Орлов Ю.Н.

Статистическое распознавание образов на основе сравнения выборочных распределений

Основные задачи

1. Определение уровня нестационарности массива данных и оптимизация объема выборки

2. Выделение типовых состояний временного ряда и построение паттернов

3. Статистическое распознавание образов

Примеры:

- распознавание разладки во временных рядах;
- определение горизонта стационарности;
- идентификация паттернов.

Проблема сравнения выборочных функций распределения

Даже для одномерного временного ряда надо уметь строить совместные выборочные распределения приращений одного-двух-трех порядков по выборкам любого объема в любой момент времени в пределах заданной совокупности.

Если данных 10^9, то для анализа изменения ВФР с точностью 1% по двум переменным, длине выборки и времени (с учетом сдвига по времени) требуется массив (10^2)^2*(10^9)^3 = 10^31 данных.

$$\rho(N;t;\tau) = \left\| f_N(x,t) - f_N(x,t+\tau) \right\|$$

Согласованный уровень стационарности

$$\rho_{1,2}(N) = \left\| f_{1,N}(x) - f_{2,N}(x) \right\|$$

или

$$\rho_{1,2}(N) = \left\| F_{1,N}(x) - F_{2,N}(x) \right\|$$

Строится функция распределения расстояний

$$G_N(\rho)$$

и определяется СУС $\rho^*(N)$ как решение уравнения

$$G_N\left(\rho^*\right) = 1 - \rho^* / \sup \rho$$

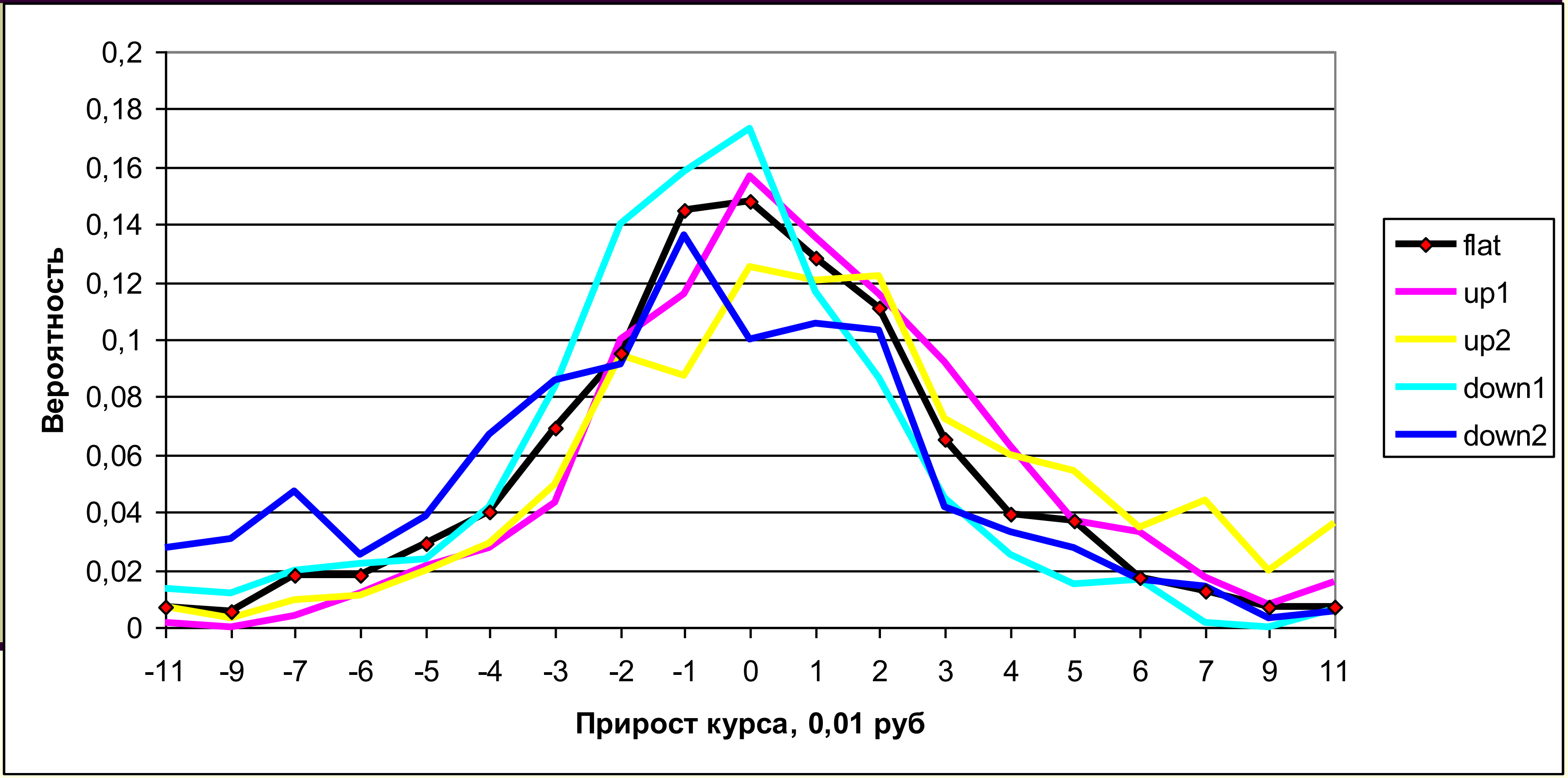
Индекс нестационарности в С

Рассматривается отношение доли расстояний, превосходящих эмпирический СУС, к доле расстояний, превосходящих согласованный уровень значимости:

$$J(N) = \frac{\rho^{*}(N)}{\varepsilon(N)}$$

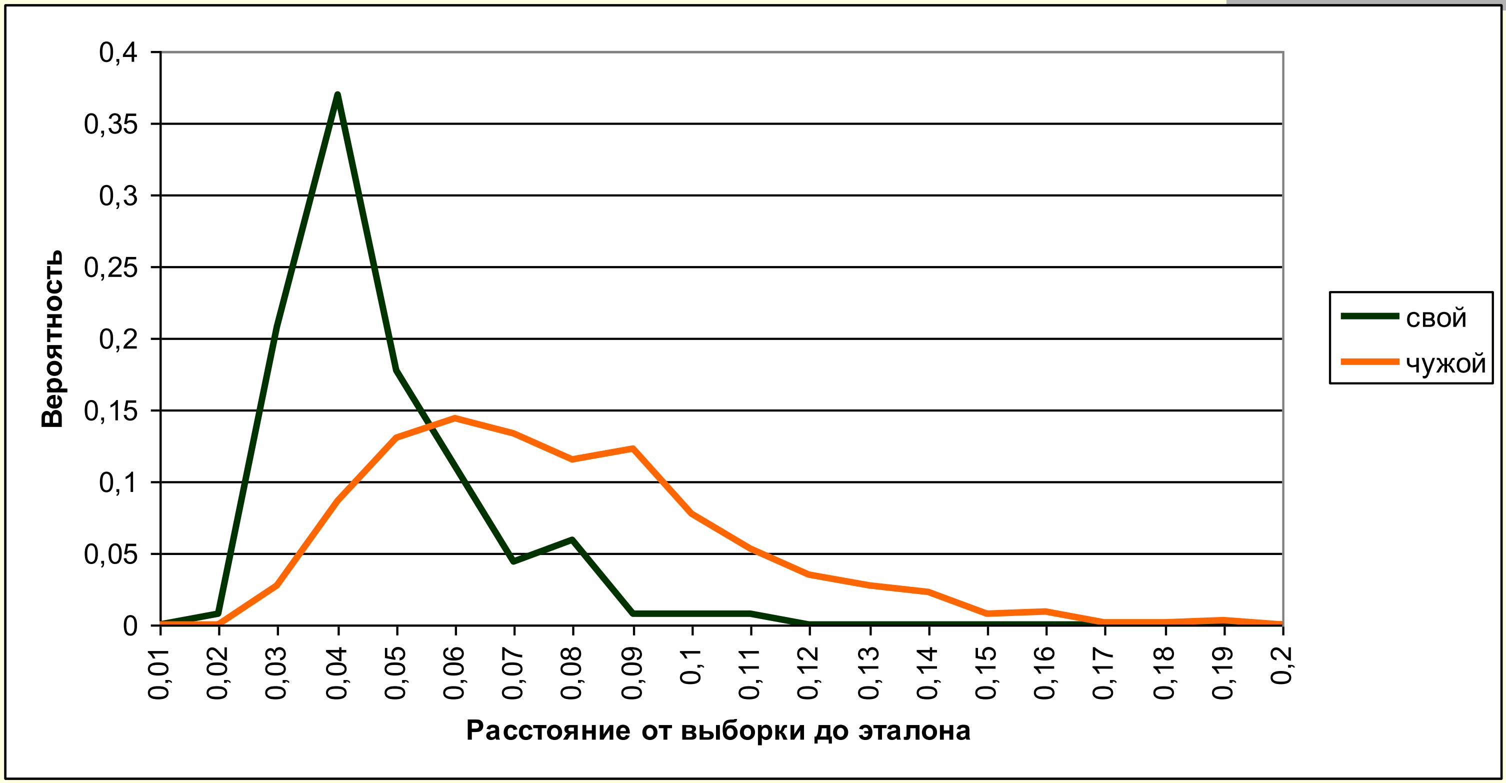
Если J>1, ряд нестационарный; если J<=1, ряд стационарный.

