

УДК 519.95

ПРИНЯТИЕ РЕШЕНИЙ В РАМКАХ СОБСТВЕННОГО ПРИЗНАКОВОГО ПРОСТРАНСТВА ОБЪЕКТА

Игнатъев Н. А., Лолаев М. Я.

ignatev@rambler.ru; musulmon.lolayev.94@mail.ru

Национальный университет Узбекистана, Ташкент

Рассматриваются формирование собственного пространства объекта из латентных признаков в задаче распознавания с непересекающимися классами. Описан метод отбора набора информативных признаков в этом пространстве. Предложен способ оценки объекта и принятия решения по нему с помощью этого набора. Разработана методика селекции обучающих выборок с помощью оценок.

Ключевые слова: логические закономерности, латентные информативные признаки, селекция обучающей выборки, нелинейное преобразование пространства

Цитирование: *Игнатъев Н. А., Лолаев М. Я.* Принятие решений в рамках собственного признакового пространства объекта // Проблемы вычислительной и прикладной математики. — 2018. — №4(16). — С. 5–13.

1 Введение

Информационные модели, основанные на знаниях, являются основным инструментом для принятия решений в слабо формализованных предметных областях. Обнаружение знаний в рамках таких моделей является основной целью интеллектуального анализа данных (ИАД). Большинство решаемых с помощью методов ИАД задач являются NP полными. Поэтому актуальной проблемой при разработке и обосновании методов ИАД является уменьшение комбинаторной сложности алгоритмов для извлечения знаний из баз и хранилищ данных.

Альтернатива использованию переборных методов для извлечения знаний была предложена в [1]. В рамках этой альтернативы рассматривается предположение, что в окрестности каждого объекта существует своя логическая закономерность. Идея реализации, предложенного в [1] метода локальной геометрии заключается в отображении описаний объектов из пространства с высокой размерностью в одно-, двух и трехмерное пространство с центром в указанном объекте. По результатам отображения выборки данных на плоскость исследователь мог производить визуальную группировку. Разбиение объектов на группы служило отправной точкой для извлечения знаний с помощью известных методов статистического анализа данных. Практическое применение метода локальной геометрии нашло отражение при решении ряда прикладных задач [2].

В качестве недостатков метода локальной геометрии отметим:

- метод главных компонент, используемый для визуализации данных, не гарантировал устойчивости решения;
- результаты визуализации использовались для статистического анализа данных, основным недостатком которого является усреднение значений показателей при расчетах;
- принятие решение производилось без учета информативных признаков объекта.

Для обнаружения любого вида логических закономерностей в ИАД имеет значение состав наборов признаков и их количество. Корректность принимаемых решений напрямую зависит от размерности признакового пространства и используемых мер

близости между объектами. Применение эвристических методов для снижения размерности признакового пространства имеет ряд существенных ограничений. Основной проблемой при поиске приемлемых решений является комбинаторная сложность алгоритмов [3].

В работе предлагается:

- формировать собственное признаковое пространство объекта за счет нелинейного отображения непересекающихся групп исходных признаков на числовую ось;
- производить выбор информативного набора из собственного признакового пространства объекта;
- оценивать объект в рамках его информативного набора признаков для принятия решения по нему.

Постановка задачи о выборе собственного пространства объекта с помощью исходных (количественных и номинальных) признаков впервые была сделана в [4]. Методика поиска собственного пространства, предложенная в [4, 5], не гарантировала единственности решения. Для решения применялись эвристические пошаговые методы отбора.

