

УДК 004.8

О МЕТОДЕ ИЗВЛЕЧЕНИЯ ПОЛЕЗНОЙ ИНФОРМАЦИИ ИЗ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДАННЫХ

Е. А. ЯКИМОВ

Научный руководитель О. М. ДЕМИДЕНКО, д-р техн. наук, проф.
Государственное учреждение высшего профессионального образования
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»
Могилев, Беларусь

В последнее время все более часто используются методы извлечения информации, основанные на искусственных нейронных сетях с использованием нечетких множеств и нечеткой логики, кластерном анализе и др. В то же время отсутствуют сведения об использовании в имитационном моделировании сингулярного спектрального анализа временных рядов, который позволяет получить дополнительную информацию разложением исходного временного ряда на трендовую, периодическую и шумовую составляющие.

Предлагается итерационный процесс M извлечения полезной информации, представленный композицией следующих операторов:

$$M = T_6 \circ T_5 \circ T_4 \circ T_3 \circ T_2 \circ T_1,$$

где оператор T_1 представляет отбор данных, накопленных в комплексных информационных системах предприятий, определение (с привлечением экспертов в исследуемой предметной области) переменных для решения поставленной задачи; T_2 – предварительная обработка данных, устранение выбросов и пропущенных данных, отличается применением методов кластерного анализа для группировки данных; T_3 – выбор данных для исследования в соответствии с решаемой задачей, отличается формированием XML-файла для автоматизированной обработки; T_4 – выбор метода поиска закономерностей, основанных в основном на применении статистических исследований, отличается использованием метода сингулярного спектрального анализа и обобщенного закона распределения для оценки структуры распределения случайных величин; T_5 – построение модели исходных данных для применения в имитационной модели на этапе ее эксплуатации; T_6 – формирование XML-файла с моделью последовательности данных для последующего применения.

Для разработки технологии извлечения знаний из накопленных баз данных исследован сингулярный спектральный метод, включающий этапы: вложение, сингулярное разложение, группировку, диагональное усреднение. Построение обобщенного закона распределения для оценки структуры распределения случайных величин обеспечивается программой BelSim2#.random в составе программно-технологического комплекса имитации сложных систем BelSim2#.