

Н. А. Игнатьев, д-р физ.-мат. наук, проф., e-mail: n\_ignatev@rambler.ru,  
Национальный университет Узбекистана им. М. Улугбека

## Выбор собственного пространства объекта с использованием нелинейных преобразований признаков

*Рассмотрено описание объектов в собственном пространстве из латентных признаков. Для формирования пространства использованы правила иерархической агломеративной группировки исходных признаков. Предложена методика отбора информативного набора признаков объекта и вычисление меры сходства его с другими объектами по этому набору. Методика позволяет относить объект к числу эталонов либо идентифицировать его как шумовой.*

**Ключевые слова:** иерархическая агломеративная группировка, оценки объекта, логические закономерности, латентные признаки объекта, нелинейные преобразования.

### Введение

Разработка методов поиска собственного признакового пространства основана на идее [1], что в окрестности каждого объекта существует своя логическая закономерность. Результаты поиска востребованы при построении информационных моделей в предметных областях со сложной структурой отношений между объектами. Особый интерес представляют методы отбора информативных латентных признаков, инвариантных к масштабам измерений данных, комбинаторная сложность реализации которых позволяет получать результаты за приемлемое время.

Примерами предметных областей со сложной структурой отношений между объектами является медицина, психология и т. д. В медицине симптомы и синдромы людей с одинаковой патологией могут сильно отличаться друг от друга. Информативный набор признаков объекта может быть использован при выборе метода лечения, определении тяжести заболевания, группы инвалидности и т. д. Получить сравнимый по значимости эффект от методов отбора наборов признаков (показателей), информативных на всей обучающей выборке, практически невозможно.

В работе предлагается собственное пространство для описания объекта латентными

признаками формировать с помощью алгоритма иерархической агломеративной группировки исходных признаков. Как отдельная задача рассматривается выбор информативных латентных признаков объекта. Краткие сведения о возможности использования латентных признаков для выбора собственного пространства объекта содержатся в работе [2]. Необходимость (потребность) применения нелинейных отображений исходных признаков в латентные по правилам иерархической агломеративной группировки объясняется следующим образом:

- начиная с некоторой (изначально неизвестной) размерности исходного признакового пространства отношения близости между объектами становятся размытыми;
- структура отношений между объектами зависит от масштабов измерений количественных признаков и используемых мер близости.

Использование правил иерархической агломеративной группировки при формировании набора латентных признаков для описания объекта позволяет:

- гарантировать единственность решения независимо от масштабов измерений и выбранных мер близости;
- упорядочивать латентные признаки по отношению информативности;
- проводить отбор информативных латентных признаков для описания объекта инва-

риантных к масштабам измерений исходных признаков.

Методика отбора индивидуальных информативных наборов разнотипных признаков с помощью локальных метрик объектов описана в работе [3]. Для отбора был использован критерий на основе максимальной разницы частот встречаемости представителей (объектов) двух классов  $K_1$  и  $K_2$  в последовательности, упорядоченной по значениям расстояний от указанного объекта по его локальной метрике. Экстремальное значение критерия, полученное по медицинским данным с описанием состояния больных и практически здоровых индивидуумов, было предложено интерпретировать как индекс здоровья. Для нахождения экстремума использовали эвристические пошаговые алгоритмы отбора. Было показано, что различные схемы отбора признаков (последовательное удаление малоинформативных либо последовательное включение наиболее информативных) не давали схожих результатов.

Свойство инвариантности к масштабам шкал измерений является атрибутом нелинейного отображения групп исходных (сырых) разнотипных (номинальных и количественных) признаков на числовую ось. Результатом отображения являются латентные признаки. Описание допустимого объекта в рамках пространства из информативных латентных признаков необходимо для нахождения индивидуальной меры сходства (различия) с другими объектами. Эта мера отражает отношения между объектами и служит средством для принятия решения. Открываются перспективы обнаружения и интерпретации скрытых закономерностей в базах данных с единых методологических позиций.

