

Заключение

Для практического использования рассмотренных моделей и алгоритмов необходимо исследование реальных процессов и получение соответствующих реальных данных из практики. Однако возможность применения динамического программирования для составления оптимальных планов реализации частично возобновляемых ресурсов сомнений не вызывает.

В данной задаче возможно существенное сокращение объема вычислений при использовании более совершенной реализации классического алго-

ритма динамического программирования за счет использования множеств Парето [3, 4].

Список литературы

1. **Беллман Р.** Динамическое программирование. М.: ИЛ, 1960.
2. **Беллман Р., Дрейфус С.** Прикладные задачи динамического программирования. М.: Наука, 1965. 458 с.
3. **Струченков В. И.** Устаревшие стереотипы и новые алгоритмы решения прикладных задач дискретной оптимизации // Информационные технологии. 2012. № 5.
4. **Struchenkov V. I.** Combined Algorithms of Optimal Resource Allocation // Applied Mathematics. Scientific Research. 2012. Vol. 3, N. 1.
5. **Косоруков О. А., Мищенко А. В.** Исследование операций: Учеб. для вузов. М.: Экзамен, 2003. 270 с.

V. I. Struchenkov, Professor, e-mail: str1942@mail.ru

Moscow State University of Radio Engineering, Electronics and Automation, Moscow, Russia

Dynamic Programming for Planning of the Partially Renewable Resources Implementation

Under study is the problem of optimal planning of the partially renewable resources implementation, such as the commercial breeding of fish, animals and so on.

The aim of this article is to study the opportunity of the optimal plan calculation using dynamic programming.

The predetermined scheduling period is divided into a number of stages, such as months or years. At each stage, the key concept of dynamic programming "system status" is formalized as the amount of available resources.

Two models of the objective function are considered. The first model is linear, and the second model is quadratic. On the basis of the dynamic programming method built computational algorithms that do not require busting options step by step. Instead of these classical algorithms of dynamical programming the recurrence formulas for optimal plan calculation are received.

Keywords: resource, the objective function, the set of states, dynamic programming, the optimal path

References

1. **Bellman R.** *Dinamicheskoe programirovanie* (Dynamic Programming), Moscow, Foreign Literature, 1960. 234 p. (in Russian).
2. **Bellman R., Drejfus S.** *Prikladnye zadachi dinamicheskogo programirovanija* (Applied Problems of Dynamic Programming), Moscow, Science, 1965, 458 p. (in Russian).

3. **Struchenkov V. I.** Ustarevshie stereotipy i novye algoritmy reshenija prikladnyh zadach diskretnoi optimizacii, *Informacionnye tehnologii*, 2012, no. 5, pp. 20–29 (in Russian).
4. **Struchenkov V. I.** Combined Algorithms of Optimal Resource Allocation. *Applied Mathematics. Scientific Research*, 2012, vol. 3, no. 1.
5. **Kosorukov O. A., Mishenko A. V.** *Issledovanie operacij: Uchebnik dlja studentov vuzov*. Moscow: Exam, 2003, 270 p. (in Russian).

УДК 004.62

Н. М. Новикова, д-р техн. наук, проф., e-mail: nov.nelly@gmail.com,

А. В. Борискин, аспирант, email: boriskinpost@gmail.com,

Воронежский государственный университет

Математическая модель поиска влиятельных объектов социальной сети на основе априорной информации

Для определенного подмножества объектов социальной сети (на примере Vkontakte) некоторые объекты обладают повышенным влиянием на другие объекты. Рассматривается задача поиска остальных влиятельных объектов на основе данной информации. Предложен алгоритм решения, приведены иллюстрации работы разработанного программного обеспечения.

Ключевые слова: виртуальные социальные сети, интеллектуальный анализ данных, нечеткие графы, методы оптимизации

Введение

В 1992 г. эксперт по бизнес-анализу Григорий Пятецкий-Шапиро ввел термин "data mining" [1]. Дословный перевод "добыча данных" вряд ли следует считать удачным. Наиболее приемлемая версия звучит как "интеллектуальный анализ данных". Под ним скрывается процесс обнаружения в "сырых" данных ранее неизвестных, нетривиальных, практически полезных, доступных интерпретаций знаний, необходимых для принятия решений в различных сферах человеческой деятельности [2]. В современном информационном мире "сырых" данных встречается достаточно много. Наиболее богатыми источниками являются виртуальные социальные сети. В них участники выкладывают на стены записи, фотографии, видео, снабжают их комментариями и оценками, помечают друг друга как друзья. Происходит практически полноценное общение, только в виртуальном формате. Целью данной статьи является обработка этих данных с последующим получением полезных результатов.

