Министерство образования и науки Российской Федерации
Российская академия наук
Российский фонд фундаментальных исследований
АО «Концерн «ЦНИИ «Электроприбор»
Институт проблем управления им. В. А. Трапезникова РАН
Институт проблем механики им. А. Ю. Ишлинского РАН
Южный федеральный университет
Научно-исследовательский институт многопроцессорных вычислительных систем им. академика А. В. Каляева Южного федерального университета
Российская Ассоциация искусственного интеллекта
Научный совет РАН по мехатронике и робототехнике
Научный совет РАН по комплексным проблемам управления и автоматизации

---

# ДЕСЯТАЯ ВСЕРОССИЙСКАЯ МУЛЬТИКОНФЕРЕНЦИЯ ПО ПРОБЛЕМАМ УПРАВЛЕНИЯ

## МКПУ-2017

Материалы
10-й Всероссийской мультиконференции
(с. Дивноморское, Геленджик, Россия,
11–16 сентября 2017 г.)

## ТОМ 1

Ростов-на-Дону – Таганрог
Издательство Южного федерального университета
2017



*Д.А. Вятченин, А.А. Дудкин*

## ОБ ИЕРАРХИЧЕСКОМ ПРЕДСТАВЛЕНИИ СТРУКТУРЫ ПРИЗНАКОВОГО ПРОСТРАНСТВА ИССЛЕДУЕМОЙ СОВОКУПНОСТИ ПОСРЕДСТВОМ ЭВРИСТИЧЕСКОЙ ВОЗМОЖНОСТНОЙ КЛАСТЕРИЗАЦИИ

*Объединенный институт проблем информатики НАН Беларуси,*
*г. Минск, Беларусь,*
*viattchenin@mail.ru*

### Введение

При решении задач классификации особую роль играет задача отбора наиболее информативных признаков, для чего традиционно используются метод главных компонент, метод многомерного шкалирования, а также метод экстремальной группировки признаков, в котором существенную роль играют методы кластерного анализа [1]. Результатом работы подобных процедур является минимальное подмножество исходного множества признаков. Однако зачастую исследователю необходимо иметь наглядное представление о стратификационной структуре исходного признакового пространства для построения классификации, отвечающей конечной цели исследования. Подобные методы исследования признакового пространства разработаны недостаточно; можно лишь указать FHDR-алгоритм, предложенный в [2].

В работе рассматриваются основы подхода к построению нечеткой иерархии признакового пространства, основанного на эвристической возможностной кластеризации.

### Эвристический подход к возможностной кластеризации

Эвристический подход к возможностной кластеризации, предложенный в [3], заключается в определении нечеткого кластера как нечеткого множества уровня $\alpha$, $\alpha \in (0,1]$, порожденного нечеткой толерантностью, описывающей геометрическую структуру исследуемой совокупности, и построении так называемого распределения $R_c^*(X)$ по априори известному или нет числу $c$ нечетких $\alpha$-кластеров. Прямые эвристические алгоритмы возможностной кластеризации, в зависимости от матрицы исходных данных, условно делятся на реляционные, матрицей исходных данных для которых служит матрица нечеткой толерантности $T=[\mu_T(x_i,x_j)]$, $i,j=1,\ldots,n$, определенной на универсуме $X=\{x_1,\ldots,x_n\}$, и алгоритмы, основанные на вычислении прототипов, матрицей исходных данных для которых служит матрица вида «объект-

признак» $\hat{X} = [\hat{x}_i^{t_1}]$, $i = 1,\ldots,n$, $t_1 = 1,\ldots,m_1$. В отличие от прямых кластер-процедур, отыскивающих единственное распределение $R_c^*(X)$, иерархический H-AFC-TC-алгоритм строит иерархию распределений $R_{\alpha_0}^*(X) \supseteq \ldots R_{\alpha_\ell}^*(X) \supseteq \ldots \supseteq R_{\alpha_Z}^*(X)$ по полностью разделенным нечетким $\alpha$ -кластерам. Следует указать, что для всех эвристических возможностных кластер-процедур, основанных на вычислении прототипов, общим, а в большинстве случаев и единственным параметром является расстояние между нечеткими множествами.

Эвристические возможностные кластер-процедуры используются в различных кластерных техниках, в том числе и при решении задачи снижения признакового пространства; в этой связи уместно упомянуть о D-PAFC-TC-FS-алгоритме [4], использующем вместо расстояния между нечеткими множествами коэффициент корреляции между нечеткими множествами и строящем так называемое главное распределение $R_P^*(X)$ по нечетким $\alpha$ -кластерам, типичные точки которых и будут формировать подмножество наиболее информативных признаков.