Выбор критериев из некоторого конечного множества для принятия решения относительно объекта является трудно формализуемой проблемой. Предлагается рассматривать эту проблему как задачу распознавания из двух классов. Объекты классов могут быть представлены в виде оппозиции: больной — практически здоровый, богатый — бедный и т. д. Индивидуальная мера сходства (различия) в данном случае является аналогом свертки критериев.

Вычисление информативных наборов признаков можно применять для отслеживания изменения (мониторинга) состояния объекта. Например, изменение степени тяжести заболевания пациента в медицине, экологического

состояния окружающей среды, уровня кредитоспособности клиента банка и т. д.

Предлагается методика выбора информативных признаков объекта, основанная на использовании трех критериев. В рамках этой методики можно оценивать степень выраженности логических закономерностей относительно объекта по его информативному набору признаков. Значения оценок позволяют обосновывать использование объектов в качестве эталонов либо рассматривать их как шумовые.

### 1. Формирование собственного пространства объекта из латентных признаков

Рассматривается задача распознавания в стандартной постановке. Объекты обучения заданы через множество  $E_0 = \{S_1, \dots, S_m\}$ , разделенное на два непересекающихся подмножества (класса)  $K_1$  и  $K_2$ ,  $E_0 = K_1 \cup K_2$ . Описание объектов выполняется с помощью набора из  $n$  разнотипных признаков  $X(n) = (x_1, \dots, x_n)$ ,  $\xi$  из которых измеряются в интервальных шкалах,  $(n - \xi)$  — в номинальной.

Выбор собственного пространства для описания объекта  $S_d \in E_0$ ,  $d = 1, \dots, m$ , набором латентных признаков  $Y(\tau) = (y_1, \dots, y_\tau)$ ,  $\tau < n$  проводится с помощью правил алгоритма иерархической агломеративной группировки. Алгоритм группировки разбивает набор признаков  $X(n)$  на непересекающиеся группы  $X(k_1), \dots, X(k_r)$ ,  $k_1 + \dots + k_r \leq n$ . Нелинейное отображение представителей каждой группы на числовую ось образует новый латентный признак в описании объекта.

Считается, что известны критерии для отбора набора информативных латентных признаков  $Y(k) \subset Y(\tau)$  объекта. Требуется для указанного объекта  $S \in E_0$  определить:

- собственное пространство из латентных признаков  $Y(\tau)$ ,  $\tau < n$ ;
- информативный набор признаков  $Y(k) \subset Y(\tau)$ ,  $k \geq 0$ ;
- оценку объекта  $S$  на информативном наборе  $Y(k)$ .

Обозначим через  $I, J$  множество индексов соответственно количественных и номинальных признаков в исходном наборе  $X(n)$ . Для выбора латентных признаков в собственном пространстве объекта  $S_d \in E_0$ ,  $S_d = (a_{d1}, \dots, a_{dn})$  проведем предобработку данных следующим образом. Значения признаков объекта  $S = (b_1, \dots, b_n)$ ,  $S \in E_0$  преобразуем как

$$b_i = \begin{cases} |a_{di} - b_i|, i \in I, \\ 1, a_{di} = b_i, i \in J, \\ -1, a_{di} \neq b_i, i \in J. \end{cases} \quad (1)$$

Преобразованные по (1) признаки считаются измеренными в количественной шкале измерений, множество номеров которых идентифицируются как  $I = \{1, \dots, n\}$ . Для вычисления значений латентных признаков используют правила иерархической агломеративной группировки. Латентные признаки, полученные на  $p$ -м шаге группировки, обозначим как  $x_j^p, j \in I, p \geq 0$ . При  $p = 0, |I| = n$ . Упорядоченное множество значений признака  $x_j^p$  объектов из  $E_0$  разделим на два интервала  $[c_1^{jp}, c_2^{jp}], (c_2^{jp}, c_3^{jp}]$ , каждый из которых рассматривается как градация номинального признака.