Постановка задачи

Пусть есть множество пользователей социальной сети, содержащее людей разных профессий. В нем могут быть подмножества врачей, программистов, писателей, кинокритиков, фотографов и многих других с иными интересами. Представим, что необходимо распространить информацию среди определенной группы пользователей, например, информацию о выходе нового объектива, который позволяет запечатлеть самые сложные моменты спортивных событий. Очевидно, что эти данные не представляют никакого интереса для людей, которые занимаются ремонтом автомобилей или строительством небоскребов. Данная информация должна дойти до некоторой группы фотографов, которая интересуется съемкой именно спортивных событий. В большинстве случаев в подобных группах существуют люди, которые пользуются авторитетом, к мнению которых прислушиваются. Через таких людей можно начинать распространять информацию. В дальнейшем будем называть подобные объекты *влиятельными*. Задача ставится следующим образом: располагая данными о некоторых влиятельных объектах, найти иные объекты, которые также пользуются повышенным влиянием.

Теоретический анализ

Социальную сеть принято представлять в виде графа, узлами которого являются люди [3]. Если объекты как-то связаны между собой (состоят в друзьях либо переписываются), то связь между этими объектами обозначается в виде дуги. Полученный в этом случае граф обладает существенным недостатком: неважно, насколько близко знакомы люди друг с другом, важно лишь, существует ли

между ними хоть какая-то связь. В конечно счете он приравняет отношения между случайными знаковыми к отношениям между людьми, которые тесно общаются уже не один десяток лет. Необходимо иметь модель, которая позволяла бы дифференцировать подобные случаи.

Для решения данной задачи может быть использовано понятие нечеткого графа [4]. В нашем случае под графом, представляющим модель данных, понимается пара (X, A) , где X — четкое множество вершин, представляющих людей в сети, A — нечеткое множество отношений между людьми. Более строго, A представляет собой множество пар (x_i, x_j) , для которых задана функция принадлежности $\mu_A(x, y)$, указывающая, в какой степени объект x напрямую влияет на объект y .

Следует заметить, что это понятие *влиятельности* не сводится только к числу связей с другими объектами. Должны учитываться степени влияния объектов, с которыми связан целевой объект (объект, для которого считается степень влияния). Таким образом, нужно так формализовать понятие влиятельности, чтобы учитывались как число связей, так и влиятельность объектов, с которыми связан целевой объект.

В статье [5] представлен один из алгоритмов решения данной задачи, который обладает вышеуказанным недостатком — бинарностью в отношениях. Ниже представлена модель, которая позволяет обойти это. Пусть $\mu_A(x, y)$ — функция принадлежности, указывающая, в какой степени объект x напрямую влияет на объект y . Принимает значения из отрезка $[0, 1]$.

В [5] вводится понятие *итерированной силы объекта i порядка k* (обозначается $p^i(k)$) через матрицу смежности графа [6]. Вместо матрицы смежности будем использовать функцию принадлежности. Получим

$$\begin{cases} p^i(k) = \sum_{j=1}^n \mu_A(x_i, x_j) p^j(k-1), i = 1, \dots, n, k \geq 1; \\ p^i(0) = 1, i = 1, \dots, n. \end{cases} \quad (1)$$

Заметим, что итерированная сила первого порядка объекта i есть сумма значений функции принадлежности данного объекта с другими. Она еще не учитывает влиятельность других объектов. Начиная со второго порядка, в данную сумму включается влиятельность иных объектов.

Возникает вопрос: до какого порядка считать вектор итерированных сил? Покажем, что из практических соображений достаточно считать до второго либо до третьего порядка. Предполагается, что итерированная сила объекта i порядка k выражает степень влияния объекта i с учетом того, что он свое влияние может распространить не более чем в радиусе k . Это вытекает из определения итерированной силы. Например, если вы просите своего друга попросить кого-то сделать что-то для вас, то

это влияние в радиусе 2, и выражается она итерированной силой второго порядка. Если учитываются цепи влияния с еще одним промежуточным участником, то нужно считать вектор итерированных сил 3-го порядка. Считать итерированные силы больших порядков кажется нецелесообразным ввиду ничтожно малой вероятности возникновения столь длинных цепочек в реальной жизни.

