

Булаенко Д.В.

Харьковский национальный университет радиозлектроники, Харьков, Украина

Об одном методе построения многофакторных моделей временных рядов

В настоящее время в результате развития информационных технологий количество данных, накопленных человечеством в электронном виде, растет быстрыми темпами. Возникает потребность автоматизации обработки данных, так как человек не в состоянии переработать такое количество сведений. Однако, средства хранения данных значительно обгоняют развитие интеллектуальных методов анализа. В рамках данной работы рассматривается задача разработки метода максимальной автоматизации поиска зависимостей в больших массивах данных. Построенные в результате работы метода модели позволяют моделировать, прогнозировать динамику различных факторов и определять взаимосвязи между наблюдаемыми переменными. Ввиду большого объема данных также необходима высокая адаптивность данных методов, чтобы с приходом новых данных не проходила полная параметрическая идентификация на всем массиве данных, а только лишь уточнение параметров модели.

С формальной точки зрения задача сводится к поиску модели вида:

$$f_s(\bar{a}) = F(x_1(t), \dots, x_1(t - \tau_1), x_2(t), \dots, x_2(t - \tau_2), \dots, x_N(t), \dots, x_N(t - \tau_N), S, \bar{a}),$$

где S – структура модели, \bar{a} – вектор параметров, x_i – независимые переменные, τ_1, \dots, τ_N – временные лаги в зависимости между зависимой переменной и соответствующей независимой.

Метод построения данной модели разбивается на несколько этапов. На первом этапе метода происходит кросс-корреляционный анализ между всеми переменными в массиве данных, в результате чего происходит редукция первоначального множества факторов и устанавливаются параметры упреждающих зависимостей. На следующем этапе определяются все возможные зависимости вида $x_i(t) = f(x_j(t - \tau_{ij}))$, при этом оставляются только модели, прошедшие проверку адекватности. Зависимости определяются на классе линейных и полиномиальных моделей, с учетом того, что между переменными может существовать упреждающая зависимость. Далее происходит составление многофакторной модели из полученных бинарных отношений и параметрическая идентификация полной модели. Одно из отличий предлагаемого метода состоит в том, что сформулирована система критериев, позволяющая в автоматическом режиме производить поиск модели. В результате действия алгоритма мы получаем систему функций, которая описывает все существующие парные закономерности между переменными.

Литература

1. Ханк Д.Э., Уичерн Д.У., Райтс А.Дж. Бизнес прогнозирование, 7-е издание. – М.: Издательский дом “Вильямс”, 2003. – 193 с.
2. Дронов С.В. Многомерный статистический анализ: Учебное пособие. Барнаул: Изд-ва Алт. гос. ун-та, 2003. – 213 с.
3. Андерсон Т. Статистический анализ временных рядов. – М.: Издательство “Мир”, 1978. – 757 с.