Пусть  $u_i^1, u_i^2$  — число значений признака  $x_j^p, j \in I, p \geq 0$  класса  $K_i, i = 1, 2$  соответственно в интервалах  $[c_1^{jp}, c_2^{jp}], (c_2^{jp}, c_3^{jp}], |K_i| > 1, v$  — порядковый номер элемента упорядоченной по возрастанию последовательности  $r_{j_1}, \dots, r_{j_v}, \dots, r_{j_m}$  значений  $x_j^p$  у объектов из  $E_0$ , определяющий границы интервалов как  $c_1^{jp} = r_{j_1}, c_2^{jp} = r_{j_v}, c_3^{jp} = r_{j_m}$ . Критерий

$$\left( \frac{\sum_{i=1}^2 u_i^1 (u_i^1 - 1) + u_i^2 (u_i^2 - 1)}{\sum_{i=1}^2 |K_i| (|K_i| - 1)} \right) \times \left( \frac{\sum_{d=1}^2 \sum_{i=1}^2 u_i^d (|K_{3-i}| - u_{3-i}^d)}{2|K_1||K_2|} \right) \rightarrow \max_{c_1^{jp} < c_2^{jp} < c_3^{jp}} \quad (2)$$

позволяет вычислять оптимальное значение границы  $c_2^{jp}$  для интервалов  $[c_1^{jp}, c_2^{jp}]$  и  $(c_2^{jp}, c_3^{jp}]$ .

Экстремум критерия (2) используется в качестве веса  $w_j^p$  ( $0 \leq w_j^p \leq 1$ ) признака  $x_j^p$ . При  $w_j^p = 1$  значения признака  $x_j^p$  у объектов из классов  $K_1$  и  $K_2$  не пересекаются между собой. Особенность использования критерия (2) на данных, полученных по (1) при  $j \in J$ , заключается в том, что число различных значений признака равно числу классов. В этом случае граница между двумя классами определяется как  $c_2^{jp} = (c_1^{jp} + c_3^{jp})/2$ .

Для вычислений без учета масштаба измерений проведем предобработку признаков объекта  $S_r = (a_{r1}, \dots, a_{rn}), S_r \in E_0$  с помощью (2) как  $a_{ri}^0 = (a_{ri} - c_2^{i0}) / (c_3^{i0} - c_1^{i0}), i = 1, \dots, n$ . Значение комбинации  $b_{rij}^p$  по паре признаков  $(x_i^p, x_j^p)$ ,

$0 \leq p < n, i, j \in I, i \neq j$  объекта  $S_r = \{a_{ru}^p\}_{u \in I}, S_r \in E_0$  вычисляется таким образом:

$$b_{rij}^p = \eta_{ij} (t_i w_i^p a_{ri}^p + t_j w_j^p a_{rj}^p) + (1 - \eta_{ij}) t_{ij} w_{ij}^p (a_{ri}^p a_{rj}^p - c_2^{ijp}) / (c_3^{ijp} - c_1^{ijp}),$$

$$i, j \in I, t_i, t_j \in \{-1, 1\}, \eta_{ij} \in [0, 1],$$

где  $w_i^p, w_j^p, w_{ij}^p$  — веса признаков, определяемые по (2) соответственно по множеству значений  $x_i^p, x_j^p$ , и их произведения  $x_i^p x_j^p$  на  $E_0$ , значения  $t_{ij}, t_i, t_j \in \{-1, 1\}, \eta_{ij} \in [0, 1]$  выбираются по экстремуму функционала

$$\varphi(p, i, j) = \frac{\min_{S_r \in K_1} b_{rij}^p - \max_{S_r \in K_2} b_{rij}^p}{\max_{S_r \in E_0} b_{rij}^p - \min_{S_r \in E_0} b_{rij}^p} = \max_{t_{ij}, t_i, t_j \in \{-1, 1\}, \eta_{ij} \in [0, 1]} b_{rij}^p. \quad (3)$$

Экстремум функционала (3) интерпретируется как отступ между объектами классов  $K_1$  и  $K_2$  по множеству значений по паре признаков  $(x_i^p, x_j^p), 0 \leq p < n, i, j \in I, i \neq j$ .

Обозначим через  $\{z_{ij}^p\}_{i, j \in I}, p \geq 0$  — квадратную матрицу размера  $(n-p) \times (n-p)$ , значение элемента  $z_{ij}^p$  которой при  $p = 0$  определяется как

$$z_{ij}^p = \begin{cases} w_i^p, i = j, \\ \text{значению (2) по } \{b_{rij}^p\}_{r=1}^m, i \neq j, \end{cases} \quad (4)$$

через  $\Gamma_\eta, \eta > 0$  — подмножество номеров признаков из  $X(n)$ . Пошаговая реализация алгоритма иерархической агломеративной группировки будет такой.