**Корреляция между нечеткими множествами**

Пусть $\hat{X} = [\hat{x}_i^{t_1}]$, $i = 1,\ldots,n$, $t_1 = 1,\ldots,m_1$ – матрица исходных данных, которые могут быть нормализованы в соответствии с формулой

$$x_i^{t_1} = \frac{\hat{x}_i^{t_1} - \min\limits_{t_1} \hat{x}_i^{t_1}}{\max\limits_{t_1} \hat{x}_i^{t_1} - \min\limits_{t_1} \hat{x}_i^{t_1}}, \; i = 1,\ldots,n, \; t_1 = 1,\ldots,m_1, \qquad (1)$$

так что каждый признак $x^{t_1}$, $t_1 = 1,\ldots,m_1$, может трактоваться как нечеткое множество на универсуме объектов исследуемой совокупности с функцией принадлежности $\mu_{x^{t_1}}(x_i)$.

Коэффициент корреляции между нечеткими множествами $x^{t_1}$ и $x^{p_1}$, предложенный в [5], для данного случая имеет вид

$$r(x^{t_1}, x^{p_1}) = 1 - \left( \sum_{i=1}^{n} \left| \frac{\mu_{x^{t_1}}(x_i)}{\left( \sum\limits_{i=1}^{n} \mu_{x^{t_1}}^{\lambda}(x_i) \right)^{\frac{1}{\lambda}}} - \frac{\mu_{x^{p_1}}(x_i)}{\left( \sum\limits_{i=1}^{n} \mu_{x^{p_1}}^{\lambda}(x_i) \right)^{\frac{1}{\lambda}}} \right|^{\lambda} \right)^{\frac{1}{\lambda}}, \qquad (2)$$

где $0 < \lambda < \infty$ является параметром. Матрица коэффициентов корреляции может быть нормализована в соответствии с формулой



$$\tilde{r}(x^{t_1}, x^{p_1}) = \frac{\left( r(x^{t_1}, x^{p_1}) - \min\limits_{t_1, k_1} r(x^{t_1}, x^{p_1}) \right)}{\max\limits_{t_1, p_1} r(x^{t_1}, x^{p_1}) - \min\limits_{t_1, p_1} r(x^{t_1}, x^{p_1})} \qquad (3)$$

и после нормализации может рассматриваться как матрица нечеткой толерантности на универсуме признаков.

Таким образом, для иерархического представления структуры признакового пространства представляется возможным использовать подход, описываемый выражениями (1) – (3) вместо расстояния между нечеткими множествами в H-AFC-TC-алгоритме.

**Заключение**

В докладе подробно рассматриваются основные определения эвристического подхода к возможностной кластеризации и схема предлагаемого H-AFC-TC-FS-алгоритма построения иерархии распределений по нечетким $\alpha$ -кластерам, описывающего структуру признакового пространства исследуемой совокупности объектов.

Таким образом, для каждого из вычисленных уровней $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_\ell < \ldots < \alpha_Z \le 1$ выбирается подмножество типичных точек нечетких $\alpha$ -кластеров, представляющих собой подмножество наиболее информативных признаков каждого уровня, что позволяет производить классификацию исследуемой совокупности для каждого из таких подмножеств с последующим сравнением результатов классификации и выбором наилучшего результата.


1. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин; под ред. С.А. Айвазяна. – Москва: Финансы и статистика, 1989. – 607 с.
2. *Dumitrescu, D.* Hierarchical pattern classification / D. Dumitrescu // Fuzzy Sets and Systems. – 1988. – Vol. 28, No. 2. – P. 145 – 162.
3. *Viattchenin, D.A.* A Heuristic Approach to Possibilistic Clustering: Algorithms and Applications / D.A. Viattchenin. – Heidelberg: Springer, 2013. – 238 p.
4. *Kacprzyk, J.* A new heuristic possibilistic clustering algorithm for feature selection / J. Kacprzyk, J.W. Owsinski, D.A. Viattchenin // Journal of Automation, Mobile Robotics and Intelligent Systems. – 2014. – Vol. 8, No. 2. – P. 40 – 46.
5. *Chaudhuri, B.B.* On correlation between two fuzzy sets / B.B. Chaudhuri, A. Bhattacharya // Fuzzy Sets and Systems. – 2001. – Vol. 118, No. 3. – P. 447 – 456.