

# Использование свойств вейвлет-преобразования в задачах поиска закономерностей

Е.А.Нелюбина

Калининградский государственный  
технический университет  
Калининград, Россия  
e-mail: e.nelubina@gmail.com

В.В.Рязанов

Федеральный исследовательский  
центр "Информатика и управление"  
Российской академии наук  
Москва, Россия  
e-mail: rvvccas@mail.ru

А.П.Виноградов

Федеральный исследовательский  
центр "Информатика и управление"  
Российской академии наук  
Москва, Россия  
e-mail: vngrccas@mail.ru

**Аннотация**—Представлен подход к проблеме поиска закономерностей в данных, основанный на использовании обобщенных прецедентов и адаптации элементов вейвлет-преобразования в повышенных размерностях, приведены примеры решения прикладных задач с этих позиций.

**Ключевые слова**— закономерность, базовый кластер, обобщенный прецедент, вейвлет, преобразование Хафа

## 1. ВВЕДЕНИЕ

В настоящее время в области анализа изображений успешно используется множество подходов и методов, которые эффективны в малых размерностях [1], [2], но их применение в случае многомерных данных затруднено [3], [4], [5]. Далее центральным объектом будет закономерность, описываемая малым числом параметров [6]. Известно, что понятие закономерности является весьма сложным. Мы не касаемся здесь общих вопросов и исследуем ниже параметрические гипотезы о закономерности, используя лишь компетенции в той или иной предметной области и в области ИТ. Ранее был выполнен ряд исследований [6], [7], [8], где при построении гипотезы о закономерности происходит выбор параметрического пространства  $Y$  нужного вида аналогично тому, как это имеет место в схеме преобразования Хафа. Возникающая в  $Y$  вторичная кластерная структура  $c^t \in C^T$  в этом случае содержит информацию о повторяемости закономерности и о типичных значениях параметров. Каждая реализация закономерности (как и представляющий её вектор в пространстве параметров) называется обобщенным прецедентом (ОП), т.е., прецедентом закономерности [6].

## 2. МОДЕЛЬ ЗАКОНОМЕРНОСТИ

Пусть  $X, X \subset R^N$  - выборка оцифрованных данных

большого объёма. Основным объектом будет служить кортеж точек  $x = \{x^1, \dots, x^M\}$ ,  $x^m \in X$ ,  $m = 1, \dots, M$ , для которого выполняются условия, сформулированные экспертом:

$$P_l(x^1, \dots, x^M) = P_l(x^1_1, \dots, x^1_N, \dots, x^M_1, \dots, x^M_N), l = 1, 2, \dots, L.$$

Набор условий  $P = \{P_1, \dots, P_L\}$  представляет собой формулировку гипотезы о наличии закономерности. Если условиями  $P$  задана некоторая пространственная форма расположения точек кортежа  $x = \{x^1, \dots, x^M\}$  в пространстве признаков, то она называется базовым кластером. Для проверки гипотезы требуется выполнить перебор всех вариантов  $x \subset X$ , если нужно обнаружить максимальное число подтверждений закономерности.

Например, если  $P = \{P_1, \dots, P_L\}$ ,  $L = N - 2$ , - система полиномиальных уравнений, то тогда в пространстве  $R^N$  определена двумерная алгебраическая поверхность  $R$ ,  $R \subset R^N$ , и проявление закономерности в точке

$x \in X$  соответствует выполнению условия  $x \in R$ . Если, к тому же, эта система приводится к виду  $x_l = F_l(x^1_1, x^1_2)$ ,  $l = 3, \dots, N$ , и функции  $F_3, \dots, F_N$  не имеют особенностей, то проекция выборки  $X$  на поверхность  $R$  может быть взаимно однозначно отображена на плоскость  $(x^1_1, x^1_2)$ , в частности, на экран монитора. Пример весьма условный, но он указывает на возможный способ задействования зрительной компетенции и/или интуиции эксперта.

## 3. ПРИМЕРЫ ПРИМЕНЕНИЯ МОДЕЛИ

Нас интересуют задачи, где комбинации параметров могут выступать в роли малой волны. В частности, это может быть сам кортеж  $x = \{x^1, \dots, x^M\}$ ,  $x^m \in X$ ,  $m = 1, \dots, M$ , описывающий форму базового кластера. Пусть  $F = (F_1, \dots, F_N)$  - список сложных объектов,  $f = (f_1, \dots, f_N)$ , - перечень вариантов их поведения. Требуется реконструировать  $f$  по имеющейся статистике, когда для наблюдения доступны лишь некоторые интегральные параметры вида  $X(F)$ . Построенная система  $c^t \in C^T$  будет содержать информацию о правильности понимания экспертом поведения объектов  $F = (F_1, \dots, F_N)$ , а также о типичных значениях параметров  $f = (f_1, \dots, f_N)$ .