**Шаг 1.**  $p = 0, \lambda c = 0, \eta = 1$ . Выполнять  $\Gamma_\eta = \{\eta\}, \text{Margin}_\eta = -2, \eta = \eta + 1$  пока  $\eta \leq n$ .

**Шаг 2.** Вычислить значения элементов матрицы  $\{z_{ij}^p\}_{i, j \in I}$  по (4).

**Шаг 3.** Выделить  $\Phi = \{z_{uv}^p | z_{uv}^p \geq \max(w_u^p, w_v^p)\}$  и  $u \neq v, u, v \in I$ . Если  $\Phi = \emptyset$ , то переход к шагу 9.

**Шаг 4.** Вычислить  $\lambda n = \max_{z_{uv}^p \in \Phi} z_{uv}^p$ . Выделить  $\Delta = \{(s, t), s, t \in I | z_{st}^p = \lambda n \text{ и } s < t\}$ . Определить пару  $\{i, j\}, i < j$  как

$$\{i, j\} = \begin{cases} \Delta, |\Delta| = 1, \\ \{s, t\}, (s, t) \in \Delta \text{ и} \\ \varphi(p, s, t) > \max_{(u, v) \in \Delta \setminus (s, t)} \varphi(p, u, v). \end{cases}$$

**Шаг 5.** Если  $\lambda n > \lambda c$  или  $\lambda n = \lambda c$  и  $\text{Margin}_i < \varphi(p, i, j)$ , то  $\Gamma_i = \Gamma_i \cup \Gamma_j, \Gamma_j = \emptyset, \text{Margin}_i = \varphi(p, i, j)$ , переход к шагу 7.

**Шаг 6.** Вывод номеров признаков из  $\Gamma_i, \Gamma_i = \emptyset, I = I \setminus \{i\}$ , переход к шагу 3.

**Шаг 7.**  $p = p + 1$ ,  $I = I \setminus \max(i, j)$ ,  $k = \min(i, j)$ ,  $\lambda c = \lambda n$ . Заменить значения признаков в описании объекта  $S_r = \{a_{ru}^{p-1}\}_{u \in I}$ ,  $r = 1, \dots, m$  на

$$a_{ru}^p = \begin{cases} a_{ru}^{p-1}, u \in I \setminus \{k\}, \\ b_{rj}^p, u = k. \end{cases}$$

**Шаг 8.** Для каждой пары  $(u, v)$ ,  $u, v \in I$  определить значение

$$z_{uv}^p = \begin{cases} z_{uv}^{p-1}, u \in I \setminus \{k\}, v \in I, \\ \text{значению (2) на } \{a_{rv}^p\}_{r=1}^m, u = k, v \in I. \end{cases}$$

Если  $n - p > 1$ , то переход к шагу 3.

**Шаг 9.** Конец.

## 2. Отбор информативных наборов признаков объекта

Формирование собственного пространства из латентных признаков по алгоритму из разд. 1 в общем случае не решает проблему обнаружения устойчивых логических закономерностей в окрестности объекта. Для решения этой проблемы необходимо использовать отбор информативного набора признаков.

Пусть для объекта  $S_d \in K_p$ ,  $p = 1, 2$  по алгоритму из разд. 1 определен его собственный набор латентных признаков  $Y(\mu)$ ,  $1 \leq \mu < n$ . Покажем, что значения, вычисленные по критерию (2), являются необходимым но недостаточным условием включения признака  $y_t \in Y(\mu)$ ,  $t = 1, \dots, \mu$  в информативный набор. Для проверки достаточности предлагается использовать два критерия.

