наиболее перспективный способ охраны авторских прав в сети Интернет с учетом выгод, издержек и рисков возможных решений. Мы не рассматриваем категорию возможностей, так как считаем их важность существенно меньшей по сравнению с другими категориями качества решений. Кроме того, формулирование возможностей порождает плохо обус-

ловленные фантазии, количественный анализ которых не имеет смысла.

Первое, что необходимо сделать, — сформировать кластер альтернатив, представляющих различные способы регулирования Интернет-контента. Многочисленные дискуссии на эту тему ведутся в двух главных направлениях: первое провозглашает *Свободный доступ к Интернет-контенту*; второе призывает к ужесточению наказаний за нарушение авторских прав [4, 10, 13]. При этом высказываются предложения ввести *наказание для пользователей* (потребителей), *владельцев сайтов* и провайдеров. Свободное использование контента привлекательно для пользователей, но вызывает вопрос: "Кто платит за создание контента?" Пытаясь ответить на этот вопрос, можно выделить следующие варианты.

- *Спонсирование создателей Интернет-контента*, например, как во время Ренессанса. Только вот где найти спонсоров? Однозначного ответа на этот вопрос нет, но существуют определенные гипотезы и практический опыт [2, 4, 13].
- Создатели контента не преследуют целей заработать на жизнь своим трудом, они имеют другие источники дохода и много свободного времени для творчества. Следует заметить, что в настоящее время таких людей не так много, и этот способ ассоциируется с утопией. Мы назвали эту альтернативу *Изменение товарно-денежной парадигмы*, предполагая, что Интернет заставит нас сделать это в будущем. Это не значит, что общество станет идеальным, но в процессе своего развития оно может выработать правовые и экономические инструменты, которые будут приемлемыми и для потребителей, и для создателей медиаконтента.
- *Добровольная оплата Интернет-контента*. Эта идея предлагается как исследователями [2, 4, 13, 14], так и некоторыми создателями (писателями, музыкантами), которые проверяют ее экспериментально.
- *Налог на медианосители* или *Налог на входящий Интернет-трафик*, который распределяется среди авторов и правообладателей. Подобные схемы уже используются в некоторых странах, в том числе в России.
- Кроме того, существует путь, связанный с развитием *Новых технологий защиты информации*, которые ограничат несанкционированное использование произведений, например, запрещая копирование или скачивание файлов. Здесь уместно привести аналогию из области вооружений, где при появлении нового вида оружия через некоторое время появляются средства защиты от него.

Итак, мы будем рассматривать десять альтернатив, которые будут включены в виде кластера *Альтернативы* в три сети для анализа *Выгод*, *Издержек* и *Рисков* исследуемых решений.

2.1. Сеть Выгод

Основными акторами сети Интернет являются *Провайдеры*, *Потребители* (пользователи) и *Создатели контента*. При построении сети *Выгод* главная цель заключалась в том, чтобы понять, какие альтернативы обеспечат наибольшие выгоды рассматриваемым акторам с учетом их взаимного влияния. Сеть *Выгод* представлена на рис. 1, демонстрирующем все кластеры и входящие в них элементы.

Кластер *Провайдеры* объединяет *Интернет-провайдеров*, *Владельцев сайтов* и *Социальные сети*. Мы полагаем, что эти акторы в большей степени, чем другие, влияют на размещение Интернет-контента и доступ пользователей к нему. Влияние провайдеров на элементы других кластеров выражает их выгоды и, наоборот, элементы, влияющие на провайдеров, затрагивают выгоды последних. В связи с этим отсутствует необходимость в специальном кластере для описания выгод провайдеров.

Кластер *Создатели контента* включает *Авторов*, *Исполнителей*, *Продюсеров* и *Правообладателей*. Такой состав обусловлен разнообразием Интернет-контента.

Кластер *Выгоды создателей* содержит следующие элементы:

- *Реклама* — создатели заинтересованы в информировании потребителей о своих произведениях.
- *Самовыражение* — многие талантливые люди испытывают потребность творить и обнародовать свои произведения, не рассчитывая на денежные вознаграждения.
- *Популярность* — этот элемент сети *Выгод* особенно значим для создателей контента массового потребления.
- *Приоритет* — кто первый и более цитируемый — вопросы, важные для всех авторов.
- *Прибыль* — этот элемент особенно важен для продюсеров и правообладателей. Он также может быть ведущим мотивом для многих авторов и исполнителей.

Вместо кластера *Потребители* мы включили в сеть кластер *Выгоды потребителей*, так как в данном исследовании нет смысла дифференцировать потребителей контента (пользователей сети Интернет). Выгоды потребителей соответствуют разным видам деятельности пользователей сети Интернет:

- *Развлечение.*
- *Общение.*
- *Образование.*
- *Поиск информации.*
- *Времяпровождение.*
- *Исследования.*

Кластер *Контент* содержит четыре категории, образованные по признакам "законность" (легальный или нелегальный) и "тип размещения" (твердая копия или Интернет). В данном исследовании мы пытались рассмотреть проблему в целом, поэтому не дифференцировали контент по функци-

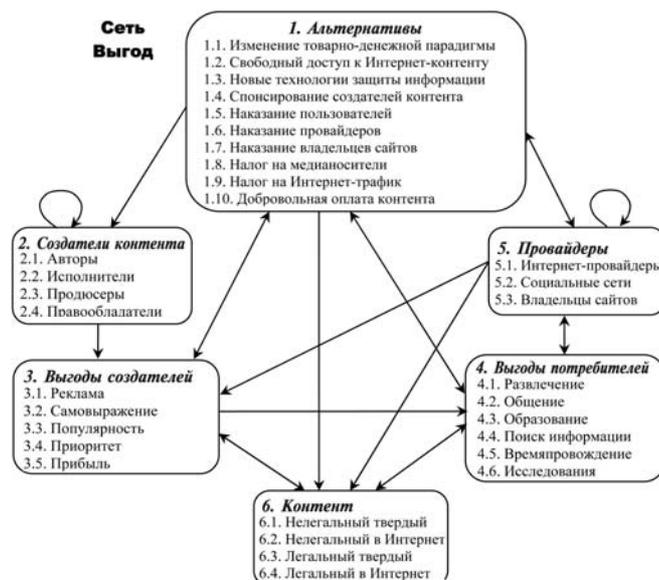


Рис. 1. Структура сети Выгод

ональным признакам, хотя планируем сделать это в будущем.

Связи между кластерами и их элементами соответствуют влиянию, предпочтению или важности в терминах выгод. Мы не детализировали выгоды набором критериев, сознательно ограничивая размерность задачи и, следовательно, число экспертных суждений. В построенной модели мы полагали, что альтернативы влияют на все остальные кластеры и их элементы. В свою очередь, на кластер альтернатив оказывают влияние *Выгоды создателей*, *Выгоды потребителей* и *Провайдеры*. Элементы кластера *Контент* влияют на выгоды создателей и потребителей и так далее в соответствии с направлением стрелок на рис. 1.

2.2. Сеть Издержек

Сеть, описывающая издержки, показана на рис. 2. Она содержит шесть кластеров: *Альтернативы*, *Контент*, *Создатели контента*, *Издержки создателей*, *Издержки потребителей* и *Провайдеры*. Новыми кластерами по сравнению с сетью выгод являются *Издержки создателей* и *Издержки потребителей*. В качестве основных затрат создателей контента мы рассматривали *Знания*, *Деньги* и *Время*. Издержки потребителей представлены следующими элементами:

- *Платные скачивания.*
- *Плата за сервисы.*
- *Затрудненный доступ к контенту*, который может быть обусловлен высокой стоимостью, затратами времени или другими обстоятельствами.
- *Плата за подключение к Интернету.*
- *Плата за трафик.*
- *Нежелательная реклама* — нематериальные издержки потребителей, связанные с агрессивным маркетингом. Следует отметить, что этот пункт

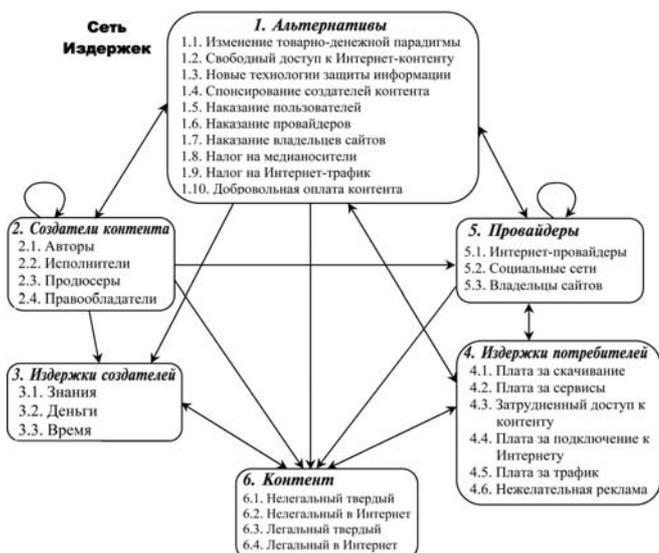


Рис. 2. Сеть Издержек

заслуживает особого внимания, так как он касается актуальной проблемы фильтрации информации в сети Интернет.

2.3. Сеть рисков

Сеть рисков рассматриваемых решений приведена на рис. 3. Она включает следующие кластеры: *Альтернативы*, *Контент*, *Создатели контента*, *Провайдеры*, *Риски создателей*, *Риски пользователей*, *Риски провайдеров*. Рассмотрим подробнее три последних кластера.

Кластер *Риски создателей* содержит следующие элементы:

- *Отсутствие стимулов*: несанкционированное использование контента может лишать создателей стимулов к творчеству.
- *Финансовые потери* имеют место, если создатели вкладывают деньги в разработку интеллектуальных продуктов, а потребители не платят за них.

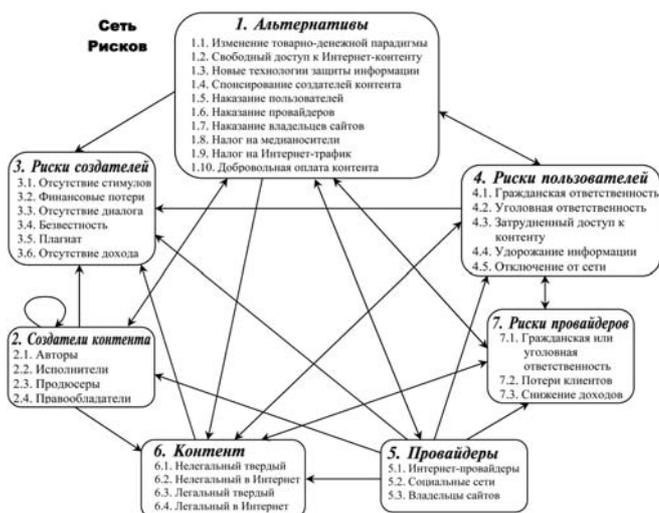


Рис. 3. Сеть Рисков

- *Отсутствие диалога*: чрезмерная защита контента от несанкционированного доступа может привести к потере связи с аудиторией пользователей.
- *Безвестность*: создатели, жестко ограничивающие область распространения своих произведений, рискуют остаться неизвестными широкой публике.
- *Плагиат*: произведения могут не только нелегально скачивать, но и обнародовать под чужим именем или безымянным.
- *Отсутствие дохода*: создатели могут лишиться дохода не только от сетевого размещения произведения, но и от твердых копий, так как их продажи могут снизиться. Именно эти обстоятельства (т. е. замещение) объясняют присутствие контента на твердых носителях в кластере *Контент*. Мы выделили следующие основные *Риски пользователей* сети Интернет:

- *Гражданская ответственность* — риск привлечения пользователей к такой ответственности.
- *Уголовная ответственность*.
- *Затрудненный доступ к контенту*.
- *Удорожание информации*.
- *Отключение от сети*.

Главные *Риски провайдеров* заключаются в привлечении к *Гражданской или уголовной ответственности*, в *Потерях клиентов* и *Снижении доходов*.

3. Обсуждение результатов

Для каждой сети были заполнены матрицы парных сравнений элементов и кластеров, вычислены главные собственные векторы этих матриц, интерпретируемые как векторы приоритетов, из векторов приоритетов были сформированы суперматрицы и получены предельные приоритеты элементов. Рассмотрим полученные результаты для каждой сети.

3.1. Выгоды

Предельные приоритеты элементов сети *Выгоды* приведены на рис. 4, который показывает, что лучшими альтернативами в свете выгод является *Свободный доступ к Интернет-контенту* и *Спонсирование создателей контента*. Этот результат можно объяснить тем, что *Выгоды потребителей* имеют самый большой приоритет, поскольку они аккумулируют влияние других элементов проблемы в наибольшей степени. Другими словами, выгоды — это прежде всего выгоды пользователей. И это не чье-то персональное суждение, а результат взаимного влияния рассматриваемых элементов проблемы, который подтверждается тем, что выгоды создателей и провайдеров возникают только тогда, когда пользователи заинтересованы в приобретении их продуктов и услуг.

Приоритеты кластеров показаны на рис. 5, где можно видеть, что *Выгоды потребителей* в наибольшей степени определяют выгоды в целом. Второе место занимает кластер *Контент*, в котором доля

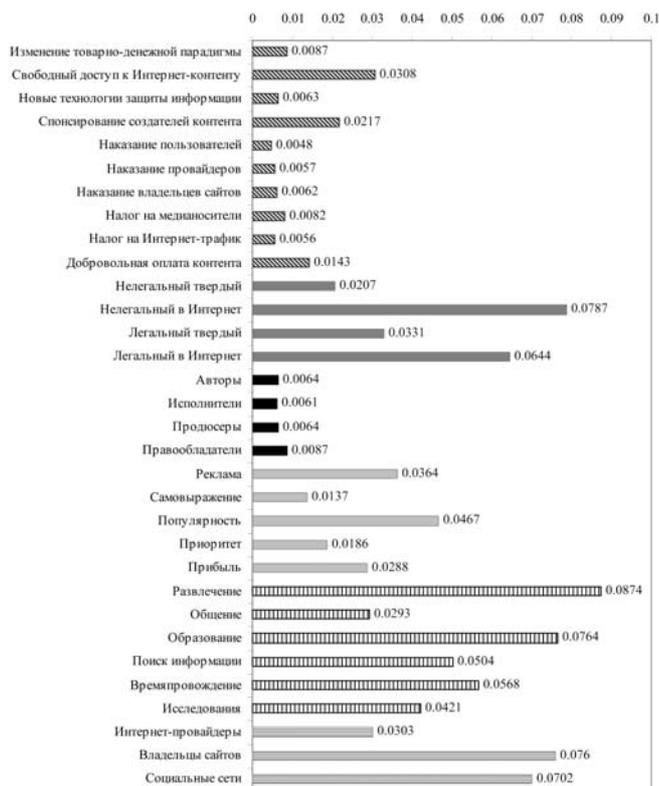


Рис. 4. Предельные приоритеты элементов Сети Выгод

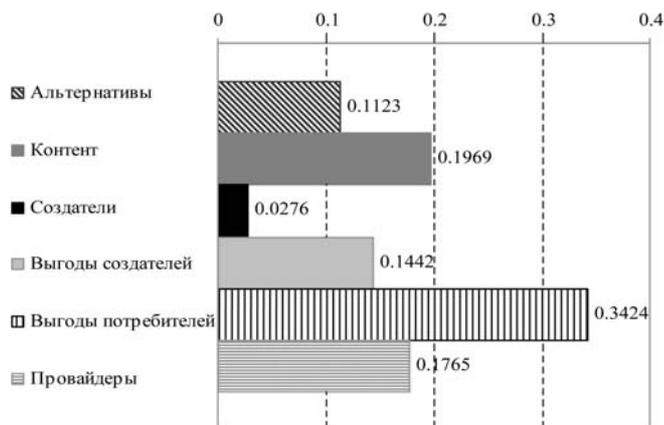


Рис. 5. Приоритеты кластеров в Сети Выгод



Рис. 6. Предпочтения пользователями типов контента

Интернет-контента составляет 73 %, из них 40 % — нелегальный контент. Это не удивляет, ибо пользователи предпочитают бесплатное скачивание из Интернета, когда это возможно. На рис. 6 показано, что для всех целей, кроме *Исследований*, потребители предпочитают нелегальный Интернет-контент.

Третий по влиянию кластер — *Провайдеры*, среди которых приоритеты *Владельцев сайтов* и *Социальных сетей* в два раза выше приоритета *Интернет-провайдеров*.

Кластер *Создатели контента* в наименьшей степени подвержен влиянию других элементов, которые влияют не на самих создателей, а на их выгоды, среди которых наибольшие приоритеты имеют *Популярность* и *Реклама*. *Прибыль* занимает третье место.

Первая тройка альтернатив — *Свободный доступ к Интернет-контенту*, *Спонсирование создателей контента* и *Добровольная оплата контента*. Именно эти альтернативы предпочтительны для пользователей. По этой же причине альтернативы, связанные с наказаниями, имеют низкие приоритеты.

3.2. Издержки

Значения предельных приоритетов кластеров и входящих в них элементов приведены на рис. 7 (см. третью сторону обложки), который показывает, что наибольший приоритет имеет кластер *Контент*, а среди его элементов — *Незаконный в Интернет*.

Это значит, что данный вид контента обуславливает более высокие издержки по сравнению с остальными. Предельные приоритеты, полученные для элементов и кластеров в данной сети (см. рис. 2) с учетом их взаимного влияния, свидетельствуют, что наибольший вклад в общие издержки вносят издержки потребителей, за ними следуют примерно равные издержки создателей и провайдеров. Основные материальные издержки потребителей связаны с оплатой скачиваний, второе место занимают нематериальные издержки от нежелательной рекламы. Главные издержки создателей контента — это материальные и интеллектуальные затраты.

Вклад кластера *Альтернативы* в общие издержки существенно меньше по сравнению с перечисленными выше кластерами. Значения предельных приоритетов альтернатив по издержкам, нормированные на кластер, приведены в табл. 1. Максимальные издержки имеет альтернатива *Налог на трафик*, за ней следуют три альтернативы, связанные с наказаниями. На рис. 8 (см. четвертую сторону обложки) показаны приоритеты альтернатив, полученные в результате обработки матриц их попарных сравнений относительно влияющих на них элементов других кластеров сети издержек.

Альтернатива *Свободный доступ к Интернет-контенту* имеет максимальные издержки для *Создателей контента*,

Таблица 1

Нормированные предельные приоритеты альтернатив

Альтернативы	Выгоды	Издержки	Риски
Изменение товарно-денежной парадигмы	0,0775	0,0597	0,0722
Свободный доступ к Интернет-контенту	0,2744	0,0607	0,1038
Новые технологии защиты информации	0,0566	0,1004	0,0856
Спонсирование создателей контента	0,1931	0,0647	0,0687
Наказание пользователей	0,0425	0,1310	0,1338
Наказание провайдеров	0,0505	0,1430	0,1392
Наказание владельцев сайтов	0,0553	0,1231	0,1329
Налог на медианосители	0,0726	0,0812	0,0803
Налог на Интернет-трафик	0,0501	0,1596	0,0956
Добровольная оплата контента	0,1276	0,0766	0,0879

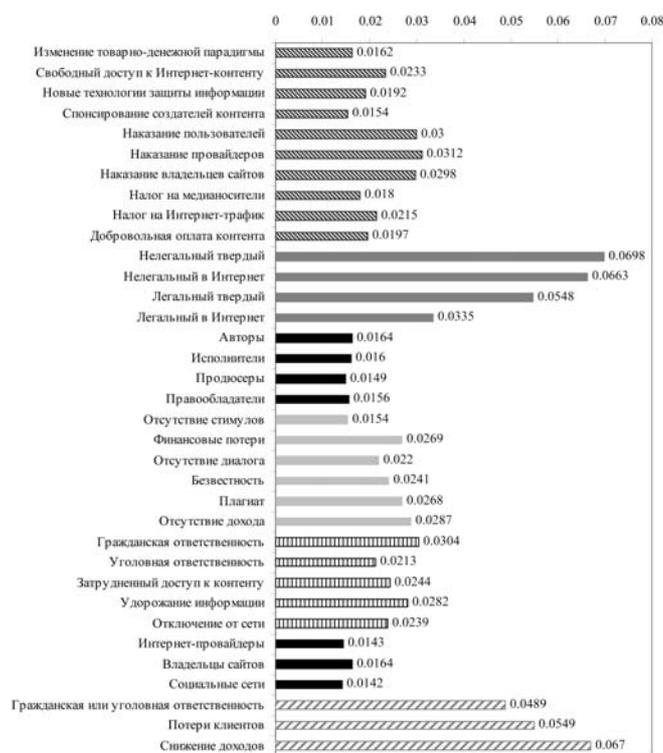


Рис. 9. Предельные приоритеты элементов Сети Рисков



Рис. 10. Нормированные предельные приоритеты альтернатив

но для *Потребителей* и *Провайдеров* издержки этой альтернативы существенно ниже по сравнению с остальными. *Налог на трафик* — наиболее затратная альтернатива для *Потребителей* и *Социальных сетей*, но не является таковой для *Создателей контента* и *Владельцев сайтов*. Альтернативы, связанные с наказаниями, имеют высокие издержки для *Потребителей* и *Провайдеров*. Поскольку влияние потребителей и провайдеров в сети издержек существенно превосходит влияние создателей контента, то лидерами по издержкам становятся *Налог на трафик* и альтернативы-наказания.

3.3. Риски

Результаты обработки сети рисков, представленной на рис. 3, приведены на рис. 9. Влияние рисков наиболее заметно аккумулируется в кластерах *Альтернативы* и *Контент*. Нелегальный контент связан с высокими рисками, наименее рискованный вид контента — *Легальный в Интернет*. Среди всех акторов провайдеры имеют наиболее высокие риски, за ними следуют создатели контента, пользователи занимают третье место. Главным риском провайдеров является *Снижение доходов*, для создателей — финансовые потери и отсутствие дохода, для пользователей — гражданская ответственность за нарушение авторских прав и удорожание информации.

Лидерами по риску среди альтернатив являются альтернативы-наказания, за ними следует *Свободный доступ к Интернет-контенту*. Минимальным риском характеризуется *Спонсирование создателей контента*.

3.4. Обобщенные результаты моделирования

Заключительный этап исследования предполагает использование процедур обобщения приоритетов альтернатив, полученных для сетей *Выгод*, *Издержек* и *Рисков*. Нормированные на единицу предельные приоритеты альтернатив по *Выгодам*, *Издержкам* и *Рискам* приведены в табл. 1 и на рис. 10. Для определения приоритетов *Выгод*, *Издержек* и *Рисков* учитывались пять стратегических критериев, которые приведены в табл. 2. При этом применялась не процедура парного сравнения, а метод лингвистических стандартов, в котором использовались три

Таблица 2

Определение приоритетов Выгод, Издержек и Рисков

Критерии	Выгоды	Издержки	Риски
Распространение знаний	Низкие	Высокие	Средние
Справедливое вознаграждение создателей	Высокие	Низкие	Высокие
Социальный прогресс	Средние	Высокие	Средние
Экономика	Низкие	Высокие	Высокие
Безопасность	Средние	Средние	Высокие
Интегральные значения приоритетов	0,357	0,264	0,379
Весовые коэффициенты критериев равной важности 0,2	Высокие (0,693); Средние (0,231); Низкие (0,077)		

оценки — Высокие, Средние, Низкие. Числовые эквиваленты оценок получены с помощью процедуры их парного сравнения. Предполагалось, что рассматриваемые критерии имеют равную важность, поэтому интегральные приоритеты *Выгод*, *Издержек* и *Рисков* были получены путем усреднения и далее использовались в качестве весовых коэффициентов при получении конечного результата.

Для получения обобщенного результата применялась формула

$$w_{A_i}^G = p_B w_{A_i}^B - p_C w_{A_i}^C - p_R w_{A_i}^R,$$

где p_B, p_C, p_R — приоритеты *Выгод*, *Издержек* и *Рисков* соответственно; $w_{A_i}^B, w_{A_i}^C, w_{A_i}^R$ — нормированные предельные приоритеты i -й альтернативы по *Выгодам*, *Издержкам* и *Рискам* соответственно; $w_{A_i}^G$ — обобщенный приоритет i -й альтернативы.

Результирующие приоритеты альтернатив показаны на рис. 11, где можно видеть, что только для двух альтернатив они имеют положительные значения — *Свободный доступ к Интернет-контенту* и *Спонсирование создателей контента*. Отрицательные значения обобщенных приоритетов остальных альтернатив свидетельствуют о том, что характерные для них издержки и риски перевешивают выгоды. Самые худшие значения имеют альтернативы, связанные с наказаниями. Полученные результаты противоречат мнению некоторых известных персон, призывающих ужесточить наказания за нарушения авторских прав. Следует заметить, что действующее в России законодательство в этой области нельзя назвать мягким, в частности, оно предусматривает уголовную ответственность и лишение свободы на срок до шести лет.

Мы объясняем полученные результаты следующими соображениями. Поскольку Интернет-контент обладает свойством неограниченной тиражируемости, сколь угодно строгие наказания не могут гарантировать его защиты от несанкционированного использования. Сеть Интернет — самоорга-

низирующаяся сетевая среда, которая не подчиняется принципам строгого иерархического управления. В связи с этим полный контроль информации в глобальной сети невозможен. Подтверждением этому являются многочисленные утечки конфиденциальных данных, публикуемых в Интернете и не защищенных авторским правом. Яркий пример — история с сайтом Wiki Leaks. Вместе с тем, Интернет предоставляет технические возможности отслеживать поведение пользователей, однако их практическая реализация связана с нарушением базовых прав человека и большими издержками. Все сказанное подтверждает мнение о невозможности полного искоренения пиратства.

Более разумный подход — исследовать причины пиратства и создать условия, в которых оно стало бы минимальным. Основные причины пиратства заключаются не только в бедности пользователей Интернета, но также в том, что медиаконтент отличается от материальных товаров в сознании потребителей. Еще не забылся советский лозунг "Искусство принадлежит народу", в котором есть немалая доля истины. Что касается создателей контента, то для многих из них деньги являются далеко не единственным стимулом творчества. Помимо денег авторы заинтересованы в самовыражении, популярности и диалоге с публикой. Поэтому самыми активными защитниками строгих мер за нарушение авторских прав являются обычно не авторы, а правообладатели. С помощью Интернета пользователи могут удовлетворить ряд своих потребностей; они регулярно платят провайдерам, но часто не заботятся о нарушении авторских прав. Эти нарушения имеют разный масштаб. Прочтение неоплаченной книги эквивалентно походу в библиотеку, но может интерпретироваться как нарушение теми, кто приравнивает этот факт к краже книги из магазина или из той же библиотеки. Очевидно, что такое приравнивание неправомерно. В определенном смысле пользователи "расплачиваются" за нужный контент своим временем и тем, что непроизвольно становятся потребителями рекламы и огромных объемов ненужной им информации. Одной из самых актуальных проблем сегодня становится разработка средств эффективного поиска и фильтрации информации в сети Интернет. Более значимые нарушения законов об авторском праве совершают те, кто размещает нелегальный контент в Интернете в целях привлечения пользователей к сайтам, где им предлагаются рекламные, коммерческие и другие услуги. Необходимость борьбы с такими нарушениями очевидна, но эффективность оставляет желать много лучшего.

Приведенные аргументы логически согласуются с результатами проведенного нами исследования, которые позволяют сделать вывод о бесперспективности ужесточения наказаний и необходимости разработки новых подходов к экономике медиа-

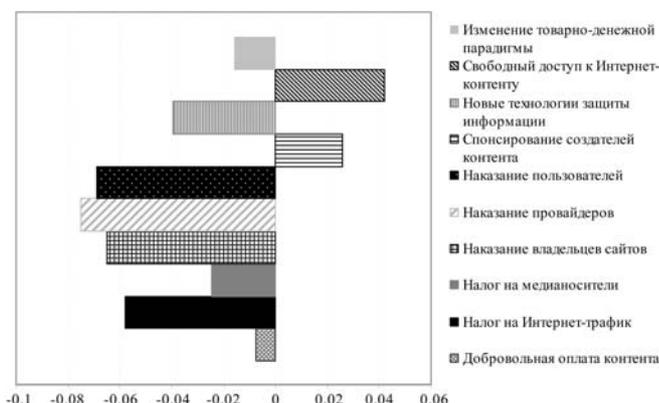


Рис. 11. Обобщенные приоритеты альтернатив

кантента. Безусловно, необходимо совершенствовать законодательство и правоприменительную практику в этой сфере. Однако будущее, на наш взгляд, за новыми представлениями об экономике медиапространства, при формировании которых необходимо рассматривать три типа проблем: технологические, экономические и этические. В инженерном проектировании есть хорошее эвристическое правило: "Если нельзя изменить объект, то попробуйте изменить его окружение". Для применения к рассмотренной в этой статье задаче его можно перефразировать: "Если невозможно обеспечить охрану авторского права в сети Интернет, то нужно изменить представления общества об охране авторского права так, чтобы необходимость в ней отпала".

