

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

ІНФОРМАТИКА, МАТЕМАТИКА,
АВТОМАТИКА

ІМА :: 2017

**МАТЕРІАЛИ
та програма**

НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

(Суми, 17–21 квітня 2017 року)



Суми
Сумський державний університет
2017

Сокращение размерности выборок данных на основе фрактального анализа

Субботин С.А., профессор

Запорожский национальный технический университет,
г. Запорожье

Сокращение размерности данных является важным этапом при построении диагностических и распознающих моделей по прецедентам. Целью работы являлось создание метода, позволяющего сокращать размерность данных на основе фрактального анализа.

Пусть задана выборка $\langle x, y \rangle$ – набор S прецедентов о зависимости $y(x)$, $x = \{x^s\}$, $y = \{y^s\}$, $s = 1, 2, \dots, S$, характеризующихся набором N входных признаков $\{x_j\}$, $j = 1, 2, \dots, N$, и выходным признаком y . Каждый s -й прецедент представим как $\langle x^s, y^s \rangle$, $x^s = \{x_j^s\}$, $y^s \in \{1, 2, \dots, K\}$, $K > 1$. Задачу сокращения размерности данных представим (1):

$$\begin{cases} \Delta f = |f - f'| \rightarrow \min; \\ \Delta S = S - S' \rightarrow \max; \\ \Delta N = N - N' \rightarrow \max; \\ E \rightarrow \min; \\ K = K', S' \geq K, N' \geq 1. \end{cases} \quad (1)$$

Здесь значком «'» обозначены параметры выборки после сокращения размерности, E – функция ошибки модели. Приведенная задача требует определения критерия f соответствия свойств формируемой подвыборки свойствам исходной выборки.

Предлагается определить f на основе фрактального анализа выборки данных в пространстве признаков по формуле (2):

$$f = \left\{ \log^{-1}(n(l)) \log(l^{-1}) \log(w^{-1}) \right\} E(\text{net}(w)) \leq \varepsilon, \quad (2)$$

где $n(l)$ – число прямоугольных гиперблоков со стороной размером l , покрывающий выборку в пространстве признаков, w – число параметров модели net , ε – допустимое значение ошибки.

Разработанный метод оперирует прямоугольными блоками одинакового размера, покрывая ними пространство признаков с учетом ошибки, точности и сложности синтезируемой модели.

Разработанный метод реализован программно и исследованы при решении задачи сокращения размерности данных. Метод рекомендуется использовать в задачах диагностики и классификации.