

МЕТОДЫ ОБРАБОТКИ ДАННЫХ ДЛЯ ПОВЫШЕНИЯ КАЧЕСТВА ПРОГНОЗИРОВАНИЯ ПРИ ИСПОЛЬЗОВАНИИ РЕГРЕССИОННЫХ МОДЕЛЕЙ

Рычка О.В.

Донецкий национальный технический университет, г. Донецк

На точность прогноза большое влияние оказывает надежность исходных данных. Поэтому необходимо решать проблему наблюдений, которые представляют собой выбросы. На сегодняшний день для проверки подозрительных значений существует ряд критериев. Однако они обладают существенными недостатками: большинство из них применяется для проверки только одного подозрительного наблюдения, а те критерии, которые позволяют проверять на аномальность несколько измерений, сложны и трудоемки. Наличие этих недостатков привело к необходимости разработки новых методов.

В работе рассматриваются два оригинальных метода обработки данных, применение которых ведет к увеличению точности прогнозных регрессионных моделей. При использовании первого метода аномальные измерения исключаются из выборки. Во втором методе аномальные и ненадежные измерения переносятся на определенный уровень. Аномальными или ненадежными наблюдениями, в данном случае, считаются данные, которые не попадают в определенную прямоугольную область.

Стороны прямоугольника равны $2k \cdot \sigma_e$ и $2k \cdot \sigma'_e$, где k – коэффициент, соответствующий вероятности попадания в заданную область (обычно $0,6 \leq k < 3$). Среднеквадратические отклонения невязок σ_e и σ'_e определяются по формулам 1 и 2.

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^n (Y_i - (A \cdot X_i + B))^2}{n - 2}} \quad (1)$$

$$\sigma'_e = \sqrt{\frac{\sum_{i=1}^n (Y_i - (A' \cdot X_i + B'))^2}{n - 2}} \quad (2)$$

где A и B – коэффициенты исходного уравнения; A' и B' – коэффициенты перпендикуляра, построенного к исходному уравнению.

На основе анализа описанных выше методов, можно сделать вывод, что их использование позволяет повысить качество прогнозирования, за счет увеличения значения коэффициента детерминации R^2 .