Список литературы

1. **Беньямин В.** Производство искусства в эпоху его технической воспроизводимости: Избранные эссе / Пер. с нем. под ред. Ю. А. Здороваго. М.: Медиум, 1996. С. 15–65.
2. **Долгин А. Б.** Экономика символического обмена. М.: Прагматика Культуры, 2007.
3. **Долгин А. Б.** Второй универсум // Логос, 2002. № 5–6. С. 243–291.

4. **Лессиг Л.** Свободная культура. М.: Прагматика культуры, 2007.
5. **Саати Т. Л.** Принятие решения при зависимостях и обратных связях: Аналитические сети / Пер. с англ. под ред. О. Н. Андрейчиковой, А. В. Андрейчикова. М.: Книжный дом "Либроком", 2011.
6. **Caves R. E.** Creative Industries: Contracts between Art and Commerce. Cambridge, MA: Harvard University Press, 2000.
7. **Colston C. & Middleton K.** Modern Intellectual Property Law. Cavendish Publishing, UK, 2005.
8. **Cowen T. & Michael F. Brown.** Who Owns Native Culture? // Journal of Cultural Economics, 2004. Vol. 28. Iss. 4. P. 317–323.
9. **Karaganis J.** (Ed.) (2011). Media Piracy in Emerging Economies. Report of Social Science Research Council, 2011. URL: <http://piracy.ssrc.org>.
10. **Liebowitz S. J.** Back to the Future: Can copyright owners appropriate revenues in the face of new technologies? The Economics of Copyright: Recent Developments and Analysis. Cheltenham: Edward Elgar Publishing, 2003.
11. **Saaty T. L.** Fundamentals of Decision Making and Priority Theory with Analytic Hierarchy Process. Pittsburg, PA: RWS Publications, 1994.
12. **Saaty T. L.** Decision making with dependence and feedback: The analytic network process. Pittsburgh, PA: RWS Publications, 2001.
13. **Towse R.** Creativity, Incentive and Reward: an economic analysis of copyright and culture in the Information Age. Cheltenham: Edward Elgar, 2001.
14. **Woodhead R.** Tipping — a method for optimizing compensation for intellectual property, 2000. URL: <http://www.tipping.self-promotion.com>.

УДК 004.825

Д. В. Баранов, аспирант,

Воронежский государственный университет,
e-mail: Denis.Baranov@dataart.com

Логические уравнения в эквациональных LP-структурах

Теория систем переписывания представляет эффективный аппарат формализации знаний в ряде важных областей искусственного интеллекта и компьютерной алгебры. В статье определяется алгебраическая система с семантикой совокупности правил условной эквациональной теории или условной системы переписывания термов. В рамках данной модели вводится и исследуется специальный класс логических уравнений. Результаты могут быть применены для верификации соответствующего множества правил.

Ключевые слова: термы, эквациональная теория, условные правила, алгебраическая модель, логические уравнения

Введение

Системы переписывания термов (СПТ) применяются в различных разделах информационных технологий. Этот формализм используется при решении таких задач, как верификация компьютерных

программ [1], автоматическое доказательство теорем [2], символьное упрощение алгебраических выражений [3], и ряда других. При определении системы переписывания отправной точкой служит эквациональная теория, множество правил которой состоит из равенств термов. Правила СПТ получают путем "ориентации" равенств и пополнения их множества для достижения свойства конfluence-энтности [4]. Аналогичный подход используется в условных СПТ [5].

Важными вопросами, связанными с СПТ, являются эквивалентные преобразования и верификация их множеств правил. Критерий эквивалентности систем переписывания предоставляют эквациональные теории. Поэтому исследование в данном направлении условных СПТ может быть основано на эквивалентности условных эквациональных теорий.

В работе [6] представлена методология исследования условных СПТ на основе "решеточных" продукционно-логических структур (LP-структур). Полученные результаты, в частности, впервые решают задачу построения логической редукции условной эквациональной теории. Они были обобщены в работе [7], где введена и исследована более сложная LP-структура, по-видимому, завершающая исследование задачи минимизации множества правил условной СПТ. В работе [8] введен класс логических уравнений в эквациональной LP-струк-

туре. Рассмотрены некоторые вопросы о разрешимости и количестве решений таких уравнений, а также методы их решения. Сформулированные результаты анонсированы без доказательств.

Настоящая статья уточняет и дополняет исследование свойств и разрешимости указанных уравнений, причем содержит доказательства основных теорем. Полученные результаты могут быть применены для верификации множества правил условной СПТ. Заметим, что основы применяемой здесь методологии были заложены в работе [9], где с аналогичных позиций рассмотрен класс логических уравнений в алгебраической системе, имеющей более простую архитектуру.

1. Основные понятия и обозначения

Напомним некоторые понятия из общей теории. Бинарное отношение R на произвольном множестве F называется *рефлексивным*, если для любого $a \in F$ справедливо $(a, a) \in R$; *транзитивным*, если для любых $a, b, c \in F$ из $(a, b), (b, c) \in R$ следует $(a, c) \in R$. Известно [10], что существует замыкание R^* произвольного отношения R относительно свойства транзитивности (*транзитивное замыкание*).

Решеткой [11] называется частично упорядоченное множество \mathbb{F} , в котором наряду с отношением \leq ("не больше", "содержится") введены две двуместные операции \wedge ("пересечение") и \vee ("объединение"), вычисляющие соответственно точную нижнюю и верхнюю грани для любых $a, b \in \mathbb{F}$. Решетка \mathbb{F} называется *ограниченной*, если она содержит общие нижнюю и верхнюю грани — такие два элемента O, I , что $O \leq a \leq I$ для любого $a \in \mathbb{F}$.

Множество всех конечных подмножеств $\lambda(U)$ универсума U образует решетку. В настоящей статье рассматривается такой вид решеток. Чтобы подчеркнуть данное обстоятельство, вместо символов \leq, \geq, \wedge и \vee будем использовать знаки теоретико-множественных операций $\subseteq, \supseteq, \cup$ и \cap .

Атомом ограниченной (снизу) решетки \mathbb{F} называется минимальный элемент ее подмножества $\mathbb{F} \setminus O$. Решетка называется *атомно-порожденной*, если каждый ее элемент представлен объединением атомов. В $\lambda(U)$ атомами являются все подмножества, состоящие из единственного элемента универсума. Для атома a элемента A будем использовать обозначение $a \in A$ (наряду с обозначением $a \subseteq A$).

Приведем базовые определения, связанные с терминами [4]. Пусть Σ — алфавит, образованный объединением следующих непересекающихся множеств: V — множество переменных; $\Sigma_n, n = 0, 1, \dots$, — множества n -арных функций (функциональных символов); 0 -арные функции называются константами. Множество термов $T(\Sigma)$ определяется рекурсивно:

- $V \subset T(\Sigma)$; $\Sigma_0 \subset T(\Sigma)$;
- если $f \in \Sigma_n$ и $t_1, \dots, t_n \in T(\Sigma)$, то $f(t_1, \dots, t_n) \in T(\Sigma)$.

Отображение $\sigma : V \rightarrow T(\Sigma)$ называется *подстановкой*. Это понятие распространяется на все $t \in T(\Sigma)$ следующим образом:

- если $t = x \in V$, то $\sigma(t) = \sigma(x)$;
- если $t = f \in \Sigma_0$, то $\sigma(t) = f$;
- если $f \in \Sigma_n, t_1, \dots, t_n \in T(\Sigma)$ и $t = f(t_1, \dots, t_n)$, то $\sigma(t) = f\sigma(t_1), \dots, \sigma(t_n)$.

Эквациональной теорией называется пара (Σ, E) , где Σ — алфавит, состоящий из счетного множества переменных и непустого множества функциональных символов (сигнатура); $E \subseteq T(\Sigma) \times T(\Sigma)$ — множество равенств вида $s = t$ ($s, t \in T(\Sigma)$).

Определяется понятие выводимости равенства $s = t$ из E ($(\Sigma, E) \vdash s = t$) на основе следующих правил.

1. Если $s = t \in E$, то $(\Sigma, E) \vdash s = t$.
2. Если $(\Sigma, E) \vdash s = t_1, \dots, s_n = t_n$, то $(\Sigma, E) \vdash f(s_1, \dots, s_n) = f(t_1, \dots, t_n)$ для любого $f \in \Sigma_n$.
3. Если $(\Sigma, E) \vdash s = t$, то $(\Sigma, E) \vdash \sigma(s) = \sigma(t)$ для любой подстановки σ .
4. $(\Sigma, E) \vdash t = t$.
5. Если $(\Sigma, E) \vdash t_1 = t_2, t_2 = t_3$, то $(\Sigma, E) \vdash t_1 = t_3$.
6. Если $(\Sigma, E) \vdash s = t$, то $(\Sigma, E) \vdash t = s$.

Для данного множества равенств E рассмотрим решетку конечных подмножеств $\lambda(E)$. В ней заданы отношения \subseteq, \supseteq , а также "решеточные" операции \cap и \cup — теоретико-множественные пересечение и объединение. Потребуется еще три группы операций, связанных соответственно с функциями, подстановками и транзитивностью равенств термов.

1. Если $a = \{s_i = t_i \mid i = 1, \dots, n\}$ и $f \in \Sigma_n$, то $f(a) = \{f(s_1, \dots, s_n) = f(t_1, \dots, t_n)\}$.
2. Если $a = \{s_j = t_j \mid j = 1, \dots, m\}$, то $\sigma(a) = \{\sigma(s_j) = \sigma(t_j) \mid j = 1, \dots, m\}$ для подстановки σ .
3. Если $a = \{t_1 = t_2, t_2 = t_3\}, b = \{t_1 = t_3\}$, то $b = \text{Tr}(a)$.

Определение 1.1. Пусть задана эквациональная теория (Σ, E) . Эквациональной решеткой \mathbb{F} будем называть решетку, полученную пополнением $\lambda(E)$ относительно дополнительных операций 1–3.

Очевидно, что эквациональная решетка является атомно-порожденной.

Рассматривается условная эквациональная теория, содержащая условные правила вида $s_1 = t_1, \dots, s_n = t_n : u_1 = v_1, \dots, u_m = v_m$ ("если имеют место равенства термов $s_i = t_i, i = 1, \dots, n$, то выполнены и все $u_j = v_j, j = 1, \dots, m$ "). Будем называть такие соотношения (условными) эквациональными правилами, или просто правилами, там, где это не будет вызывать недоразумений. Таким образом, предпосылка и заключение правила являются элементами \mathbb{F} .

Используя в качестве основы аксиомы и правила, определенные в работе [4], сформулируем их аналоги для условной эквациональной дедукции. Аксиомы порождаются перечисленными выше правилами вывода равенств. Правило вывода 2 означает наличие условного правила (аксиомы) $a : f(a)$ для любых $a = \{s_1 = t_1, \dots, s_n = t_n\}$ и $f \in \Sigma_n$; из правила 3 следует аксиома $a : \sigma(a)$ для любых $a \in \mathbb{F}$ и подстановки σ . Правило 5 порождает аксиому $a : \text{Tr}(a)$ для каждого

подходящего $a \in \mathbb{F}$. Такие условные правила в рассматриваемой логике можно также назвать тавтологиями. Еще одной очевидной тавтологией является правило $a:b, a, b \in \mathbb{F}$ при $a \supseteq b$. Правила вывода в условной эквациональной логике таковы:

1. $a : b \vdash \sigma(a) : \sigma(b)$ ($a, b \in \mathbb{F}$) для любой подстановки σ (см. аналогичное правило в [4] с исходными обозначениями);

2. $a : b, a : c \vdash a : b \cup c$ ($a, b, c \in \mathbb{F}$) (возможность вывода по частям);

3. $a : b, b : c \vdash a : c$ ($a, b, c \in \mathbb{F}$) (транзитивность).

2. Эквациональная LP-структура

Согласно работам [6, 7] LP-структурой (*Lattice Production structure*) называется решетка с дополнительным бинарным отношением, которое имеет ряд продукционно-логических свойств. В этом разделе рассматриваются отношения на эквациональной решетке. Вводится понятие логического отношения, которое соответствует множеству правил условной эквациональной теории. Свойства такого отношения отражают сформулированные в п. 1 правила условной эквациональной дедукции.

Во-первых, логическое отношение R содержит все тавтологии. Введем для них общее обозначение: $a \succcurlyeq b$, если $a \supseteq b$, $b = \sigma(a)$, $b = f(a)$ или $b = \text{Tr}(a)$. Обозначим также \succ отношение $\{(a, b) \mid a \succcurlyeq b, a \neq b\}$, \prec — обратное к нему. Таким образом, для логического отношения R справедливо $\succ \subseteq R$. Другие свойства логического отношения вытекают из правил дедукции.

Определение 2.1. Отношение R назовем применимым, если для любой подстановки σ из $(a, b) \in R$ следует $(\sigma(a), \sigma(b)) \in R$.

Определение 2.2. Отношение R назовем \cup -дистрибутивным, если для любых $(a, b_1), (a, b_2) \in R$ выполнено $(a, b_1 \cup b_2) \in R$.

Следующее определение суммирует рассмотренные свойства.

Определение 2.3. Бинарное отношение на эквациональной решетке называется логическим, если оно содержит тавтологии, а также является применимым, \cup -дистрибутивным и транзитивным. Логическим замыканием R^L произвольного отношения R называется наименьшее логическое отношение, содержащее R .

Два отношения R_1 и R_2 , определенные на общей эквациональной решетке, называются эквивалентными, если их логические замыкания совпадают. Этот факт будем обозначать $R_1 \sim R_2$. Логической редукцией данного отношения R называется минимальное эквивалентное ему отношение R_0 .

Определение 2.4. Пусть задано некоторое отношение R на эквациональной решетке \mathbb{F} . Будем говорить, что упорядоченная пара $a, b \in \mathbb{F}$ логически связана отношением R (обозначим этот факт $a \xrightarrow{R} b$), если выполнено одно из следующих условий:

1) $(a, b) \in R$;

2) $a \succcurlyeq b$:

2.1) $a \supseteq b$ или 2.2) $b = \sigma(a)$ или 2.3) $b = f(a)$ или 2.4) $b = \text{Tr}(a)$;

3) существуют такие $a_1, b_1 \in \mathbb{F}$ и подстановка σ ,

что $a = \sigma(a_1)$, $b = \sigma(b_1)$, причем $a_1 \xrightarrow{R} b_1$;

4) существуют такие $b_1, b_2 \in \mathbb{F}$, что $b_1 \cup b_2 = b$, причем $a \xrightarrow{R} b_1$, $a \xrightarrow{R} b_2$;

5) существует элемент $c \in \mathbb{F}$, такой, что $a \xrightarrow{R} c$ и $c \xrightarrow{R} b$.

В работе [7] доказана теорема о существовании логического замыкания R^L для произвольного бинарного отношения R на эквациональной решетке.

Показано, что оно совпадает с отношением \xrightarrow{R} .

Под эквациональной LP-структурой будем подразумевать эквациональную решетку с определенным на ней логическим отношением.

Определение 2.5. Отношение на эквациональной решетке называется каноническим, если оно состоит из пар "хорновского" типа: правая часть пары представляет собой атом решетки, левая часть — объединение атомов.

Можно показать, что любое отношение на атомно-порожденной решетке эквивалентно некоторому каноническому отношению.

В работе [7] рассмотрен также вопрос о возможности в процессе построения логического замыкания бинарного отношения выделить этап, соответствующий транзитивному замыканию. Положительный ответ позволяет свести изучение некоторых важных вопросов, касающихся логических отношений, к соответствующим проблемам транзитивных отношений. В частности, построение логического замыкания или редукции можно осуществлять с помощью быстрых алгоритмов (типа Уоршолла [10]).

Для отношения R на эквациональной решетке рассмотрим отношение \bar{R} , построенное последовательным выполнением следующих шагов.

Шаг 1. Добавить к R все пары (a, b) , для которых $b = \sigma(a)$, $b = f(a)$ либо $b = \text{Tr}(a)$, и обозначить новое отношение R_1 .

Шаг 2. Добавить к R_1 всевозможные пары вида $(\sigma(a), \sigma(b))$, для которых $(a, b) \in R_1$, и обозначить новое отношение R_2 .

Шаг 3. Добавить к R_2 всевозможные пары вида $(a_1 \cup a_2 \cup \dots \cup a_m, b_1 \cup b_2 \cup \dots \cup b_m)$, где $(a_i, b_i) \in R_2$, и обозначить новое отношение R_3 .

Шаг 4. Объединить R_3 с отношением \supseteq .

Заметим, что шаги 1 и 4 относятся ко всем теоретически возможным элементам эквациональной решетки \mathbb{F} и, соответственно, не зависят от данного R . В силу бесконечности множества \mathbb{F} описанный процесс построения отношения \bar{R} носит теоретический характер. В приложениях вместо \mathbb{F} можно

взять конечное подмножество эквациональной решетки, построенное при ограничении уровня вложенности термов.

В работе [7] доказана теорема, которая потребуется в разд. 3.

Теорема 2.1. Логическое замыкание R^L отношения R на эквациональной решетке совпадает с транзитивным замыканием \tilde{R}^* соответствующего отношения \tilde{R} .

Из этой теоремы выводится следующее утверждение.

Следствие 2.1. Если $R_1 \subseteq R_2$, то $\xrightarrow{R_1} \subseteq \xrightarrow{R_2}$.

Доказательство. Во-первых, из описания процесса построения \tilde{R} легко заметить, что если $R_1 \subseteq R_2$, то $\tilde{R}_1 \subseteq \tilde{R}_2$. Аналогичное вложение имеет место и для транзитивных замыканий этих отношений, что в силу теоремы 2.1 означает доказываемый факт. \square

В работе [7] доказана также теорема о существовании логической редукции бинарного отношения на эквациональной решетке.

3. Логические уравнения в эквациональной LP-структуре

В данном разделе исследуется связанный с эквациональными LP-структурами класс уравнений. Рассматриваются вопросы о разрешимости и количестве решений таких уравнений, а также методы их решения. Нахождение решения логического уравнения соответствует обратному выводу на эквациональной решетке.

Данный вид решеток порождает более сложную в сравнении с работой [9] LP-структуру, содержащую расширенный набор операций и свойств логических отношений. Принятый в связи с данным обстоятельством (в отличие от работы [9]) рекурсивный подход к определению отношения \xrightarrow{R} существенно влияет на технику исследования логических уравнений.

Напомним (разд. 1), что эквациональная решетка является атомно-порожденной, т. е. каждый ее элемент представим в виде конечного объединения атомов. В настоящем разделе часто будем обозначать атомы решетки строчными буквами, остальные элементы — прописными.

Пусть задано конечное бинарное отношение R на эквациональной решетке \mathbb{F} . Согласно результатам разд. 2 без ограничения общности будем предполагать, что R представлено логической редукцией в ее канонической форме. Рассмотрим множество всех атомов решетки, образующих элементы пар отношения R , и обозначим его \mathbb{F}_R^0 . Оно конечно в силу конечности R . Обозначим далее \mathbb{F}_R — под-

множество решетки \mathbb{F} , построенное на \mathbb{F}_R^0 как системе образующих с помощью операции "решеточного" объединения. Очевидно, \mathbb{F}_R также конечно и содержит все элементы пар отношения R .

Расширим отношение R множеством нерелективных ($a \neq b$) пар, выводимых из R с помощью правил 2.2—2.4 и 3 определения 2.4. Точнее, последовательно выполним следующие шаги.

Шаг 1. Добавить к R все пары (a, b) , для которых $(a, b) \in \mathbb{F}_R$, $b = \sigma(a)$, и обозначить новое отношение R_1 .

Шаг 2. Добавить к R_1 все пары (a, b) , для которых $a \in \mathbb{F}_R$, $b \in \mathbb{F}_R$, причем $b = f(a)$ либо $b = \text{Tr}(a)$, и обозначить новое отношение R_2 .

Шаг 3. Добавить к R_2 всевозможные пары вида $(\sigma(a), \sigma(b))$, для которых $(a, b) \in R_2$, $\sigma(b) \in \mathbb{F}_R$, и обозначить новое отношение \bar{R} .

В результате получим отношение $\bar{R} \sim R$, очевидно, являющееся подмножеством отношения \tilde{R} (см. разд. 2). При выполнении шага 3 множество используемых в новом отношении атомов решетки может расширяться. В связи с данным обстоятельством, наряду с \mathbb{F}_R^0 и \mathbb{F}_R , введем для отношения \bar{R} аналогичные множества $\mathbb{F}_{\bar{R}}^0$ и $\mathbb{F}_{\bar{R}}$.

Из результатов работы [7] следует, что логическое замыкание исходного отношения R может быть получено применениями к \bar{R} лишь правил 4, 5 определения 2.4, дополненных отношением \supseteq . Данный факт наглядно иллюстрируется также последовательностью шагов построения отношения \tilde{R} (разд. 2) и теоремой 2.1. Таким образом, $R \sim \bar{R} \sim \tilde{R}$, причем отношение \bar{R} включает в виде пар элементов результаты всех дополнительных операций эквациональной LP-структуры, отличающих ее от модели работы [9].

Пусть есть построенное выше отношение \bar{R} на эквациональной решетке \mathbb{F} и имеет место $A \xrightarrow{\bar{R}} B$. Тогда B является образом A , а A — прообразом B при отношении $\xrightarrow{\bar{R}}$. Каждый элемент может иметь много образов и прообразов. Более того, в данном случае в силу определения логического отношения, любой $B_1 \subset B$ является образом A и каждый $A_1 \supset A$ является прообразом B . По этой причине при изучении образов и прообразов логических отношений целесообразно уточнение рассматриваемых понятий.

Для данного $B \in \mathbb{F}$ минимальным прообразом при отношении $\xrightarrow{\bar{R}}$ называется такой элемент $A \in \mathbb{F}$, что $A \xrightarrow{\bar{R}} B$ и A является минимальным, т. е. не содержит никакого другого $A_1 \in \mathbb{F}$, для которого $A_1 \xrightarrow{\bar{R}} B$.

Определение 3.1. Атом $x \in \mathbb{F}_{\bar{R}}^0$ называется начальным при отношении \bar{R} , если в \bar{R} нет ни одной пары вида (A, x) . Элемент $X \in \mathbb{F}_{\bar{R}}$ называется начальным, если все его атомы начальные (при отношении \bar{R}). Подмножество $\mathbb{S}(\bar{R}) \subseteq \mathbb{F}_{\bar{R}}$ (будем обозначать \mathbb{S} , если это не вызовет неоднозначностей), состоящее из всех начальных элементов $\mathbb{F}_{\bar{R}}$, называется начальным множеством решетки (при отношении \bar{R}).

Рассмотрим уравнение

$$\bar{R}^L(X) = B, \quad (1)$$

где B — заданный элемент решетки; X — неизвестный элемент.

Определение 3.2. Точным (частным) решением X уравнения (1) называется любой минимальный прообраз элемента B , содержащийся в \mathbb{S} . Приближенным (частным) решением X уравнения (1) называется любой прообраз элемента B , содержащийся в \mathbb{S} . Общим решением уравнения (1) называется совокупность всех его частных решений $\{X_p\}$, $p \in P$.

По определению, точное решение является и приближенным. Очевидно, приближенное решение всегда содержит хотя бы одно точное решение. Уравнения вида (1) будем называть логическими уравнениями на эквациональных решетках.

Рассмотрим вопрос о том, как меняется общее решение уравнений вида (1) при объединении их правых частей. Точнее, нельзя ли для одного и того же отношения R вместо исходного уравнения решить несколько уравнений с правыми частями, объединение которых совпадает с B ?

Лемма 3.1. Пусть X_1 — частное решение уравнения вида (1) с правой частью B_1 , а Y_1 — частное решение уравнения того же вида с правой частью B_2 . Тогда $X_1 \cup Y_1$ является приближенным решением уравнения

$$\bar{R}^L(X) = B_1 \cup B_2. \quad (2)$$

Доказательство. Действительно, поскольку $\bar{R} \rightarrow X_1 \cup Y_1$ содержит отношение включения, то $X_1 \cup Y_1 \xrightarrow{\bar{R}} X_1$ и $X_1 \cup Y_1 \xrightarrow{\bar{R}} Y_1$. Отсюда, так как по условию леммы $X_1 \xrightarrow{\bar{R}} B_1$ и $Y_1 \xrightarrow{\bar{R}} B_2$, в силу транзитивности $\xrightarrow{\bar{R}}$ имеем $X_1 \cup Y_1 \xrightarrow{\bar{R}} B_1$ и $X_1 \cup Y_1 \xrightarrow{\bar{R}} B_2$. Наконец, из последних двух соотношений, пользуясь дистрибутивностью $\xrightarrow{\bar{R}}$, получим $X_1 \cup Y_1 \xrightarrow{\bar{R}} B_1 \cup B_2$. \square

Теорема 3.1. Пусть $\{X_p\}$, $p \in P$ — общее решение уравнения вида (1) с правой частью B_1 , а $\{Y_q\}$, $q \in Q$, — общее решение уравнения того же вида с правой частью B_2 . Тогда общее решение уравнения (2) представляет собой множество всех элементов вида $X_p \cup Y_q$, из которого исключены элементы, содержащие другие элементы этого же множества.

Доказательство. По лемме 3.1, каждый элемент $X_p \cup Y_q$ является прообразом $B_1 \cup B_2$, т. е. содержит хотя бы одно частное решение (2). Покажем теперь, что уравнение (2) не имеет частных решений, отличных от вида $X_p \cup Y_q$. Предположим противное — пусть некоторый $z \in \mathbb{S}$, являясь частным решением (2), не совпадает ни с одним элементом вида $X_t \cup Y_r$. При этом Z не может содержать ни одного другого элемента вида $X_p \cup Y_q$, иначе он не был бы частным решением (оно по определению минимально). Отсюда следует, что Z не содержит ни одного X_p (либо ни одного Y_q , что симметрично). Вместе с тем, по-

скольку $Z \xrightarrow{\bar{R}} B_1 \cup B_2$, то $Z \xrightarrow{\bar{R}} B_1$ и $Z \xrightarrow{\bar{R}} B_2$. Этот факт означает, что Z содержит хотя бы по одному частному решению уравнений с правыми частями B_1 и B_2 , а все эти решения содержатся соответственно в множествах $\{X_p\}$ и $\{Y_q\}$. Таким образом, приходим к противоречию. \square

Рассмотрим методы решения логических уравнений вида (1). Предполагаем, что \bar{R} — построенное выше конечное каноническое отношение на подмножестве эквациональной решетки $\mathbb{F}_{\bar{R}}$, а правая часть B уравнения (1) представляет собой конечное объединение атомов из $\mathbb{F}_{\bar{R}}^0$.

Введем понятие структурного расслоения отношения \bar{R} на виртуальные слои (частичные отношения) $\{\bar{R}_t \mid t \in T\}$. Оно позволит упростить изучение

свойств отношения $\xrightarrow{\bar{R}}$, а также облегчит построение и исследование ряда алгоритмов, связанных с решением соответствующих логических уравнений. С этой целью вначале разобьем \bar{R} на непересекающиеся подмножества, каждое из которых образовано парами вида (A, x_p) с одним и тем же атомом x_p в качестве правой части. Такое разбиение имеет смысл благодаря тому, что \bar{R} является каноническим.

Обозначим эти подмножества \bar{R}^p соответственно их элементу x_p , $p \in P$.

Определение 3.3. Слоем \bar{R}_t в отношении \bar{R} называется подмножество \bar{R} , образованное упорядоченными парами, взятыми по одной из каждого непустого \bar{R}^p , $p \in P$. Два слоя, отличающихся хотя бы одной парой, считаются различными.