Таким образом, если анализируется влияние объектов в масштабе крупного города и выше, то нужно считать до 3-го порядка. Для меньших масштабов (в том числе если необходимо провести анализ в какой-либо группе по профессиональному интересу) целесообразно использовать 2-й порядок.

Следует заметить, что важно не числовое значение итерированной силы, а то, как силы для разных объектов соотносятся между собой. Поэтому после расчета вектора итерированной силы очередного порядка целесообразно нормировать данный вектор.

Остался открытым вопрос о вычислении функции принадлежности $\mu_A(x, y)$. Ее можно определить по-разному в зависимости от количества предоставляемой социальной сетью информации. Ключевым моментом при определении должно стать следующее правило: чем выше значение функции, тем большее прямое влияние оказывает объект x на объект y .

Социальная сеть VK (Vkontakte) предоставляет следующую информацию о взаимодействии объектов x и y :

- являются ли объекты x и y друзьями;
- число комментариев, которые пользователь y оставил на фотографиях пользователя x ;
- число комментариев, которые пользователь y оставил на видеозаписях пользователя x ;
- число комментариев, которые пользователь y оставил на заметках пользователя x ;
- число лайков, которые пользователь y оставил на фотографиях/видеозаписях/заметках пользователя x ;

и др.

Пусть $f_i(x, y)$ характеризует степень влияния i -го фактора объекта x на объект y . Например, в случае $i = 1$ это может быть число месяцев, прошедших с момента, когда объект x пометил объект y как друга. В случае $i = 2$ это будет число фотографий, которые объект x пометил у объекта y как "мне нравится" и т. д. Обозначим число факторов через m . Тогда функцию принадлежности можно определить следующим образом:

$$\mu_A(x, y) = \sum_{i=1}^m c_i F_i(x, y),$$

где $F_i(x, y) = 1 - e^{-f(x, y)}$, а c_i — некоторые вещественные коэффициенты, которые удовлетворяют следующим ограничениям:

$$0 \leq \sum_i c_i \leq 1, \quad 0 \leq c_i \leq 1.$$

Следует заметить, что при фиксированных x и y функция принадлежности будет представлять собой линейную функцию относительно коэффициентов c_j . Принимая во внимание формулу (1), получим, что итерированная сила k -го порядка является членом k -го порядка относительно данных коэффициентов.

Пусть из n объектов социальной сети нам известны объекты, которые должны обладать значительным влиянием. Обозначим множество таких объектов через A . Необходимо таким образом подобрать коэффициенты c_j , чтобы максимизировать их суммарную влияние в сети, что сводится к следующей задаче:

$$\sum_{j \in A} p^j(c, 2) \rightarrow \max;$$

$$\begin{cases} 0 \leq \sum_i c_i \leq 1; \\ 0 \leq c_i \leq 1. \end{cases}$$

Это задача условной оптимизации, которая может быть решена с помощью метода внутренней штрафной функции [7]. Его суть заключается в преобразовании задачи с ограничениями в последовательность задач безусловной оптимизации некоторых вспомогательных функций. Последние получаются путем модификации целевой функции с помощью функций-ограничений таким образом, чтобы данные ограничения не фигурировали в задаче оптимизации в явном виде. В нашем случае штрафной функцией могла бы быть

$$\Phi(c, a) = a \left[\ln \left(\sum_i c_i \right) + \ln \left(1 - \sum_i c_i \right) + \sum_i \ln(c_i) + \sum_i \ln(1 - c_i) \right],$$

где a — положительный параметр, определяющий масштаб штрафа.

Алгоритм решения состоит в следующем. В качестве начальной точки $c[0]$ выбирается произвольная точка допустимого множества (например, $c[0]_i = 1/m$). Задается некоторая монотонно убывающая к нулю последовательность $\{a_k\}$, $k = 1, 2, \dots$. Берется первый элемент этой последовательности и решается задача безусловной оптимизации (например, по методу наискорейшего спуска)

$$F(c, a) = \sum_{j \in A} p^j(c, 2) + \Phi(c, a) \rightarrow \max.$$

Решением этой задачи будет точка $c[a_1]$. Эта точка используется в качестве начальной для решения задачи поиска максимума $F(c, a_2)$, где $a_2 \leq a_1$. Последовательность полученных таким образом то-

чек $c[a_k]$ сходится к оптимальному решению задачи. Вычисления прекращают при выполнении условий:

$$\left| \sum_{j \in A} p^j(c[k], 2) - \sum_{j \in A} p^j(c[k-1], 2) \right| \leq \varepsilon;$$
$$\|c[k] - c[k-1]\| \leq \beta,$$

где ε, β — числа, определяющие точность вычислений.

