

УДК 007:62.50

МЕТРИЗАЦИЯ НЕКОЛИЧЕСТВЕННЫХ ОПИСАНИЙ ПРИ АНАЛИЗЕ РАЗНОТИПНЫХ ДАННЫХ

А.М.Никифоров

При статистическом анализе неколичественных данных широко используются методы, основанные на определении метрики для градаций категоризованных переменных. При этом в результате анализа таблиц сопряженности ищутся такие числовые метки для градаций неколичественных показателей, при которых достигается экстремальных значений некоторый функционал ковариационной матрицы "оцифрованных" переменных. Нами предложена методика метризации неколичественных описаний для некоторых задач статистического анализа информации с целью выявления отражаемых этими данными закономерностей в случае, когда пространство описания выборки наряду с качественными и номинальными признаками включает также и количественные показатели. Методика предусматривает также возможность сохранения отношения порядка при оцифровке качественных признаков. Для задач корреляционного, регрессионного, факторного и дискриминантного анализов получены критерии поиска оптимальных наборов числовых меток, представимые в виде минимизации отношения квадратичных форм от искомым наборов числовых меток с матрицами, элементы которых характеризуют взаимозависимость разнотипных признаков, формирующих пространство описания выборки.

Приведение положительно-определенных квадратичных форм к главным осям позволяет найти оптимальный набор меток на множестве действительных чисел. Общность критерия для различных статистических методов позволяет построить единый алгоритм метризации неколичественных описаний. Для номинальных признаков разработана итерационная процедура поиска меток на множестве натуральных чисел,

сводящая по сути дела задачу оцифровки номинальных признаков к индуцированию отношения порядка для шкалы наименований. При определении метрики для качественных признаков, при условии сохранения отношения порядка градаций, переменные квадратичных форм подвергаются преобразованию, приводящему к задаче выпуклого программирования с ограничениями в форме неравенств и равенств.

Таким образом, предлагаемая методика дает возможность осуществить переход от бедных шкал к более богатым, индуцируя сначала для шкал наименований отношение порядка, а затем сохраняя порядок градаций при оцифровке качественных признаков, что позволяет наиболее полно использовать априорную информацию при реализации диалогового режима работы процедур метризации. При программной реализации данного подхода процедуры метризации описаний подключаются для решения задач статистической обработки в следующей последовательности: корреляционный анализ, регрессионный, факторный, дискриминантный анализ. Эта последовательность обусловлена постепенным усложнением интерпретируемости результатов оцифровки для указанных задач, а также возможностью использовать результаты работы процедур на ранних этапах в качестве начальных приближений для последующих. Рассмотрена диалоговая система статистической обработки разнотипных данных, сочетающая возможности формальных математических методов в рамках предлагаемого подхода, с преимуществами человеко-машинного взаимодействия.