Замечание 3.1. Каждый слой содержит максимально возможное подмножество пар отношения \bar{R} с уникальными правыми частями. Добавление к слою еще одной пары нарушило бы это условие.

Замечание 3.2. Любое подмножество пар в отношении \bar{R} с уникальными правыми частями принадлежит некоторому слою.

Замечание 3.3. Различные слои могут иметь непустые пересечения. Объединение всех слоев равно \bar{R} .

Замечание 3.4. Нетрудно заметить, что общее число слоев N определяется равенством $N = \prod_{p \in P} N_p$, где N_p — мощность подмножества пар отношения \bar{R} с правой частью x_p .

Пусть $A \in \mathbb{F}_{\bar{R}}$. Будем говорить, что элемент $B \in \mathbb{F}_{\bar{R}}$ получен из A применением пары $(X, z) \in \bar{R}$, если $X \subseteq A$ и $B = A \cup z$. Очевидно, что $(A, B) \in \bar{R}$.

Рассмотрим некоторые свойства построенного выше расслоения отношения \bar{R} . Вначале докажем следующее полезное утверждение.

Лемма 3.2. Пусть $A \xrightarrow{\bar{R}} B$. Тогда существует кортеж (C_1, \dots, C_m) элементов множества $\mathbb{F}_{\bar{R}}$, такой, что $A = C_0 \subset C_1 \subset \dots \subset C_m \supseteq B$, причем каждый элемент $C_j, j > 0$, получен из C_{j-1} применением некоторой пары $(X_j, z_j) \in \bar{R}$. При $(A, B) \in \bar{R}$ полагаем $m = 0$.

Доказательство. При условии леммы $(A \xrightarrow{\bar{R}} B)$ по теореме 2.1 имеется упорядоченный конечный набор элементов (B_1, \dots, B_n) , такой, что в последовательности $(B_0, B_1), (B_1, B_2), \dots, (B_n, B_{n+1})$, где $B_0 = A, B_{n+1} = B$, все пары принадлежат отношению \bar{R} (см. разд. 2). При $(A, B) \in \bar{R}$ можно считать $m = 0$.

Далее заметим, что, по построению данного отношения, из $(B_j, B_{j+1}) \in \bar{R}$ следует $(B_j, B_j \cup B_{j+1}) \in \bar{R}$ для всех $j = 0, \dots, n$. Применяя такое свойство последовательно по $j = 0, \dots, n$, получим новый набор $(\tilde{B}_1, \dots, \tilde{B}_n)$, где $\tilde{B}_k = \bigcup_{j=0, \dots, k} B_j, k = 0, \dots, n$. Данный кортеж, наряду с исходным (B_1, \dots, B_n) , также по определению 2.4 реализует логическую связь $A \xrightarrow{\bar{R}} B$, при этом обладая дополнительным свойством монотонности $A = \tilde{B}_0 \subseteq \tilde{B}_1 \subseteq \dots \subseteq \tilde{B}_n \supseteq B_{n+1} = B$.

Рассмотрим далее структуру пар $(\tilde{B}_{j-1}, \tilde{B}_j), j = 1, \dots, n$. Учитывая содержание процессов построения отношений \bar{R} и $\tilde{\bar{R}}$, нетрудно заметить следующий факт. Множество пар \bar{R} , дополненное рефлексивным отношением над используемыми атомами, образует базис при построении $\tilde{\bar{R}}$ на основе операции объединения. Это означает, что каждая пара $(\tilde{B}_{j-1}, \tilde{B}_j) \in \tilde{\bar{R}}$ может быть получена из некоторого (рефлексивно-дополненного) подмножества \bar{R}_j пар отношения \bar{R} путем объединения их левых

(для \tilde{B}_{j-1}) и правых (для \tilde{B}_j) частей. Отсюда следует, что элемент \tilde{B}_j , включающий \tilde{B}_{j-1} , может быть получен из \tilde{B}_{j-1} последовательным применением пар подмножества $\bar{R}_j \subseteq \bar{R}$. Распространяя данное свойство на каждую пару $(\tilde{B}_{j-1}, \tilde{B}_j), j = 1, \dots, n$, а также исключая все избыточные применения (рефлексивно-дополненных) пар \bar{R} , приходим к утверждению леммы при некотором $m \geq n$. \square

Замечание 3.5. Построенный в лемме кортеж (C_0, \dots, C_m) можно выбрать *точным* в следующем смысле. Правый атом z_j каждой примененной в кортеже для перехода от C_{j-1} к C_j пары (X_j, z_j) должен содержаться в B либо в левой части X_{j+k} некоторой применяемой позднее пары (X_{j+k}, z_{j+k}) , и не может содержаться в C_{j-1} . Переходы, не удовлетворяющие такому свойству, не являются необходимыми для получения B . Поэтому они могут быть исключены из кортежа вместе с C_j .

Докажем свойства, характеризующие расслоение $\{\bar{R}_t\}$ отношения \bar{R} с точки зрения решения логических уравнений вида (1).

Лемма 3.3. Если $A \xrightarrow{\bar{R}} B$, то в \bar{R} существует слой \bar{R}_t , такой, что $A \xrightarrow{\bar{R}_t} B$.

Доказательство. Пусть $A \xrightarrow{\bar{R}} B$. Тогда по лемме 3.2 имеется упорядоченный "монотонно возрастающий" набор элементов (C_1, \dots, C_m) , такой, что в последовательности $(C_0, C_1), (C_1, C_2), \dots, (C_{m-1}, C_m)$, где $C_0 = A, C_m \supseteq B$, каждый элемент $C_j, j > 0$, получается из C_{j-1} применением пары $(X_j, z_j) \in \bar{R}$.

Из монотонности $C_0 \subset C_1 \subset \dots \subset C_m$ следует, что все переходы вида (C_{j-1}, C_j) могут быть получены на основе пар отношения \bar{R} с уникальными правыми частями. Действительно, любая пара (X_j, z_j) канонического отношения \bar{R} , впервые (слева направо) примененная к C_{j-1} для получения C_j , исключает возможность еще одного применения любой пары с тем же атомом z_j в правой части, поскольку z_j уже содержится во всех элементах $(C_j, C_{j+1}, \dots, C_m)$. Отсюда, по замечанию 3.2 и определению 2.4, получаем $A \xrightarrow{\bar{R}_t} B$. \square

Следствие 3.1. Логическое замыкание отношения \bar{R} равно объединению логических замыканий его слоев, то есть $\bar{R} \rightarrow = \bigcup_{t \in T} \bar{R}_t \rightarrow$.

Доказательство. Поскольку $\bar{R}_t \subseteq \bar{R}$, то согласно следствию 2.1 имеем $\bar{R}_t \rightarrow \subseteq \bar{R} \rightarrow (\forall t \in T)$. Обратно,

если $A \xrightarrow{\bar{R}} B$, то по лемме 3.3 справедливо $A \xrightarrow{\bar{R}_t} B$ при некотором $t \in T$. Следовательно, $(A, B) \in \bigcup_{t \in T} \xrightarrow{\bar{R}_t}$. □

Будем говорить, что решение X уравнения (1) (точное или приближенное) порождается в \bar{R} некоторым слоем \bar{R}_t , если $X \xrightarrow{\bar{R}_t} B$.

Замечание 3.6. Согласно лемме 3.3 и следствию 3.1 любое частное решение уравнения (1) порождается в \bar{R} некоторым слоем \bar{R}_t .

Замечание 3.7. Из следствия 3.1 вытекает, что для нахождения решения уравнения (1) в некотором слое \bar{R}_t достаточно вместо (1) решить аналогичное уравнение с отношением \bar{R}_t .

Последние замечания не гарантируют, что два различных слоя не могут порождать одного и того же решения. Кроме того, могут существовать слои, дающие точное решение для \bar{R}_t , но приближенное для \bar{R} . Некоторые слои могут вообще не давать решений. Однако справедливо утверждение о том, что слой не может порождать более одного точного решения.

Лемма 3.4. Ни один слой \bar{R}_t отношения \bar{R} не может порождать двух различных точных решений уравнения (1).

Доказательство. Предположим противное, что некоторый слой \bar{R}_t порождает два точных решения $X_1 \neq X_2$. Пусть, для определенности, существует начальный атом $a_0 \in X_1$ и $a_0 \notin X_2$. Согласно леммам

3.2 и 3.3 для логической связи $X_1 \xrightarrow{\bar{R}_t} B$ существует "монотонно возрастающий" кортеж вывода $(X_1 = C_0, C_1, \dots, C_m)$, каждый переход в котором осуществляется соответствующей парой отношения \bar{R}_t . Выберем кортеж точным в соответствии с замечанием 3.5.

Далее будем просматривать (C_0, \dots, C_m) слева направо и при этом строить некоторую специальную последовательность пар слоя \bar{R}_t . Поскольку рассматривается точное решение X_1 , то его атом a_0 , как и любой атом точного решения содержится: а) в левой части хотя бы одной пары (A_1, a_1) , осуществляющей переход в кортеже (C_0, \dots, C_m) либо б) в B — правой части уравнения (1). Если верно а), добавим соответствующую пару (A_1, a_1) к накапливаемой последовательности пар, в противном случае завершим просмотр. На следующем шаге (переместившись в кортеже вправо до перехода (A_1, a_1)) ищем аналогичную пару (A_2, a_2) для атома a_1 . Для него имеем ту же альтернативу: а) либо б), поскольку просматриваемый кортеж является точным (см. замечание 3.5). Рано или поздно, закончив просмотр кортежа, получим последовательность пар-переходов $\{(A_i, a_i) \mid i = 1, \dots, n\}$, причем $a_i \in A_{i+1}$, $(0 \leq i \leq n-1)$, $a_n \in B$. Если эта последовательность

окажется пустой ($n = 0$), то, соответственно, имеем $a_0 \in B$.

Рассмотрим теперь *точный* кортеж вывода $(\tilde{C}_0, \dots, \tilde{C}_k)$, соответствующий другому решению — X_2 . Его будем просматривать справа налево, пытаясь в переходах в обратном порядке выделить ту же самую последовательность $\{(A_i, a_i)\}$. Поскольку $a_n \in B$, то по построению кортежа имеем и $a_n \in \tilde{C}_k$. Далее циклически выполняем однотипные шаги. Для атома a_n возможны два варианта: а) существует пара слоя \bar{R}_t с правой частью a_n , применение которой вызвало некоторый переход в кортеже $(\tilde{C}_0, \dots, \tilde{C}_k)$; б) a_n содержится в $\tilde{C}_0 = X_2$. В случае варианта а) упомянутой парой может быть лишь (A_n, a_n) , так как слой \bar{R}_t не может содержать двух пар с одинаковой правой частью a_n . В этом случае переходим к предыдущей паре в выделяемой последовательности $\{(A_i, a_i)\}$, исходя из отправной точки $a_{n-1} \in A_n$ — правой части очередной пары. В случае б) получим противоречие: если $n > 0$, то атом a_n не является начальным и, соответственно, не может содержаться в решении X_2 ; вариант $n = 0$ противоречит сделанному в начале доказательства предположению $a_0 \notin X_2$. Продолжая данный процесс, в силу конечности последовательности пар $\{(A_i, a_i)\}$ рано или поздно придем к ситуации б), противоречащей сделанным предположениям, что и доказывает лемму. □

Объединяя полученные выше результаты, можно сформулировать следующее утверждение.

Теорема 3.2. Для нахождения общего решения уравнения (1) достаточно найти частное решение X_t в каждом слое \bar{R}_t , если оно существует. Далее из полученного множества решений необходимо исключить элементы, содержащие другие элементы этого же множества.

Утверждение теоремы сразу следует из замечания 3.6 и леммы 3.4. □

Следствие 3.2. Число частных решений уравнения (1) оценивается сверху выражением $N = \prod_{p \in P} N_p$,

где N_p — мощность подмножества пар отношения \bar{R} с правой частью x_p , а P определяет все такие подмножества (см. замечание 3.4). □

Теорема 3.2 и замечание 3.7 позволяют свести вопрос о решении уравнения (1) к задаче нахождения решения уравнения

$$\bar{R}_t^L(X) = B, \quad (3)$$

где B — не начальный элемент $\mathbb{F}_{\bar{R}}$; \bar{R}_t — произвольный слой в \bar{R} .

Согласно лемме 3.4 обратное к $\xrightarrow{\bar{R}_t}$ отношение является однозначным отображением $f_t: \mathbb{F}_{\bar{R}} \rightarrow \mathbb{F}_{\bar{R}}$

с некоторой областью определения $D(f_i) \subseteq \mathbb{F}_{\bar{R}}$. Рассмотрим его некоторые свойства.

Лемма 3.5. Отображение f_i является \cup -гомоморфизмом [11], т. е. $f_i(B_1 \cup B_2) = f_i(B_1) \cup f_i(B_2)$, $\forall B_1, B_2 \in D(f_i)$.

Доказательство. Обозначим $X_1 = f_i(B_1)$, $X_2 = f_i(B_2)$. Как и в доказательстве леммы 3.4, для логической связи $X_1 \xrightarrow{\bar{R}_i} B_1$ рассмотрим точный монотонный кортеж вывода $C = (C_0, \dots, C_m)$, каждый переход в котором осуществляется одной парой отношения \bar{R}_i . Для связи $X_2 \xrightarrow{\bar{R}_i} B_2$ существует собственный набор $\tilde{C} = (\tilde{C}_0, \dots, \tilde{C}_k)$ с аналогичными свойствами. Предположим для определенности, что $m \geq k$. На основе этих двух кортежей построим еще один упорядоченный набор элементов $\mathbb{F}_{\bar{R}}$ следующим образом:

$$\hat{C} = (C_0 \cup \tilde{C}_0, C_1 \cup \tilde{C}_0, C_1 \cup \tilde{C}_1, C_2 \cup \tilde{C}_1, \\ C_2 \cup \tilde{C}_2, \dots, C_k \cup \tilde{C}_k, C_{k+1} \cup \tilde{C}_k, \dots, C_m \cup \tilde{C}_k).$$

Эта монотонная последовательность, очевидно, реализует логическую связь $X_1 \cup X_2 \xrightarrow{\bar{R}_i} B_1 \cup B_2$.

В соответствии с замечанием 3.5 набор \hat{C} может быть сокращен до кортежа, не содержащего избыточных переходов. При этом элемент $X_1 \cup X_2$ представляет именно точное решение уравнения (3) с правой частью $B_1 \cup B_2$. Данный факт справедлив в силу точности решений X_1 и X_2 , а также детерминированности процесса обратного вывода в слое \bar{R}_i для каждого атома из $B_1 \cup B_2$ (см. доказательство леммы 3.4). \square

Основываясь на лемме 3.5, для решения уравнения (3) достаточно решить уравнение с каждым атомом элемента B в качестве правой части. Итак, рассмотрим задачу нахождения решения следующего уравнения:

$$\bar{R}_i^L(X) = b, \quad (4)$$

где b — не начальный атом из $\mathbb{F}_{\bar{R}}^0$, \bar{R}_i — произвольный слой в \bar{R} .

Покажем, что эта задача эквивалентна задаче на ориентированном графе перечисления всех входных вершин, из которых достижима данная вершина. Вначале построим такой граф $G_{\bar{R}_i}$. Каждому атому из множества $\mathbb{F}_{\bar{R}}^0$, т. е. участвующему в отношении \bar{R}_i , поставим в соответствие вершину гра-

фа. Далее для каждой пары (X, z) рассматриваемого слоя \bar{R}_i построим дуги, ведущие из всех вершин, соответствующих атомам элемента X , в вершину, соответствующую правой части пары z . В полученном графе $G_{\bar{R}_i}$ выберем вершину b , соответствующую правой части уравнения (4). Рассмотрим подграф $G_{\bar{R}_i, b} \subseteq G_{\bar{R}_i}$, содержащий все вершины, из которых достижима вершина b (включая саму b), и все дуги, соединяющие такие вершины.

Путем исследования графа $G_{\bar{R}_i, b}$ нетрудно получить следующую теорему, доказательство которой не приводится в целях сокращения изложения.

Теорема 3.3. Уравнение (4) имеет не более одного решения. Если граф $G_{\bar{R}_i, b}$ не содержит ориентированных циклов, то единственное решение уравнения состоит из всех атомов, соответствующих входным вершинам графа. Если $G_{\bar{R}_i, b}$ содержит ориентированный цикл, решений нет. \square

Данная теорема завершает обоснование пошагового процесса решения исходного логического уравнения (1).

Список литературы

1. **Воробьев С. Г.** Условные системы подстановок термов и их применение в проблемно-ориентированной верификации программ: автореф. дис. ... к. ф.-м. н. Новосибирск: ВЦ СО АН СССР, 1987.
2. **Hsiang J.** Refutational theorem proving using term-rewriting systems // *Artif. Intell.* 1985. V. 25. P. 255–300.
3. **Buchberger B., Loos R.** Algebraic Simplification // *Computer Algebra — Symbolic and Algebraic Computation* / eds. B. Buchberger, G. E. Collins. Vienna, New York: Springer-Verlag, 1982. P. 11–43.
4. **Klop J. W.** Term rewriting systems // *Handbook of Logic in Computer Science* (Vol. 2): Background: Computational Structures / eds S. Abramsky, D. M. Gabbay and S. E. Maibaum // *Osborne Handbooks of Logic In Computer Science*. New York: Oxford University Press. 1992. Vol. 2. P. 1–116.
5. **Dershowitz N., Okada M., Sivakumar G.** Canonical Conditional Rewrite Systems // *Proc. of the 9th international Conference on Automated Deduction* (May 23–26, 1988) / eds E. L. Lusk, R. A. Overbeek // *Lecture Notes In Computer Science*. London : Springer-Verlag. 1988. Vol. 310. P. 538–549.
6. **Махортов С. Д.** Основанный на решетках подход к исследованию и оптимизации множества правил условной системы переписывания термов // *Интеллектуальные системы*. 2009. Т. 13. Вып. 1–4. С. 51–68.
7. **Баранов Д. В.** Алгебраическая интерпретация условных систем переписывания на основе LP-структур // *Вестник ВГУ. Серия Системный анализ и информационные технологии*. Воронеж. 2010. № 2. С. 131–138.
8. **Баранов Д. В.** Эквациональные LP-структуры представления знаний и связанные с ними логические уравнения // Третья Всероссийская конференция с международным участием "Знания — Онтологии — Теории" (ЗОНТ-2011). Новосибирск, 3–5 октября 2011 г.: Материалы конференции. Новосибирск: Институт математики им. С.Л. Соболева СО РАН, 2011. Т. 1. С. 24–29.
9. **Махортов С. Д.** Логические уравнения на решетках // *Вестник ВГУ. Серия Физика, математика*. 2004. № 2. С. 170–178.
10. **Aho A. V., Garey M. R., Ulman J. D.** The transitive reduction of a directed graph. *SIAM // J. Computing* 1 : 2. 1972. P. 131–137.
11. **Биркгоф Г.** Теория решеток: пер. с англ. М.: Наука, 1984. 568 с.

УДК 004.03

А. Б. Барский, д-р техн. наук, проф.,
Московский государственный университет
путей сообщения (МИИТ),
e-mail: arkbarsk@mail.ru

Двух- и трехкомпьютерная кластеризация *Grid*-ресурсов для оптимизированной сборки результатов распределенных вычислений по *SPMD*-технологии

Для схемы Grid-вычислений "раздача заданий — вычисления — сборка результатов" исследуется проблема параллельно-последовательного, "древесного" обмена ресурсных процессоров с головным на основе их элементарной кластеризации по два или по три процессора. Вычисления проводятся по SPMD-технологии.

Ключевые слова: *Grid-вычисления, SPMD-технология, обмен информацией, элементарный кластер*

Введение

Основными вычислительными ресурсами при построении *Grid*-систем являются компьютерные сети. Предполагается возможность использования как глобальной сети Интернет, так и локальных сетей, а также построенных на их основе суперкомпьютеров.

Вычисления, обусловленные *SPMD*-технологией организации счета по схеме "раздача одинаковых заданий ресурсным компьютерам (процессорам) — вычисления — сборка результатов на головном процессоре", требуют "максимального" распараллеливания работ по обмену в компьютерной сети [1–3]. Под "максимальным" на прагматическом уровне будем понимать распараллеливание, удовлетворяющее критерию минимума времени решения задачи, а точнее — минимуму времени сборки всех результатов счета.

Очевидно, что при сборке данных должны быть использованы все возможности параллельного обмена между парами процессоров. Желательно исключить элемент последовательного обмена, когда один процессор ждет, когда некоторая пара закончит обмен. Он должен, в свою очередь, стремиться участвовать в параллельном обмене.

Сказанное выше обуславливает "древесную" структуру сборки результатов вычислений на го-

ловном процессоре. Необходимость такой структуры породила теорию кластеризации процессоров для формирования вершин дерева: внутри каждого кластера нижнего уровня (яруса) результаты объединяются на процессоре — лидере. Затем лидеры каждого такого кластера объединяются в кластеры следующего яруса и т. д.

При этом оставляется открытым вопрос о порядке обмена внутри кластера. В работе [2] рассматривается проблема последовательного направленного обмена между парами процессоров так, чтобы сконцентрировать результаты на процессоре-лидере. Такой порядок обмена интерпретируется задачей о коммивояжере, последовательно обходящем промежуточные пункты — процессоры — по пути минимальной длины для концентрации результатов на лидере. При этом основной метрикой в сетевых технологиях является географическое расстояние.

Вполне логично и внутри кластера применять параллельный обмен. Но тогда, по-видимому, внутри кластера необходимо провести кластеризацию! Значит, возникает проблема построения кластеров с минимальным числом процессоров — элементарных кластеров. Именно такие кластеры, в конце концов, лежат в основе дерева сборки данных.

Так какая кластеризация целесообразна на глобальном уровне сборки данных? Каково минимальное число процессоров, объединяемых в элементарный кластер? В какой степени внутри кластера допустим последовательный обмен, т. е. ожидание одних процессоров элементарного кластера, пока закончится обмен между другими?

Кроме того, *SPMD*-технология, которая в наибольшей степени адекватна схеме решения задачи, должна самым простым образом отображать и схему обмена. *SPMD*-программа должна содержать единый, универсальный для всех процессоров модуль, предусматривающий участие каждого процессора в сборке результатов вычислений.

1. Элементарные кластеры

Правомочна постановка задачи минимизации времени T сборки данных на головном процессоре, зависящего от объема S элементарного кластера, числа k ярусов в дереве сборки результатов на головном процессоре и времени T_S сборки результатов на лидере кластера:

$$T = T_S k. \quad (1)$$

Очевидно,

$$k = \lceil \log_S n \rceil, \quad (2)$$

где n — число ресурсных процессоров.

Будем считать, что элементарные кластеры, на которые разбиваются ресурсные процессоры в соответствии с древовидной структурой, одинаковы и объединяют ровно S процессоров. Тогда время T_S следует рассматривать постоянным для данного способа кластеризации.

Не следует забывать, что кластеризация не должна приводить к большим накладным расходам на организацию и должна удовлетворять требованиям *SPMD*-технологии. В соответствии с этими требованиями работы по обмену, включая сборку данных, должны быть включены в единую *SPMD*-программу, копии которой розданы всем ресурсным процессорам. Программа инвариантна относительно размерности задачи и числа n процессоров. Однако она активно использует номер процессора в системе для указания на конкретные обрабатываемые данные и на порядок участия процессора в обмене.

С учетом (2) преобразуем выражение для времени сборки данных:

$$T = T_S \lceil \log_S n \rceil. \quad (3)$$

Задача оптимизации выбора S

$$T \rightarrow \min_S$$

не может быть решена аналитически. Выберем логически экспериментальный путь решения этой задачи.

Несложные рассуждения наводят на мысль, что оптимальное значение S равно либо двум, либо трем. Действительно, выбор любого целого основания логарифма в выражении (3), превышающего три, приводит к появлению возможностей распараллеливания обмена внутри кластера, и вновь обнаруживается целесообразность кластеризации по 2 или по 3. Ведь любое целое, большее трех, может быть представлено композицией "двойки" и "тройки". Тогда попарно или по трое процессоры могут вступать в обмен независимо, и все дерево обмена образует сложную структуру, резко усложняющую *SPMD*-программу.

2. Двухпроцессорные элементарные кластеры

Рассмотрим пример сборки результатов вычислений при решении тестовой задачи решения системы линейных уравнений методом Крамера (рис. 1). Этот тест исследовался в работах [2, 3]. Пусть размерность задачи $N = 11$. Для исключения громоздкости предположим, что по *SPMD*-технологии один процессор находит один корень, т. е. $n = 11$.

Стрелки на рисунке соответствуют направлению передачи всех уже имеющихся на процессоре корней для их накопления, т. е. их объединения с корнями процессора-адресата.

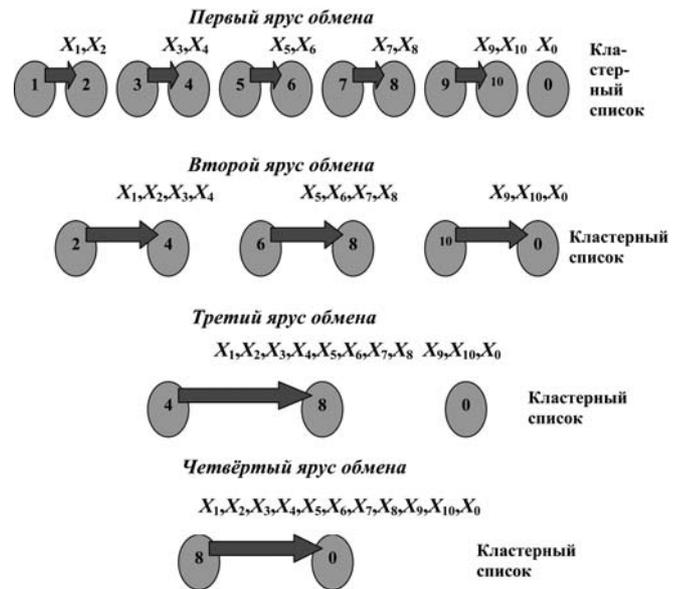


Рис. 1. Сборка результатов при двухпроцессорной кластеризации

Число ярусов сборки $k = \lceil \log_2 11 \rceil = 4$. Каждый ярус на рис. 1 отображает кластерный список — список процессоров, между которыми формируется обмен на данном уровне, т. е. образуется один кластер. "Нулевой" процессор является головным. Показано накопление корней.

Пусть $t_{\text{обм}}$ — время единичного обмена в сети, отражающее все необходимые операции согласно протоколу, маршрутизации и пр. Это время требуется для объединения данных внутри кластера, т. е. $T_{S=2} = t_{\text{обм}}$. Тогда полное время T сборки данных на головном процессоре

$$T = \lceil \log_2 11 \rceil t_{\text{обм}} = 4t_{\text{обм}}$$

Представим алгоритм, являющийся основой модуля сборки результатов *SPMD*-программы. Предполагаем, что каждый i -й процессор ($i = 0, \dots, n-1$) имеет и изменяет по ярусам обмена список (кластерный список) тех процессоров, на основе которых составляются двухпроцессорные кластеры и проводится обмен внутри кластера, которому он принадлежит. Если i -й процессор не принадлежит кластеру (ему нет "пары"), обмен на текущем ярусе будет им пропущен.

Алгоритм 1

1. Проверка: занимает ли номер i процессора нечетную позицию в кластерном списке? Если да, выполняется шаг 2; если нет — шаг 4.
2. Позиция справа от исследуемой занята каким-либо номером j другого процессора? Если да, выполняется следующий шаг, если нет — *Выход*.
3. Организуется обмен: "процессор $i \rightarrow$ процессор j ", *Выход*.

4. Процессор i настраивается на ожидание прихода данных от процессора, занимающего нечетную позицию слева.

5. Корректируется кластерный список i -го процессора исключением всех процессоров, для которых существовал ранее "правый" процессор, т. е. которым они на данном ярусе передали свои результаты. Так процессор i готовится к участию в следующем ярусе обмена. Если этот список выролдился в единственную позицию, сборка прошла успешно, формируется *Выход*. В противном случае — переход к выполнению шага 1.

3. Трехпроцессорные элементарные кластеры

Объединение в одном кластере трех процессоров основано на соглашении о том, что результаты счета собираются и передаются лидеру последовательно двумя другими процессорами, как показано на рис. 2.

