# Об оптимальном выборе закономерностей, составляющих плавно меняющуюся закономерность

***Филипенков Н. В.***
filipenkov@mail.ru
Москва, Вычислительный Центр РАН

В работе [2] был предложен подход к поиску плавно меняющихся закономерностей в пучках временных рядов. Идея подхода состоит в разбиении исходного пучка временных рядов на отрезки, на каждом из которых применяется алгоритм поиска постоянных закономерностей. Наиболее близкие (в смысле определённой в работе [2] меры сходства) закономерности, полученные на различных отрезках, «склеиваются» в плавно меняющуюся закономерность. Однако в упомянутой работе достаточно слабо был освещён вопрос выбора постоянных закономерностей при построении плавно меняющейся закономерности. Настоящая работа ставит своей целью заполнить этот пробел.

*Пучком временных рядов* $\mathfrak{S}$ называется совокупность взаимосвязанных временных рядов $S_i,\ i = 1,\ldots,N$. Каждый ряд $S_i$ представляет собой последовательность чисел конечнозначной логики $E_{k_i}$. Значение ряда $S_i$ в момент времени $t \in \{1,\ldots,T\}$ обозначим $a(i,t)$. *Маской* $\omega$ на прямоугольнике $N \times \Delta$ называется булева матрица размерности $N \times \Delta$ (здесь параметр $\Delta$ определяет максимальный отступ по времени). Число единиц в маске $\omega$ называется *мощностью* маски и обозначается $\|\omega\|$. Элемент маски, находящийся в $i$-ой строке и $j$-ом столбце обозначается $\omega(i,j)$. *Закономерностью* $R$ (постоянной) называется набор $(p,\omega,f)$, где число $p \in \{1,\ldots,N\}$ указывает на целевой ряд (то есть ряд, значения которого определяются закономерностью $R$); маска $\omega$ указывает на значения рядов, являющиеся аргументами функции $f$; частично-определённая функция $f$ задаёт зависимость значений целевого ряда от переменных, на которые указывает маска $\omega$.

$$f\colon E_{k_{i_1}} \times \ldots \times E_{k_{i_{\|\omega\|}}} \to E_{k_p} \cup \{\lambda\},$$

где $\omega(i_1,j_1),\ldots,\omega(i_{\|\omega\|},j_{\|\omega\|})$ — единичные элементы матрицы $\omega$, $p$ — номер целевого ряда, символ $\lambda$ обозначает, что функция $f$ не определена на соответствующем наборе значений переменных.

*Отрезком* $\mathfrak{S}_1$ с *началом* $t_b \in \{0,\ldots,T\}$ и *концом* $t_e \in \{0,\ldots,T\}$, $(t_b < t_e)$ на пучке временных рядов $\mathfrak{S}$ (обозначается $\mathfrak{S}_1 \subset \mathfrak{S}$) назовём матрицу $N \times \theta$, составленную из последовательных столбцов матрицы $\mathfrak{S}$, первым из которых является столбец с номером $t_b$, последним — столбец с номером $t_e$, где $\theta = t_e - t_b + 1$ называется *длиной* отрезка $\mathfrak{S}_1$.

Определим следующие показатели качества постоянных закономерностей. Их названия совпадают с принятыми в нечёткой логике [1] показателями качества правил, так как несут сходный смысл. *Репрезентативностью* $\mathrm{rep}(R)$ $(0 \leqslant \mathrm{rep}(R) \leqslant 1)$ закономерности $R = (p, \omega, f)$ назовём следующую величину:

$$\mathrm{rep}(R) = \frac{|D(\omega)|}{k_{i_1} \cdots k_{i_{\|\omega\|}}},$$

где $|D(\omega)|$ — число наборов из $E_{k_{i_1}} \times \ldots \times E_{k_{i_{\|\omega\|}}}$, на которых значение функции $f$ отлично от $\lambda$, а $\omega(i_1, j_1), \ldots, \omega(i_{\|\omega\|}, j_{\|\omega\|})$ — единичные элементы матрицы $\omega$.

*Локальной эффективностью* $\mathrm{eff}_\varepsilon(R, t)$ $(0 \leqslant \mathrm{eff}(R) \leqslant 1)$ закономерности $R$ в точке $t$ назовём следующую величину:

$$\mathrm{eff}_\varepsilon(R, t) = 1 - \frac{1}{2\varepsilon + 1} \sum_{i=-\varepsilon}^{\varepsilon} \Big( \frac{\hat{a}(t+i) - a(p, t+i)}{k_p} \Big)^2,$$

где $\hat{a}(t)$ — прогноз закономерности $R$ для значения $a(p, t)$ пучка временных рядов.