Нелинейное отображение многомерных данных [6, 7] использовалось для представления геометрической структуры матрицы исходных данных в пространстве малой размерности (как правило, двух- или трехмерном) с сохранением, по возможности, расстояний между ними. С помощью методов отображения пытались минимизировать некоторую функцию потерь (меру искажения), характеризующую величину 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 нелинейных методах метрического многомерного шкалирования (МШ) предполагалось, что различия (близость) измеряются в ординальной шкале, так как важен только ранговый порядок, а сами их численные значения не так важны [6]. Процедуры неметрического МШ стремятся построить такую геометрическую конфигурацию точек в определяемом q -мерном пространстве, чтобы ранговый порядок попарных расстояний между ними совпадал по возможности с ранговым порядком различий, т.е. отобразить неметрическую (ранговую) информацию в метрической шкале.

В данном исследовании речь не идет о сохранении геометрической близости между объектами. Близость представляет интерес с позиций гипотезы о компактности классов в теории распознавания образов. Методы обнаружения логических закономерностей очень чувствительны к размерности пространства. Отношения между объектами при относительно небольшой (изначально неизвестной) размерности становятся размытыми, нечеткими. Для исследования этой проблемы [8] (известной как проклятие размерности) предложено использовать меру компактности классов и выборки в целом. Значение меры компактности зависит от разбиения объектов классов на непересекающиеся группы. При разбиении использовалось свойство связанности объектов по оболочкам классов. Оболочки определялись как подмножество граничных объектов классов по заданной метрике.

Для последовательного формирования групп исходных признаков и нелинейного отображения их на числовую ось в работе предлагается использовать правила иерархической агломеративной группировки. Выбор агломеративного иерархического метода группировки гарантирует:

- единственность выбора собственного пространства из латентных признаков за счет использования принципа динамического программирования;
- инвариантность латентных признаков по масштабам измерения количественных признаков.

На примере выборки медицинских данных демонстрируется отбор индивидуальных информативных признаков объектов. Показано как по оценкам объектов на информативных признаках разделять их на эталонные и шумовые.

2 Постановка задачи и метод выбора собственного пространства объекта

Рассматривается задача распознавания в стандартной постановке. Объекты обучения заданы через множество $E_0 = \{S_1, \dots, S_m\}$, разделённое на два непересекающихся подмножества (класса) K_1 и K_2 , $E_0 = K_1 \cup K_2$. Описание объектов производится с помощью набора из n разнотипных признаков $X(n) = (x_1, \dots, x_n)$, ξ 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 $Y(\tau) = (y_1, \dots, y_\tau)$, $\tau < n$ производится с помощью правил алгоритма иерархической агломеративной группировки. Алгоритм группировки разбивает набор признаков $X(n)$ на непересекающиеся группы $X(k_1), \dots, X(k_\tau)$, $k_1 + \dots + k_\tau \leq n$. Нелинейное отображение представителей каждой группы на числовую ось образуют 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 $S_d \in E_0$, $S_d = (a_{d1}, \dots, a_{dn})$ произведём преобработку 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 используются 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 – количество 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

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

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, полученным 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 так $c_2^{jp} = \frac{c_1^{jp} + c_3^{jp}}{2}$.

Для вычислений без учета масштаба измерений произведем предобработку признаков объекта $S_r = (a_{r1}, \dots, a_{rn})$, $S_r \in E_0$ с помощью (2) как $a_{ri}^0 = \frac{a_{ri} - c_2^{i0}}{c_3^{i0} - c_1^{i0}}$, $i = 1, \dots, n$. Значение комбинации b_{rij}^p по паре признаков (x_i^p, x_j^p) , $0 \leq p < n$, $i, j \in J$, $i \neq j$ объекта $S_r = \{a_{ru}^p\}_{u \in I}$, $S_r \in E_0$ вычислим таким образом

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

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

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 $t_i, t_j, t_{ij} \in \{-1, 1\}$, $\eta_{ij} \in [0, 1]$ выбираются по экстремуму функционала

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

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) , $0 \leq p < n$, $i, j \in J$, $i \neq j$.

