

Выделение информации и анализ данных для слабоформализованных предметных областей

М.А. Гузев

Институт прикладной математики ДВО РАН

guzev@iam.dvo.ru

Е.Ю. Никитина

Дальневосточный федеральный университет

nikitina@imcs.dvfu.ru

Во многих слабо-формализованных предметных областях выделение информации из накопленных данных часто проводится с целью определения влияющих факторов, а также изучения поведения моделируемых объектов предметной области во времени. Трудность решения этой проблемы требует разработки новых методов и подходов для анализа накопленных данных, результаты которых могут использоваться при построении моделей предметных областей с целью получения информации и поддержки процесса принятия решений. Для этого в анализируемой предметной области выделим классы объектов, каждый из которых обладает набором характеристик, зависящих от набора параметров и факторы, которые могут влиять на значения характеристик объектов предметной области. Основная идея предлагаемого нами подхода состоит в построении ранговых распределений для функций, задающих значения характеристик и факторов и сравнение полученных распределений. При анализе информации рассматривается проекция каждой построенной функции на некоторую ось, задающую значения параметра характеристики или фактора. Предполагается, что фактор влияет на характеристику, если распределение его ранговых значений совпадает с распределением значений характеристики. Приложение сформулированного выше подхода было выполнено для различных предметных областей: биология (выявление факторов, влияющих на структуру популяций лососевых), криминология (выявление влияющих факторов на состояние преступности, анализ изменения Уголовного Кодекса РФ), библиометрия (изучение эффективности научных публикаций). Проведенный анализ структуры распределений наборов данных позволяет выделить лакунарные области в наборах характеристик, выдать рекомендации по корректировке и получению новых данных.