



ПРИМЕНЕНИЕ ФОРМАЛЬНЫХ МЕТОДОВ ПРИ РАЗРАБОТКЕ И
СОПРОВОЖДЕНИИ ПРОГРАММНОЙ СИСТЕМЫ АНАЛИЗА ДАННЫХ
НА АНОМАЛЬНОСТЬ

Дударь З.В., Кобзев В.Г., Шубин И.Ю.

Харьковский национальный университет радиоэлектроники

Важной составляющей современных систем интеллектуального анализа данных DataMining [1], а также анализа и прогнозирования больших наборов данных BigData являются подсистемы проверки анализируемых наборов на наличие в них выбросов (на аномальность отдельных значений).

Программная система анализа данных на аномальность предназначена для проверки присутствия в наборах входных данных различной природы отдельных элементов, несоответствующих известным или скрытым закономерностям, которые свойственны конечным совокупностям имеющихся данных. Процесс проверки базируется на известных или предполагаемых общих закономерностях образования (поведения) данных и возможных видах их нарушения, имеет определенную последовательность шагов, которые соответствуют переходам анализируемого набора данных по конечному множеству взаимосвязанных состояний валидации. Обеспечение корректности и полноты анализа данных на аномальность требует соблюдения формальных методов разработки и сопровождения такой системы в соответствии с известными признаками происхождения самих данных.

Для наборов данных, которые не характеризуют развитие каких-либо процессов во времени или в пространстве, описание последовательностей этапов проведения анализа необходимо проводить путем разработки и использования соответствующих формальных грамматик. Анализ должен предусматривать последовательную проверку на аномальность как одиночных элементов, так и их групп среди односторонних или двусторонних экстремальных значений в конкретном наборе.

Для анализа наборов данных, которые по своей сути являются временными рядами, необходимо применять средства известных разновидностей темпоральных логик. При этом должна быть обеспечена возможность проверки новых по времени поступления данных на соответствие как изначально предполагаемым закономерностям, так и закономерностям, установленным по результатам обработки ранее полученных наборов данных.

В обоих случаях тщательно обоснованный выбор формального аппарата описания логических последовательностей действий по анализу данных в рамках знаний существующих априорных и скрытых закономерностей позволяет достичь достаточного уровня уверенности в выводах о наличии аномальных данных.

1. Han, Jiawei. Data mining: concepts and techniques / Jiawei Han, Micheline Kamber, Jian Pei. – 3rd ed. - Morgan Kaufmann Publishers is an imprint of Elsevier. 2012. – 740p.