Процессор-лидер становится владельцем результатов всех трех процессоров.

Данная схема обмена рассмотрена в работах [2, 3] и действительно соответствует задаче коммивояжера (без возвращения в исходный пункт) в элементарной постановке.

Однако правомочно стремление к другой, параллельной, схеме накопления результатов вычислений на лидере, при которой хотя бы подготовительные работы по обмену, включая маршрутизацию, выполняются параллельно (рис. 3). Эта схема представляется двоичным деревом и соответствует распараллеливанию операции "вектор — скаляр" методом "пирамиды" [5].

По этой схеме процессоры-нелидеры пытаются одновременно связаться с лидером. Конечно же, прием результатов им будет выполнен последовательно. Тем не менее за счет ординарности действий соответствующий модуль обмена *SPMD*-программы может быть упрощен. Хотя и в том, и в другом случае следует "честно" положить $T_{S=3} = 2t_{обм}$.

Необходимо отметить важность использования коммутаторов с очередями [4], способных сохранять кадры в буфере на то время, пока освободится канал связи с нужным процессором.

Зададимся целью не вовлекать привилегированный "нулевой" процессор в промежуточные обмены. Позволим ему участвовать только в единственной, окончательной сборке результатов. Тогда для той же тестовой задачи решения системы линейных уравнений методом Крамера и для $n = 12$ двоичное дерево сборки корней на "нулевом" головном процессоре с применением трехпроцессорных кластеров будет иметь вид, представленный на рис. 4. Здесь каждый ярус отображает и соответствующий адресный список.

На первом уровне накопительного обмена с учетом того, что кластеры составляются из процессоров с последовательными номерами, получаем

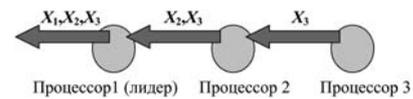


Рис. 2. Схема последовательного обмена в трехпроцессорном кластере

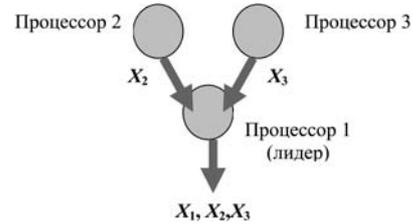


Рис. 3. "Параллельный" обмен в трехпроцессорном кластере

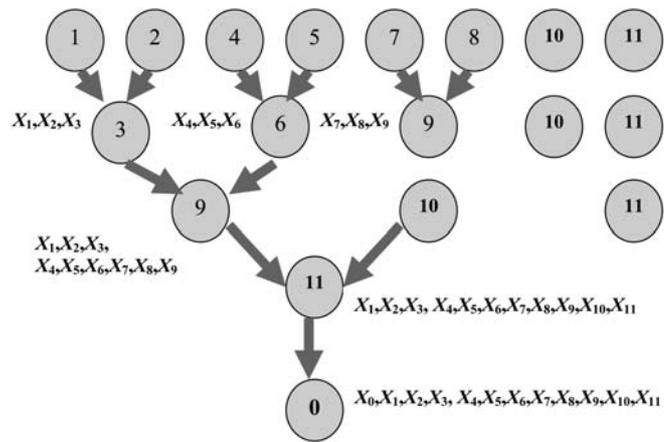


Рис. 4. Сборка данных при трехпроцессорной кластеризации без участия головного процессора

кластерный список 3, 6, 9, 10, 11. Два последних номера соответствуют процессорам, которым не хватило третьего для формирования кластера. На основе полученного списка формируется единственный кластер, внутри которого выполняется накопительный обмен.

После выполнения этого обмена кластерный список имеет вид: 9, 10, 11. На его основе формируется кластер. Далее все результаты с процессора 11 направляются на головной процессор, дополняя найденный им корень X_0 .

Поскольку головной процессор не участвовал в промежуточном обмене и потребовал дополнительного обмена $11 \rightarrow 0$, то общее время T сборки корней составило:

$$T = (2 \lceil \log_3 11 \rceil + 1) t_{обм} = 7 t_{обм}$$

Рассмотрим аналогичное дерево обмена (рис. 5) для случая равноправного непривилегированного участия головного процессора.

Число уровней сборки $k = \lceil \log_3 12 \rceil = 3$. Однако активное участие головного процессора обеспечило ускорение передачи, такое, что только на двух

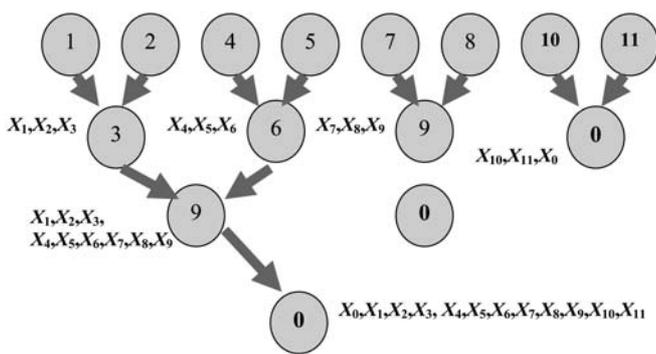


Рис. 5. Сборка данных при трехпроцессорной кластеризации с участием головного процессора

уровнях время обмена $T_3 = 2t_{обм}$. На третьем уровне оно равно $t_{обм}$. Таким образом,

$$T = (2\lceil \log_3 12 \rceil - 1)t_{обм} = 5t_{обм}.$$

Простые и регулярные действия, демонстрируемые выше, позволяют составить несложный алгоритм, являющийся основой модуля сборки результатов в составе *SPMD*-программы с участием головного процессора.

Предполагается, что каждый i -й процессор, $i = 0, \dots, n - 1$, после выполнения очередного яруса обмена располагает и изменяет (только для собственного применения) кластерный список — список тех процессоров, на основе которых составляются кластеры и проводится обмен внутри них.

Алгоритм 2

1. Проверка: занимает ли номер i в кластерном списке позицию P , кратную трем, $P_{\text{mod } 3} = 0$? Если да, выполняется следующий шаг. При отрицательном результате анализа — переход к выполнению шага 5.

2. Организуется ожидание прихода данных от процессоров, занимающих позиции $P - 2$ и $P - 1$.

3. Корректируется кластерный список i -го процессора исключением тех пар процессоров, для которых в этом списке есть третий, т. е. процессор, занимающий позицию с номером P , таким, что $P_{\text{mod } 3} = 0$. Так процессор i готовится к следующему ярусу обмена.

4. Проверка: вырожден ли список в два элемента, соответствующих позициям процессоров i и 0 ? Если да, организуется обмен $i \rightarrow 0$. Выход.

5. Если $P_{\text{mod } 3} = 1$, проверяется, есть ли в кластерном списке позиция с номером $P + 2$? Если такая позиция занята номером j процессора, организуется обмен $i \rightarrow j$. Выход.

6. Если $P_{\text{mod } 3} = 2$, проверяется, есть ли в кластерном списке позиция с номером $P + 1$? Если такая позиция занята номером l процессора, организуется обмен $i \rightarrow l$. Выход.

Заключение

На рассмотренных маленьких примерах двухпроцессорные кластеры, требующие единственного обмена внутри себя, показывают лучшие результаты по сравнению с трехпроцессорными кластерами, требующими двойного последовательного обмена.

Однако не следует забывать, что с ростом числа n процессоров $\lceil \log_2 n \rceil$ растет быстрее, чем $\lceil \log_3 n \rceil$. Следовательно, при больших значениях n даже двойной обмен внутри кластера может оказаться выгоднее, так как он компенсируется меньшим числом ярусов полной сборки данных.

В то же время одновременный множественный обмен по одному адресу должен совершенствоваться на основе буферизации коммутаторов так, чтобы только самая завершающая и быстрая часть его выполнялась последовательно. Это послужило бы целесообразности применения не только трехпроцессорных кластеров, но и кластеров, объединяющих большее число процессоров.

В целом необходимо признать, что в сетевых технологиях проблема множественного обмена пока решена недостаточно. Проработана быстрая широкополосная и множественная передача по всем или многим адресам. Сборка данных, когда многие абоненты передают одному адресату, практически не поддерживается ни алгоритмически, ни аппаратно. Но именно ее быстроедействие обусловлено широко распространенной схемой вычислений по *SPMD*-технологии уже сегодня. Тем более данное направление исследований важно для организации эффективных *Grid*-вычислений.

Список литературы

1. Барский А. Б. Оптимизационные задачи в основе пакета параллельных прикладных программ и системы информационного обслуживания Центра *GRID*-технологий // Информационные технологии. Приложение. 2010. № 10. 32 с.
2. Амиршахи Бита. Кластеризация *GRID*-ресурсов для оптимизации информационного обмена при совместной обработке результатов распределенных вычислений // Информационные технологии. 2011. № 2. С. 22–28.
3. Амиршахи Бита. *GRID*-технологии решения больших систем линейных уравнений на вычислительной сети и на суперкомпьютере кластерного типа // Информационные технологии. 2011. № 6. С. 17–22.
4. Смирнова Е. В., Пролетарский А. В., Баскаков И. В., Федотов Р. А. Управление коммутируемой средой. М.: РУСАКИ, 2011. 335 с.
5. Барский А. Б. Параллельные информационные технологии. М.: ИНТУИТ; БИНОМ, Лаборатория знаний. 2007. 503 с.

УДК 519.1

Р. А. Сологуб, аспирант,
Вычислительный Центр РАН
им. А. А. Дородницына

Порождение регрессионных моделей поверхности волатильности биржевых опционов¹

Приведено решение задачи порождения модели оптимальной структуры при восстановлении нелинейной регрессии. Порождается набор конкурирующих моделей, регрессионная выборка разбивается на обучающую и тестовую, параметры моделей оцениваются на обучающей выборке. Выбирается модель, максимизирующая коэффициент детерминации на тестовой выборке. Исследуются методы порождения моделей различных классов: линейных, нейросетей и существенно-нелинейных. Предлагается единый алгоритм их порождения. В качестве иллюстрации предложенного подхода рассматривается задача восстановления поверхности волатильности биржевых опционов.

Ключевые слова: нелинейная регрессия, символьная регрессия, индуктивное порождение, сложность моделей, биржевые опционы, волатильность

Введение

Регрессионной моделью называется параметрическое семейство функций, а каждая из порождаемых моделей является суперпозицией функций из экспертно-заданного множества. Эти функции называются порождающими. Модель задана суперпозицией порождающих функций. В работе исследуются методы порождения моделей различных классов.

Использование нелинейной регрессии для решения прикладных задач широко представлено в работах Дж. Себера [1, 2]. В них описываются построение и оценка параметров нелинейных моделей. Для оценки моделей используется алгоритм Левенберга—Марквардта [3].

Для индуктивного порождения моделей в работах Дж. Козы [4] и Н. Зелинки [5], связанных с генетическим программированием, используется символьная регрессия — метод построения регрессионных моделей путем перебора различных произвольных суперпозиций функций из заданного набора.

Индуктивное порождение моделей рассматривается в приложении к задаче определения оптимальной формы антенны [6]. При этом авторы ставят ряд нерешенных вопросов: появление моделей с ненастраиваемыми параметрами, деление на ноль, возникновение комплексных аргументов. Часть этих проблем разрешима с помощью алгоритма, предложенного в данной статье. В работах В. В. Стрижова [7, 8] идеи индуктивного порождения регрессионных моделей находят свое развитие в применении методов байесовского вывода к процессу порождения и оценки параметров моделей.

В данной работе развитие идей индуктивного порождения моделей заключается в создании процедуры порождения моделей вышеназванных классов с помощью единообразного подхода к их записи и порождению. В работе ставится задача поиска модели и набора параметров, минимизирующих сумму квадратов невязок, доставляемых построенной моделью на тестовой выборке. Для выбора оптимальной модели, принадлежащей к определенному классу, необходимо осуществлять ее поиск среди всего множества моделей, принадлежащих к данному классу. В связи с этим требуется показать возможность построения всех моделей этого класса.

Для недопущения эффекта переобученности сложность моделей должна ограничиваться. Сложность моделей в рамках данной работы оценивается по методу, предложенному К. Владиславлевой [9]. Для модели, представленной в виде дерева, ее сложность определяется числом элементов во всех поддеревьях данного дерева.

В качестве иллюстрации предложенного подхода рассматривается задача поиска формулы поверхности волатильности [10] биржевых опционов. Данная задача является актуальной, так как позволяет уточнить оценку опционов дальних цен и времен исполнения. Подобные опционы являются одним из основных инструментов страхования биржевых рисков для институциональных инвесторов, а их справедливая оценка необходима для обеспечения минимальной разницы между спросом и предложением на рынке ценных бумаг.

1. Задача многомерной нелинейной регрессии

Задана выборка — множество пар $D = \{(x_n, y_n)\}_{n=1}^N$, в котором $x \in \mathbf{R}^P$ — свободная переменная и $y \in \mathbf{R}^1$ — зависимая переменная. Строится отображение

¹ Работа поддержана РФФИ, проект 10-07-00422.

$\varphi(\mathbf{x}, \mathbf{w}) \rightarrow \mathbf{R}^1$. Требуется определить модель f — отображение из декартова произведения множества свободных переменных $\mathbf{x} \in \mathbf{R}^n$ и множества параметров $\mathbf{w} \in \mathbf{R}^m$ в \mathbf{R}^1 . Для модели определяется набор параметров \mathbf{w}_0 , доставляющих минимум функции квадратичной ошибки:

$$S(\mathbf{w} | D, f) = \sum_{n=1}^N (f(\mathbf{x}_n, \mathbf{w}) - y_n)^2.$$

Выражение $S(\mathbf{w} | D, f)$ означает значение S , соответствующее минимизирующему эту ошибку набору параметров \mathbf{w} при заданной выборке D и модели f . Такая модель будет называться оптимальной при условии, что ее сложность $S(f)$ не превышает заданной. Сложность определяется как число элементов во всех поддеревьях, которые можно выделить из дерева, представляющего модель.

Задано множество G порождающих функций $g(\mathbf{w}, \mathbf{x})$. Для каждого элемента данного множества g_i определены области аргументов $\mathbf{w} \in \mathbf{R}^m$, $\mathbf{x} \in \mathbf{R}^n$ и значений, при этом область значений принадлежит \mathbf{R}^1 . В множество порождающих функций обязательно входит не имеющая аргументов функция $id(\mathbf{x})$, значение которой тождественно значению свободной переменной. Порождается множество моделей $f \in F$ — допустимых суперпозиций, состоящих из функций $g_i \in G$. Требуется выбрать модель, доставляющую минимум $S(f | \mathbf{w}^*, D)$ при условии, накладываемом на сложность $C(f) < C^*$.

Для описания процедуры порождения моделей необходим язык описания моделей, легко интерпретируемый как пользователем — экспертом в предметной области, так и программной системой. Каждой модели поставим в соответствие направленный граф — дерево $\Gamma = \langle V, E \rangle$. Каждой вершине $v_i \in V$ соответствует порождающая функция g_i . Число ветвей $e_i \in E$, выходящих из каждой вершины v_i , будет равно числу аргументов порождающей функции g_i , соответствующей данной вершине. Листьями дерева Γ являются порождающие функции, не имеющие аргументов, — константы $id(C)$ и свободные переменные $id(\mathbf{x})$.

2. Конструктивное порождение суперпозиций

Рассмотрим построение всех допустимых моделей для различных классов, а также задание ограничений на граф, представляющий модель, так что любой граф описанной структуры будет описывать модель, принадлежащую к заданному классу.

Линейные модели. Термин "линейная модель" в контексте данной работы означает модель $f(\mathbf{w}, \mathbf{x})$, являющуюся суммой входных переменных с настраиваемыми коэффициентами — параметрами $\mathbf{w} = [c_0, c_1, \dots, c_n]$:

$$f(\mathbf{w}, \mathbf{x}) = \mathbf{w}^T \mathbf{x} = \sum_i c_i x_i + c_0. \quad (1)$$

Для порождения множества линейных моделей требуются две порождающие функции: функция умножения на константу (c) и функция сложения (+). Также к множеству порождающих функций добавляются функции аргумента ($id(x_i)$) и константы ($id(C)$): $G = \{+, c, id(x_i), id(C)\}$. Вводятся следующие правила порождения.

1. В случае если вершине v_i дерева Γ соответствует порождающая функция $+$, родительской вершине может соответствовать только $+$.

2. В случае если вершине v_i соответствует порождающая функция c , родительской вершине должна соответствовать порождающая функция $+$.

3. В случае если вершине соответствует порождающая функция C , родительской вершине должна соответствовать порождающая функция $+$.

Алгоритм 1. Модели порождаются рекурсивно (в каждую заново создаваемую вершину v_i подставляются все возможные для нее порождающие функции $g_j \in G$), при этом порождение идет от вершины дерева v_0 , и каждый следующий элемент g_{j+1} выбирается с учетом приведенных выше правил. Его реализация в псевдокоде показана на рис. 1.

Теорема 1. Любая линейная модель вида (1) порождается при использовании данных правил порождения моделей. Любая порожденная алгоритмом 1 модель будет линейной.

Доказательство. Линейная модель имеет вид

$$y = \sum_i^n c_i x_i + c_0. \text{ Данная запись может быть пред-}$$

ставлена в польской нотации:

$$y = + \times c_1 x_1 + \times c_2 x_2 \dots + \times c_n x_n c_0.$$

При замене $\times c_i$ на c_i запись принимает вид:

$$y = + c_1 x_1 + c_2 x_2 \dots + c_n x_n c_0.$$

Данная замена эквивалентна опусканию знака умножения в традиционной записи формул.

Нотацию справа от знака равенства можно рассматривать как запись дерева при проходе в глубину. Построенное таким образом дерево будет удовлетворять всем условиям, наложенным выше, по построению.

Докажем, что любое дерево, удовлетворяющее изложенным выше условиям, будет задавать линейную модель. Структура данного дерева такова, что родительская вершина для вершин с порождающей

```

ПРОЦЕДУРА ПостроениеДерева(ТекущееДерево,Примитивы,Родитель)
для j = 1, ..., size(Примитивы)
НовыйЭлемент:=Примитивы(j);
ВременноеДерево:={ТекущееДерево;НовыйЭлемент};
если ПроверкаПравилПорождения(ВременноеДерево)==1 то
ТекущееДерево:=ВременноеДерево;
НовыеЭлементы:=Примитивы(НовыйЭлемент).КоличествоДетей;
для i = 1, ..., НовыеЭлементы
ТекущееДерево:=ПостроениеДерева(ТекущееДерево,Примитивы,НовыйЭлемент);
return ТекущееДерево
выход

```

Рис. 1. Процедура порождения всех допустимых деревьев заданного класса

функцией x_i или C может быть или функцией $+$, или c_i (правило 3). Над порождающей функцией c_i может располагаться только порождающая функция $+$ (правила 1,2). Над порождающей функцией $+$ располагаются только функции $+$ (правило 1), поэтому можно заменить все порождающие функции $+$ суммой n элементов. В этой сумме будут элементы типа $c_i * x_j$, x_j или C . Таким образом, собрав элементы по свободным переменным, получим сумму свободных переменных с коэффициентами, которые состоят из сумм настраиваемых переменных. Это возможно потому, что любая порождающая функция c_i при объединении попадет только внутрь одной скобки, так как у порождающей функции c_i может быть только один потомок. Данная структура полностью соответствует описанию линейной модели.

Обобщенно-линейные модели. Класс обобщенно-линейных моделей расширяет класс линейных моделей включением функций связи:

$$f(\mathbf{w}, \mathbf{x}) = \sum_i^n c_i \mu_i(x_i) + c_0. \quad (2)$$

Для порождения множества обобщенно-линейных моделей требуются две порождающие функции класса линейных моделей (c , $+$), функции связи μ_i и порождающая функция аргумента ($id(x_i)$). Накладываются следующие правила порождения.

1. В случае если вершине v_i соответствует порождающая функция $+$, родительской вершине v_j может соответствовать только $+$.

2. В случае если вершине v_i соответствует порождающая функция c , родительской вершине v_j должна соответствовать порождающая функция $+$.

3. В случае если вершине v_i соответствует порождающая функция связи, родительской вершине v_j должна соответствовать порождающая функция c .

4. В случае если вершине v_i соответствует порождающая функция C , родительской вершине v_j должна соответствовать порождающая функция $+$.

Модели порождаются рекурсивно (в каждую заново создаваемую вершину v_i подставляются все возможные для нее порождающие функции $g_j \in G$), при этом порождение идет от вершины дерева v_0 , и каждый следующий элемент g_{j+1} выбирается с учетом приведенных выше правил. Алгоритм построения моделей аналогичен алгоритму 1.

Теорема 2. Любая обобщенно-линейная модель (2) может быть порождена при использовании данных правил порождения моделей.

Доказательство. Доказательство теоремы 2 аналогично доказательству теоремы 1. Обобщенно-линейная модель имеет вид $y = \sum_i^n c_i \mu_i(x_i) + c_0$. Данная запись может быть представлена в польской нотации:

$$f(\mathbf{w}, \mathbf{x}) = +c_1 \mu_1(x_1) + c_2 \mu_2(x_2) \dots + c_n \mu_n(x_n) c_0.$$

При переобозначении $\times c_i$ на c_i запись принимает вид:

$$f(\mathbf{w}, \mathbf{x}) = +c_1 \mu_1(x_1) + c_2 \mu_2(x_2) \dots + c_n \mu_n(x_n) c_0.$$

Нотацию справа от знака равенства можно рассматривать как запись дерева при проходе в глубину (при этом следует рассматривать $\mu_i(x_i)$ как пару вершин). Построенное таким образом дерево будет удовлетворять всем условиям, наложенным выше, по построению.

Радиальные базисные функции. Класс моделей, построенных на радиально-базисных функциях,

$$f(x) = \mathbf{w}^T \Phi \left(\frac{x^2}{\sigma^2} \right) = \sum_{i=1}^U w_i r(x_i), \quad (3)$$

является подклассом класса нейронных сетей. Однако в силу его простого устройства он также является подклассом класса обобщенно-линейных моделей, в связи с чем доказательства возможности построения всех моделей данного класса не требуется.

Для порождения множества моделей, построенных на радиально-базисных функциях, требуются две порождающие функции: функция суммы ($+$) и радиально базисные функции (r_i). Также, аналогично классу линейных моделей, к множеству порождающих функций добавляются функции аргумента (x_i) и константы (C). Накладываются следующие правила порождения.

1. Вершине дерева v_0 ставится в соответствие порождающая функция $+$.

2. В случае если вершине v_i соответствует порождающая функция $+$, родительской вершине v_j может соответствовать только $+$.

3. В случае если вершине v_i соответствует порождающая функция радиальной базисной функции r_i , родительской вершине v_j должна соответствовать порождающая функция $+$.

Модели порождаются рекурсивно (в каждую заново создаваемую вершину v_i подставляются все возможные для нее порождающие функции $g_j \in G$), при этом порождение идет от вершины дерева v_0 , и каждый следующий элемент g_{j+1} выбирается с учетом приведенных выше правил. Алгоритм построения моделей аналогичен алгоритму 1.

Теорема 3. Любая модель, построенная на радиально базисных-функциях (3), может быть порождена при использовании данных правил порождения моделей.

Доказательство теоремы 3 полностью аналогично доказательству теоремы 2.

Нейронные сети. Основное отличие обобщенных нейронных сетей от класса радиальных базисных функций состоит в существовании в нейронных сетях общего вида скрытых слоев нейронов. Фор-

мула нейронной сети может быть записана следующим образом:

$$f(\mathbf{w}, \mathbf{x}) = \mathbf{v} \sum_{i=1}^N u(S_i(\mathbf{x}, \mathbf{w})),$$

где $S_i(\mathbf{x}, \mathbf{w}) = \sum_j w_j u(A_j(\mathbf{x}, \mathbf{w}));$ (4)

$$A_j(\mathbf{x}, \mathbf{w}) = \mathbf{x}^T \mathbf{w}_j.$$

Данная модель не может быть представлена деревом, так как у одного элемента A_j может быть несколько родителей. Однако эта проблема может быть обойдена с помощью добавления правила построения, согласно которому всем вершинам, соответствующим одному A_j , будут присвоены одни и те же веса.

Модели порождаются рекурсивно (в каждую заново создаваемую вершину v_i подставляются все возможные для нее порождающие функции $g_j \in G$), при этом порождение идет от вершины дерева v_0 , и каждый следующий элемент g_{j+1} выбирается с учетом приведенных выше правил. Алгоритм построения моделей аналогичен алгоритму 1, однако добавляется нестандартное правило, что поддерева, являющиеся потомками различных S -вершин, полностью идентичны.

Теорема 4. Любая модель, построенная на радиально базисных-функциях (3), может быть порождена при использовании данных правил порождения моделей.

Доказательство теоремы 4 аналогично доказательству теоремы 2.

Существенно-нелинейные модели. Класс существенно-нелинейных моделей строится как множество моделей, которые могут быть записаны формулой, состоящей из заранее известных элементов. Данные элементы следует отнести к множеству порождающих функций и переписать формулу в префиксном виде. Таким образом, оказывается, что любая существенно-нелинейная модель может быть построена в виде дерева и что любое дерево описывает существенно-нелинейную модель.

Множество всех существенно-нелинейных моделей, являющихся суперпозициями некоторого набора порождающих функций G , может быть построено с помощью алгоритма 1, при этом множество правил будет пустым.

3. Задача восстановления поверхности волатильности

Для иллюстрации алгоритма порождения моделей рассматривается задача восстановления регрессии поверхности волатильности. Для решения задачи порождались линейные, обобщенно-линейные, существенно-нелинейные модели, нейронные сети. Сложность моделей ограничивалась числом 80. Полученные модели при этом сравнивались с создан-

ными ранее [19] для решения схожей задачи. Для улучшения работы алгоритма для каждого класса моделей использовались следующие спецификации.

1. Для класса линейных моделей было запрещено использование в качестве элемента суперпозиции одной входной переменной более одного раза для получения корректной оценки параметров моделей.

2. Для нейронных сетей количество S -вершин было ограничено числом 10, использовалась настройка нейронной сети с помощью метода обратного распространения ошибки.

3. Для обобщенно-линейных моделей использовались полиномиальные функции, функции $\frac{1}{x}$, $\frac{1}{\sqrt{x}}$,

e^x и $\ln x$ как наиболее часто встречающиеся в работах, посвященных финансовой математике.

4. Тот же набор функций использовался для существенно нелинейных моделей, при этом для упрощения алгоритма поиска модель имела вид суммы произведения двух других моделей и константы.

Порождаемые модели настраивались с помощью алгоритма Левенберга—Марквардта, после чего для каждого класса моделей была выбрана модель с наилучшим значением MSE — средней квадратичной ошибки.

Поверхность волатильности. Чтобы дать определение подразумеваемой волатильности, сначала необходимо привести описание некоторых терминов из области финансовой математики. Опционом европейского типа называется производная ценная бумага (контракт), дающая ее обладателю право купить или продать актив по указанной в опционе цене (цене исполнения опциона) в указанное в опционе время (момент исполнения опциона). Опцион, дающий право купить активы, называется опционом колл. Опцион, дающий право продать активы, называется опционом пут. Таким образом, выигрыш владельца европейского опциона равен

$$d = \max(P(T) - K; 0)$$

где T — момент исполнения опциона; $P(T)$ — цена базового актива; K — цена исполнения опциона.

Сами опционы, будучи ценными бумагами, также являются объектами торговли. Согласно известной формуле Блэка—Шоулза [12], текущая справедливая цена $c = c(S)$ европейского опциона колл в момент времени t_0 имеет вид

$$c = P(t_0)\Phi(d_1(\sigma)) - K \exp(-r(T - t_0))\Phi(d_2(\sigma)),$$

$$d_1(\sigma) = \frac{\log P(t_0) - \log K + (T - t_0)(r + \sigma^2/2)}{\sigma \sqrt{T - t_0}};$$

$$d_2(\sigma) = d_1(\sigma) - \sigma \sqrt{T - t_0},$$

где $\Phi(\cdot)$ стандартная нормальная функция распределения; r — безрисковая ставка доходности; σ — волатильность, т. е. стандартное отклонение цены базового актива за календарный год.

