МЕТОД МАКСИМИЗАЦИИ СОГЛАСОВАНИЯ ДЛЯ АНАЛИЗА ДАННЫХ В УСЛОВИЯХ ИНТЕРВАЛЬНОЙ НЕОПРЕДЕЛЁННОСТИ

С.П. Шарый

Институт вычислительных технологий, Новосибирск, Россия
e-mail: shary@ict.nsc.ru

Для линейной регрессионной модели $b = a_1x_1 + a_2x_2 + \ldots + a_nx_n$ мы рассматриваем задачу восстановления зависимости при интервальной неопределённости в данных. Пусть интервальная $m \times n$-матрица $\boldsymbol{A} = (\boldsymbol{a}_{ij})$ и интервальный m-вектор $\boldsymbol{b} = (\boldsymbol{b}_i)$ представляют входные воздействия и выходные отклики модели, такие что $a_1 \in \boldsymbol{a}_{i1}$, $a_2 \in \boldsymbol{a}_{i2}$, …, $a_n \in \boldsymbol{a}_{in}$, $b \in \boldsymbol{b}_i$ в i-ом эксперименте, $i = 1, 2, \ldots, m$. Необходимо найти коэффициенты x_i, которые задают линейную зависимость, наилучшим образом приближающую рассматриваемые данные.

Говорят, что семейство значений параметров x_i *согласуется* с интервальными данными $(\boldsymbol{a}_{i1}, \boldsymbol{a}_{i2}, \ldots, \boldsymbol{a}_{in})$, $\boldsymbol{b}_i$, $i = 1, 2, \ldots, m$, если для каждого индекса i существуют такие точечные представители $a_{i1} \in \boldsymbol{a}_{i1}$, $a_{i2} \in \boldsymbol{a}_{i2}$, …, $a_{in} \in \boldsymbol{a}_{in}$, $b_i \in \boldsymbol{b}_i$ что $a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i$. Множество всех значений параметров, согласующихся с данными, образует *множество неопределённости параметров* модели. Оценкой параметров имеет смысл взять точку из этого множества, если оно непусто.

Множество неопределённости параметров, определённое выше, является ничем иным как множеством решений $\varXi(\boldsymbol{A}, \boldsymbol{b}) = \{\, x \in \mathbb{R}^n \mid Ax = b \text{ для каких-то } A \text{ из } \boldsymbol{A} \text{ и } b \text{ из } \boldsymbol{b} \,\}$ интервальной системы линейных уравнений $\boldsymbol{A}x = \boldsymbol{b}$. Для задачи восстановления зависимостей в условиях интервальной неопределённости предлагается в качестве меры согласования брать значения *распознающего функционала* множества решений $\varXi(\boldsymbol{A}, \boldsymbol{b})$, т. е.

$$\mathrm{Uss}(x, \boldsymbol{A}, \boldsymbol{b}) = \min_{1 \le i \le m} \left\{ \,\mathrm{rad}\, \boldsymbol{b}_i + \sum_{j=1}^{n} (\mathrm{rad}\, \boldsymbol{a}_{ij})\, |x_j| - \left| \,\mathrm{mid}\, \boldsymbol{b}_i - \sum_{j=1}^{n} (\mathrm{mid}\, \boldsymbol{a}_{ij})\, x_j \,\right| \right\},$$

где «mid» и «rad» — середина и радиус интервала. Функционал Uss «распознаёт» точки из $\varXi(\boldsymbol{A}, \boldsymbol{b})$ зна́ком своих значений: $x \in \varXi(\boldsymbol{A}, \boldsymbol{b})$ тогда и только тогда, когда $\mathrm{Uss}\,(x, \boldsymbol{A}, \boldsymbol{b}) \ge 0$. Кроме того, Uss имееет неплохие свойства, как функция от x и $\boldsymbol{A}$, $\boldsymbol{b}$.

В качестве оценки параметров линейной зависимости мы берём значение $x = (x_1, x_2, \ldots, x_n)$, которое обеспечивает максимум распознающего функционала Uss (метод максимизации согласования). Тогда, если множество неопределённости параметров непусто, мы получаем точку из него. Напротив, если множество неопределённости параметров пусто, мы получаем точку, которая обеспечивает максимальное возможное согласование с данными (определяемое функционалом Uss).

Мы рассматриваем свойства распознающего функционала Uss, его интерпретацию и свойства оценок, получаемых с помощью метода максимума согласования. Обсуждается также взаимоотношение с другуми методами анализа данных.

ЛИТЕРАТУРА

1. С.П. Шарый *Разрешимость интервальных линейных уравнений и анализ данных с неопределённостями.* – Автоматика и Телемеханика, 2012, №2, с. 111–125.
2. С.П. Шарый, И.А. Шарая *Распознавание разрешимости интервальных уравнений и его приложения к анализу данных.* – Вычислительные Технологии, 2013, т. 18, №3, с. 80–109.