

# Работа с реализациями скрытых закономерностей в данных как с независимыми обобщёнными прецедентами

В.В. Рязанов<sup>1</sup>, А.П. Виноградов<sup>1</sup>

<sup>1</sup>Федеральный исследовательский центр "Информатика и управление" РАН, Вавилова 44, Москва, Россия, 119333

## Аннотация

Исследуется подход к анализу данных на основе обобщённых прецедентов как прецедентов проявления скрытых частичных закономерностей, в частности, варианты обеспечения независимости их реализаций.

## Ключевые слова

Скрытая закономерность, обобщённый прецедент, преобразование Хафа, параметрическая гипотеза, многомерная текстура

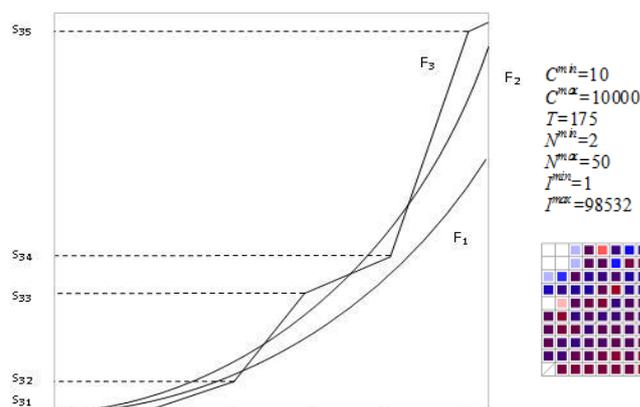
## 1. Введение

Рассматривается подход к анализу данных на основе обобщённых прецедентов (ОП), как прецедентов проявления объективных частичных закономерностей. Ранее в ВЦ РАН В.В. Рязановым и В.А. Ворончихиным был предложен видеологический подход, который предполагает применение быстрых методов обработки изображений (ОИ) к различным срезам и/или проекциям исходной выборки на различные двумерные гиперповерхности в пространстве признаков. Подход на основе ОП можно рассматривать как его обобщение и развитие [1], [2], [3], прослеживаются также параллели с классическим преобразованием Хафа из области ОИ. Мы исследуем ситуации, когда первичная кластеризация служит аналогом пространственного дифференцирования и даёт исходный материал для последующего многоаспектного исследования выборки [4], [5], [6].

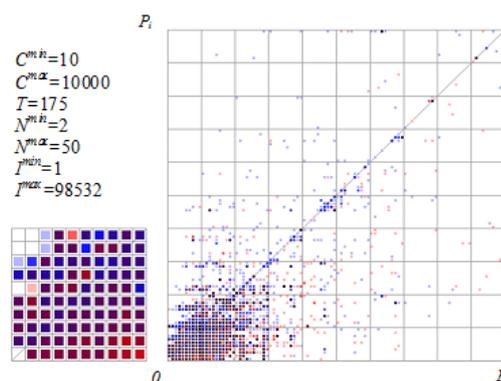
## 2. Кластеризация в роли пространственного дифференцирования и многомерные текстуры реализаций ОП

Пусть из каких-либо соображений задан набор  $\mathbf{B}_s$ ,  $s=1,2,\dots,S$ , параметрических описаний кластеров (Рисунок 1), среди которых далее выбираются базовые [7], и  $\mathbf{Q}_z$ ,  $z = 1,2,\dots,Z$ , - некоторый список критериев качества представления класса  $X_\lambda$ ,  $\lambda=1,2,\dots,l$ , с использованием кластеров формы  $\mathbf{B}_s$ . В обозначениях присутствуют только две переменные, необходимые для задания критерия  $\mathbf{Q}_z = \mathbf{Q}_z(s, \lambda)$ ,  $s=1,2,\dots,S$ ,  $\lambda = 1,2,\dots,l$ , с помощью которого вычисляется  $S \times Z$  матрица голосов за выбор той или иной формы кластера в качестве базовой для класса  $X_\lambda$ . Применяя набор форм  $\mathbf{B}_s$ ,  $s=1,2,\dots,S$ , и критериев  $\mathbf{Q}_z(s, \lambda)$ ,  $s=1,2,\dots,S$ , к  $\lambda$ -му классу  $X_\lambda \subset X$  выборки  $X$ , получим набор матриц  $q_{sz}(\lambda)$ ,  $\lambda=1,2,\dots,l$ , содержащий голоса за базовую форму  $\mathbf{B}_s$  для класса  $\lambda$  по  $z$ -му критерию  $\mathbf{Q}_z$ . Набор  $q_{sz}(\lambda)$ ,  $\lambda=1,2,\dots,l$ , служит основанием для выбора тех или иных форм кластеров в качестве базовых. При этом используется минимум априорных знаний или субъективных предпочтений.