Данная постановка рассматривалась в [8]. Список  $F$  представлял закономерности стока в регионах бассейна реки, где измерение стока затруднено или невозможно. Выборка представляла собой временной ряд  $x(t) = (x_0, x_1(t), \dots, x_N(t))$ , признаковое пространство  $R^{N+1}$  содержало вычисленные по осадкам уровни влаги в регионах  $(x_1(t), \dots, x_N(t))$ , а также объем стока реки в устье  $x_0(t)$  как единственный измеряемый параметр.

Закономерность стока  $i$ -го региона определяется как функция зависимости расхода от уровня влаги  $x_i$ . Её аппроксимацией может служить вектор  $f_i = (f_i^1, f_i^2, \dots, f_i^N)$ , построенный из дискретных значений расхода в фиксированных точках на шкале уровня влаги. Этот способ не единственный (Рис.1), но предполагается, что он одинаков для всех индексов  $i$ , и  $f$  теперь представляет собой  $N \times D$  матрицу  $f = \{f_i^d\}$ , где  $D$  - число параметров объекта  $F_i$ , в нашем случае, позиций на шкале уровня. Проверять гипотезу  $f = \{f_i^d\}$  необходимо во всех точках ряда  $x(t) = (x_0, x_1(t), \dots, x_N(t))$ , где  $\lambda$  - некоторый порог подтверждения:  $\sum (f_0(t) - \sum f_i(t))^2 < \lambda$ .

Форма функции стока служит аналогом малой волны, которая реализуется в различные моменты  $t$ . Эта форма может проявиться в чистом виде на временной шкале,

например, когда наблюдается одномоментное большое повышение уровня  $x_i(i^*)$  и последующий период свободного стока. Как было отмечено, даже в таких случаях прямое измерение стока не всегда возможно, и адекватность представления функции стока  $i$ -го региона вектором  $f_i$  будет проявляться в виде кластера плотности реализаций  $f$  как ОП в окрестности гиперплоскости  $f_i$ . Заметим, что тем самым определяются не только точки реализаций, но и правильные варианты формы волны. Например, если пользователя интересуют лишь различия формы стока с точки зрения паводковой опасности, то могут быть приемлемыми 2-параметрическое описание

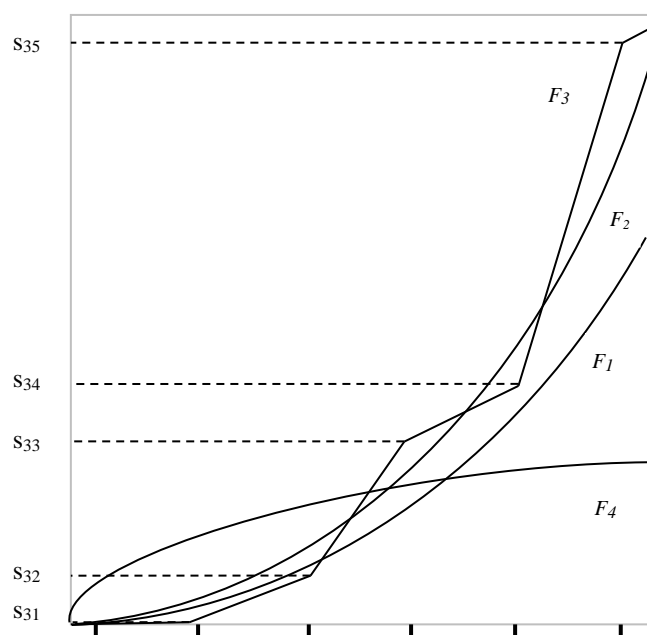


Рис. 1. Параметрическое представление формы числовой зависимости.  $F_1$  и  $F_2$  соответствуют представлению вида  $ax^b$  для выпуклой функции,  $b>1$ ; для  $F_4$  – также  $ax^b$ , но для вогнутой функции,  $b<1$ ; вариант  $F_3$  – кусочно-линейная аппроксимация в виде вектора значений  $s_{id}$  функции в фиксированных точках

вида  $ax^b$  и классификация регионов лишь по знаку неравенства:  $b<1$  или  $b>1$ .

