

Изложенные алгоритмы могут служить основой автоматизированной обработки нестационарных временных рядов с нелинейным трендом.

ЛИТЕРАТУРА

1. Тюкин И. Ю., Терехов В. А. Адаптация в нелинейных динамических системах. СПб.: ЛКИ, 2008. 384 с.
2. Колесникова С. И., Лаходынов В. С., Цой Ю. Р. Исследование качества распознавания состояний стохастической системы // Информационные технологии. 2010. № 6. С. 21–31.
3. Волченко Е. В. Модифицированный метод потенциальных функций // Бионика интеллекта. 2006. № 1. С. 86–92.
4. Zagoruiko N. G., Borisova I. A., Dyubanov V. V., Kutnenko O. A. Methods of Recognition Based on the Function of Rival Similarity // Pattern Recognition and Image Analysis. 2008. No. 1. P. 1–16.
5. Тырсин А. Н. Идентификация нестационарных экономических процессов на основе дискретно-совпадающих моделей авторегрессии // Изв. Уральского государственного экономического университета. 2004. № 9. С. 44–51.

УДК 007.52:681.518.2;681.327.12

ЛОГИЧЕСКИЙ ВЫВОД НА ОСНОВЕ ОПТИМАЛЬНОГО ПОДМНОЖЕСТВА СМЕШАННЫХ ДИАГНОСТИЧЕСКИХ ТЕСТОВ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ¹

А. Е. Янковская, А. И. Гедике

Существует широкий круг проблемных областей, для которых характерно последовательное уточнение решений, принимаемых относительно исследуемых объектов, где описания объектов задаются сотнями и даже тысячами признаков. Поэтому весьма актуальной является разработка алгоритмов сокращения признакового пространства, необходимого для принятия решений, и поэтапного логического вывода решений в ориентированных на подобные задачи интеллектуальных системах. Существенными при этом являются такие параметры, как повышение быстродействия и уменьшение трудоемкости (стоимости) принятия решений, естественно, без потери качества принимаемых решений.

В рамках развиваемого логико-комбинаторного подхода к принятию решений, базирующегося на матричной модели представления данных и знаний и построении логических тестов [1], по каждому из которых осуществляется логический вывод, сокращение признакового пространства достигается построением на основе многокритериального выбора оптимального подмножества (ОП) безызбыточных безусловных диагностических тестов (ББДТ) [2].

В [3] обоснована целесообразность использования смешанных диагностических тестов (СДТ), представляющих собой оптимальное сочетание безусловных и условных составляющих [4], что позволяет одновременно с построением СДТ принимать решения при меньших вычислительных и стоимостных затратах.

В [5] предложен алгоритм построения СДТ на основе ОП ББДТ.

Для представления данных и знаний используются матрицы описаний (**Q**) в пространстве характеристических признаков и матрицы различий (**R**) в пространстве

¹Работа поддержана грантом РФФИ, проект № 10-01-00462а.

классификационных признаков, которым сопоставлены различные механизмы классификации.

Элемент q_{ij} матрицы $\mathbf{Q}$ принимает значения из множества $\{0, 1, -\}$, где единица (ноль) означает, что j -й признак присущ (не присущ) i -му объекту, а прочерк означает, что значение признака безразлично для данного объекта, т. е. оно может равняться и 0, и 1. Элементы целочисленной матрицы $\mathbf{R}$ интерпретируются как номера классов, которым принадлежат объекты по соответствующим механизмам классификации. Считается, что объекты, для которых заданы одинаковые решения, сопоставленные строкам матрицы $\mathbf{R}$, принадлежат одному и тому же образу, а число образов равно числу различных решений.

ОП ББДТ представляется двоичной матрицей $\mathbf{T}_o$, строки которой сопоставляются тестам, столбцы — признакам, входящим во все тесты (единица означает, что признак входит в соответствующий этой строке тест). Отметим, что в процессе анализа обучающих объектов вычисляются весовые коэффициенты признаков (ВКП), характеризующие их вклад в различимость объектов из разных образов.

Будем называть ядром ($\mathbf{G}$) подмножество признаков, входящих в каждый тест из ОП ББДТ. Ядро образует безусловную составляющую всех СДТ, которые могут быть построены по ОП ББДТ.

Описание исследуемого объекта формируется путем последовательного ввода значений запрашиваемых признаков, входящих в очередной тест, из множества $\{0, 1, -\}$, где прочерк (в отличие от матрицы $\mathbf{Q}$) означает, что значение признака неизвестно, т. е. равно 0, или 1.

Реакцией матрицы $\mathbf{Q}$ на i -й тест называется подмножество строк матрицы $\mathbf{Q}$, в которых значение 0 (или 1) каждого признака, входящего в тест, совпадает со значением 0 (или 1) вводимых признаков. Считается также, что значение «—» соответствующего элемента матрицы $\mathbf{Q}$ совпадает со значением 0 (или 1) вводимого признака. Если же значение вводимого признака равно «—», то нужно строить две реакции: и по значению 0, и по значению 1. Реакция матрицы $\mathbf{Q}$ на тест строится в виде последовательности реакций матрицы $\mathbf{Q}$ на значения уже введенных на текущий момент признаков.

Предлагается осуществлять логический вывод в процессе построения дерева СДТ.

Корню дерева сопоставляется матрица $\mathbf{Q}$ и приписываются номера признаков, входящих в ядро $\mathbf{G}$. Если $\mathbf{G} = \emptyset$, то корню дерева СДТ приписывается номер признака, выбранного первым для ввода. Ребрам, ведущим к первому ярусу, приписываются значения всех введенных на текущий момент признаков.