Подразумеваемой волатильностью σ_{implied} принято называть такое значение волатильности исходных активов, при котором теоретически справедливая цена $c = c(S)$ европейского опциона, вычисленная по формуле Блэка—Шоулза, совпадает с его рыночной ценой c_m . Другими словами, σ_{implied} — это решение уравнения Блэка—Шоулза относительно σ [12].

В связи с понятием неявной волатильности необходимо упомянуть два термина: улыбка волатильности и поверхность волатильности. Улыбка волатильности получается при построении графика неявной волатильности как функции относительной цены исполнения $M = P/K$ при фиксированном времени исполнения опциона T . При этом функция $\sigma_{\text{implied}} = \sigma_{\text{implied}}(M, T)$ оказывается выпуклой вниз, а ее график напоминает улыбку.

Поверхность волатильности получается при построении графика функции $\sigma_{\text{implied}} = \sigma_{\text{implied}}(M, T)$ как функции двух переменных M и T .

В данной работе рассматривается прикладная задача восстановления зависимости значения σ_{implied} от значений S/K , T . Для этого проводится построение моделей различных классов и отбор среди данных моделей лучших по критериям средней квадратичной ошибки (MSE), коэффициента детерминации и информационного критерия Акаике (AIC). Полученная модель сравнивается с рассмотренными ранее [19].

4. Вычислительный эксперимент

Для анализа выбраны данные торгов опционами колл на *Brent Crude Oil* — опционы *Chicago Mercantile Exchange* на сырую нефть марки *Brent*. Фактически использованы все ликвидные по состоянию на 11 ноября 2011 года опционы — с датами исполнения от 15 ноября 2011 года до 15 декабря 2018 года. Регрессионная выборка $\{(x_n, y_n)\} = \{(M_m, t_n, \sigma_n)\}$ была построена на основе данных системы *Bloomberg* и дополнена с помощью исходных данных — исторических цен опциона $S_{K, t}$ и базового инструмента P_p , где $K \in \mathbf{K}$, $t \in T$, следующим образом. Для каждого желаемого значения M и $t \in T$ вычисляется значение предполагаемой волатильности как аргумент минимума

$$\sigma_{K, t}^{\text{imp}} = \arg \min_{\sigma \in [0, 1.5]} (C_{K, t} - C(\sigma, P_p, B, K, t)),$$

по значению волатильности. Здесь справедливая цена опциона C вычислена по формуле Блэка—Шоулза. Время t , выраженное в годах до момента исполнения опциона, рассчитывается по формуле $t = \tau/365$, где τ — число дней, оставшихся до исполнения опциона. Для индексации выборки задана биекция $(t, M) \mapsto n$. Безрисковая ставка доходности $B = 0,025$, что соответствует ставке доходности по облигациям казначейства США.

Результаты вычислительного эксперимента

Класс моделей	Число параметров	$C(f)$	MSE_{learn}	MSE_{test}	$R_{\text{adj}}^2, \%$	AIC
Линейная	3	19	46,98	51,53	63	192,09
Нейронная сеть	10	81	20,43	25,21	89	178,45
Обобщенно-линейная	6	48	27,06	30,11	78	133,43
Нелинейная	4	50	11,28	13,76	90	69,27
Экспертная модель	5	66	27,78	30,85	77	37,95

Результаты вычислительного эксперимента. Результаты вычислительного эксперимента и графики, отображающие наилучшие модели, представлены в таблице. Из этой таблицы можно видеть, что линейная модель выглядит недостаточно хорошо приближающей данные (плохие значения AIC, MSE) и дает низкий коэффициент детерминации R_{adj}^2 . Следует заметить, что при более низкой средней квадратичной ошибке нейронная сеть по параметру AIC оказывается хуже обобщенно-линейной модели. Наилучшим образом показывает себя нелинейная модель. При этом число настраиваемых параметров в ней меньше, чем в обобщенно-линейной модели. Модели, полученные в ходе вычислительного эксперимента, оказываются лучше моделей, которые были предложены экспертами ранее [19], — нелинейные модели дают лучшую оценку волатильности при меньшем числе параметров, чем полиномиальные модели значительно большей сложности.

На графиках, представленных на рис. 2 (см. вторую сторону обложки), справа по горизонтальным осям отложены значения относительной цены исполнения M и времени до исполнения опциона t в годах. По вертикальным осям отложены предполагаемые волатильности σ_{implied} , соответствующие реально торгуемому опциону с параметрами M и t . Соответствие рисунков моделям:

- рис. 2, а — линейная модель: $\sigma_{\text{imp}} = c_1 M + c_2 T + C$;
- рис. 2, б — нейронная сеть (в виде дерева изображена нейронная сеть меньшего размера в качестве примера): $\sigma_{\text{imp}} = \sum_{i=1}^{10} S_i(\Sigma_j A_j(M, T))$;
- рис. 2, в — обобщенно-линейная модель: $\sigma_{\text{imp}} = c_1 M + c_2 M^2 + c_3(\sqrt{T})^{-1} + C$;
- рис. 2, г — экспертная модель: $\sigma_{\text{imp}} = c_1 M + c_2 M^2 + c_3 T + c_4 MT + C$;
- рис. 2, д — существенно-нелинейная модель: $\sigma_{\text{imp}} = \frac{c_1 M + c_2 M + C_1}{\sqrt{T}} + C_2$.

На рис. 2, слева, изображены деревья Γ_p , отображающие данные модели при представлении моделей в виде деревьев.

Заключение

В работе предложен алгоритм описания и конструктивного порождения регрессионных моделей. Описан способ представления моделей в виде суперпозиций заданных параметрических функций и способ порождения всех моделей заданных классов с помощью алгоритма. Для иллюстрации работы алгоритма рассмотрена модель зависимости волатильности биржевого опциона от цены исполнения и времени до исполнения. Модели, полученные в ходе вычислительного эксперимента, оказываются предпочтительнее моделей, которые были предложены экспертами ранее: нелинейные модели дают лучшую оценку волатильности при меньшем числе настраиваемых параметров, чем полиномиальные модели значительно большей сложности.

Список литературы

1. Seber G. A. F., Wild C. J. Nonlinear Regression. Wiley—IEEE, 2003.
2. Seber G. A. F., Schwarz C. J. Estimating Animal Abundance: Review III. Stat Sci. 1999. Vol. 14, N 4. P. 427—456.
3. Levenberg K. A Method for the Solution of Certain Non-Linear Problems in Least Squares // The Quarterly of Applied Mathematics. 1944. Vol. 2. P. 164—168.
4. Koza J. R., Keane M. A., Rice J. P. Performance improvement of machine learning via automatic discovery of facilitating functions as applied to a problem of symbolic system identification // IEEE International Conference on Neural Networks I. 1993. San Francisco, USA. 1993. P. 191—198.
5. Zelinka I., Nolle L., Oplatkova Z. Analytic Programming — Symbolic Regression by Means of Arbitrary Evolutionary Algorithms // Journal of Simulation. 2012. Vol. 6, N 9. P. 44—56.
6. Comisky W., Yu J., Koza J. R. Automatic synthesis of a wire antenna using genetic programming // Late Breaking Papers at the 2000 Genetic and Evolutionary Computation Conference, Las Vegas, Nevada. 2000. P. 179—186.
7. Стрижов В. В. Поиск параметрической регрессионной модели в индуктивно заданном множестве // Журнал вычислительных технологий. 2007. N 1. С. 93—102.
8. Стрижов В. В., Сологуб П. А. Индуктивное построение регрессионных моделей волатильности опционов // Журнал вычислительных технологий. 2009. N 5. С. 102—113.
9. Vladislavleva E., Smith G., Hertog D. Order of nonlinearity as a complexity measure for models generated by symbolic regression via pareto genetic programming // IEEE Transactions on Evolutionary Computation. 2009. Vol. 13 (2). P. 333—349.
10. Dalglish T., Hull J., Suo W. Volatility Surfaces: Theory, Rules of Thumb, and Empirical Evidence // Quantitative Finance. 2007. Vol. 7, No. 5. P. 507—524.
11. Jackwerth J., Rubenstein M. Recovering Probability Distributions from Option Prices // Journal of finance. December 1996. Vol. 51, No. 5.
12. Alentorn A. Modelling the implied volatility surface: an empirical study for FTSE options. May 2004. URL: http://www.amadeo.name/working_papers/volatility_surface_may04.pdf

УДК 519.865.5

Д. Р. Галиев, студент,
e-mail: damir.galiev@mail.ru,
А. Г. Исавнин, д-р физ.-мат. наук, проф.,
e-mail: isavnin@mail.ru,
Филиал Казанского (Приволжского)
федерального университета,
г. Набережные Челны

Модели выбора оптимального инвестиционного портфеля с нечетко-множественными доходностями

Рассмотрена проблема выбора оптимального инвестиционного портфеля с нечетко-множественными входными параметрами. Подобный подход позволяет учитывать широкий спектр экспертных оценок. Представлены две модели выбора оптимального портфеля с нечетко-множественными входными параметрами, на базе которых предложена новая модель, учитывающая несколько мер риска. Приведены результаты численных экспериментов по данным российского фондового рынка. В рамках исследования разработано программное обеспечение, в котором реализованы описанные в статье методы и модели.

Ключевые слова: оптимальный портфель, Value-at-Risk, Drawdown-at-risk, нечеткие множества, программирование

Введение

В 1952 г. Гарри Марковиц опубликовал свою пионерскую статью [1], которая послужила толчком для развития современной портфельной теории. Mean-variance-модель Марковица явилась основой для развития современной теории портфельного инвестирования в последующие десятилетия. Однако несмотря на свою популярность и математическую обоснованность, mean-variance-подход стал и предметом обширной критики. Например, доходности активов необязательно должны быть распределены "нормально", распределение доходности портфеля может обладать "тяжелыми хвостами" или быть асимметричным. Кроме математического аппарата критике также подвергается теория "эффективного рынка". Альтернативные подходы позволяют более корректно учитывать реальное положение вещей. Например, можно использовать такие меры риска, как Value-at-Risk (VaR), ожидаемая просадка (Drawdown-at-Risk), среднее абсолютное отклонение (Mean Absolute Deviation), полудисперсия (Semi-variance) и т. д. [2].

В конце прошлого века становится популярной теория нечетких множеств, основателями которой являются Беллман и Заде [3]. Все большее число исследователей начинают применять новую теорию к широкому спектру задач, в том числе и к задачам

финансового риск-менеджмента. В случае задачи выбора оптимального инвестиционного портфеля теория нечетких множеств позволяет учитывать знания и опыт инвесторов. Впервые теорию нечетких множеств к проблеме выбора оптимального портфеля облигаций применил Рамасвами [4]. Аналогичный подход был предложен Леоном и др. [5]. Остермаск [6] и Ватада [7] предложили свои модели, которые основывались на методе принятия решений с использованием нечетких чисел. В этих моделях риск и доходность описываются с помощью логистической функции принадлежности. Карлсон и др. [8] исследовали модель выбора оптимального портфеля, в которой в качестве входных данных использовалась информация о возможном распределении доходности активов.

Некоторые модели формирования оптимального портфеля

Рассмотрим некоторые модели формирования оптимального портфеля. Вопрос о выборе меры риска является ключевым при построении инвестиционного портфеля. Одним из недостатков стандартной меры риска — среднего квадратического отклонения (СКО) — является то, что эта мера отражает изменение доходности в обе стороны: как в положительную, так и в отрицательную. Инвесторов же, как правило, интересует минимизация риска снижения доходности (*Drawdown-at-risk*, *Drawdown*). Для учета этого вида риска можно использовать меру среднего полуабсолютного отклонения (*mean semi-absolute deviation* [9]):

$$MSAD(R) = M(|\min\{0, R - M(R)\}|), \quad (1)$$

где R — доходность портфеля; $M(x)$ — математическое ожидание от x .

Также довольно популярной мерой, отражающей риск изменения доходности портфеля ниже заданного уровня, является величина *Value-at-Risk* (*VaR*) [10]:

$$P(R < VaR_\beta) \leq \beta. \quad (2)$$

Другими словами, эта мера риска характеризует то, что доходность портфеля не опустится ниже величины VaR_β с вероятностью $p \geq 1 - \beta$.

В классическом подходе для вычисления доходности используются оценки средних доходностей $M(R)$. При этом предполагается, что исторические средние сохраняют свое значение в будущем. Практика показывает, что подобная ситуация встречается редко [11]. По этой причине предпочтительным является использование экспертных оценок относительно будущей доходности или же результатов прогноза по нейросетевым или эконометрическим моделям. В рассматриваемых в статье моделях мы также будем учитывать транзакционные издержки. Авторы работы [12] показывают в своих исследованиях, что пренебрежение транзакцион-

ными издержками может привести к формированию неэффективного портфеля. К подобным результатам, проведя эмпирические исследования, пришел и Йошимото [13].

Пусть имеются n активов, из которых инвестор может сформировать портфель. Капитал распределяется между активами долями x_i , $0 \leq x_i \leq 1$. Рассмотрим несколько моделей, в которых будут использоваться описанные меры риска.

Пусть инвестор в качестве базовой меры риска выбрал *VaR*. Тогда модель будет выглядеть следующим образом:

$$\begin{aligned} \max_x & \left(M(R) - \sum_{i=1}^n c_i x_i \right); \\ P(R < VaR_\beta) & \leq \beta; \\ \sum_{i=1}^n x_i & = 1; \\ l_i \leq x_i \leq u_i, & \quad i = \overline{1, n}, \end{aligned} \quad (3)$$

где l_i и u_i отражают дополнительные ограничения в портфеле на актив i ; c_i — транзакционные издержки по активу i .

Если же в качестве базовой меры риска выбрано среднее полуабсолютное отклонение, то модель выбора оптимального портфеля может быть сформирована следующим способом:

$$\begin{aligned} \min_x & M(|\min\{0, R - M(R)\}|); \\ M(R) - \sum_{i=1}^n c_i x_i & \geq R_0; \\ \sum_{i=1}^n x_i & = 1; \\ l_i \leq x_i \leq u_i, & \quad i = \overline{1, n}, \end{aligned} \quad (4)$$

где R_0 задается пользователем модели и означает нижний предельный уровень доходности портфеля.

Аналогичным образом сформируем модель, в которой учитывались бы несколько мер риска:

$$\begin{aligned} \min_x & M(|\min\{0, R - M(R)\}|); \\ P(R < VaR_\beta) & \leq \beta; \\ M(R) - \sum_{i=1}^n c_i x_i & \geq R_0; \\ \sum_{i=1}^n x_i & = 1; \\ l_i \leq x_i \leq u_i, & \quad i = \overline{1, n}. \end{aligned} \quad (5)$$

Таким образом, в модели (5) минимизируется просадка портфеля *MSAD* и в то же время учитываются потери по *VaR*.

Нечетко-множественная постановка задачи

Мы уже отметили, что вместо использования исторических средних доходностей предпочтительным является использование экспертных оценок относительно будущей доходности или сторонних моделей прогнозирования. Часто можно наблюдать, что мнения экспертов относительно будущей доходности активов расходятся, образуя целый интервал возможных значений. Именно эту ситуацию можно рассматривать в качестве предпосылки для использования нечетко-множественных входных данных, с помощью которых можно учесть предельно негативное, позитивное и наиболее вероятное развитие событий.

Рассмотрим некоторые базовые понятия, которые необходимы для нечетко-множественной постановки задач. Пусть доходность актива выражается трапециевидным нечетким числом $\tilde{r} = (r_1, r_2, r_3, r_4)$, где $r_1 < r_2 \leq r_3 < r_4$. Функция принадлежности нечеткого числа \tilde{r} определяется следующим образом:

$$\mu(x) = \begin{cases} \frac{x-r_1}{r_2-r_1}, & r_1 \leq x \leq r_2; \\ 1, & r_2 \leq x \leq r_3; \\ \frac{x-r_4}{r_3-r_4}, & r_3 \leq x \leq r_4; \\ 0, & \text{в других случаях.} \end{cases} \quad (6)$$

Графически $\tilde{r} = (r_1, r_2, r_3, r_4)$ можно представить в виде рис. 1.

Теперь рассмотрим два нечетких трапециевидных числа $\tilde{r} = (r_1, r_2, r_3, r_4)$ и $\tilde{b} = (b_1, b_2, b_3, b_4)$. Согласно [14] имеем:

1. Сумма (разность) двух трапециевидных чисел также является трапециевидным числом: $\tilde{r} \pm \tilde{b} = (r_1 \pm b_1, r_2 \pm b_2, r_3 \pm b_3, r_4 \pm b_4)$.

2. Произведение скалярного числа и нечеткого трапециевидного числа также является трапециевидным числом: $k\tilde{r} = \begin{cases} (kr_1, kr_2, kr_3, kr_4), & k \geq 0; \\ (kr_4, kr_3, kr_2, kr_1), & k < 0. \end{cases}$

3. Вероятность того, что нечеткое трапециевидное число \tilde{r} меньше либо равно нечеткому трапециевидному числу \tilde{b} , определяется следующим образом:

$$P(\tilde{r} \leq \tilde{b}) = \begin{cases} 1, & r_2 \leq b_3; \\ \frac{b_4 - r_1}{b_4 - b_3 + r_2 - r_1}, & r_2 \geq b_3, r_1 \leq b_4; \\ 0, & r_1 \leq b_4. \end{cases}$$

Лемма [15]. Пусть $\tilde{r} = (r_1, r_2, r_3, r_4)$ — трапециевидное нечеткое число, тогда для любого $\beta \in [0; 1]$ неравенство $P(\tilde{r} \leq 0) \leq \beta$ справедливо тогда и только тогда, когда $(1 - \beta)r_1 + \beta r_2 \geq 0$. Эта Лемма необходима для формирования VaR-ограничений в рассматриваемых моделях. Величину требуемого уровня VaR обозначим посредством нечеткого трапециевидного числа $\tilde{b} = (b_1, b_2, b_3, b_4)$. Тогда, со-

гласно Лемме, ограничение (2) эквивалентно следующему условию:

$$(1 - \beta) \left(\sum_{i=1}^n r_{i1} x_i - b_4 \right) + \beta \left(\sum_{i=1}^n r_{i2} x_i - b_3 \right) \geq 0. \quad (7)$$

В работе [9] продемонстрирован способ, в котором ограничение (1) по MSAD преобразуется в линейное в случае использования трапециевидных нечетких чисел. В этой работе было использовано L—R-представление нечеткого числа. В нашем случае при использовании классического определения нечеткого трапециевидного числа ограничение (1) будет выглядеть так:

$$MSAD(\tilde{r}) = \frac{3}{2} \sum_{i=1}^n r_{i3} x_i - \frac{1}{2} \left(\sum_{i=1}^n r_{i1} x_i + \sum_{i=1}^n r_{i2} x_i + \sum_{i=1}^n r_{i4} x_i \right). \quad (8)$$

Множеством α -уровня нечеткого числа $\tilde{r} = (r_1, r_2, r_3, r_4)$ называется множество вещественных чисел, определенных следующим образом:

$$[\tilde{r}]^\alpha = \{x \mid \mu(x) > \alpha, x \in R\} = [r_1 + \alpha(r_2 - r_1), r_4 - \alpha(r_4 - r_3)]. \quad (9)$$

Понятие множества α -уровня необходимо для вычисления вещественного среднего значения нечеткого числа, определенного Карлсоном [16]. Вещественное среднее значение нечеткого числа будет использоваться в рассматриваемых моделях как оценка потенциальной доходности. Пусть $[\tilde{r}]^\alpha = [a_1(\alpha), a_2(\alpha)]$, где $a_1(\alpha)$ и $a_2(\alpha)$ — некоторые функции от α , тогда вещественное среднее значение вычисляется так:

$$\tilde{M}(\tilde{r}) = \int_0^1 \alpha(a_1(\alpha) + a_2(\alpha)) d\alpha. \quad (10)$$

Для нашего нечеткого числа $\tilde{r} = (r_1, r_2, r_3, r_4)$ формула вещественного среднего будет выглядеть следующим образом:

$$\begin{aligned} \tilde{M}(\tilde{r}) &= \int_0^1 \alpha(r_1 + \alpha(r_2 - r_1) + r_4 - \alpha(r_4 - r_3)) d\alpha = \\ &= \frac{r_2 + r_3}{3} + \frac{r_1 + r_4}{6}. \end{aligned} \quad (11)$$

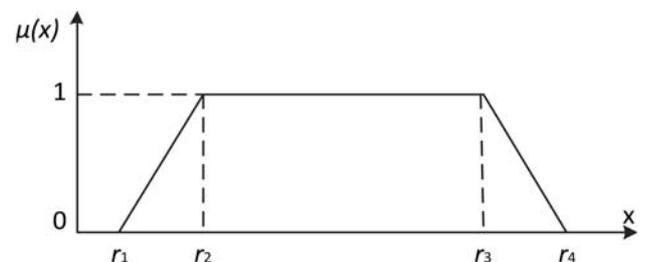


Рис. 1. Графическое представление нечеткого трапециевидного числа

Таким образом, у нас есть все необходимые правила и преобразования для формирования моделей в нечетко-множественной постановке. Например, модель (3) в нечетко-множественной постановке примет вид

$$\max_x \left(\frac{\sum_{i=1}^n r_{i2}x_i + \sum_{i=1}^n r_{i3}x_i}{3} + \frac{\sum_{i=1}^n r_{i1}x_i + \sum_{i=1}^n r_{i4}x_i}{6} - \sum_{i=1}^n c_i x_i \right);$$

$$(1 - \beta) \left(\sum_{i=1}^n r_{i1}x_i - b_4 \right) + \beta \left(\sum_{i=1}^n r_{i2}x_i - b_3 \right) \geq 0; \quad (12)$$

$$\sum_{i=1}^n x_i = 1;$$

$$l_i \leq x_i \leq u_i, \quad i = \overline{1, n}.$$

Аналогично, модель (4) с ограничением по VaR в нечетко-множественной постановке примет вид

$$\min_x \left(\frac{3}{2} \sum_{i=1}^n r_{i3}x_i - \frac{1}{2} \left(\sum_{i=1}^n r_{i1}x_i + \sum_{i=1}^n r_{i2}x_i + \sum_{i=1}^n r_{i4}x_i \right) \right); \quad (13)$$

$$\frac{\sum_{i=1}^n r_{i2}x_i + \sum_{i=1}^n r_{i3}x_i}{3} + \frac{\sum_{i=1}^n r_{i1}x_i + \sum_{i=1}^n r_{i4}x_i}{6} - \sum_{i=1}^n c_i x_i \geq R_0;$$

$$\sum_{i=1}^n x_i = 1;$$

$$l_i \leq x_i \leq u_i, \quad i = \overline{1, n}.$$

Модель (5), в которой учитываются несколько мер риска, в нечетко-множественной постановке примет вид

$$\min_x \left(\frac{3}{2} \sum_{i=1}^n r_{i3}x_i - \frac{1}{2} \left(\sum_{i=1}^n r_{i1}x_i + \sum_{i=1}^n r_{i2}x_i + \sum_{i=1}^n r_{i4}x_i \right) \right);$$

$$(1 - \beta) \left(\sum_{i=1}^n r_{i1}x_i - b_4 \right) + \beta \left(\sum_{i=1}^n r_{i2}x_i - b_3 \right) \geq 0; \quad (14)$$

$$\frac{\sum_{i=1}^n r_{i2}x_i + \sum_{i=1}^n r_{i3}x_i}{3} + \frac{\sum_{i=1}^n r_{i1}x_i + \sum_{i=1}^n r_{i4}x_i}{6} - \sum_{i=1}^n c_i x_i \geq R_0;$$

$$\sum_{i=1}^n x_i = 1;$$

$$l_i \leq x_i \leq u_i, \quad i = \overline{1, n}.$$

Все полученные модели являются стандартными задачами линейного программирования. Для решения задач был запрограммирован улучшенный симплекс-метод, разработанный Данцигом [17].

Разработка программного обеспечения

Разработанное программное обеспечение позволяет рассчитывать описанные в статье модели выбора оптимального инвестиционного портфеля в нечетко-множественной постановке. Программа состоит из нескольких основных частей: область ввода данных, окно ввода настроек и область вывода информации. В настройках программы можно изменять следующие параметры: нечетко-множественное описание параметра VaR (b1, b2, b3, b4, Beta), погрешность для модифицированного симплекс-метода (Eps), безрисковую ставку (Rf) и нижнюю предельную ставку для ограничения по доходности (рис. 2). После определения настроек пользователь должен последовательно заполнить информацию об активах:

- полное название актива;
- сокращенное название (тикер);
- необходимость учета актива в портфеле;
- оценку распределения доходности в виде нечетко-множественного трапециевидного числа;
- транзакционные издержки;
- дополнительные линейные ограничения на структуру портфеля, если таковые имеются.



Рис. 2. Настройки программы

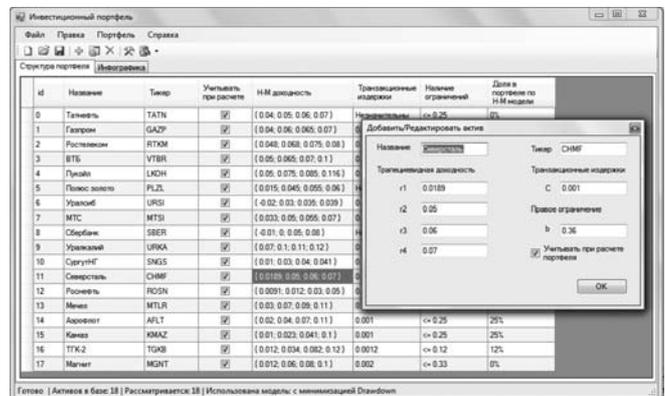


Рис. 3. Область ввода данных

На рис. 3 представлена область программы, в которой осуществляется ввод информации об используемых активах. При этом можно заполнить информацию обо всех необходимых активах, но при построении портфеля выбрать только некоторые.

На рис. 4 (см. третью сторону обложки) представлена область программы, в которой реализован вывод результирующей информации. Информация выводится в виде гистограммы сравнения доходностей и в виде круговой диаграммы структуры портфеля.

В гистограмме сравнения доходностей отражены следующие показатели: безрисковая доходность, ожидаемая доходность, доходность при негативном развитии ситуации, доходность при позитивном развитии ситуации. Последние показатели вычисляются по экспертным оценкам распределения доходностей. Также в область вывода результирующей информации выводятся значения показателей *VaR* и *Drawdown-at-Risk* — основных мер риска, которые использовались в данном исследовании.

Программа разрабатывалась в среде Microsoft Visual Studio 2010 на языке C# 4.0. Продукт требует установленной среды Microsoft .Net 4, которая, в случае отсутствия, загружается автоматически с узла Microsoft во время установки. Приложение тестировалось и корректно работало на ОС Windows XP, Vista, Seven.

Эксперименты на российском фондовом рынке

Нами были проведены эксперименты с описанными в статье моделями. Эксперименты проводились с несколькими высоколиквидными активами, торгующимися на Московской межбанковской валютной бирже: обыкновенные акции ОАО "Татнефть" (TATN), ОАО "Газпром" (GAZP), ОАО "Ростелеком" (RTKM), ОАО "ВТБ" (VTB), ОАО "Лукойл" (LKOH), ОАО "Полюс-Золото" (PLZL), ОАО "Сбербанк" (SBER). Временной интервал проведения экспериментов: 01.11.2010—01.06.2011. Для подготовки входных данных (нечетко-множественных оценок доходностей) была использована информация с официальных порталов одного из ведущих рейтинговых агентств [18] и брокерских компаний [19]. Для оценки соотношения риск—доходность использован коэффициент Шарпа [20]. Этот показатель иначе называют "коэффициент доходность—разброс" *RVAR* (*reward-to-variability ratio*):

$$RVAR = \frac{r_p - r_f}{\sigma_p}, \quad (15)$$

где r_p — средняя доходность портфеля за рассматриваемый интервал времени; r_f — среднее значение безрисковой ставки; σ_p — стандартное отклонение доходности портфеля (общий риск). В качестве безрисковой ставки использовалась доходность государственных облигаций. Хотя мы уже отмечали

недостатки СКО как меры риска, для сравнения эффективности портфелей было решено использовать общепринятую меру (*RVAR*). Значения коэффициентов Шарпа для рассмотренных моделей приведены в таблице.