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:

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

2 шаг: Вычислить значения элементов матрицы $\{z_{ij}^p\}_{i, j \in I}$ по (4);

3 шаг: Выделить $\Phi = \{z_{uv}^p | z_{uv}^p \geq \max(w_u^p, w_v^p) \text{ and } u \neq v, u, v \in I\}$. Если $\Phi = \emptyset$, то идти 9;

4 шаг: Вычислить $\lambda n = \max_{z_{uv}^p \in \Phi} z_{uv}^p$. Выделить $\Delta = \{(s, t), s, t \in i | z_{st}^p = \lambda n \text{ and } s < t\}$.

Определить пару $\{i, j\}, i < j$ как

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

5 шаг: Если $\lambda n > \lambda c$ или $\lambda n = \lambda c$ и $\text{Margin}_i < \varphi(p, i, j)$, то $\Gamma_i = \Gamma_i \cup \Gamma_j$, $\Gamma_j = \emptyset$, $\text{Margin}_i = \varphi(p, i, j)$, идти 7;

6 шаг: Вывод номеров признаков из Γ_i , $\Gamma_i = \emptyset$, $I = I \setminus \{i\}$, идти 3;

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

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

8 шаг: 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 идти 3;

9 шаг: Конец.

3 Отбор информативных наборов признаков объекта

Формирование собственного пространства из латентных признаков по алгоритму из п.2 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 п.2 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) соответственно из класса K_p и K_{3-p} .

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

$$\theta_1 = \frac{|\{S_i \in K_p | y_{it} \in [c_1, c_2]\}|}{|K_p|}, \theta_2 = \frac{|\{S_i \in K_{3-p} | y_{it} \in [c_1, c_2]\}|}{|K_{3-p}|}, S_i = (y_{i1}, \dots, y_{i\mu}).$$

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). Оценкой объекта $S_d \in K_p$ считается максимальное значение критерия (1) на признаках из набора $Y(k)$. Одной из форм использования оценок является селекция обучающих выборок [9].

Целью селекции обучающих выборок является повышение обобщающей способности алгоритмов распознавания. Состав множества шумовых объектов зависит от используемых методов распознавания и алгоритмов отбора. Например, один и тот же объект может быть отнесен или не отнесен к числу шумовых в зависимости от используемой меры близости. Специфика, рассматриваемого в рамках работы способа отбора такова, что вопрос об отнесении или не отнесении объекта к числу шумовых решается в рамках его индивидуального информативного набора признаков.

4 Вычислительный эксперимент

Для вычислительного эксперимента использовалась выборка данных по пневмококковому и серозному менингиту [10]. Каждый из 64 объектов выборки описывался 3 количественными и 18 номинальными признаками. Первый класс K_1 (пневмококковый менингит) представлен 35-ю объектами, второй класс K_2 (серозный менингит) – 29-ю объектами. Выбор собственного признакового пространства ряда объектов представлено в табл. 1.

Информативный набор латентных признаков (см. табл. 1) объекта $S_{26} \in K_1$ представлен 2-мя нелинейными комбинациями $((((x_4, x_{10}), x_9), x_2), x_{17})$ и $((((x_{11}, x_{18}), x_{21}), x_{19}), x_{13})$. Согласно правила из п. 2 оценка объекта S_{26} равна 0,6049 на латентном признаке $((((x_{11}, x_{18}), x_{21}), x_{19}), x_{13})$, состав которого сформирован в такой последовательности $(((($ иммунодефицитные состояния, ДВС), сыпь), гипертестезия), пневмония). Пример упорядочения (ранжирования) логических закономерностей по значениям оценок объектов на их информативных наборах признаков приводится в табл. 2.