Вершинам первого яруса сопоставляются реакции матрицы $\mathbf{Q}$ на значения ведущих к ним ребер и приписываются значения, равные номерам признаков, выбранных для ввода. Ребрам, выходящим из этих вершин, приписываются последовательности значений введенных на текущий момент признаков.

Аналогично вершинам следующих ярусов сопоставляются реакции матрицы $\mathbf{Q}$ на значения ведущих к ним ребер и приписываются (кроме концевых вершин) значения, равные номерам признаков, выбранных для ввода. Ребрам, выходящим из этих вершин, приписываются последовательности значений введенных на текущий момент признаков.

Каждой концевой вершине соответствует реакция, содержащая строки, сопоставленные только одному образу, номер которого приписывается этой вершине. Отметим, что концевые вершины могут достигаться на подмножествах признаков, входящих в тест.

При построении дерева СДТ существенное значение имеет очередность рассмотрения как ББДТ, так и признаков, входящих в каждый из них. Для выбора очередного ББДТ из матрицы T_o и очередного признака в ББДТ предлагаются следующие критерии.

Критерий 1. Выбирается неотмеченная строка матрицы T_o , имеющая наименьшее число неотмеченных признаков. Если таковых несколько, то выбирается строка с наибольшей суммой весовых коэффициентов неотмеченных признаков. Если же таковых тоже несколько, то выбирается строка с наименьшим номером.

Критерий 2. Выбирается неотмеченный признак в строке матрицы T_o , для которого в соответствующем столбце имеется наибольшее число единиц в неотмеченных строках. Если таковых несколько, то выбирается признак с наибольшим весовым коэффициентом. Если же таковых тоже несколько, то выбирается признак с наименьшим номером.

Критерии 1 и 2 позволяют сокращать перебор и, как правило, строить дерево СДТ не по всем ББДТ из ОП, что снижает вычислительные затраты.

Построение каждого СДТ выполняется в интерактивном режиме одновременно с принятием решения относительно исследуемого объекта на основе голосов, отданных соответствующим тестом за образы. Если для некоторых ББДТ из ОП значения всех признаков уже известны, но эти ББДТ не участвовали в построении дерева СДТ, то на их основе тоже принимаются решения.

Сначала строится одна ветвь дерева СДТ по одному из ББДТ, представленному строкой матрицы T_o . При этом все запрашиваемые для ввода признаки отмечаются. Затем последовательно строятся все возможные ветви по другим ББДТ с учетом значений ранее введенных и отмеченных признаков. Строки матрицы T_o , по которым строятся СДТ, тоже отмечаются.

Если при формировании описания исследуемого объекта ни один вводимый признак не принимает значение «—», то каждый СДТ задает единственный путь от корня к концевой вершине, иначе появляются разветвления. В первом случае голос, отданный тестом за образ, принимается равным единице. Во втором случае голос, отданный тестом за каждый образ, делится пропорционально количеству концевых вершин для данного СДТ.

Процесс построения дерева СДТ заканчивается, когда будут отмечены все признаки, сопоставленные столбцам матрицы T_o . Если при этом в матрице T_o останутся неотмеченные строки, то для соответствующих этим строкам ББДТ ищутся пути в уже построенном дереве СДТ с одновременным подсчетом голосов, отданных этими тестами за образы. Итоговым решением является принадлежность исследуемого объекта образу, получившему наибольшее число голосов.

Результаты испытаний алгоритма на тестовых примерах позволяют сделать вывод, что предлагаемый алгоритм построения ОП СДТ является эффективным по быстродействию, так как позволяет сократить перебор (поскольку длина СДТ не больше длины ББДТ и принятие решения осуществляется одновременно с построением СДТ) [3], а также дает возможность строить дерево СДТ не по всем ББДТ из ОП, что существенно снижает вычислительные затраты. В настоящее время ведется работа по программной реализации алгоритма построения ОП СДТ с одновременным принятием решений в интеллектуальном инструментальном средстве ИМСЛОГ [6], являющемся интегрированной средой разработки прикладных интеллектуальных систем.

ЛИТЕРАТУРА

1. Янковская А. Е. Логические тесты и средства когнитивной графики в интеллектуальной системе // Новые информационные технологии в исследовании дискретных структур. Докл. 3-й Всерос. конф. с международным участием. Томск: Изд-во СО РАН, 2000. С. 163–168.
2. Янковская А. Е. Критерии оптимизации выбора безызбыточных диагностических тестов для принятия решений в интеллектуальных диагностических системах // Математические методы распознавания образов: сб. докл. 13-й Всерос. конф. (ММРО-13). М.: МАКС Пресс, 2007. С. 73–76.
3. Янковская А. Е., Ильинский Н. Н., Черногорюк Г. Э., Кузоваткин А. Н. Результаты исследований алгоритма принятия решений непосредственно в процессе построения смешанных диагностических тестов, реализованного в интеллектуальных биомедицинских системах // Естествознание и гуманизм: сб. науч. трудов. Т. 1. № 3. Томск: Издание СибГМУ, 2004. С. 95–99.
4. Yankovskaya A. E. Design of Optimal Mixed Diagnostic Test With Reference to the Problems of Evolutionary Computation // Proc. of the First International Conference on Evolutionary Computation and Its Applications (EVCA'96). Moscow, 1996. P. 292–297.
5. Yankovskaya A. E., Gedike A. I. Mixed Diagnostic Tests Building from an Optimal Unconditional Test Subset in Intelligent Pattern Recognition // Pattern Recognition and Image Analysis. 2009. V. 19. No. 4. P. 575–582.
6. Yankovskaya A. E., Gedike A. I., Ametov R. V., Bleikher A. M. IMSLOG-2002 Software Tool for Supporting Information Technologies of Test Pattern Recognition // Pattern Recognition and Image Analysis. 2003. V. 13. No. 4. P. 650–657.