Также целесообразно уделить внимание динамике доходностей различных активов и сравнить их с бенчмарком, в качестве которого может выступить индекс ММВБ (MICEX) — Московской Межбанковской Валютной Биржи (Moscow Interbank Currency Exchange). На рис. 5 и 6 представлена динамика доходностей сформированных в рамках экспериментов портфелей.

Из динамика доходностей видно, что сформированные на основе рассмотренных методик портфели часто демонстрируют результаты лучше рынка, представленного в наших экспериментах индексом ММВБ.

Средние значения коэффициента Шарпа для различных моделей

Название модели	Среднее значение коэффициента Шарпа
Модель Марковица	1,25
Модель с нечетко-множественными доходностями и мерой риска <i>Value-at-Risk</i>	1,27
Модель с нечетко-множественными доходностями и мерой риска <i>Drawdown-at-Risk</i>	1,31
Модель с нечетко-множественными доходностями и с комбинированными мерами риска (<i>VaR</i> и <i>Drawdown-at-Risk</i>)	1,28



Рис. 5. Динамика доходностей для рассмотренных моделей (*VaR*, *VaR+Drawdown*)



Рис. 6. Динамика доходностей для рассмотренных моделей (*Drawdown*)

Выводы

Модели с нечетко-множественными доходностями позволяют учитывать экспертные оценки. В качестве прогноза формируется ожидаемая доходность портфеля, вычисляемая как вещественное среднее нечеткого трапециевидного числа, а также доходности при негативном и позитивном развитии ситуации. Сформированная задача является более понятной для конечного пользователя. Результаты экспериментов показывают, что модели с нечетко-множественными входными параметрами могут применяться на практике. На рассмотренном временном промежутке такие модели показали более качественные результаты по сравнению с классическим методом формирования портфеля и по сравнению с рынком. Однако результаты напрямую зависят от качества экспертных оценок. В настоящем исследовании использовались экспертные оценки рейтинговых агентств [18] и брокерских компаний [19]. Предложенная модель, в которой учитываются несколько мер риска VaR и $MSAD$, позволяет повысить уровень диверсификации портфеля без ввода дополнительных ограничений на структуру портфеля, нивелируя этим основной недостаток первых двух моделей. Разработанное в рамках исследования программное обеспечение позволяет вычислять структуру инвестиционного портфеля согласно описанным в статье моделям и в дальнейшем может быть модернизировано путем добавления новых моделей и развитием своего функционала.

Список литературы

1. **Markowitz H.** Portfolio Selection // Journal of Finance. 1952. N 7. P. 77–91.

2. **Исавнин А.Г., Галиев Д.Р.** Модели портфельного инвестирования с применением асимметричных мер риска и генетических алгоритмов // Финансовая аналитика. 2011. N 48. С. 32–38.
3. **Bellman R., Zadeh L. A.** Decision making in a fuzzy environment, Management Science. 1970. N 17. P. 141–164.
4. **Ramaswamy S.** Portfolio selection using fuzzy decision theory. URL: <http://www.bis.org/publ/work59.pdf>
5. **Leon T., Liern V., Vercher E.** Viability of infeasible portfolio selection problem: a fuzzy approach // European Journal of Operational Research. 2002. N 139. P. 178–189.
6. **Ostermark R.** A fuzzy control model for dynamic portfolio management // Fuzzy sets and Systems. 1998. N 78. P. 243–254.
7. **Watada J.** Fuzzy portfolio selection for decision making in investment / Y. Yoshida ed. // Dynamical Aspects in fuzzy decision making — Heidelberg: New York: Physica Verlag. 2001. P. 141–162.
8. **Carlsson C., Fuller R., Majlender P.** A possibilistic approach to selecting portfolios with highest utility score // Fuzzy sets and systems. 2002. N 131. P. 13–21.
9. **Vercher E., Bermudez J. D., Segura J. V.** Fuzzy portfolio optimization under downside risk measures // Fuzzy sets and systems. 2007. N 158. P. 769–782.
10. **Галиев Д. Р., Исавнин А. Г.** Использование VaR-ограничений в модели Блэка—Литтермана // В мире научных открытий. 2011. N 18. С. 261–270.
11. **Анализ портфельных инвестиций / М. Д. Миссаров, А. Г. Исавнин, И. И. Махмутов, Д. Р. Галиев.** Набережные Челны: Лаб. операт. Полиграфии, 2011. 239 с.
12. **Arnott R. D., Wanger W. H.** The measurement and control of trading costs // Financial Analysts Journal. 1990. N 46 (6). P. 73–80.
13. **Yoshimoto A.** The mean-variance approach to portfolio optimization subject to transaction costs // Journal of the Operational Research Society of Japan. 1996. N 39. P. 99–117.
14. **Liu B., Iwamura K.** Chance constrained programming with fuzzy parameters // Fuzzy sets and systems. 1998. N 94. P. 227–237.
15. **Chen G., Chen S., Fang Y., Wang S.** A possibilistic Mean VaR Model for Portfolio Selection, Advanced modeling and optimization. 2006. Vol. 8, N 1. P. 99–107.
16. **Carlsson C., Fuller R.** On possibilistic mean value and variance of fuzzy number // Fuzzy sets and Systems. 2001. N 122. P. 315–326.
17. **Банди Б.** Основы линейного программирования / Под ред. В. А. Волынского. М.: Радио и связь, 1989. 176 с.
18. **Аналитика, прогнозы, котировки от РБК.** URL: <http://quote.rbc.ru>
19. **Комментарии и прогнозы по акциям.** URL: <http://finam.ru>
20. **Le Sourd V.** Performance Measurement for Traditional Investment // EDHEC : 2007. 65 p. URL: http://www.edhec-risk.com/performance_and_style_analysis/perf_measurement

Информация

22–25 октября 2012 года
в Санкт-Петербурге, на территории ВК "Ленэкспо" состоится

3-я Международная специализированная выставка

"ИНФОРМАЦИЯ: ТЕХНИКА И ТЕХНОЛОГИИ ЗАЩИТЫ".

Выставка пройдет в рамках 21-го Международного Форума "Охрана и безопасность — SFITEX".
Выставка — единственная в России площадка, на которой представлен весь спектр современных технических средств обеспечения информационной безопасности: программные, программно-аппаратные, аппаратные средства защиты информации.

Разделы выставки:

- Сетевая безопасность
- Технические средства защиты информации
- Криптография и ограничение доступа
- Специальные технические средства
- Услуги
- Интеллектуальные карточные системы
- NEW: Безопасность электронного документооборота

Одной из основных составляющих успеха выставки является

научно-практическая конференция

"Информационная безопасность: Невский диалог — 2012"

УДК 621.391

С. Н. Агиевич, канд. техн. наук,
ст. науч. сотр., вед. специалист,
ООО "Специальный Технологический Центр",
г. Санкт-Петербург,
e-mail: aserzhnic@mail.ru

Теоретические основы сплайн-алгебраического гармонического анализа сигналов систем радиосвязи

Рассматриваются возможности использования функций сплайн-характеров для анализа и синтеза сигналов. Представляется математический аппарат теоретических основ сплайн-алгебраического гармонического анализа. Определяются понятия линейности, сдвига, свертки, корреляции, инвариантности, преобразования Фурье для сплайн-функций. Доказывается их применимость для базиса функций сплайн-характеров.

Ключевые слова: сплайн-характеры, базисные функции, формирование сигналов, теоретические основы

Введение

В настоящее время развитие систем радиосвязи ограничивается рамками существующего математического аппарата синтеза и обработки сигналов, основанного на гармонических базисах их представлений. Применение методологии цифровой обработки предполагает, с одной стороны, дискретную форму представления сигналов, а с другой — континуальную, вызванную необходимостью их излучения. В результате возникает необходимость в разработке математического аппарата, объединяющего в себе преимущества дискретного и непрерывного описания сигналов. В этом смысле естественным образом приходим к необходимости использования сплайнов, являющихся по определению непрерывными функциями, но построенных по дискретным данным [1]. Именно поэтому была разработана теория сплайн-гармонического анализа [2—5], частными случаями которой являются классический дискретный и непрерывный анализ Фурье. Однако многообразие дискретных представлений сигналов существенно шире классических экспоненциальных функций, которые в настоящее время широко используются в радиотехнике. В связи с этим появились работы [6—8], которые существ-

венно расширяют круг систем базисных функций, пригодных для синтеза сигналов. Дальнейшие исследования в этом направлении привели к разработке теории, названной автором *сплайн-алгебраическим гармоническим анализом*. Для формирования и обработки сигналов здесь используются системы базисных функций сплайн-характеров. В настоящей работе представлены теоретические основы сплайн-алгебраического гармонического анализа.

Определение и свойства базисов функций сплайн-характеров

Пусть имеется пространство гладких функций — периодических сплайнов дефекта 1 порядка p [1]. Тогда любой сигнал $S^p(t)$ из этого пространства может быть построен из своего рода "кирпичиков" $M^p(t)$ — B -сплайнов заданной степени гладкости $p - 1$:

$$S^p(t) = \frac{1}{N} \sum_k q_k M^p(t - t_k), \quad (1)$$

где q_k — коэффициенты сплайна; t_k — дискретные отсчеты времени; k — текущий номер; t — текущая координата времени; N — число отсчетов сигнала.

Согласно [1] сплайн-функции обладают следующими свойствами, существенными для обработки сигналов:

- относительно легко реализуются цифровыми методами;
- обеспечивают минимальную среднеквадратичскую ошибку аппроксимации;
- позволяют сглаживать различного рода биения и разрывы.

При этом сплайны, являясь гладкими функциями с конечным носителем, строятся по дискретным отсчетам. Указанный факт имеет принципиальное значение для разработки математического аппарата, объединяющего в себе преимущества дискретного и непрерывного представления сигналов.

Учитывая, что класс дискретных функций значительно шире непрерывного, базирующегося на экспоненциальных описаниях, очевиден интерес к указанным представлениям, в том числе и для формирования сигналов систем радиосвязи.

Рассмотрим возможность реализации преобразования Фурье в пространстве дискретных функций $L(H, K)$, областью определения которых является абелева группа H , а областью значений — кольцо K . Согласно [8], аналогами комплексных экспонент в $L(H, K)$, образующих ортонормированный

базис, являются характеры $\chi(n, k)$, которые при определении на конечном отрезке называют χ -функциями, где n, k — переменные частоты и времени соответственно.

Между тем, несмотря на указанные достоинства дискретная природа характеров не позволяет непосредственно использовать их для формирования сигналов систем радиосвязи. Следовательно, необходимы теоретические основы сглаживания $\chi(n, k)$ базисов при сохранении их фундаментальных свойств.

Введем пространство G_n^p периодических сплайнов (здесь и далее верхние дуги указывают на принадлежность к пространству периодических сплайнов, полученному на основе пространства $L(H, K)$ дискретных функций) и в соответствии с формулой (1) сформируем в этом пространстве сигнал $\widehat{S}^p(t)$:

$$\begin{aligned} \widehat{S}^p(t) &= \frac{1}{N} \sum_k \widehat{q}_k M^p(t \ominus t_k) = \\ &= \frac{1}{N} \sum_k M^p(t \ominus t_k) \sum_n \chi^*(n, k) \widehat{F}_n(q), \end{aligned} \quad (2)$$

где μ — модуль представления чисел; \ominus — сдвиг по модулю μ ; $\chi^*(n, k)$ — комплексно-сопряженное $\chi(n, k)$; $\widehat{F}_n(*)$ — преобразование Фурье в базисе характеров, значение которого можно представить как $\widehat{F}_n(q) = \frac{1}{N} \sum_k \chi(n, k) \widehat{q}_k$. Учитывая, что $\widehat{m}_n^p(t) = \frac{1}{N} \sum_k \chi^*(n, k) M^p(t \ominus t_k)$ есть базисная функция, полученная сверткой характеров $\chi(n, k)$ и B -сплайнов, а $t_k = ((p/2) + k)/N$, то в конечном счете получим спектр сигнала в базисе сплайн-характеров в виде следующего выражения:

$$\begin{aligned} \sum_n \widehat{F}_n(q) \frac{1}{N} \sum_k \chi^*(n, k) M^p(t \ominus t_k) = \\ = \sum_n \widehat{\xi}_n \widehat{m}_n^p(t) = \sum_n \widehat{c}_n \widehat{U}_n^p(t), \end{aligned} \quad (3)$$

где $\widehat{\xi}_n = \widehat{F}_n(z)/\widehat{u}_n^p$; $\widehat{u}_n^p = \widehat{F}_n(M^p) = \frac{1}{N} \sum_k \chi^*(n, k) M^p(t_k)$;

$$\widehat{c}_n = \widehat{F}_n(z); \quad \widehat{U}_n^p(t) = \widehat{m}_n^p(t)/\widehat{u}_n^p.$$

Таким образом, в соответствии с формулой (3) произвольный сигнал $\widehat{S}^p(t)$ в пространстве G_n^p можно представить с использованием функций $\widehat{m}_n^p(t)$ или $\widehat{U}_n^p(t)$.

Для того чтобы функции $\widehat{m}_n^p(t)$ можно было использовать в качестве базисов формирования сигналов, они должны удовлетворять основным свойствам преобразований. Докажем их выполнимость для рассматриваемых условий.

Свойство 1 (инвариантность относительно группового сдвига):

$$\widehat{m}_n^p(t \ominus l/N) = \chi^*(n, l) \widehat{m}_n^p(t).$$

Доказательство:

$$\begin{aligned} \widehat{m}_n^p(t \ominus l/N) &= \frac{1}{N} \sum_k \chi^*(n, l) M^p\left(t \ominus \frac{k \ominus l}{N}\right) = \\ &= \chi^*(n, l) \frac{1}{N} \sum_k \chi^*(n, l) M^p\left(t \ominus \frac{k}{N}\right) = \chi^*(n, l) \widehat{m}_n^p(t). \end{aligned}$$

Свойство 2.

Функции $\widehat{m}_n^p(t)$ — N -периодические по отношению к N .

Доказательство. При $l = N$, исходя из доказательства выполнения свойства 1, имеем $\widehat{m}_n^p(t \ominus N/N) = \chi^*(n, N) \widehat{m}_n^p(t) = \chi^*(n, 0) \widehat{m}_n^p(t) = \widehat{m}_n^p(t)$.

Свойство 3.

Для функций $\widehat{m}_n^p(t)$ справедливы выражения

$$a) \int_0^1 \widehat{m}_n^{*p}(t) \widehat{m}_r^b(t) dt = \delta_r^i \widehat{u}_n^{p+b};$$

$$б) \frac{1}{N} \sum_k \widehat{m}_n^{*p}(t_k) \widehat{m}_r^b(t_k) = \delta_r^i \widehat{u}_n^p \widehat{u}_r^b.$$

Доказательство а):

$$\begin{aligned} \int_0^1 \widehat{m}_n^{*p}(t) \widehat{m}_r^b(t) dt &= \delta_r^i \widehat{u}_n^{p+b} = \sum_{n=-\infty}^{\infty} \widehat{c}_n(\widehat{m}_n^{*p}) \widehat{c}_n(m_r^b) = \\ &= \sum_{n=-\infty}^{\infty} \left\{ \int_0^1 \chi(n, t) \widehat{m}_n^{*p}(t) dt \right\} \times \left\{ \int_0^1 \chi(n, t) \widehat{m}_r^b(t) dt \right\} = \\ &= \sum_{n=-\infty}^{\infty} \left\{ \left(\delta_n^i \int_0^1 \chi(n, t) M^p(t \ominus t_k) dt \right) \times \right. \\ &\times \left. \left(\delta_n^r \int_0^1 \chi(n, t) M^b(t \ominus t_k) dt \right) \right\} = \sum_{n=-\infty}^{\infty} (\delta_n^i \widehat{V}_n^p) (\delta_n^r \widehat{V}_n^b) = \\ &= \sum_{n=-\infty}^{\infty} \delta_r^n \sum_{j=-\infty}^{\infty} \widehat{V}_{N+jN}^{p+b} = \delta_r^i \widehat{u}_n^{p+b}, \end{aligned}$$

где \widehat{V}_n^p — коэффициенты Фурье в базисе $\chi(n, t)$; δ_r^i — символ Кронекера для r -й и i -й функций в рассматриваемом базисе.

Из доказательства следует: сплайны $\widehat{m}_n^p(t)$ образуют ортогональный базис пространства G_n^p ; реализуется свертка $\widehat{m}_n^p(t) * \widehat{m}_r^b(t) = \widehat{m}_n^{p+b} b \delta_n^r$.

Доказательство б):

$$\begin{aligned} \frac{1}{N} \sum_k \widehat{m}_n^{*p}(t_k) \widehat{m}_r^b(t_k) &= \sum_k \widehat{F}_n(m_n^p(t_k)) \widehat{F}_n(m_r^b(t_k)) = \\ &= \delta_n^r \widehat{u}_n^p \widehat{u}_r^b. \end{aligned}$$

Свойство 4.

Сплайны $\widehat{m}_n^p(t \ominus p/(2N))/\widehat{u}_n^p$ интерполируют $\chi(n, t)$,

а именно: $\widehat{m}_n^p(l/N) = \chi(n, l/N)$.

Доказательство:

$$\begin{aligned} \widehat{m}_n^p \left(\frac{p}{2N} p / (2N) \ominus l / N / N \right) &= \widehat{m}_n^p (p / 2 \ominus l / N) = \\ &= \chi^*(n, l / N) \widehat{m}_n^p (p / (2N)) = \chi^*(n, l / N) \widehat{n}_n^p. \end{aligned}$$

Доказанные выше свойства показывают, что выражение (3) описывает разложение сигнала $S^p(t)$ по базисным функциям $\widehat{m}_n^p(t)$ или $\widehat{U}_n^p(t)$.

Отметим еще одно свойство функций $\widehat{m}_n^p(t)$.

Свойство 5 (мультипликативности).

Произведение двух базисных функций $\widehat{m}_n^p(t_k)$ и $\widehat{m}_r^b(t_k)$ дает новую базисную функцию $\widehat{m}_n^p(t_k) \widehat{m}_r^b(t_k) = \widehat{m}_{n \ominus r}^{p+b}(t_k)$.

Доказательство:

$$\begin{aligned} \widehat{m}_n^p(t_k) \widehat{m}_r^b(t_k) &= \frac{1}{N} \sum_k \chi(n, k) M^p(t_k) \frac{1}{N} \sum_k \chi(r, k) M^b(t_k) = \\ &= \frac{1}{N} \sum_k \chi(n, k) M^p(t_k) \chi(r, k) M^b(t_k) = \\ &= \frac{1}{N} \sum_k \chi(n \ominus r, k) M^{p+b}(t_k) = \widehat{m}_{n \ominus r}^{p+b}(t_k). \end{aligned}$$

Теорема 1.

Сплайны $\widehat{\chi}_n^p(t) = \widehat{m}_n^p(t) / \sqrt{\widehat{u}_n^{2p}}$ образуют ортонормированный базис пространства \widehat{G}_n^p .

Доказательство: из свойства 3а следует, что

$$\int_0^1 (\widehat{m}_n^{*p}(t) / \sqrt{\widehat{u}_n^{2p}}) (\widehat{m}_r^{*p}(t) / \sqrt{\widehat{u}_n^{2p}}) dt = \delta_n^r.$$

Базисные функции $\widehat{\chi}_n^p(t)$, $\widehat{m}_n^p(t)$, $\widehat{U}_n^p(t)$ будем называть сплайн-характерами. Ниже приведем некоторые следствия из теоремы 1.

Следствие 1.

При $p = 1$ и $t = k$ $\widehat{\chi}_n^p(t) = \chi(n, k)$.

Доказательство:

$$\begin{aligned} \widehat{\chi}_n^p(t) &= \widehat{m}_n^p(t) / \sqrt{\widehat{u}_n^{2p}} = \widehat{m}_n^1(k) / \sqrt{\widehat{u}_n^2} = \\ &= \chi(n, k) / \sqrt{\widehat{u}_n^2} = \chi(n, k), \end{aligned}$$

т. е. теорема 3.1, рассмотренная в [8], есть частный случай доказанной теоремы 1.

Следует отметить, что теорема 1 позволяет получить новые пространства. Например, пространство комплексных чисел $L(H_{\mu_1} + H_{\mu_2} + \dots + H_{\mu_j}, C)$, где $H := H_{\mu_1} + H_{\mu_2} + \dots + H_{\mu_j}$ и $K = C$ — поле комплексных чисел, а $\mu_1, \mu_2, \dots, \mu_j$ — произвольные целые числа (нижний индекс при μ указывает число возможных вариантов модуля числа). Базисные функции $\widehat{\chi}_n^p(t)$ сплайн-характеров, сформирован-

ных в данном пространстве, определим как ортонормированный базис сплайн-Понтрягина — Виленкина — Крестенсона. Пространство $L(H_{2^l}, GF(r^v))$, где $H := H_{2^l}$ и $K = GF(r^v)$ — поле Галуа, где $v = 1, 2, 4, l = 1, 2, \dots$, позволяет сформировать ортонормированный базис сплайн-Галуа. Пространство $L(H_{2^l}, GF(r^v))$, где $H := H_{2^l}$ и $K = GF(r^v)$ — поле Галуа, $v = 1, l = 1, 2, \dots$, — ортонормированный базис сплайн-Рейдера. Пространство $L(\Gamma, Z_{\mu}[*])$, где $H = \Gamma$ — группа целых чисел, $K = Z_{\mu}[*]$ — кольцо целых комплексных чисел, $\mu = a + jb$ — целое комплексное число — ортонормированный сплайн-базис Гаусса.

Таким образом, теорему 1 можно рассматривать как методологическое ядро теоретических основ формирования базисных функций в пространстве сплайн-характеров.

Базовые теоремы сплайн-алгебраического гармонического анализа

Представленная теоретическая основа базисных функций сплайн-характеров (БФСХ) позволяет представить базовые теоремы для цифровой обработки в следующей трактовке.

Теорема 2 (линейности).

Пусть $\widehat{S}^p(t) = \frac{1}{N} \sum_{k=0}^N \widehat{q}(k) M^p(t \oplus t_k)$. Тогда, если

$$\widehat{S}^p(k) = \sum_{l=1}^r \widehat{c}_l \widehat{S}^p(l, k), \text{ то } \widehat{F}_n(\widehat{S}^p(k)) = \sum_{l=1}^r \widehat{F}_n(\widehat{S}^p(l, k)),$$

где $l = 1, 2, \dots, r$ — текущее значение суммы для r суммируемых слагаемых.

Следствие теоремы 2: при $p = 1$ $\widehat{F}_n(\widehat{S}^p(k)) =$

$$= \sum_{l=1}^r \widehat{F}_n(\widehat{q}(l, k)), \text{ т. е. приходим к частному случаю — теореме 3.2, доказанной в работе [8].}$$

Теорема 3.

Пусть $\widehat{S}^p(t) = \frac{1}{N} \sum_{k=0}^N \widehat{q}_k M^p(t \oplus t_k)$, тогда для

сдвинутых оригинала $\widehat{S}^p(k \oplus a)$ и изображения

$\widehat{F}_{n \oplus b}$, где a, b — соответственно сдвиг временной и частотной координаты, имеют место равенства:

$$\text{а) } \widehat{F}_n(\widehat{S}^p(k \oplus a)) = \chi^*(n, a) \widehat{F}_n(\widehat{S}^p(k)) =$$

$$= \widehat{m}_n^{*p}(a) \widehat{F}_n(\widehat{q}_k);$$

$$\text{б) } \widehat{F}_{n \oplus j}(\widehat{S}^p(k)) = \widehat{F}_n(\widehat{S}^p(k) \chi^*(b, k)) =$$

$$= \widehat{F}_n(\widehat{q}_k \widehat{m}_n^{*p}(b, k)).$$

Следствие 1 теоремы 3: при $p = 1$

$$\widehat{F}_n(\widehat{S}^p(k \oplus_a)) = \chi^*(b, k) \widehat{F}_n(\widehat{q}_k).$$

Следствие 2 теоремы 3: при $p = 1$

$$\widehat{F}_{k \oplus_a}(\widehat{S}^p(k)) = \widehat{F}_n(q_k \chi^*(b, k)),$$

т. е. приходим к частному случаю — теореме 3.3, доказанной в работе [8].

Пусть $\widehat{S}^p(t) = \frac{1}{N} \sum_{k=0}^N \widehat{q}_k M^p(t \oplus_\mu t_k)$. Тогда разностью функции $\widehat{S}^p(k)$ в направлении a будет называться функция $\Delta_a \widehat{S}^p(k) = \widehat{S}^p(k \oplus_a) - \widehat{S}^p(k)$.

Теорема 4.

Разность функции $\widehat{S}^p(k)$ имеет изображение

$$\begin{aligned} \widehat{F}_n(\Delta_a) &= [\chi(n, a) - 1] \widehat{F}_n(\widehat{S}^p(k)) = \\ &= [\widehat{m}_n^p(a) - 1] \widehat{F}_n(\widehat{q}_k). \end{aligned}$$

Разности более высокого порядка определяются следующим образом:

$$\Delta_{a_1 a_2 \dots a_l}^l = \Delta_{a_l} \Delta_{a_{l-1}} (\dots (\Delta_{a_1} \Delta_{a_2} \dots \Delta_{a_{l-1}}) \widehat{S}^p(k) \dots).$$

Нетрудно видеть, что

$$\begin{aligned} \widehat{F}_n(\Delta^l) &= \left\{ \prod_{a=1}^l [\chi(n, a) - 1] \right\} \widehat{F}_n(\widehat{S}^p(k)) = \\ &= \left\{ \prod_{a=1}^l [\widehat{m}_n^p(a) - 1] \right\} \widehat{F}_n(\widehat{q}_k). \end{aligned}$$

Следствие 1 теоремы 4: при $p = 1$

$$\widehat{F}_n(\Delta_a) = [\chi(n, a) - 1] \widehat{F}_n(\widehat{q}_k).$$

Для разностей более высокого порядка:

$$\widehat{F}_n(\Delta^l) = \left\{ \prod_{a=1}^l [\chi(n, a) - 1] \right\} \widehat{F}_n(\widehat{q}_k).$$

Замечание: при соответствующем выборе H, K, p можно получить следствия теорем 2...4, как это было получено для теоремы 1.

Теперь рассмотрим возможность реализации процедур свертки и корреляции, а также выполнение обобщенной теоремы Винера — Хинчина в БФСХ.

Пусть

$$\widehat{S}_1^p(t) = \frac{1}{N} \sum_{k=0}^{N-1} \widehat{q}_1(k) M^p(t \oplus_\mu a) = \frac{1}{N} \sum_{k=0}^{N-1} z_k \widehat{L}^p(t \oplus_\mu t_k) \in \sigma^p;$$

$$\widehat{S}_1^v(t) = \frac{1}{N} \sum_{k=0}^{N-1} \widehat{q}_1(k) M^v(t \oplus_\mu a) = \frac{1}{N} \sum_{k=0}^{N-1} z_k \widehat{L}^v(t \oplus_\mu t_k) \in \sigma^v,$$

где L — ядро глобального базисного сплайна, σ^v — пространство сплайнов гладкости v .

Определение. Если $\widehat{S}_1^p(k), \widehat{S}_2^v(k) \in L(H, K)$, то функция

$$\widehat{B}_{\widehat{S}_1^p, \widehat{S}_2^v}(i) = \frac{1}{N} \sum_k \widehat{S}_1^p(k) \widehat{S}_2^{*v}(k \oplus_\mu a)$$

называется НК-взаимной корреляционной функцией сигналов $\widehat{S}_1^p(k)$ и $\widehat{S}_2^v(k)$. Если $\widehat{S}_1^p(k) = \widehat{S}_2^v(k)$, то $\widehat{B}_{\widehat{S}_1^p, \widehat{S}_2^v}(a) = \widehat{B}(a)$ есть НК-автокорреляционная функция сигнала $\widehat{S}_1^p(k)$.

Теорема 5.

Изображение НК-взаимно корреляционной функции сигналов $\widehat{S}_1^p(k)$ и $\widehat{S}_2^v(k)$ есть произведение спектров этих сигналов:

$$\begin{aligned} \widehat{F}_n[\widehat{B}_{\widehat{S}_1^p, \widehat{S}_2^v}(i)] &= \widehat{F}_n(\widehat{S}_1^p(k)) \widehat{F}_n^*(\widehat{S}_2^v(k)) = \\ &= \widehat{u}_n^p \widehat{F}_n(\widehat{q}_1(k)) \widehat{u}_n^{*v} \widehat{F}_n^*(\widehat{d}_2(k)), \end{aligned}$$

где $\widehat{d}_2(k)$ — коэффициенты сплайна $\widehat{S}_2^v(k)$.