*Плавно меняющейся закономерностью* $\tilde{R} \in \mathfrak{R}^{T-\Delta}$ на пучке временных рядов $\mathfrak{S}$ называется последовательность постоянных закономерностей $R_{\Delta+1}, \ldots, R_T$ такая, что элементы $a(p, \Delta+1), \ldots, a(p, T)$ пучка временных рядов $\mathfrak{S}$ прогнозируются соответственно постоянными закономерностями $R_{\Delta+1}, \ldots, R_T$. При этом $R_{\Delta+1}, \ldots, R_T$ могут представлять собой одни и те же закономерности.

Определим три основных показателя качества плавно меняющейся закономерности в точке пучка временных рядов.

*Локальной репрезентативностью* $\widetilde{\mathrm{rep}}_\varepsilon(\tilde{R}, t)$ и *локальной эффективностью* $\widetilde{\mathrm{eff}}_\varepsilon(\tilde{R}, t)$ плавно меняющейся закономерности $\tilde{R}$ в $\varepsilon$-окрестности точки $t$ пучка временных рядов $\mathfrak{S}$ называются следующие величины:

$$\widetilde{\mathrm{rep}}(\tilde{R}, t) = \mathrm{rep}(R_t), \qquad \widetilde{\mathrm{eff}}_\varepsilon(\tilde{R}, t) = \mathrm{eff}_\varepsilon(R_t, t).$$

Третьим показателем качества плавно меняющейся закономерности в точке $t$ пучка временных рядов $\mathfrak{S}$ является значение $\rho(R_t, R_{t+1})$ — меры сходства закономерностей $R_t$ и $R_{t+1}$. Для точки $T$ данный показатель принимается равным нулю. Определение меры сходства закономерностей подробно рассмотрено в работе [2]. Заметим, что значение меры сходства закономерностей несёт смысл штрафа за смену закономерностей, и этот штраф тем больше, чем больше различие между закономерностями.

На основании приведённых выше локальных показателей качества плавно меняющейся закономерности $\tilde{R}$ определяются показатели качества $\tilde{R}$ на всём пучке временных рядов. *Средней репрезентативностью* $\widetilde{\mathrm{rep}}(\tilde{R})$ и *средней эффективностью* $\widetilde{\mathrm{eff}}(\tilde{R})$ закономерности $\tilde{R}$ называются следующие показатели:

$$\widetilde{\mathrm{rep}}(\tilde{R}) = \frac{\sum_{i=\Delta+1}^{T} \widetilde{\mathrm{rep}}(\tilde{R}, t)}{T - \Delta}, \qquad \widetilde{\mathrm{eff}}(\tilde{R}) = \frac{\sum_{i=\Delta+1}^{T} \widetilde{\mathrm{eff}}_0(\tilde{R}, t)}{T - \Delta}.$$

Здесь $\widetilde{\mathrm{eff}}_0(\tilde{R}, t)$ — локальная эффективность $\widetilde{eff}_\varepsilon(\tilde{R}, t)$ при $\varepsilon = 0$.

*Длиной* $l(\tilde{R})$ изменяющейся закономерности $\tilde{R}$ называется следующий показатель качества плавно меняющейся закономерности $\tilde{R}$:

$$l(\tilde{R}) = \sum_{t=\Delta+1}^{T-1} \rho(R_t, R_{t+1}).$$

Поиск наилучшей плавно меняющейся закономерности $\tilde{R}$ сводится к задаче многокритериальной оптимизации:

$$\begin{cases} \widetilde{\mathrm{rep}}(\tilde{R}) \to \max\limits_{\tilde{R} \in \mathfrak{R}^{T-\Delta}}; \\ \widetilde{\mathrm{eff}}(\tilde{R}) \to \max\limits_{\tilde{R} \in \mathfrak{R}^{T-\Delta}}; \\ l(\tilde{R}) \to \min\limits_{\tilde{R} \in \mathfrak{R}^{T-\Delta}}. \end{cases}$$

Здесь оптимизация происходит по всем возможным закономерностям $\tilde{R}$, т. е. по всем последовательностям постоянных закономерностей $R_{\Delta+1}, \ldots, R_T$.

## Литература


[1] *Барсегян А. А., Куприянов М. С., Степаненко В. В., Холод И. И.* Методы и модели анализа данных: OLAP и Data Mining. — СПб.: БХВ-Петербург, 2004.

[2] *Филипенков Н. В.* О задачах анализа пучков временных рядов с изменяющимися закономерностями // Искусственный интеллект — Донецк, 2006. — № 2. — С. 125–129.