Практические эталоны



Построение нормированных паттернов

Распределения расстояний до эталонов в норме L1-F, длина выборки 30



Выборки длины 30 различаются с ошибкой 20%, длины 60 – с ошибкой 10%, длины 120 – с ошибкой 1%.

Байесовское распознавание

$$\mathbf{f} \in R^n \qquad \left\|\mathbf{f} - \Phi\mathbf{c}\right\|_{L2} \to \min \qquad \Phi = \left(\varphi_1, \ldots, \varphi_p\right),\ \varphi_i \in R^n$$

Делается QR-разложение матрицы Ф эталонов:

$$\mathbf{f} - \Phi\mathbf{c} = I\mathbf{f} - QR\mathbf{c} = \left(I - QQ^T + QQ^T\right)\mathbf{f} - QR\mathbf{c} = Q\left(Q^T\mathbf{f} - R\mathbf{c}\right) + \left(I - QQ^T\right)\mathbf{f}$$

$$\left(R\mathbf{c} - Q^T\mathbf{f}\right)^T Q^T\left(I - QQ^T\right)\mathbf{f} = \left(R\mathbf{c} - Q^T\mathbf{f}\right)^T\left(Q^T{}_{p\times n} I_{n\times n} - I_{p\times p} Q^T{}_{p\times n}\right)\mathbf{f} =$$

$$= \left(R\mathbf{c} - Q^T\mathbf{f}\right)^T O_{p\times n}\mathbf{f} = \mathbf{o}.$$

$$\mathbf{c}_{opt} = R^{-1}Q^T\mathbf{f}$$

$$\sum_{i=1}^{n} f_i = 1, \quad \forall k \sum_{i=1}^{n} \varphi_{k,i} = 1, \ f_i = \sum_{k=1}^{p} c_k \varphi_{k,i} \quad \Rightarrow \sum_{k=1}^{p} c_k = 1$$

$$\mathbf{f} \equiv \varphi_j, \quad j = \arg\max c_k \quad \Rightarrow \quad j = \arg\min \left\|\mathbf{f} - \varphi_k\right\|$$

Используемые нормы

Норма L1 для ВФР или ВПФР:

$$\rho_{L1}^{f}(n)=\int_0^1\left|f_1(x;n)-f_2(x,n)\right|dx;\quad \rho_{L1}^{F}(n)=\int_0^1\left|F_1(x;n)-F_2(x,n)\right|dx$$

Норма C для ВФР или ВПФР:

$$\rho_{C}^{f}(n)=\max_{x\in[0;1]}\left|f_1(x;n)-f_2(x,n)\right|;\quad \rho_{C}^{F}(n)=\max_{x\in[0;1]}\left|F_1(x;n)-F_2(x,n)\right|$$

Норма Хеллингера для ВПФР:

$$\rho_{HE}^{f}(n)=2-2\int_0^1\sqrt{f_1(x;n)\,f_2(x,n)}\,dx$$

Норма Кульбака-Лейблера для ВПФР:

$$\rho_{KL}^{f}(n)=\int_0^1 f_1(x;n)\ln\left(\frac{f_1(x;n)}{f_2(x;n)}\right)dx$$

Практические ограничения Байесовского распознавания

Решающее правило $$\rho_a^0 = \left\| f_0 - f_a \right\|, \quad a^0 = \arg\min_a \rho_a^0$$

корректно, если существует вероятностная интерпретация разложения неизвестного состояния по заданному базису:

$$\left\| \mathbf{f}_0 - \sum_{a=1}^{A} c_a f_a \right\| \le \varepsilon, \quad \sum_{a=1}^{A} c_a = 1, \quad c_a \ge 0; \quad a^0 = \arg\max c_a$$

При распознавании состояния фрагмента разложение по базису из типовых эталонов не имеет места, но решающее правило «ближайшего эталона» остается эффективным, хотя и не обоснованным, инструментом

Среднеквадратичное нарушение вероятностной интерпретации:

Число фрагментов	**2**	**3**	**4**	**10**
1-ПФР	**0,15**	**0,26**	**0,54**	**2,50**
2-ПФР	**0,12**	**0,20**	**0,35**	**1,70**
3-ПФР	**0,09**	**0,18**	**0,29**	**1,25**

СПАСИБО ЗА ВНИМАНИЕ