Следствие 1 теоремы 5: при $v = p = 1$

$$\widehat{F}_n[\widehat{B}_{\widehat{S}_1^p, \widehat{S}_2^v}(a)] = \widehat{F}_n[\widehat{q}_1(k)] \widehat{F}_n^*[\widehat{q}_2(k)],$$

т. е. приходим к частному случаю — теореме 3.5, доказанной в работе [8].

Следствие 2 теоремы 5: изображение \widehat{D}_n НК-автокорреляционной функции определяется следующим выражением:

$$\begin{aligned} \widehat{F}_n[\widehat{B}_{\widehat{S}_1^p, \widehat{S}_1^p}(a)] &= \widehat{F}_n[\widehat{S}_1^p(k)] \widehat{F}_n^*[\widehat{S}_1^p(k)] = \\ &= |\widehat{F}_n[\widehat{S}^p(k)]|^2 = |\widehat{u}_n^p \widehat{F}_n[\widehat{q}(k)]|^2 = \widehat{D}_n. \end{aligned}$$

Определение. Если $\widehat{S}_2^p(k) \in \sigma^p, \widehat{S}_1^v(k) \in \sigma^v$, то

$\widehat{Y}(k) = \frac{1}{N} \sum_a \widehat{S}_1^v(k \oplus_\mu a) \widehat{S}_2^p(a)$ называется НК-сверткой сигналов $\widehat{S}_2^p(k), \widehat{S}_1^v(k)$.

Теорема 6.

$$\text{Спектр свертки } \widehat{Y}(k) = \frac{1}{N} \sum_a \widehat{S}_1^v(k \oplus_\mu a) \widehat{S}_2^p(a)$$

есть произведение спектров сигналов $\widehat{S}_1^p(k), \widehat{S}_2^v(k)$.

Следствие 1 теоремы 6: при $v = p = 1$ $\widehat{F}_n[\widehat{Y}(k)] =$

$= \widehat{F}_n(\widehat{q}_1(k)) \widehat{F}_n(\widehat{d}_2(k))$, т. е. приходим к частному случаю — теореме 3.6, доказанной в работе [8].

Теорема 7.

Энергетический спектр инвариантен относительно H -сдвигов.

Замечание. Для теорем 5...7 естественны следствия, аналогичные следствиям теоремы 1.

Представленные теоремы носят доказательный характер, что свидетельствует о достоверности рассматриваемого научного направления.

Теперь рассмотрим возможность выполнения теоремы Котельникова в БФСХ. Обозначим \widehat{G}_f^p пространство непериодических сплайнов. Напомним, что для $\widehat{G}_n^p N = T\widehat{f}_{\max}$, где $T=1$ (однопериодический случай) — период сигнала; \widehat{f}_{\max} — максимальная частота анализа сигнала в базисе $\widehat{m}_n^p(t)$ ($\widehat{U}_n^p(t)$).

Теорема 8.

Любой сигнал $\widehat{S}^p(t)$ из пространства \widehat{G}_f^p с ограниченной спектральной плотностью может быть представлен выражением $\widehat{S}^p(t) = \sum_{k=-\infty}^{\infty} \widehat{L}^p(t \ominus_{\mu} t_k)$.

Доказательство: для сигналов из $\widehat{S}^p(t)$ имеем:

$$\begin{aligned} \widehat{S}^p(t) &= \frac{1}{N} \sum_k q_k M^p(t \ominus_{\mu} t_k) = \\ &= \frac{1}{N} \sum_k M^p(t \ominus_{\mu} t_k) \sum_n \chi^*(n, k) \widehat{F}_n(q) = \\ &= \sum_k \widehat{F}_n(q) \frac{1}{N} \sum_k \chi^*(n, k) M^p(t \ominus_{\mu} t_k) = \sum_n \widehat{F}_n(q) \widehat{m}_n^p(t) = \\ &= \sum_n \widehat{\xi}_n \widehat{m}_n^p(t) = \sum_n \widehat{F}_n(z) \frac{\widehat{m}_n^p(t)}{\widehat{u}_n^p} = \sum_n \widehat{c}_n \widehat{U}_n^p(t) = \\ &= \sum_n \frac{1}{N} \sum_k z_k \chi(n, k) \frac{\widehat{m}_n^p(t)}{\widehat{u}_n^p} = \sum_n \frac{1}{N} \sum_k z_k \chi(n, k) U_n^p(t) = \\ &= \sum_k z_k \widehat{L}^p(t \ominus_{\mu} t_k). \end{aligned} \quad (4)$$

Перейдем из \widehat{G}_n^p в \widehat{G}_f^p . Для этого устремим $T \rightarrow \infty$. Тогда:

$$\widehat{S}^p(t) = \sum_{k=-\infty}^{\infty} z_k \widehat{L}^p(t \ominus_{\mu} t_k); \quad (5)$$

$$\begin{aligned} \lim_{T \rightarrow \infty} n/T = f; \quad \lim_{T \rightarrow \infty} 1/T = df; \quad \lim_{T \rightarrow \infty} \widehat{L}^p(t \ominus_{\mu} t_k) = \\ = \lim_{T \rightarrow \infty} \frac{1}{N} \sum_{n=-N/2}^{N/2-1} \chi(n, k) \widehat{U}_n^p(t) = \\ \lim_{T \rightarrow \infty} \frac{1}{f_{\max} T} \sum_{n/T=-f_{\max}/2}^{f_{\max}/2} \chi(n/T, k/f_{\max}) \widehat{U}_{n/T}^p(t) = \\ = 1/f_{\max} \int_{-f_{\max}/2}^{f_{\max}/2} \chi(f, k/f_{\max}) \widehat{U}_f^p(t) df, \end{aligned} \quad (6)$$

где

$$\begin{aligned} \widehat{U}_f^p(t) &= \widehat{m}_f^p(t) / \widehat{u}_f^p, \quad \widehat{m}_f^p(t) = \frac{1}{N} \sum_{k=-\infty}^{\infty} \chi^*(f, \\ & \quad k/f_{\max}) M^p(t \ominus_{\mu} t_k); \\ \widehat{u}_f^p &= \frac{1}{N} \sum_{k=-\infty}^{\infty} \chi^*(f, k/f_{\max}) M^p(t_k). \end{aligned}$$

Здесь значения f и f_{\max} — текущая и максимальная частоты в используемых базисах. Таким образом, выражение (5) с ядром (6) доказывает теорему 8.

Замечание. Континуальный случай теоремы Котельникова является частным по отношению к теореме 8, что позволяет рассматривать ее как обобщающую. Кроме того, для теоремы 8 естественны следствия для случаев, аналогичных следствиям теоремы 1.

Мощным инструментом анализа являются процедуры преобразования Фурье. В целях выяснения их применимости для БФСХ рассмотрим обобщенный ряд Фурье.

Теорема 9.

Обобщенный ряд Фурье сигнала из пространства \widehat{G}_n^p представляется в следующем виде:

$$\widehat{S}^p(t) = \sum_{n=-\infty}^{\infty} \widehat{c}_n \widehat{U}_n^p(t) = \sum_{n=-\infty}^{\infty} \widehat{c}_n \frac{\widehat{m}_n^p(t)}{\widehat{u}_n^p}.$$

Доказательство: Из (4) имеем:

$$\begin{aligned} \widehat{S}^p(t) &= \sum_k z_k \frac{1}{N} \sum_n \chi(n, k) \widehat{U}_n^p(t) = \\ &= \sum_n \widehat{U}_n^p(t) \frac{1}{N} \sum_k z_k \chi(n, k) = \sum_{n=-N/2}^{N/2-1} \widehat{c}_n \widehat{U}_n^p(t). \end{aligned}$$

Устремим $f_{\max} \rightarrow \infty$, тогда, учитывая, что $N = T\widehat{f}_{\max}$, получаем:

$$\lim_{f_{\max} \rightarrow \infty} S^p(t) = \sum_{n=-\infty}^{\infty} \widehat{c}_n \widehat{U}_n^p(t),$$

что и требовалось доказать.

Замечание. Для теоремы 9 естественны следствия для случаев, аналогичных следствиям теоремы 1.

Заключение

Разработанные теоретические основы сплайн-алгебраического гармонического анализа представляют математический аппарат, в своей совокупности позволяющий проводить синтез и обработку сигналов в БФСХ. Разработанные базисы сочетают преимущества дискретной и континуальной форм представлений, определяемых двойственной природой сплайнов, которые являются непрерывными функциями, построенными по дискретным данным.

Данное обстоятельство позволяет наилучшим образом учитывать сложившийся порядок синтеза и обработки сигналов современных систем радиосвязи.

Разработанные теоретические основы являются обобщающими по отношению к непрерывному гармоническому анализу, широко используемому в настоящее время в радиотехнике. Данный факт подтверждается представленной доказательной базой и выполнимостью фундаментальных процедур свертки, корреляции, преобразований Фурье и пр.

Дальнейшее направление исследований автору видится в практических реализациях разработанного математического аппарата.

Список литературы

1. Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л. Методы сплайн-функций. М.: Наука, 1980. 352 с.
2. Zheludev V. A. Periodic splines, harmonic analysis and wavelets in Signal and image representation in combined spaces, wavelet // Anal. Appl., 7 / Eds Y. Y. Zeevi and R. Coifman. San Diego: Academic Press, 1998. P. 477—509.
3. Желудев В. А. Периодические сплайны и быстрое преобразование Фурье // Вычисл. матем. и матем. физ. 1992. Т. 32, № 2. С. 179—198.
4. Агиевич С. Н., Алексеев А. А., Глушанков Е. И. Модели сигналов в базисах сплайнов дефекта 1 и оценивание парамет-

ров радиоизлучений // Радиоэлектроника (Известия вузов). 1995. Т. 38, № 4. С. 3—16.

5. Агиевич С. Н. Сплайн-Виленкина — Крестенсона функции в представлении сигналов // Научное приборостроение. 2002. Т. 12, № 1. С. 79—89.
6. Агиевич С. Н., Дворников С. В., Гусельников А. С. Описание сигналов в базисах функций сплайн-Виленкина—Крестенсона // Контроль—Диагностика. 2009. № 3. С. 52—57.
7. Агиевич С. Н., Беспалов В. Л. Цифровая обработка сигнала: скорость и экономия ресурса // Мобильные системы. 2007. № 3. С. 26—29.
8. Вариченко Л. В., Лабунец В. Г., Раков М. А. Абстрактные алгебраические системы и цифровая обработка сигналов. Киев: Наукова Думка, 1986. 247 с.

УДК 004.043

- А. В. Захаров**, канд. физ.-мат. наук, доц.,
e-mail: ralzax@yandex.ru,
М. Р. Богданов, канд. биол. наук, доц.,
e-mail: bogdanov_marat@mail.ru,
А. Р. Мухамедьянов, ассистент,
e-mail: amazing-unit@mail.ru,
БГПУ им. М. Акмуллы, г. Уфа

Аналитический метод решения задачи позиционирования

Обсуждается аналитический метод решения задачи позиционирования (ЗП) применительно к системам ГЛОНАСС/GPS как альтернатива итерационного подхода. Выявляются этапы решения ЗП и предлагаются оптимальные методы их преодоления. Указываются преимущества аналитического метода решения ЗП по сравнению с итерационным методом.

Ключевые слова: навигационная задача, позиционирование, спутниковая радионавигационная система, ГЛОНАСС, GPS, геометрический фактор снижения точности

Постановка задачи

Предлагаемая работа направлена на исследование математических методов решения задачи позиционирования (ЗП) в спутниковых радионавигационных системах (СРНС) ГЛОНАСС/GPS, а также на разработку оптимальных программных алгоритмов их решения. Под позиционированием, как обычно, понимается задача определения координат и вектора скорости навигационной аппаратуры потребителя (НАП). На первом этапе это геоцентрические декартовы координаты $\{x, y, z\}$ и координаты вектора скорости в этой же системе координат $\{V_x, V_y, V_z\}$, отнесенные к моменту времени t_i в i -й такт позиционирования. Нами рассматриваются аналитические методы решения навигационной задачи как

альтернатива итерационных методов, утвержденных в соответствующих ГОСТ Р [1].

Аналитические расчеты таят в себе ряд сложностей, которые в явном виде отсутствуют в итерационных методах, что предопределило предпочтение последних.

Цель нашей работы — выявление этих затруднений и разработка оптимальных методов их преодоления.

Определение координат потребителя

Определение координат потребителя с помощью спутниковых радионавигационных систем выполняется в несколько этапов. На первом этапе по данным альманаха СРНС определяют геоцентрические координаты навигационных космических аппаратов (НКА) в i -й момент времени $\{X_n(t_i), Y_n(t_i), Z_n(t_i)\}$, $n = \{1...24\}$, где n — номер НКА, псевдодальности $S_n(t_i)$ — расстояния от НАП до НКА с номером n , угол места $\alpha_n(t_i)$ — угол с вершиной в конкретном НАП между плоскостью горизонта и лучом, направленным в n -й НКА. На этом завершается первый этап позиционирования — получение необходимых данных НАП от СРНС.

Второй этап заключается в выборе оптимального рабочего созвездия из числа НКА, для которых доступным является уверенный прием радиосигнала конкретным НАП.

В качестве критерия отбора используется угол места. Если он превышает $+5^\circ$, то НКА является наблюдаемым данным НАП. Однако, на наш взгляд, более удобным может быть критерий наблюдаемости n -го НКА по его псевдодальности: если псевдодальность удовлетворяет условию

$$S_{\min} > S_n(t_i) < S_{\max}, \quad (1)$$

то он наблюдаем. Преимущество заключается в том, что псевдодальности $S_n(t_i)$ — основные параметры ЗП, а угол места — вспомогательная величина и надобность в ней отпадает. Кроме того, критерий (1) позволяет отфильтровывать ряд выставляемых

помех с теми же углами места, но с иными псевдодальностями.

Обычно НАП принимает устойчивый сигнал от девяти до трех НКА из 24 в СРНС. Временами СРНС может насчитывать меньшее или большее число активно работающих спутников. Тогда и число наблюдаемых спутников может выходить из названных границ.

Третий этап состоит в выборе трех или четырех НКА из числа наблюдаемых, которые могут обеспечить решение ЗП с наименьшими погрешностями. Значения погрешностей, которые дает расчетный метод, определяются фактором геометрического снижения точности GDOP (*geometric dilution of precision*). Для наиболее точного определения координат НАП подбирают такие четыре или три спутника, пространственное расположение которых относительно потребителя дает наименьший GDOP.

Можно показать, что геометрический фактор снижения точности GDOP обратно пропорционален объему тетраэдра в случае рабочего созвездия, состоящего из четырех НКА (рис. 1, а), и обратно пропорционален площади треугольника в случае с тремя спутниками (рис. 1, б). Чем больше объем или площадь, тем точнее расчеты.

Расчет объема тетраэдра и площади треугольника проводится по известным формулам. Вычисляем векторы, начало которых находится в точке расположения НКА с номером a и конец — в точке расположения НКА с номерами b, c, d :

$$\begin{aligned}\vec{A} ab &= [X_a - X_b, Y_a - Y_b, Z_a - Z_b]; \\ \vec{B} ac &= [X_a - X_c, Y_a - Y_c, Z_a - Z_c]; \\ \vec{C} ad &= [X_a - X_d, Y_a - Y_d, Z_a - Z_d].\end{aligned}\quad (2)$$

Если вычисления ведутся по трем спутникам, то находим векторные произведения и их модули, перебирая номера a, b, c среди номеров видимых НКА:

$$\vec{\Sigma} abc = \frac{1}{2} \vec{A} ab \times \vec{B} ac; \quad (3)$$

$$|\vec{\Sigma} abc| = \frac{1}{2} |\vec{A} ab \times \vec{B} ac|. \quad (4)$$

Наибольшее значение площади, вычисляемой по формуле (4), и дадут нам номера трех НКА, по которым мы будем решать ЗП с наибольшей точностью.

Если задача решается по четырем спутникам, как в итерационном методе, то вычисляем объем

$$V_{abcd} = \begin{vmatrix} X_a - X_b & Y_a - Y_b & Z_a - Z_b \\ X_a - X_c & Y_a - Y_c & Z_a - Z_c \\ X_a - X_d & Y_a - Y_d & Z_a - Z_d \end{vmatrix}. \quad (5)$$

Путем перебора и выбора наибольшего объема определяем рабочее созвездие четырех НКА из числа наблюдаемых.

Однако нередко встречается ситуация, когда видны всего четыре НКА и все они находятся практи-

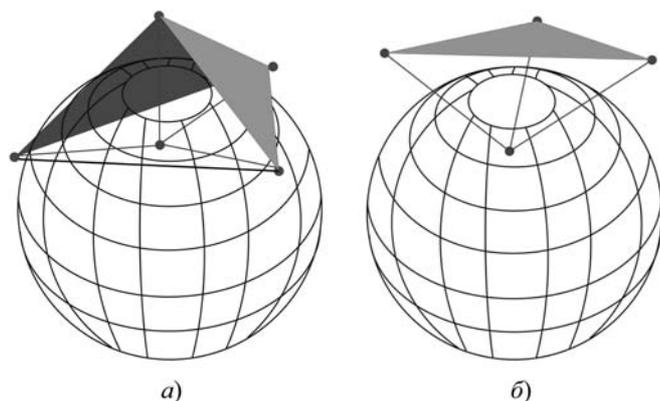


Рис. 1. Конфигурация оптимального рабочего созвездия из четырех спутников (а) и трех спутников (б)

чески в одной плоскости. Тогда объем становится малым и погрешности вычислений становятся большими. Если число активных НКА в СРНС по тем или иным причинам окажется менее 24, то возникают случаи трех НКА в зоне уверенного приема радионавигационного сигнала. Тогда решение ЗП по четырем НКА также становится невозможным. Но эта задача решается по трем НКА.

Алгебраический метод решения навигационной задачи

Суть алгебраического метода заключается в следующем. Предполагается, что с некоторой периодичностью, в определенные такты времени t_i , где i — номер такта с начала периода СРНС, синхронно каждый НКА передает сигналы, содержащие информацию о времени отправки цифрового пакета и о номере n отправившего его НКА. На основании информации о местоположении спутников, заложенной в альманахе системы, приемник выделяет видимые спутники. После этого выполняется решение системы дифференциальных уравнений в целях нахождения точных геоцентрических координат видимых спутников $X_n(t_i), Y_n(t_i), Z_n(t_i)$. Здесь n — номер видимого спутника из спутниковой радионавигационной системы. Сигналы от каждого спутника доходят до потребителя с некоторой задержкой $\Delta t_{i,n}$. С поправками на скорость распространения сигнала по времени задержки вычисляется псевдодальность от приемника до видимого спутника $S_n(t_i)$.

Для аналитического решения задачи позиционирования требуется найти решение системы уравнений, которые приравнивают квадраты расстояний от НАП до спутника рабочего созвездия к квадрату соответствующей псевдодальности. Нам потребуются данные только от трех НКА. Допустим, что приемник принимает устойчивый сигнал от N спутников, тогда выполняем вычисления по формулам (2), (3), (4), перебирая комбинации сочетаний трех спутников из N , выбираем номера трех

спутников, дающих наибольшее значение площади (4) и, соответственно, наименьшее значение коэффициента погрешности GDOP.

После чего индекс n принимает три разных значения из N . Обозначим их индексами a, b, c . Получим систему алгебраических уравнений для каждого i -го момента времени:

$$\begin{cases} (X_a - x)^2 + (Y_a - y)^2 + (Z_a - z)^2 = S_a^2, \\ (X_b - x)^2 + (Y_b - y)^2 + (Z_b - z)^2 = S_b^2, \\ (X_c - x)^2 + (Y_c - y)^2 + (Z_c - z)^2 = S_c^2. \end{cases} \quad (6)$$

По техническим условиям позиционирования допускается расположение приемника на высоте до 2000 км от поверхности Земли. При этом система уравнений (6) будет иметь два решения. Добавление четвертого уравнения может исключать совместность системы при определенных расположениях спутников созвездия и погрешностях в определении псевдодалности. Поэтому возникает задача выбора решения, максимально близкого к истинному положению НАП, и исключения второго решения — изображения. Эту процедуру позволяют осуществить явные формулы решения ЗП алгебраическим методом.

Составим разности уравнений системы (6) $\{eq[1] - eq[2], eq[1] - eq[3]\}$, получим линейную систему относительно неизвестных координат приемника $\{x, y, z\}$:

$$\begin{cases} (A_1) \cdot x + (A_2) \cdot y + (A_3) \cdot z = (b_1), \\ (B_1) \cdot x + (B_2) \cdot y + (B_3) \cdot z = (b_2). \end{cases} \quad (7)$$

Коэффициентами служат координаты векторов, определенные в формуле (2), — стороны треугольника. Линия пересечения плоскостей (7) определяется вектором $\vec{a} = \vec{A}ab \times \vec{B}ac$, вычисляемого по формуле (3).

Решение системы (7) в параметрическом виде содержит множитель $|\vec{a}|^{-2}$, который и определяет погрешность вычислений.

Однако для уменьшения объема вычислений следует в качестве независимого параметра взять одну из координат НАП, а две другие зависимые координаты вычислить из системы (4).

Для выбора независимой координаты существует простой критерий: за независимую координату выбираем ту, на которую модуль проекции вектора \vec{a} наибольший.

Оставшуюся после решения системы (7) независимую переменную вычисляем, используя одно уравнение системы (6). Выбор уравнения из трех проводится на основе аналогичного критерия: проекция какого из трех радиус-векторов $\mathbf{R}_n\{X_n(t_i), Y_n(t_i), Z_n(t_i)\}$ выбранных НКА на направление вектора нормали к площади треугольника \vec{a} максимальна по моду-

лю, уравнение с таким радиус-вектором и выбираем для решения.

При решении выбранного по указанному критерию квадратного уравнения из системы (6) появляются два корня. Определить истинное и отсеять дуальное изображение можно следующим образом. Легко заметить из технических условий применения системы ГЛОНАСС, что нам нужно определять положения НАП, которые расположены ниже плоскости треугольника, с вершинами в выбранных трех НКА. Явные формулы решения квадратного уравнения легко позволяют реализовать этот критерий.

На рис. 2 изображены три сферы с центрами в трех НКА рабочего созвездия и радиусами, равными их псевдодалностям. Малая сфера изображает земной шар. Стрелками указаны две точки пересечения трех сфер. Плоскость, проходящая через три центра сфер, разделяет эти точки. Одна из них всегда ближе к центру масс Земли, чем вторая.

Решение в явном виде имеет вид:

$$D = \sqrt{b^2 - 4ac}; x_1 = \frac{-b + D}{2a}; x_2 = \frac{-b - D}{2a}. \quad (8)$$

Так как в качестве независимого параметра мы выбрали ту координату, проекция которой на соответствующую ось вектора нормали к плоскости треугольника максимальна, то истинному решению соответствует наименьшее из двух значений модуля $|x_1|$ или $|x_2|$. Это будет то решение, в котором в числителе стоит разность. Если окажется $b > 0$, то выбираем в качестве решения x_1 . Если $b \leq 0$, то $-x_2$.

Выводы

Существование решения в итерационном методе обеспечивается выпуклостью функционала — сумма квадратов расстояний от НКА до приемника должна наименьшим образом отличаться от суммы квадратов псевдодалностей. Однако, как показывает математическое моделирование, проведенное нами, в ряде случаев итерационные методы дают неоправданно большие погрешности, иногда метод дает расходящиеся итерации, в то время как аналитические методы позволяют успешно решать задачу позиционирования. В пользу аналитических методов можно отнести следующие преимущества.

1. Навигационную задачу можно решать, используя рабочее созвездие, состоящее из трех спутников, а не четырех. Это может оказаться важным

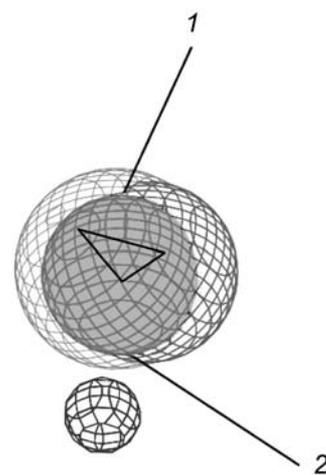


Рис. 2. Иллюстрация точек пересечения трех сфер

в случае выхода одного или нескольких спутников группировки из строя.

2. Значения факторов геометрического снижения точности при аналитическом методе решения навигационной задачи обратно пропорциональны площади треугольника (см. рис. 1, б), в то время как при использовании итерационного подхода значения геометрического снижения точности обратно пропорциональны объему четырехвершинной пирамиды (см. рис. 1, а). Подобрать треугольник наибольшей площади гораздо проще, чем пирамиду наибольшего объема. В случае с треугольником неблагоприятной является ситуация, когда все три спутника находятся на одной прямой (нулевая высота треугольника), в случае с пирамидой неблагоприятной будет ситуация, когда спутники располагаются в одной плоскости — пирамида будет иметь

нулевой объем вследствие нулевой высоты. Моделирование системы ГЛОНАСС показывает, что ситуация с большой погрешностью вычислений по четырем спутникам возникает гораздо чаще, чем аналогичная ситуация с тремя спутниками.

3. Алгебраический метод позволяет вести вычисления с контролируемой точностью и надежностью. Его можно применить для решения задачи позиционирования быстро движущихся приемников, когда прием осуществляется не в одной точке траектории.

Список литературы

1. ГОСТ Р 52865—2009. Глобальная навигационная спутниковая система. Параметры радионавигационного поля. Технические требования и методы испытаний. URL: <http://protect.gost.ru/document.aspx?control=7&baseC=6&page=0&month=7&year=-1&search=&id=176110>

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

УДК 519.68:15:681.5

В. М. Гриняк, канд. техн. наук, зав. каф.,
Владивостокский государственный
университет экономики и сервиса,
e-mail: Victor.Grinyak@vvsu.ru,

А. С. Девятисильный,
д-р техн. наук, проф., гл. науч. сотр.,
Институт автоматизации и процессов управления
ДВО РАН,
e-mail: devyatish@iacp.dvo.ru

Нечеткая система экспертных оценок характера движения судна*

Рассматривается проблема обнаружения маневра объекта в современных системах управления движением на море при обработке навигационных данных двухкоординатной РЛС кругового обзора. Обсуждается задача оценки интенсивности маневрирования, связанная с выработкой тревожных сигналов и принятием управленческих решений в целях обеспечения безопасности коллективного движения судов. Предлагается нечеткая модельная интерпретация традиционных алгоритмов оптимальной фильтрации, основанная на машине нечеткого вывода типа Мамдани, предоставляющая возможность вербальной обобщенной оценки маневренности судна операторами СУДС и судоводителями.

Ключевые слова: управление движением судов, обнаружитель маневра, сопровождение объекта, нечеткая система типа Мамдани, оценка траектории движения

Введение

Обеспечение навигационной безопасности движения судов является актуальной проблемой эксплуатации водных транспортных путей. В зонах высокой интенсивности движения решение этой проблемы возложено на особые информационные средства — бортовые и береговые системы управления движением судов (СУДС) [1, 2]. В соответствии с общепринятой концепцией построения таких систем их задачи реализуются с использованием измерительной информации, доставляемой радарными и/или спутниковыми средствами траекторных измерений — транспондерами.

Сложившаяся судоводительская практика, правила которой прямо регламентируют, что управление судном есть исключительное право его капитана, отводит СУДС (как береговому, так и бортовому) роль особого инструмента информирования судоводителя о возможном наступлении опасной ситуации (столкновении) [3]. Генерация тревожного сигнала по какому-либо объекту или их группе служит указанием судоводителю (и/или оператору СУДС), на основании которого он принимает решение об изменении курса и скорости движения.

Обращение к автоматизированным средствам информационного обеспечения требует предельно формализованных представлений понятия "опас-

* Работа выполнена в рамках Государственного задания высшим учебным заведениям в части проведения научно-исследовательских работ, проект № 7.2104.2011.

