

**С. А. Горбатов, Д. В. Полупанов** (Уфа, Филиал ГОУ ВПО ВЗФЭИ, ГОУ ВПО БашГУ). **Регуляризация нейросетевых моделей стохастических объектов с существенным зашумлением данных.**

Нами исследовались проблемы построения эффективных нейросетевых моделей (НСМ) для задач налогового контроля [1], коррозионного разрушения магистральных трубопроводов [2] и др. в условиях сильных зашумлений базы данных (БД). Составной частью в эти задачи входит подзадача восстановления с помощью НСМ нелинейной функции многих переменных, т. е. обратная задача, некорректная по Адамару. Эффективным инструментом ее решения является метод регуляризации по Тихонову [3].

Пусть оператор  $Az = u$ ,  $z \in Z$ ,  $u \in U$ , ставит в соответствие элементу метрического пространства  $Z$  искомых в обратной задаче восстановления характеристик объекта или процесса элемент пространства характеристик объекта  $U$ . Пусть  $Z^*$  — такой компакт, что  $z \in Z^*$ . Естественной является следующая постановка задачи восстановления [3]:

$$z_\delta: \rho_{\tilde{u}}(Az, u) \leq \delta, \quad z \in Z^*, \quad z_\delta \in Z_\delta, \quad (1)$$

где  $\tilde{u}$  и  $\delta$  известны,  $\rho_{\tilde{u}}(u, \tilde{u}) \leq \delta$  — расстояние между  $u, \tilde{u} \in U$ . Всякое решение (1)  $z_\delta$  является устойчивым приближением к искомому решению  $z$  [3]. Проблему состоятельности задачи (1) можно ослабить за счет улучшения аппроксимативных свойств отображения  $A$  и уменьшения погрешности наблюдений  $\delta$ .

В [1] нами была сформулирована и подтверждена вычислительными экспериментами концепция многоуровневого иерархического структурирования НСМ, включающая итерационные процедуры предобработки БД. Согласно ей, НСМ можно представить в виде

$$\hat{y} = [F_3 \circ F_2 \circ F_1](\mathbf{x}, W(\mathbf{x}, y)), \quad (2)$$

$\mathbf{x}$  — вектор входных факторов,  $y$  — выходная величина,  $\hat{y}$  — ее расчетное значение,  $\{W\}$  — множество синаптических весов,  $Z \equiv \langle \mathbf{x}, y \rangle$ ,  $i = 1, \dots, N$  ( $N$  — общее число наблюдений) — исходная БД,  $U \equiv \hat{Y}$  — множество значений  $\hat{y}$ ,  $A$  — оператор НСМ, состоящий из композиции  $F_3 \circ F_2 \circ F_1$ .

Здесь  $F_1: Z \rightarrow \cup_{q=1}^{Q^*} Z_q$  — оператор оптимальной кластеризации,  $Z_q = \langle \mathbf{x}, y \rangle$  ( $i = 1, \dots, N_q$ ) — БД  $q$ -го кластера;

$$Q^*: \left[ \left( \min_{d_{i,k}} \sum_{q=1}^Q \sum_{i,k} d_{i,k}^2 \right) \cap \left( \min_q \max_q \{ \Phi^{(q)}(Q) \} \right) \right]$$

есть оптимальное число кластеров;  $\Phi^{(q)} = E^{(q)} S^{(q)} R^{(q)}$  — обобщенный критерий качества НСМ;  $E^{(q)} = \|\hat{y}^{(q)} - y\|/\|y\|$  характеризует точность субмоделей  $\hat{y}^{(q)} = f(\mathbf{x}, W(\mathbf{x}, y))$  по возмущению БД;  $S = |\hat{y}_\alpha^{(q)} - \hat{y}_\beta^{(q)}|/\|\mathbf{x}_\alpha - \mathbf{x}_\beta\|$  характеризует устойчивость субмоделей, векторы  $\mathbf{x}_\alpha, \mathbf{x}_\beta$  близки по норме в  $\mathfrak{R}^n$ ,  $\hat{y}_\alpha^{(q)} = f(\mathbf{x}_\alpha, W)$ ,  $\hat{y}_\beta^{(q)} = f(\mathbf{x}_\beta, W)$ ;  $R = 1 - (r_{y, \hat{y}^{(q)}})^2$  определен через коэффициент корреляции между  $y$  и  $\hat{y}^{(q)}$ ;  $d_{i,k}$  — расстояния между элементами в кластере;  $F_2: Z_q \rightarrow Z_q^*$  — оператор очистки кластера, т. е. устранения аномальных наблюдений в  $q$ -м кластере,  $Z_q^* = \langle \mathbf{x}, y \rangle_i$  ( $i = 1, \dots, N_k$ ) — БД очищенного кластера,  $k^*: ((\min_k \Phi^{(k)}) \cap (|(\hat{y}_i(k) - y_i)/y_i| < \varepsilon(k), i = 1, \dots, N_k))$  — номер оптимальной итерации очистки;  $F_3: Z_q \rightarrow \hat{Y}$  — оператор рабочей НСМ,  $\hat{y} = f^*(x, W(x, y)) \in \hat{Y}$  — расчетное значение выходной величины на основе рабочей НСМ.

Процедуры предобработки БД в значительной степени улучшают качество отображения  $A$  по векторному критерию  $\Phi$ .

## СПИСОК ЛИТЕРАТУРЫ

1. Горбатков С. А., Полупанов Д. В. Методы нейроматематики в налоговом контроле. Уфа: РИЦ БашГУ, 2008.
2. Беслебнова Г. А., Горбатков С. А. Итерационный алгоритм формирования «русел» для прогнозной нейросетевой модели коррозионных повреждений магистральных трубопроводов. — Информационные технологии, 2005, № 12.
3. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. М.: Наука, Физматлит, 1979.