Упорядочим множество объектов  $E_0$  по возрастанию значений признака  $y_t$  при описании их по набору  $Y(\mu)$  как

$$S_{d_1}, S_{d_2}, \dots, S_{d_m}. \quad (5)$$

Для оценки объекта  $S_d \in K_p$  по (5) применим критерий

$$F(S_d, y_t) = \max_{1 \leq i \leq m} \left( \frac{z_p(i)}{|K_p \cap E_0|} - \frac{z_{3-p}(i)}{|K_{3-p} \cap E_0|} \right), \quad (6)$$

где  $z_p(i)$ ,  $z_{3-p}(i)$  — число объектов в  $\{S_{d_1}, \dots, S_{d_i}\} \subset E_0$  из (5) из класса  $K_p$  и  $K_{3-p}$  соответственно.

Пусть по критерию (2) на (5) определены границы интервалов  $[c_1, c_2]$ ,  $(c_2, c_3]$ . Значение

критерия устойчивости объекта  $S_d \in K_p$  по латентному признаку  $y_t \in Y(\mu)$  вычисляется как

$$\Omega(y_t) = \theta_1(1 - \theta_2), \quad (7)$$

где

$$\theta_1 = \frac{|\{S_i \in K_p \mid y_t \in [c_1, c_2]\}|}{|K_p|},$$

$$\theta_2 = \frac{|\{S_i \in K_{3-p} \mid y_t \in [c_1, c_2]\}|}{|K_{3-p}|}.$$

Критерий (7) является мультипликативным, критерий (6) — аддитивным. Множество допустимых значений (6) и (7) принадлежат интервалу  $[0, 1]$ . Интерес представляет анализ значений этих критериев и использование их для отбора информативного набора признаков объекта. Методологически критерии (2), (6) и (7) связаны друг с другом. Так как с помощью критерия (2) формируется набор латентных признаков  $Y(\mu)$ ,  $\mu < n$ , то его использование предшествует применению критериев (6) и (7).

Рассмотрим условия отбора информативных признаков на  $Y(\mu)$  с использованием критериев (2), (6) и (7). Упорядочим возможные варианты на включение (не включение) признака  $y_t \in Y(\mu)$  в информативный набор по их приоритетности следующим образом:

- а) значения по (6) и (7) больше значения по (2);
- б) максимальное значение из (6) и (7) больше значения по (2);
- в) значение по (2) больше максимального значения по (6) и (7).

Истинность условия по варианту в) по всем латентным признакам указывает на то, что объект не отличается от объектов противоположного класса и может считаться шумовым.

Для формирования информативного набора  $Y(k) \subset Y(\mu)$  рекомендуется использовать условия вариантов а) и б). В качестве оценки объекта  $S_d \in K_p$  предлагается выбирать максимальное значение критерия (2) на признаках из набора  $Y(k)$ . Одной из форм использования оценок является селекция обучающих выборок.

Для мониторинга состояний объекта рекомендуется формировать обучающие выборки в виде оппозиции, например, класс практически здоровых людей и больных гипертонией в медицине. На практике выбор собственного пространства объекта при числе непересекающихся классов больше двух можно свести к задаче с двумя классами. С помощью информатив-

ных признаков можно визуализировать отношения объектов в одно-, двух- и трехмерном пространстве. Такое представление позволяет проводить анализ данных аналогично предложенному в методе локальной геометрии [1].

*О нелинейном отображении многомерных данных в пространство низкой размерности.* Нелинейное отображение многомерных данных [4] использовалось для представления геометрической структуры матрицы исходных данных в пространстве малой размерности (как правило, двух- или трехмерное) с сохранением, по возможности, расстояний между ними. С помощью методов отображения пытались минимизировать некоторую функцию потерь (меру искажения), характеризующую значение рассогласования расстояний между первоначальными и полученными векторами в пространстве малой размерности. Меры искажения, как правило, основаны на сравнении попарных расстояний между точками в исходном пространстве и пространстве отображения.

В данной работе близость между объектами представляет интерес с точки зрения ее влияния на степень истинности гипотезы о компактности классов. Методы обнаружения логических закономерностей очень чувствительны к размерности пространства, в котором вычисляется значение близости. Для исследования этой проблемы в работе [5] предложено использовать меру компактности классов и выборки в целом. Значение компактности связано с результатом разбиения объектов классов на непересекающиеся группы. При разбиении использовали свойство связанности объектов по оболочкам классов. Оболочки определяли как подмножество граничных объектов классов по заданной метрике.