Как и в случае преобразования Хафа, мы отыскиваем проявления типичных ОП в виде особенностей функций плотности вторичных распределений в параметрических пространствах описаний базовых кластеров. В простых случаях, когда базовые кластеры проявляются независимо друг от друга, схема Хафа может быть применена напрямую. Так, на Рисунке 2 приведён пример распределения ОП в пространстве 3-х параметров для большой выборки записей о покупках в некоторой ритейлерской сети. Базовым кластером служит чек, объединяющий несколько покупок.



**Рисунок 1:** Варианты параметризации закономерности



**Рисунок 2:** Пример распределения ОП в пространстве параметров

Поскольку чек отражает прежде всего предпочтения конкретного покупателя, то в некотором приближении перекрёстным влиянием наполнений чеков можно пренебречь. Как и в схеме Хафа, вторичное дискретное распределение здесь формируется из независимых реализаций и поэтому объективно отражает форму обнаруженной закономерности. Задача усложняется, если процедура выбора нелокальна, и реализация любого базового кластера влияет на возможность реализации соседних. Тем не менее, работоспособность подхода на основе ОП сохраняется и здесь. В самом деле, набор матриц  $q_{sz}(\lambda)$ ,  $\lambda=1,2,\dots,l$ , определяет зависимость  $s^*(\lambda) = \min(q_{sz}(\lambda))$ ,  $\lambda=1,2,\dots,l$ , в которой представлены «текстурные» характеристики классов выборки, обеспечивающие преимущества индексов  $s^*$  при выборе типов кластеризации. Конечно, списки  $B_s$ ,  $s=1,2,\dots,S$ , и  $Q_z$ ,  $z=1,2,\dots,Z$ , выбраны субъективно, но решение оптимизационной задачи содержит уже вполне объективную информацию о том, что базовая форма  $s^*(\lambda)$  лучше других подходит для описания класса  $X_j$ . Если данный аспект поведения выборки зафиксирован, то каждая реализация базовой формы  $B_{s^*}$  в представлении класса  $X_j$  может рассматриваться в качестве независимой реализации его текстурных особенностей в целом. Тогда, как и в схеме Хафа, вариации плотности вторичного распределения будут формироваться из независимых проявлений соответствующей частичной закономерности и определять типичные значения параметров базовой формы  $B_{s^*}$ , но – лишь в предположении использования именно данного типа кластеризации.

### 3. Заключение

Изложен вариант подхода к анализу данных сложной природы в ситуациях дефицита априорных оснований для задания оптимального способа кластеризации. Предложен вычислительный инструментарий, позволяющий использовать для этого объективные скрытые закономерности. Подход успешно применялся в нескольких практических приложениях, полученные результаты позволяют говорить о его перспективности и необходимости дальнейшей разработки.

### 4. Благодарности

Работа выполнена при частичной поддержке грантов РФФИ 18-29-03151, 20-01-00609.

### 5. Литература

- [1] Vinogradov, A. Using Bit Representation for Generalized Precedents / A. Vinogradov, Yu. Laptin // Proceedings of International Workshop OGRW-9 (Koblenz, Germany). – 2015. – P. 281-283.

- [2] Ryazanov, V.V. Using generalized precedents for big data sample compression at learning / V.V. Ryazanov, A.P. Vinogradov, Yu.P. Laptin // *Journal of Machine Learning and Data Analysis*. – 2015. – Vol. 1(1). – P. 1910-1918.
- [3] Ryazanov, V. Assembling Decision Rule on the Base of Generalized Precedents / V. Ryazanov, A. Vinogradov, Yu. Laptin // *Information Theories and Applications*. – 2016. – Vol. 23(3). – P. 264-272.
- [4] Nelyubina, E. Transforms of Hough Type in Abstract Feature Space: Generalized Precedents / E. Nelyubina, V. Ryazanov, A. Vinogradov // *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications VISIGRAPP*. – 2017. – Vol. 4. – P. 651-656.
- [5] Nelyubina, E. Shape of Basic Clusters: Finding Coherent ELR-2s via Hough-type Transform / E. Nelyubina, V. Ryazanov, A. Vinogradov // *Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence ICPRAI, Concordia University, Montréal, Canada. CENPARMI, Concordia University, 2018*. – P. 702-706.
- [6] Laptin, Yu.P. Shape of Basic Clusters: Using Analogues of Hough Transform in Higher Dimensions / Yu.P. Laptin, E.A. Nelyubina, V.V. Ryazanov, A.P. Vinogradov // *Pattern Recognition and Image Analysis*. – 2018. – Vol. 28(4). – P. 653-658.
- [7] Naumov, V.A. Analysis and prediction of hydrological series based on generalized precedents / V.A. Naumov, E.A. Nelyubina, V.V. Ryazanov, A.P. Vinogradov // *Book of abstracts of the 12-th International Conference Intelligent Data Processing (IDP-12), Gaeta, Italy, 2018*. – P. 178-179.