Таблица 1 Выбор собственного признакового пространства объектов

№ объекта (класс)	Порядок формирования групп признаков	Значение критерия		
		(2)	(6)	(7)
26 (1)	$(((((x_3, x_{16}), x_7), x_{12}), x_{20}), x_6, x_{15}, x_8, x_5)$	0,8266	0,0	0,0020
	$((((x_4, x_{10}), x_9), x_2), x_{17})$	0,5361	0,6788	0,7025
	$((((x_{11}, x_{18}), x_{21}), x_{19}), x_{13})$	0,6049	0,7478	0,7635
	(x_1, x_{14})	0,2332	0,0759	0,2128
40 (2)	$(((((x_6, x_7), x_{12}), x_{20}), x_{14}), x_5, x_9, x_{17})$	0,7761	0,9310	0,8778
	$((((x_4, x_{16}), x_{15}), x_{10}), x_{11})$	0,6434	0,0	0,0118
	$((((x_{18}, x_{21}), x_{13}), x_1)$	0,5052	0,0	0,0296
	(x_2, x_8)	0,3420	0,0571	0,0552
	(x_3, x_{19})	0,2440	0,4335	0,2759

Таблица 2 Ранжирования объектов по оценкам

№ объекта (класс)	Комбинация признаков	Оценка (2)
42(2)	$(((((x_6, x_7), x_2), x_4), x_{12}), x_5, x_{20}), x_9, x_{17})$	0,8808
29(1)	$(((((x_6, x_7), x_2), x_{12}), x_{20}), x_5, x_9)$	0,8267
3(1)	$((x_2, x_8), x_{19})$	0,3183
41(2)	(x_{10}, x_{19})	0,2461

Результаты ранжирования позволяют направленно отбирать объекты в качестве эталонов или рассматривать их как шумовые. Максимальное значение оценки объектов (см. табл. 2) на наборе информативных признаков равно 0.8808, минимальное – 0,2461. Объекты $S_8, S_{11}, S_{36}, S_{45}, S_{57}, S_{58}, S_{64}$ не имеют индивидуальных наборов информативных признаков и относятся к числу шумовых.

5 Заключение

Предложена методика поиска информативного набора латентных признаков объекта. Информативный набор инвариантен к масштабам измерений исходных признаков. Свойство инвариантности позволяет с единых методологических позиций осуществлять поиск и интерпретацию логических закономерностей из разных предметных областей. Логические закономерности востребованы в информационных моделях в качестве источника новых знаний.

Литература

- [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] Берестнева О. Г., Муратова Е. А., Янковская А. Е. Анализ структуры многомерных данных методом локальной геометрии // Известия Томского политехнического университета. 2003. Т. 306. № 3. С. 19–23.
- [3] Гордеев Э. Н. Задачи выбора и их решение // Компьютер и задачи выбора. - М.: Наука, 1989. С. 5–48.
- [4] Ignat'ev N. A., Mirzaev A. I. The Intelligent Health Index Calculation System // Pattern Recognition and Image Analysis. 2016. V. 26. №1. P. 73–77.

- [5] *Игнатъев Н. А.* Индексирование объектов по индивидуальным наборам информативных признаков // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. №4(37). 2016. С. 27–35.
- [6] *Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика. Классификация и снижение размерности. // Москва. Финансы и статистика. 1989. 607 с.
- [7] *Chalmers M.* A Linear Iteration Time Layout Algorithm for Visualising High-Dimensional Data // Proc. IEEE visualization, San Francisco, 1996. P. 127–132.
- [8] *Игнатъев Н. А.* Кластерный анализ данных и выбор объектов-эталонов в задачах распознавания с учителем // Вычислитель. технологии. 2015. Т. 20. № 6. С. 34–43.
- [9] *Борисова И. А., Кутненко О. А.* Цензурирование ошибочно классифицированных объектов выборки // 17-я Всероссийская конференция “Математические методы распознавания образов – 2015” 19–25 сентября 2015, Россия, г. Светлогорск.
- [10] *Касымова Р. И.* Клинико-лабораторные особенности острых гнойных и серозных менингитов в зависимости от этиологии: Дис. ... канд. мед. наук. / Ташкент, 2009. 145 с.