### 3. Вычислительный эксперимент

Для эксперимента была взята выборка данных German из работы [6]. Выборка представлена 1000 объектами, разделенными на два класса  $K_1$  и  $K_2$ . В качестве объектов рассмотрены заемщики банка из Германии. Показатели заемщиков описываются 20 признаками, 7 из которых измеряются в количественных шкалах, 13 — в номинальных. В табл. 1 представлена последовательность выбора собственного пространства из латентных признаков для ряда объектов по результатам иерархической агломеративной группировки.

Как видно из табл. 1, объект 778 имеет оценку 0,2794 по критерию (2) на латентном признаке, полученном из комбинации  $((x_3, x_{15}), x_9, x_7)$ . Выбор этого признака в качестве информативного следует из того, что значение 0,3330 по критерию (7) больше 0,2794. Для объекта 702 в состав информативного набора входят латентные признаки  $(((((x_1, x_5), x_3), x_2), x_6), x_{13}), x_{15}), x_{20})$  и  $((x_{12}, x_{14}), x_{11}), x_4)$ . Оценка объекта 0,3479 является нелинейной комбинацией объединения исходных признаков в латентный в такой последовательности  $(((((проверка статуса, сумма кредита), кредитная история), продолжительность), состояние сбережений), возраст), корпус), иностранец)$ .

В табл. 2 приведены значения оценок для ряда объектов и последовательность формирования латентных признаков, на которых они получены.

Анализ табл. 2 показывает, что информативные латентные признаки объектов отлича-

Таблица 1

Выбор собственного признакового пространства объектов

№ объекта (класс)	Формирования групп признаков	Значение критерия		
		(2)	(6)	(7)
778(1)	$(((((x_1, x_5), x_{13}), x_{20}), x_{10}), x_8)$	0,3522	0,0	0,0848
	$(x_2, x_{14})$	0,2875	0,0243	0,1403
	$((x_4, x_6), x_{12}), x_{16})$	0,2936	0,0043	0,1318
	$((x_3, x_{15}), x_9, x_7)$	0,2794	0,1786	0,3330
	$(x_{17}, x_{19})$	0,2535	0,0438	0,2167
	$(x_{11}, x_{18})$	0,2502	0,0324	0,2440
702(2)	$(((((x_1, x_5), x_3), x_2), x_6), x_{13}), x_{15}), x_{20})$	0,3479	0,3467	0,4248
	$((x_{12}, x_{14}), x_{11}), x_4)$	0,2807	0,1881	0,3050
	$(x_7, x_8)$	0,2650	0,0057	0,1737
	$((x_9, x_{16}), x_{10}), x_{18}), x_{19})$	0,2635	0,0171	0,1780
	$(x_{17})$	0,2499	0,0686	0,2410

Таблица 2

Оценки объектов по латентным признакам

№ объекта (класс)	Формирование латентного признака	Оценка объекта
100(1)	$(((((про продолжительность, другие способы платежей), состояние сбережений), кредитная история), существующие кредиты)$	0,2934
490(1)	$((кредитная история, собственный телефон), личный статус), работа)$	0,2607
16(2)	$((проверка статуса, сумма кредита), иностранец), возраст)$	0,3352
918(2)	$((проверка статуса, сумма кредита), продолжительность), иностранец)$	0,3450

ются как по составу, так и по порядку включения исходных признаков.

### Заключение

Разработан алгоритм иерархической агломеративной группировки для формирования собственного пространства объекта из латентных признаков. Описаны условия выбора информативных признаков из собственного пространства. Результаты вычислений востребованы для наполнения баз знаний и построения информационных моделей в слабо формализованных предметных областях. Например, при медицинской компьютерной диагностике можно будет определять индивидуальный набор информативных показателей больного и рекомендовать ему соответствующий курс лечения.