Пусть теперь  $X, X \subset R^N$  – выборка записей параметров индивидов в популяции, например, антропометрических показателей для жителей страны, города, и т.д. Если эксперта-практика интересует отношение «предок-потомок» для  $x^1, x^2$  из  $X, x^1 \neq x^2$ , то определим меру близости кортежа к носителю закономерности сходства:

$$\rho(x, R) = \sum_1^N \sigma_n, \quad \sigma_n = \begin{cases} 1, & |x_n^1 - x_n^2| \leq \varepsilon_n \\ 0, & |x_n^1 - x_n^2| > \varepsilon_n \end{cases}$$

где  $\varepsilon_n$  – набор границ на шкалах антропометрических показателей. Включим условие  $x^1 \neq x^2$  в список  $P$  и будем отображать отметки о наличии закономерности точками на плоскости  $(\rho, \chi)$ , где  $\chi = |x_n^1 - x_n^2|, n'$  – параметр «возраста». Тогда в окрестности некоторых точек  $(\rho=w, \chi \approx 25), 1 < w \leq N$ , на данной плоскости будут наблюдаться вторичные кластеры с повышенной плотностью отметок, поскольку антропометрические

показатели в паре «предок-потомок» часто близки. Представим теперь в виде малой волны гипотезу о возрастных изменениях, приспособив определение меры близости для сравнения двух кортежей:

$$\rho(x, x^*) = \sum_1^N \sigma_n, \quad \sigma_n = \begin{cases} 1, & |x_n^1 + \alpha_n(x) - x_n^2| \leq \varepsilon_n \\ 0, & |x_n^1 + \alpha_n(x) - x_n^2| > \varepsilon_n \end{cases}$$

где вектор-функция  $\alpha(x)$  выступает в роли малой волны, задающей коррекцию эталонного кортежа  $x^*$  согласно возрасту. Постановка  $\alpha(x)=0$  соответствовала ситуации идеального сходства предка и потомка.

Отметим, что в  $x$  содержатся абсолютные координаты объектов в пространстве  $R^N$ , и они могут использоваться для формирования  $Y$  как подходящего параметрического пространства, в котором представлены также различные координатные аспекты поведения ОП.

#### 4. ЗАКЛЮЧЕНИЕ

В работе представлен новый подход к проблеме поиска закономерностей в прикладных данных, в основе которого лежит понятие ОП. Показано, что при этом полезным также оказывается использование некоторых свойств вейвлет-преобразования. В целом, при работе со сложными данными, центральным моментом и в этом случае оказывается адекватное встраивание прикладных компетенций в числовые модели.

#### БЛАГОДАРНОСТИ

Работа выполнена при частичной поддержке РФФИ, проект 20-01-00609.

#### ЛИТЕРАТУРА

- [1] Davies, E.R. Advanced Methods and Deep Learning in Computer Vision / E.R. Davies, M. A. Turk – Oxford: Elsevier, 2022. –562 p. doi.org/10.1016/B978-0-12-822109-9.00002-3.
- [2] Nixon, M.S. Feature Extraction for Image Processing and Computer Vision / M.S. Nixon, A.S. Aguado – Elsevier Ltd., 2020. – 626 p. DOI: 10.1016/C2017-0-02153-5.
- [3] Аникеев, Ф.А. Эффективная реализация быстрого преобразования Хафа с использованием сопроцессора СРСА / Ф.А. Аникеев, Г.О. Райко, Е.Е. Лимонова, М.А. Алиев, Д.П. Николаев // Программирование. – 2021. – Т. 5. – С. 3-11. DOI: 10.31857/S0132347421050022.
- [4] Rinoshika, A. Application of multi-dimensional wavelet transform to fluid mechanics / A. Rinoshika, H. Rinoshika // Theoretical and Applied Mechanics Letters. – 2020. – Vol. 10(2). – P. 98-115. DOI: 10.1016/j.taml.2020.01.017
- [5] De Mauro, A. Understanding Big Data Through a Systematic Literature Review: The ITMI Model / A. De Mauro, M. Greco, M. Grimaldi // International Journal of Information Technology & Decision Making. – 2019. – Vol. 18(4), – P. 1433-1461
- [6] Ryazanov, V. Analogues of Image Analysis Tools in the Problems of Finding Latent Regularities in Big Applied Data / V. Ryazanov, A. Vinogradov // Pattern Recognition and Image Analysis. – 2022. – Vol.32(3). – P. 639-644. DOI: 10.1134/S105466182203035X.
- [7] Ryazanov, V. Dealing with Realizations of Hidden Regularities in Data as Independent Generalized Precedents," / V. Ryazanov, A. Vinogradov // IEEE Xplore Proceedings of 2021 International Conference on Information Technology and Nanotechnology (ITNT). – 2021. – P. 1-3. DOI: 10.1109/ITNT52450.2021.
- [8] 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 Int. Conf. Intelligent Data Processing (IDP-12). – 2018. – P.178-179.