Поступила в редакцию 28.04.2018

UDC 519.95

DECISION-MAKING WITHIN THE FRAMEWORK OF AN OBJECT'S OWN FEATURES SPACE

Ignatev N. A., Lolaev M. Y.

ignatev@rambler.ru; musulmon.lolayev.94@mail.ru

National University of Uzbekistan named after Mirzo Ulugbek, 4 Universitet st., Tashkent, Uzbekistan

Formation of an object's own space from latent features in the recognition problem with disjoint classes is considered. A method for selecting a set of informative features in this space is described. A method for evaluating an object and making a decision on it using this set are proposed. The technique of selection of training samples is developed. Values of estimates are used for separating objects into standards and noise.

Keywords: logical regularities, latent informative features, selection of training sample, non-linear space transformation

Citation: Ignatev N. A., Lolaev M. Y. 2018. Decision-making within the framework of an object's own features space. *Problems of Computational and Applied Mathematics*. 4(16): 5–13.

References

- [1] Dyuk, V. A. 2005. Metodologiya poiska logicheskikh zakonomernostey 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. tekhn. nauk.* [Tekhn. science dokt. ... diss]. 309 p. (In Russian)

- [2] Berestneva, O. G., E. A. Muratova, and A. E. Yankovskaya. 2003. Analiz struktury mnogomernykh dannykh metodom lokal'noy geometrii. [Multidimensional data structure analysis using local geometry]. *Izvestiya Tomskogo politekhnicheskogo universiteta*. [News of Tomsk Polytechnic University]. 306(1):19–23. (In Russian)
- [3] Gordeev, E. N. 1989. *Zadachi vybora i ikh reshenie*. [Choice and their decision]. *Kompyuter i zadachi vybora*. [Computer and task selection].:5–48. (In Russian)
- [4] Ignat'ev, N. A., A. I. Mirzaev. 2016. The Intelligent Health Index Calculation System. *Pattern Recognition and Image Analysis*. 26(1): 73–77.
- [5] Ignat'ev, N. A. 2016. Indeksirovanie ob'ektov po individual'nym naboram informativnykh priznakov. [Indexing objects for individual sets of informative features]. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika*. [Tomsk State University Journal of Control and Computer Science]. 4(33):27–35. (In Russian).
- [6] Ayvazyan, S. A., V. M. Buhstaber, I. S. Enyukov, and L. D. Meshalkin. 1989. *Prikladnaya statistika. Klassifikatsiya i snizhenie razmernosti*. [Applied statistics classification and reduction of dimensionality]. Moskva. Finansy i statistika. [Moscow: Finances and statistik]. 607 p. (In Russian)
- [7] Chalmers, M. 1996. *A Linear Iteration Time Layout Algorithm for Visualising High-Dimensional Data*. *Proc. IEEE visualization*. P. 127–132.
- [8] Ignat'ev, N. A. 2015. Klasterny analiz dannykh i izbor ob'ektov–etalonov v zadachah raspoznavaniya s uchitelem. [Clustered data analysis and selection objects–standards recognition tasks with the teacher]. *Vychislitel. tehnologii*. [Comput. technology]. 20(6):34–43. (In Russian)
- [9] Borisova, I. A., and O. A. Kutnenko. 2015. Censurirovanie oshibochno klassifitsirovannykh ob'ektov vyborki. [Censurirovanie incorrectly classified the sample object]. *17-ya Vserossiyskaya konferentsiya "Matematicheskie metody raspoznavaniya obrazov – 2015"*. [17th all-Russian Conference "Mathematical methods of pattern recognition-2015]. t. Svetlogorsk. (In Russian)
- [10] Kasymova, R. I. 2009. Kliniko–laboratornye osobennosti ostrykh gnoynykh i seroznykh meningitov v zavisimosti ot etiologii. [Clinical and laboratory features of serous and purulent meningitis, depending on etiology]. *Dis. ... kand. med. nauk*. [Medicine science cand. ... diss]. 145 p. (In Russian)

Received April 28, 2018