После нахождения индексов функция принадлежности задана полностью. На последнем шаге необходимо рассчитать вектор итерированных сил 2-го порядка согласно формулам (1), (2). Отсортированный по убыванию список итерированных сил будет представлять наиболее влиятельные объекты в вершине данного списка.

Методика эксперимента

Для проведения эксперимента использовалась группа социальной сети "Vkontakte" известного Интернет-портала. Приложение выводит 100 объектов данной группы, обладающих наибольшим влиянием. В качестве априорной информации были выбраны пять человек, которые являются авторами известных статей.

Анализ результатов показал, что данные пять человек находятся в первой двадцатке. Приложение предоставило большое число людей, которые постоянно и регулярно публикуют новые статьи, в результате чего пользуются повышенным вниманием других пользователей.

Выводы

В работе предложена математическая модель поиска наиболее влиятельных объектов социальной сети на основе априорной информации. Поставлена задача, разработан алгоритм ее решения, продемонстрировано применение данного алгоритма к известной социальной сети "ВКонтакте". Анализ результатов показывает, что расчет влиятельностей объектов в крупных социальных сетях может предоставить информацию, которая была бы полезной в различных сферах человеческой деятельности.

Список литературы

1. **Frawley W. J., Piatetsky-Shapiro G., Matheus C. J.** Knowledge Discovery in Databases: An Overview // *AI Magazine*. 1992. Vol. 13, N. 3. P. 57–70.
2. **Интеллектуальный анализ данных** [сайт]. URL: http://www.machinelearning.ru/wiki/index.php?title=Data_Mining (дата обращения 22.07.2015).
3. **Давыдов А. А.** Системная социология. М.: ИС РАН, 2009. 192 с.
4. **Ибрагимов В. А.** Элементы нечеткой математики. Баку: АГНА, 2010. 394 с.
5. **Web и Social Mining**: [сайт]. URL: http://www.basegroup.ru/library/web_mining/ (дата обращения 23.06.2015).
6. **Берж К.** Теория графов и ее применения. М.: Изд. иностр. лит., 1962. 318 с.
7. **Васильев Ф. П.** Методы оптимизации. М.: Факториал Пресс, 2002. 824 с.

N. M. Novikova, Professor, e-mail: nov.nelly@gmail.com,
A. V. Boriskin, Graduate Student, e-mail: boriskinpost@gmail.com
Voronezh State University, Voronezh, Russia

Mathematical Model of the Powerful Object Search Based on a Known Information

The modern virtual social networks contain a lot of information. Most of this data is provided in a raw view. It's reasonable to process it for finding the information that can be considered as useful. For example, information about powerful objects can be used by many PR companies. The problem can be described this way: there is a subset of virtual social network objects. It's known that some of them can be considered as powerful objects in the group. The goal of the article is to use this information for searching other powerful objects. The algorithm of solving the problem is provided. The authors consider other approaches and compare the current one with them. The screenshot of the implemented software is shown.

Keywords: virtual social networks, data mining, fuzzy graphs, optimization methods

References

1. **Frawley W. J., Piatetsky-Shapiro G., Matheus C. J.** Knowledge Discovery in Databases: An Overview, *AI Magazine*, 1992, vol. 13, no. 3, pp. 57–70.
2. **Интеллектуальный анализ данных** [web-site]. (URL: http://www.machinelearning.ru/wiki/index.php?title=Data_Mining (access date 07/22/2015).
3. **Davydov A. A.** *Sistemnaya sociologiya*. Moscow: IS RAN, 2009, 192 p.
4. **Ibragimov V. A.** *Jelementy nechetkoj matematiki*. Baku: AGNA, 2010. 394 p.
5. **Web i Social Mining**: [web-site]. URL: http://www.basegroup.ru/library/web_mining/ (access date 06/23/2015).
6. **Berzh K.** *Teorija grafov i ee primenenija*. Moscow: Izd-vo inostr. lit., 1962, 318 p.
7. **Vasil'ev F. P.** *Metody optimizacii*. Moscow: Faktorial Press, 2002, 824 p.