

И. А. Карлов, В. Д. Кошур (Красноярск, СФУ). **Сравнительный анализ нейросетевых методов восстановления пропущенных значений в данных.**

В докладе представлен сравнительный анализ нейросетевых методов восстановления пропущенных значений. Проблема пропущенных значений в массивах данных, используемых в различных системах управления, существенно осложняет процесс анализа данных и принятия решений.

Наиболее простым методом восстановления является использование в качестве подстановки среднего, вычисленного по всем известным значениям, но данный метод имеет ряд существенных недостатков. Более эффективными методами восстановления являются метод ближайших соседей и EM-алгоритм [1].

Один из вариантов методов, использующих нейронные сети [2], заключается в предсказании пропущенных значений при помощи многослойной нейронной сети, в которой на вход подаются все известные компоненты вектора.

В основе алгоритма TEST [3] лежит метод обратного распространения ошибки. На нейронную сеть, обученную на полных данных, подается вектор, пропущенные значения в котором заменены средними. В результате вычисляется ошибка между выходом сети и выходным значением, затем при помощи обратного распространения ошибки находятся улучшенные значения пропущенных элементов.

В [4] приводится ряд модификаций алгоритма, осуществляющего построение автоассоциативной нейросети на полных данных. Затем на вход системы подается вектор, содержащий пропущенные значения, и строится функционал ошибки между выходом и входом системы, который минимизируется с использованием генетических алгоритмов.

Эффективность использования описанных методов проверена на конкретных случайных выборках. Выявлено влияние ряда факторов на качество восстановления данных.

СПИСОК ЛИТЕРАТУРЫ

1. *Литтл Р. Дж. А., Рубин Д. Б.* Статистический анализ данных с пропусками. М.: Финансы и статистика, 1991.
2. *Gupta A., Lam M. S.* Estimation missing values using neural networks. — *Journal of Operational Research Society*, 1996, v. 47 (2), p. 229–239.
3. *Yoon S.-Y., Lee S.-Y.* Training algorithm with incomplete data for feed-forward neural networks. — *Neural Processing Letters*, 1999, v. 10, p. 171–179.
4. *Mohamed A. K., Nelwamondo F. V., Marwala T.* Estimating missing data using neural network techniques, principal component analysis and genetic algorithms. — In: *Proceedings of the Eighteenth Annual Symposium of the Pattern Recognition Association of South Africa*, 2007.