ная ситуация", и здесь следует обратиться к анализу опыта практического судовождения, показывающего, что главным условием безопасного движения является недопущение чрезмерного сближения судов. При таком подходе к интерпретации опасности ее формальным критерием служит уменьшение расстояния между объектами до некоторого критического, определяющего своего рода "зону безопасности" вокруг судна (корабельный домен) [4].

Маневрирующие и неманеврирующие объекты с точки зрения оценки безопасности имеют ряд принципиальных различий [5, 6]. Во-первых, при внешнем наблюдении полностью достоверный прогноз траектории маневрирующего объекта невозможен. Во-вторых, если исходить из принятого на практике положения, что маневрирование судна, как правило, свидетельствует о попытке судоводителя придать движению безопасный характер и о его контроле над ситуацией, то для маневрирующих объектов вербальный уровень опасности заведомо ниже, чем для неманеврирующих. Это является побудительным мотивом разработки таких информационных моделей выработки тревожных сигналов, которые выделяли бы различные уровни опасности ситуации типа "очень опасная", "опасная", "почти безопасная" и т. п., учитывая при этом маневренные характеристики траектории движения судна.

Специальные алгоритмы обнаружения маневра нередко реализуются на практике при решении задач сопровождения траектории и наблюдения [7]. Они используются для "переключения" параметров системы на участках маневрирования объекта, характеризующихся несоответствием между моделируемым и реальным движением. Вместе с тем, известные обнаружители маневра предназначены для автоматизированных систем и неудобны для поддержки вербального принятия решений операторами СУДС и судоводителями.

В настоящей работе рассматривается новый подход к обнаружению маневра, связанный с классификацией наблюдаемых объектов по степени интенсивности маневрирования с помощью аппарата нечеткой логики. Получаемые таким образом обобщенные представления о характере движения того или иного судна используются при распознавании опасных ситуаций бортовой или береговой СУДС.

Основные модельные представления и постановка задачи

Пусть движение объекта описывается следующими уравнениями:

$$\begin{aligned} x(t_{k+1}) &= x(t_k) + v_x(t_k)\tau + q_x(t_k); \\ y(t_{k+1}) &= y(t_k) + v_y(t_k)\tau + q_y(t_k), \end{aligned} \quad (1)$$

где k — идентификатор (порядковый номер) момента времени, $x(t_k)$, $y(t_k)$ — координаты объекта в момент времени t_k ; $v_x(t_k)$, $v_y(t_k)$ — компоненты век-

тора скорости объекта, $q_x(t_k)$, $q_y(t_k)$ — компоненты вектора случайных немоделируемых параметров движения; $\tau = t_{k+1} - t_k$.

Пусть измеряемыми параметрами являются декартовы координаты объекта. Тогда модель рассматриваемой задачи можно представить следующим дискретным матричным уравнением "состояние—измерение":

$$\begin{aligned} s(t_{k+1}) &= \Phi s(t_k) + q(t_k), \\ z(t_k) &= Hs(t_k) + r(t_k). \end{aligned} \quad (2)$$

Здесь $s(t_k) = (x(t_k), v_x(t_k), y(t_k), v_y(t_k))^T$ — вектор состояния объекта, включающий его координаты и их производные (t — символ транспонирования); $q(t_k)$ — вектор немоделируемых параметров движения; $z(t_k)$ — вектор измерений; $r(t_k)$ — вектор погрешностей измерений. С учетом (1) матричные коэффициенты Φ и H системы уравнений (2) определяются формулами

$$\Phi = \begin{bmatrix} 1 & \tau & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Модель оценивания вектора состояния $s(t_k)$ по измерениям $z(t_k)$ может быть представлена следующим уравнением:

$$\hat{s}(t_{k+1}) = \Phi \hat{s}(t_k) + K(z(t_k) - H\Phi \hat{s}(t_k)), \quad (3)$$

где $\hat{s}(t_k)$ — оценка вектора состояния; K — матричный коэффициент.

Пусть матрица K определяется схемой классического $\alpha - \beta$ -алгоритма [8] и имеет вид

$$K = \begin{bmatrix} \alpha & 0 \\ \beta/\tau & 0 \\ 0 & \alpha \\ 0 & \beta/\tau \end{bmatrix}, \quad (4)$$

а коэффициенты α и β выбираются по следующему правилу:

$$\alpha_k = \frac{2(2k+1)}{(k+2)(k+1)}; \quad \beta_k = \frac{6}{(k+2)(k+1)},$$

где k — порядковый номер момента времени в формуле (3).

Пусть J — число измерений (и, соответственно, итераций), участвующих в оценке вектора состояния $s(t_k)$ итерационной процедурой (3), так что $k = 1, \dots, J$. При увеличении J коэффициенты α и β асимптотически уменьшаются до 0. Поэтому алгоритм (3), реализованный с большим J , будет успешно оценивать координаты и скорости объектов, движущихся прямолинейно и равномерно, а для маневрирующих объектов погрешность оценки вектора состояния будет довольно высокой.

Пусть $\hat{s}_J(t_i)$ — оценка вектора состояния системы в момент времени t_i , полученная итерационным алгоритмом (3) при обработке J последних измерений. Если при этом задача одновременно решается при $J, J-1, J-2, \dots$, и, наконец, только при двух измерениях (минимально возможном их числе), то тогда в момент времени t_i будем иметь кортеж векторов оценки

$$\hat{S}_J(t_i) = \{\hat{s}_2(t_i), \hat{s}_3(t_i), \hat{s}_4(t_i), \dots, \hat{s}_J(t_i)\}. \quad (5)$$

Введем вектор $\delta z(t_{k+1}) = z(t_{k+1}) - H\hat{s}(t_{k+1})$, характеризующий невязку измерения при оценке вектора состояния уравнением (3). Пусть $\|\delta z_j(t_i)\|$ — евклидова норма вектора невязки $\delta z(t_i)$, полученного в момент времени t_i при реализации итерационного алгоритма (3), обрабатывающего J последних измерений. Тогда при оценке вектора состояния в каждый момент времени t_i наряду с кортежем векторов оценки (5) будем иметь кортеж норм векторов невязок

$$\delta_j(t_i) = \{\|\delta z_2(t_i)\|, \|\delta z_3(t_i)\|, \|\delta z_4(t_i)\|, \dots, \|\delta z_J(t_i)\|\}. \quad (6)$$

Элементы кортежа (6) являются, по сути, основным информативным признаком, характеризующим качество оценки вектора состояния алгоритмом (3)—(4) с тем или иным значением J . Задача об обнаружении маневра объекта и оценки его интенсивности сводится, таким образом, к анализу свойств кортежа (6).

Метод решения задачи

Перейдем от формулы (6) к кортежу относительных величин

$$\Delta_j(t_i) = \{L_2(t_i), L_3(t_i), L_4(t_i), \dots, L_J(t_i)\}, \quad (7)$$

где $L_j(t_i) = \frac{\|\delta z_j(t_i)\|}{\sigma}$, σ — величина, характеризующая среднеквадратичное отклонение погрешности измерений $r(t_k)$ в системе (2).

Введем лингвистическую переменную $Q_j(t_i)$, $j = 2, \bar{J}$, "Качество оценки вектора состояния алгоритмом (3)—(4) в момент времени t_i по j последним измерениям" с термами "good" (g, "хорошее") и "bad" (b, "плохое"). Пусть термы имеют следующие функции принадлежности типа "дополнение", определенные на универсальном множестве $u \in [0, 3]$:

$$\begin{aligned} \mu_g(u) &= 1 - \frac{1}{1 + \exp(-a_1(u - c_1))}; \\ \mu_b(u) &= \frac{1}{1 + \exp(-a_2(u - c_2))}, \end{aligned} \quad (8)$$

где a_1, a_2, c_1, c_2 — настраиваемые параметры.

Введем лингвистическую переменную $P(t_i)$ "Характер движения судна в момент времени t_i " с термами "high-high-maneuverable" (hhm, "очень высокоманевренное"), "high-maneuverable" (hm, "высоко-

маневренное"), "low-maneuverable" (lm, "низкоманевренное") и "low-low-maneuverable" (llm, "очень низкоманевренное"). Пусть термы имеют следующие функции принадлежности типа "кластер", определенные на универсальном множестве $v \in [2, J]$:

$$\begin{aligned} \mu_{hhm}(v) &= 1 - \frac{1}{1 + \exp(-a_3(v - c_3))}; \\ \mu_{hm}(v) &= \exp\left(-\frac{(v - c_4)^2}{a_4}\right); \\ \mu_{lm}(v) &= \exp\left(-\frac{(v - c_5)^2}{a_5}\right); \\ \mu_{llm}(v) &= \frac{1}{1 + \exp(-a_6(v - c_6))}, \end{aligned} \quad (9)$$

где a_3-a_6, c_3-c_6 — настраиваемые параметры.

Пусть переменные $Q_j(t_i)$ обрабатываются машинной нечеткой выводом типа Мамдани [9], на вход которой подается кортеж величин (7), а на выходе формируется числовое значение $m(t_i)$ — вещественное число, характеризующее степень интенсивности маневрирования судна. Машина нечеткого вывода работает согласно системе правил, представленной в табл. 1.

Работу нечеткого алгоритма оценки степени интенсивности маневрирования судна можно, таким образом, окончательно представить двухуровневой иерархической схемой, показанной на рис. 1. Здесь

Таблица 1
Система правил машины нечеткого вывода типа Мамдани (уровень 1)

№	$Q_2(t_i)$	$Q_3(t_i)$	$Q_4(t_i)$...	$Q_{J-2}(t_i)$	$Q_{J-1}(t_i)$	$Q_J(t_i)$	$P(t_i)$
1	g	g	g	...	g	g	g	P_1
2	g	g	g	...	g	g	b	P_2
3	g	g	g	...	g	b	b	P_3
...
$J-1$	g	b	b	...	b	b	b	P_{J-1}
J	b	b	b	...	b	b	b	P_J

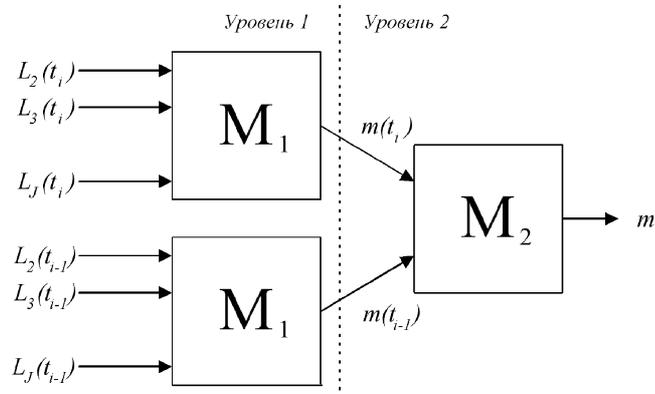


Рис. 1. Схема работы нечеткого алгоритма оценки степени интенсивности маневрирования судна

Таблица 2
Система правил машины нечеткого вывода типа Мамдани (уровень 2)

№	$P(t_{i-1})$ (вход 1)	$P(t_i)$ (вход 2)	$P(t_i)$ (выход)
1	hbm	hbm	hbm
2	hm	hbm	hm
3	lm	hbm	lm
4	llm	hbm	llm
5	hbm	hm	hm
6	hm	hm	hm
7	lm	hm	lm
8	llm	hm	llm
9	hbm	lm	lm
10	hm	lm	lm
11	lm	lm	lm
12	llm	lm	llm
13	hbm	llm	llm
14	hm	llm	llm
15	lm	llm	llm
16	llm	llm	llm

$L_j(t_i)$ — величины кортежа (7) в момент времени t_i (вход), если $L_j(t_i) > 3$, то вход принимается равным 3; $m(t_i)$ — определенная системой типа Мамдани M_1 (уровень 1) в момент времени t_i степень интенсивности маневрирования судна.

Величины $m(t_i)$ и $m(t_{i-1})$ подаются на вход системы типа Мамдани M_2 (уровень 2). На выходе этой системы формируется числовое значение m , которое и принимается за окончательное значение степени интенсивности маневрирования судна. Машина нечеткого вывода M_2 работает согласно системе правил, представленной в табл. 2.

Обработка двух соседних величин $m(t_{i-1})$ и $m(t_i)$ машиной нечеткого вывода M_2 необходима для повышения устойчивости работы системы при больших ошибках измерений (для фильтрации случайных выбросов).

Настройка описанной системы состоит в задании максимального числа измерений J , параметров функций принадлежности $a_1, \dots, a_6, c_1, \dots, c_6$, значений лингвистической переменной P_i и величины σ , характеризующей погрешность измерений.

Результаты численного моделирования

При моделировании задачи было принято, что информационной базой СУДС является двухкоординатный радар кругового обзора (например, типа Raytheon) с периодом обращения 3с и разрешением по углу и дальности соответственно $\Delta\varphi = 0,03^\circ$ и $\Delta r = 6$ м. Максимальное число измерений было принято равным $J = 10$. Принятые значения лингвистической переменной P_i приведены в табл. 3.

Заданные значения параметров функций принадлежности (8) и (9) приведены в табл. 4 (в данном случае параметры задаются экспертом, система не подвергается настройке на обучающей выборке, см. также рис. 2 и 3).

Если вероятностные характеристики ошибок измерений r_k хорошо известны, то величина σ может задаваться априорно. В тех случаях, когда r_k можно оценить только приблизительно, с точностью до порядка величин, величина σ может быть оценена формулой:

$$\hat{\sigma}_k = \frac{\sum_{i=1}^k \|\delta z_2(t_i)\|}{k},$$

где k — порядковый номер момента времени, прошедшего от начала наблюдения судна (в настоящей работе рассматривается именно этот случай).

Таблица 3
Значения лингвистической переменной P_i

i	1	2	3	4	5	6	7	8	9	10
P_i	llm	llm	lm	lm	lm	hm	hm	hm	hbm	hbm

Таблица 4
Значения параметров функций принадлежности

i	a_i	c_i
1	5,0	1,5
2	5,0	1,5
3	6,0	3,0
4	2,0	4,5
5	2,0	7,5
6	6,0	9,0

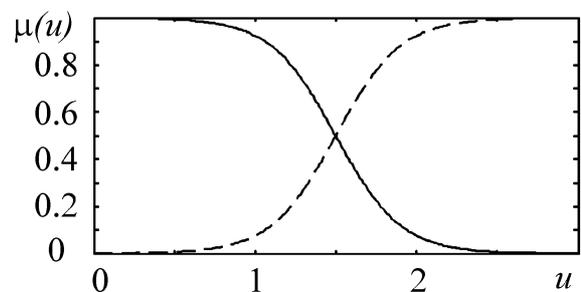


Рис. 2. Функции принадлежности термов "good" (сплошная линия) и "bad" (штриховая линия)

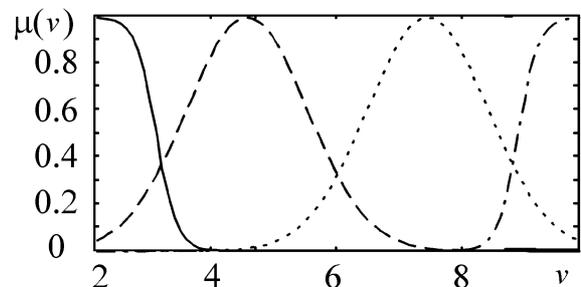


Рис. 3. Функции принадлежности термов "hbm" (сплошная линия), "hm" (штриховая линия), "lm" (пунктирная линия) и "llm" (штрих-пунктирная линия)

На рис. 4 показана моделируемая траектория движения судна. Вначале судно движется прямолинейно и равномерно, а затем совершает маневр — поворот с радиусом 300 м (такие кинематические свойства вполне характерны для современных судов небольшой размерности).

На рис. 5 показан результат решения задачи оценки степени интенсивности маневрирования для судна, движущегося по изображенной траектории со скоростью 10 м/с (левая колонка рисунков) и 20 м/с (правая колонка рисунков). Здесь t — время, прошедшее от начала работы алгоритма, m — определенное по мере движения судна значение степени интенсивности маневрирования судна (рис. 5, а и 5, б). В данном случае m близко к максимальному значению (около 10) на прямолинейном участке траектории и уменьшается до значений $\approx 3,5 \dots 5$ при повороте на скорости 10 м/с и до $\approx 2,5$ при повороте на скорости 20 м/с. При этом алгоритм достаточно быстро реагирует на изменение характера движения судна (не более 15 с, участок [63, 78] с на рис. 5, б).

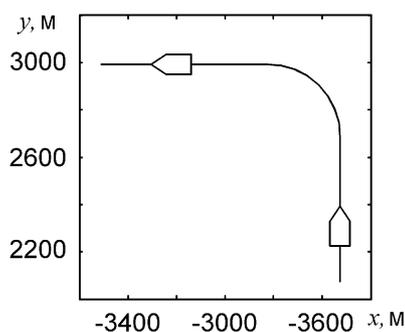


Рис. 4. Траектория движения судна

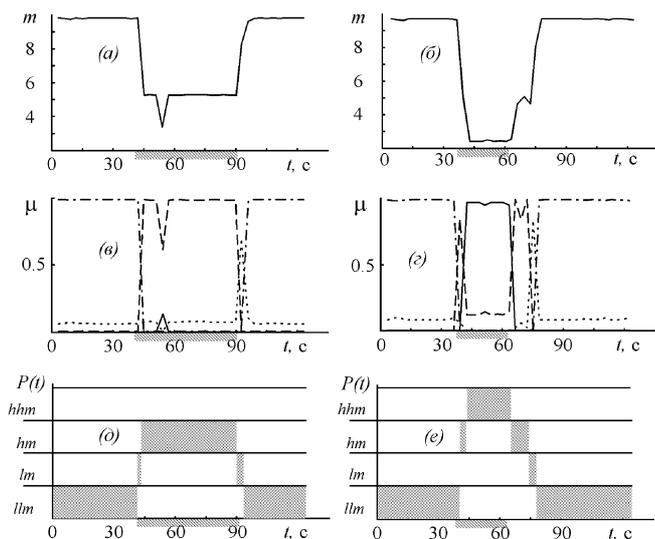


Рис. 5. Работа алгоритма нечеткой оценки степени маневрирования судна; штриховой областью и снизу оси абсцисс показан участок маневрирования судна

Рис. 5, в и 5, г показывают значения функций принадлежности термов "hhm" (сплошная линия), "hm" (штриховая линия), "lm" (пунктирная линия) и "llm" (штрих-пунктирная линия) по мере движения судна. Так, при движении со скоростью 10 м/с на участке [0, 40] с движение с наивысшей степенью оценено как "очень низкоманевренное", на участке (40, 43) с — как "низкоманевренное", на участке [43, 90] с — как "высокоманевренное" и т. д. (рис. 5, в). Соответствующие термы с максимальными значениями функции принадлежности приведены на рис. 5, д и 5, е, которые иллюстрируют обобщенную вербальную интерпретацию траекторных свойств движения в разрезе значений "очень высокоманевренное", "высокоманевренное", "низкоманевренное" и "очень низкоманевренное".

Заключение

В заключение перечислим основные результаты настоящей статьи. В работе сформулированы методологические основы и дана концепция решения задачи оценки интенсивности маневрирования. Рассмотрена нечеткая модель классификации характера движения судна, ориентированная на реализацию в качестве обнаружителя маневра. Предложен нечеткий алгоритм оценки степени маневра, основанный на сопровождении траектории судна классическим $\alpha - \beta$ -алгоритмом с различной степенью фильтрации и обработке параметров сопровождения машиной нечеткого вывода типа Мамдани. На модельном примере продемонстрирована конструктивность и эффективность предлагаемой методики.

Результаты исследования ориентированы на расширение функциональных возможностей существующих береговых и бортовых систем управления движением судов.

Список литературы

1. **ОАО Норфес.** URL: <http://www.norfes.ru/>
2. **Группа компаний ТРАНЗАС.** URL: <http://www.transas.ru/>
3. **Huges T.** When is a VTS not a VTS // The J. of Navigation. 2009. Vol. 62, N 3. P. 439—442.
4. **Pietrzykowski Z. and Uriasz J.** The Ship Domain — a Criterion of Navigational Safety Assessment in an Open Sea Area // The J. of Navigation. 2009. Vol. 62, N 1. P. 93—108.
5. **Девятисильный А. С., Гриняк В. М.** Прогнозирование опасных ситуаций при управлении движением на море // Изв. РАН. Теория и системы управления. 2004. № 3. С. 127—136.
6. **Девятисильный А. С., Дорожко В. М., Гриняк В. М.** и др. Система экспертных оценок состояния безопасности на морских акваториях // Информационные технологии. 2004. № 11. С. 48—53.
7. **Бакулев П. А., Сычев М. И., Нгуен Чонг Лью.** Многомодельный алгоритм сопровождения траектории маневрирующей цели по данным обзорной РЛС // Радиотехника. 2004. № 1.
8. **Benedict T. R., Bordner G. R.** Synthesis of an optimal set of radar track-while-scan smoothing equations // IRE Trans, on AC-1, July 1962. P. 27—32.
9. **Круглов В. В., Дли М. И., Голунов Р. Ю.** Нечеткая логика и искусственные нейронные сети. М.: Физматлит, 2001. 224 с.

CONTENTS

Kazakov P. V. *Performance Assessment of the Genetic Algorithms for Multi-Objective Optimization. Part 1.* 2

The manners for performance assessment of the genetic algorithms for multi-objective optimization are considered. The set of special indicators for quantitative assessment of quality Pareto sets is noted.

Keywords: multi-objective optimization, Pareto's principles, Pareto front, performance indicators, multi-objective genetic algorithms

Eltarenko E. A. *Conjunctive Operators of Aggregation in Multicriteria Problems* 7

The class of conjunctive preferences is considered. For it axioms of preferences are formulated. Questions of check of axioms in preferences of the person, the making decision are analyzed. Questions of generation of operators for the description of conjunctive preferences are stated. Examples of conjunctive operators are resulted.

Keywords: decision-making, multicriteria problems, aggregation of criteria, the description of preferences, conjunctive operators

Zack Yu. A. *The Distribution of Multiple Tasks and to Determine the Optimal Sequence of Their Execution on Parallel Machines Using Dynamic Programming.* 14

A mathematical model and study the properties of the tasks of building the schedule for the parallel and different specifications vehicles in allowance for losses on the conversion machines, restrictions on the execution times of tasks and allowable time of the machine. Operators are constructed excluding from consideration of subsets of schedules that do not contain valid and non-optimal solutions. On the basis of established properties considered algorithms for solving these problems using dynamic programming.

Keywords: scheduling on parallel machines, the losses on conversion, restrictions on deadlines and time machines, dynamic programming

Trudonoshin V. A., Fedoruk V. G. *Comparison of Mathematical Models Hinge* 20

In the article there are considered three variants of construction of mathematical model of the hinge when modeling of two-dimensional mechanical systems. The peculiarities of these models have been met, the numerical experiments, recommendations are given for application.

Keywords: mathematical simulation, basis, model, hinge, machinery, dynamic

Norenkov I. P. Uvarov M. Yu., Grushin M. A., Oveshnicov A. A. *Document Classification by Ontology Clasterization* 24

The method of ontology clasterization requires an application of particular ontologies, corresponded to some applications. The paper is devoted to automatic creating particular ontologies on the base of a general ontology of knowledge field. There are experimental data concerned to efficiency of document classification by ontology clasterization method and forming particular ontologies.

Keywords: ontology, document classification, ontology clasterization

Andreichicova O. N., Andreichicov A. V. *What will be with Copyright in Internet? The Search of Decision with Analytic Network Process* 28

The Analytic Network Process (ANP) is a powerful tool for examining difficult decisions under uncertainty. This paper explores the urgent decision connected with the copyrights on Internet. This theme is being widely discussed, but nobody knows, what to do. We have attempted to do a multicriteria analysis of the problem, taking into consideration benefits, costs and risks. As alternatives we examined different measures for the regulation of Internet-content, and as main actors — content creators, customers and providers. The resulting outcomes show that free use of Internet-content is the best alternative. Measures related to punishment turn out to have negative prospects.

Keywords: ANP, copyright, Internet, content, priority

Baranov D. V. *Logical Equations on the Equational LP-Structures.* 35

The theory of rewrite systems is an effective tool for knowledge formalization in a number of important fields of artificial intelligence and computer algebra. This paper defines an algebraic system with the semantics of a set of rules of conditional equational theory, or a conditional term rewriting system. For the given model a special class of logical equations is introduced and studied. The results can be applied to verification of the corresponding set of rules.

Keywords: terms, equational theory, conditional rules, algebraic model, logical equations

Barsky A. B. *Two- and Three-Computer Clusterization of Grid Resources for the Optimized Results Aggregation of the Distributed Calculations on Base of SPMD Technology* 43

The aggregation of the results of Grid computing on the head processor is studied on the basis of the cluster tree. The problem is reduced to the examination of clusters, containing two or three of the processor. The inclusion of the simplest modules exchange in the SPMD program is discussed.

Keywords: Grid-calculations, SPMD-technology, information exchange, elementary clusters

Sologub R. A. *Volatility Surface Regression Model Generation* 47

The problem of optimal structure model generation in nonlinear regression is proposed. The set of competing models is generated, the data set is divided into learning set and testing set, the parameters of the models are estimated using the learning set. The model, that maximises the determination coefficient on the testing set is chosen. The models of the different classes are generated and researched: linear models, neural nets, nonlinear models. The uniform model generation algorithm is proposed. As an illustration of the algorithm, the regression problem of the volatility surface is solved.

Keywords: nonlinear regression, symbolic regression, inductive generation, model complexity, options

Galiev D. R., Isavnin A. G. *Optimal Portfolio Selection with Fuzzy Returns* 52

Problems of optimal portfolio selection with fuzzy returns have been investigated. That approach allows using wide range of expert views. First we consider two portfolio selection models with fuzzy returns and different risk measures. At the next stage we describe our model with fuzzy returns and combination of risk measures. The results of numerical examples on Russian stock exchange markets are presented. We also depict the software, developed during the study, which implements methods and models described in the paper.

Keywords: optimal portfolio, Value-at-Risk, Drawdown-at-risk, fuzzy sets, programming

Agievich S. N. *Theoretical Bases of Signals Spline-Algebraic-Harmonic Analysis of Radio Communication Systems* 58

Possibility of usage of functions of splines-characters for the analysis and synthesis of signals is considered. The mathematical apparatus of theoretical bases a harmonic analysis spline-algebraic is represented. Concepts of linearity, shift, convolution, correlation, invariancy, Fourier transform for splines-functions are defined. Their applicability for base of functions of splines-characters is proved.

Keywords: spline-character, base functions, shaping signals, theoretical bases

Sakharov A. V., Bogdanov M. R., Muckhamedjanov A. R. *Analytical Approach for Solving of GLONASS/GPS Navigation Problem*. 63

Analytical approach for solving of GLONASS/GPS Navigation Problem is discussed. Hidden problems of iterative approach for solving of positioning problem are discovered. Methods for solving above mentioned problems are given. Algebraic method for solution of the navigating problem is given. The algorithms were realized with Maple 13.

Keywords: navigating problem, GLONASS, GPS, positioning, geometric dilution of precision

Grinyak V. M., Devyatitsilny A. S. *Fussy Logic Expert System of Vessels Maneuverability* 66

The problem of maneuver detection in modern vessel traffic systems based on two-coordinate radar is watched. Maneuver intensity estimation for marine traffic alarm signals generation is discussed. Fuzzy logic model based on optimal filtering and Mamdani fuzzy logic system giving possibility of a verbal generalized vessels maneuverability estimation by VTS operators and navigators is offered.

Keywords: vessel traffic control, maneuver detection, target tracking, Mamdani fuzzy logic system, trace estimation

Адрес редакции:

107076, Москва, Стромьинский пер., 4

Телефон редакции журнала **(499) 269-5510**

E-mail: it@novtex.ru

Дизайнер *Т.Н. Погорелова*. Технический редактор *Е. В. Конова*.

Корректор *З.В. Наумова*.

Сдано в набор 04.06.2012. Подписано в печать 20.07.2012. Формат 60×88 1/8. Бумага офсетная.

Усл. печ. л. 8,86. Заказ ИТ812. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Оригинал-макет ООО "Авансед солюшнз". Отпечатано в ООО "Авансед солюшнз".

105120, г. Москва, ул. Нижняя Сыромятническая, д. 5/7, стр. 2, офис 2.