### Список литературы

1. Дюк В. А. Методология поиска логических закономерностей в предметной области с нечеткой системологией: На примере клинико-экспериментальных исследований: Дисс. ... докт. тех. наук: Санкт-Петербург. 2005. 309 с.
2. Игнат'ев Н. А. Индексирование объектов по индивидуальным наборам информативных признаков // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2016. № 4 (37). С. 27–35.
3. Ignat'ev N. A., Mirzaev A. I. The Intelligent Health Index Calculation System // Pattern Recognition and Image Analysis. 2016. Vol. 26. N. 1. P. 73–77.
4. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.
5. Игнат'ев Н. А. О мере компактности объектов классов // Труды конференции "Проблемы современной топологии и ее приложения". Ташкент. 11–12 мая 2017 г. С. 283–285.
6. Machine Learning Repository. URL: <http://archive.ics.uci.edu/ml/datasets>.

N. A. Ignatiev, D. Sc., Professor, e-mail: [n\\_ignatiev@rambler.ru](mailto:n_ignatiev@rambler.ru),  
National University of Uzbekistan named after M. Ulugbek, Tashkent

## Selecting an Object's Own Space Using Nonlinear Transformations of Features

*The description of objects in their own space from latent features is considered. For the formation of the space, the rules of a hierarchical agglomeration grouping of initial features are used. The rules provide the property of invariance to the scales of measurement for non-linear mapping of groups of raw polytypical (nominal and quantitative) features onto the numerical axis. The need to apply non-linear displays of initial features to latent ones according to the rules of the hierarchical agglomeration group is explained as follows:*

- from with some (initially unknown) dimension of the source features space, the relations of proximity between the objects of classes become blurred;
- the structure of relations between objects depends on the scale of measurements of quantitative features and on the adjacency measures used.

*As a separate problem, the choice of informative latent features of the object is considered. The description of an admissible object within a space of informative latent features is necessary for finding an individual measure of similarity (difference) with other objects. This measure reflects the relationship between the objects and serves as a means for decision-making. It is practically impossible to obtain a comparable effect from the methods of selecting sets of features that are informative on the whole training sample. The technique of selection of an informative set of features of an object by three criteria is described. Measure of similarity in this set allows relate the object to the number of standards or to identify it as noisy. The prospects of discovering and interpreting hidden regularities in databases from unified methodological positions are revealed. For example, in the field of medical computer diagnostics, one can choose an individual course of treatment for a patient.*

**Keywords:** hierarchical agglomerative grouping, object estimates, logical regularities, latent features of the object, nonlinear transformations, class compactness

DOI: 10.17587/it.24.665-670

### References

1. Dyuk V. A. Metodologiya poiska logicheskikh zakonomenostey v predmetnoy oblasti s nechetkoy sistemologiyey: Na primere kliniko-eksperimental'nykh issledovaniy: (Methodology of the search for logical regularities in the subject domain with fuzzy systemology: On the example of clinical and experimental studies), Diss. ... dokt. tekh. nauk: Sankt-Peterburg, 2005, 309 p. (in Russian).
2. Ignat'yev N. A. Indeksirovaniye ob'yektov po individual'nym naboram informativnykh priznakov (Indexing of objects on individual sets of informative features), *Vestnik Tomskogo gosudarstvennogo universiteta. Upravleniye, vychislitel'naya tekhnika i informatika*, 2016, no. 4 (37), pp. 27–35 (in Russian).

3. Ignat'ev N. A., Mirzaev A. I. The Intelligent Health Index Calculation System. *Pattern Recognition and Image Analysis*, 2016, vol. 26. no. 1, pp. 73–77.
4. Ayvazyan S. A., Bukhshtaber V. M., Yenyukov I. S., Meshalkin L. D. *Prikladnaya statistika. Klassifikatsiya i snizheniye razmernosti* (Applied statistics. Classification and reduction of dimensionality), Moscow, Finansy i statistika, 1989, 607 p. (in Russian).
5. Ignat'yev N. A. O mere kompaktnosti ob'yektov klassov (About measure of compactness of objects of classes), *Trudy konferentsii "Problemy sovremennoy topologii i yeyo prilozheniya"*, Tashkent, 2017, pp. 283–285 (in Russian).
6. Machine Learning Repository, available at: <http://archive.ics.uci.edu/ml/datasets>