

Выбор функции расстояния при сравнении фазовых траекторий

Пташко Г.О. Стрижов В.В. Шакин В.В.
(Москва)

При решении задач медицинской диагностики возникает проблема сравнения фазовых траекторий историй болезни пациентов. Предполагается, что пациенты с одинаковым диагнозом имеют сходные траектории [1]. Требуется найти функцию расстояния между траекториями, которая бы удовлетворяла внешнему критерию, задаваемому экспертом. С помощью этой функции создается матрица парных расстояний между траекториями для последующей классификации пациентов по типам болезней.

Задано X – множество многомерных временных рядов $\{X_1, \dots, X_N\}$, где $X_i = \langle X_i^{(1)}, \dots, X_i^{(n_i)} \rangle$ – вектор длины n_i и C – внешний критерий качества классификации, заданный экспертом. Требуется найти такую функцию ρ , которая бы давала наилучшую классификацию кривых из X , минимизирующую критерий $C(\rho_{ij})$. Аргументы функции $\rho(w(X_i, X_j), \varphi)$ – путь минимальной стоимости w , определенный двумя временными рядами, и модельная кривая φ из заданного класса кривых B , приближающая w .

Рассмотрим временные ряды X_i и X_j из множества X . Найдем расстояния μ между всеми парами элементов этих рядов. Элемент матрицы $A = \{a_{kl}\}_{k,l=1}^{m,n}$ равен расстоянию между парой элементов временных рядов X_i и X_j , то есть $a_{kl} = \mu(X_i^{(k)}, X_j^{(l)})$. Путь $w = \{w_l\}$ длины K в матрице A есть последовательность элементов $\{w_l\}$, $l = \overline{1, K}$ таких, что $w_1 = a_{11}$, $w_K = a_{mn}$, а для $w_l = a_{ij}$ элемент $w_{l+1} = a_{i+\xi, j+\zeta}$, $\xi + \zeta \in \{1, 2\}$. Назовем стоимостью пути число $P(w) = \frac{1}{K} \sum_{l=1}^K w_l$. Определим функцию расстояния между двумя временными рядами DTW (Dynamic Time Warping [2]), как минимальную стоимость пути w , то есть $DTW(X_1, X_2) = \min_w \{P(w)\}$. Аппроксимируем путь минимальной

стоимости w модельной кривой $\varphi \in B$. Выберем такую φ , которая минимизирует функционал $C(\rho_{ij})$ для всех данных временных рядов X .

Предложенный метод протестирован на историях болезни пациентов гематологического центра РАМН.

Данная работа поддержана грантом РФФИ 04-01-00401-а.

Литература

1. V. Shakin, V., Ptashko, G. et. al. Decision support system using multimedia case history: quantitative comparison and multivariate statistical analysis // IEEE CBMS 2005 proceedings. Dublin, 2005. pp. 128-133.
2. Selina Chu, Eammon Keogh, David Hart, and Michael Pazzani. Iterative Deepening Dynamic Time Warping for Time Series // Proceedings of the Second SIAM International Conference on Data Mining. 2002