Российская академия наук
Отделение математических наук РАН
Вычислительный центр им. А. А. Дородницына РАН
Национальный комитет РАН по распознаванию образов и анализу изображений

Национальная академия наук Украины
Крымский научный центр НАНУ
Крымская академия наук
Таврический национальный университет им. В. И. Вернадского

Национальная академия наук Беларуси

при поддержке

Российского фонда фундаментальных исследований
Компаний Форексис и ЦСПиР

Интеллектуализация обработки информации

ИОИ-2010

Республика Кипр, г. Пафос, 17–24 октября 2010

Доклады 8-й международной конференции

Москва, 2010

УДК 004.85+004.89+004.93+519.2+519.25+519.7
ББК 22.1:32.973.26-018.2
И73

Интеллектуализация обработки информации: 8-я международная конференция.
И73 Республика Кипр, г. Пафос, 17–24 октября 2010 г.: Сборник докладов. — М.: МАКС Пресс, 2010. — 556 с.
ISBN 978-5-317-03409-2

В сборнике представлены доклады 8-й Международной конференции «Интеллектуализация обработки информации–2010», проводимой Вычислительным центром им. А. А. Дородницына РАН, Таврическим национальным университетом им. В. И. Вернадского Национальной академии наук Украины и Национальной академией наук Беларуси при финансовой и организационной поддержке РФФИ и компаний Форексис и ЦСПиР.

Конференция проводится с 1989 года; начиная с 2000 года — регулярно один раз в два года, и является представительным научным форумом в области интеллектуального анализа данных, машинного обучения, распознавания образов и анализа изображений, обработки сигналов, дискретного анализа.

УДК 004.85+004.89+004.93+519.2+519.25+519.7
ББК 22.1:32.973.26-018.2

ISBN 978-5-317-03409-2

Интеллектуальный анализ k-значных данных*

Янковская А. Е., Китлер С. В.
ayyankov@gmail.com, kitsv@sibmail.com
Томск, Томский государственный архитектурно-строительный университет

Данная статья посвящена описанию интеллектуального анализа данных (ИАД) в пространстве k-значных признаков. Излагается алгоритм ИАД, заключающийся в построении диагностических тестов одновременно с выявлением закономерностей. Приводится иллюстративный пример. Алгоритм реализован в виде динамически подключаемого модуля к интеллектуальному инструментальному средству ИМСЛОГ.

Intelligent analysis of k-valued data*

Yankovskaya A. Y., Kitler S. V.
Tomsk State University of Architecture and Building, Tomsk, Russian Federation

Authors describe algorithm of intelligent data analysis in k-valued features space. This algorithm constructs diagnostic tests with simultaneous revealing of regularities. An illustrative example is given. Software realization of the algorithm is made in form of plug-in module for the intelligent instrumental software IMSLOG described.

Введение

Анализ данных при решении задач большой размерности является довольно сложным и трудоёмким процессом. В связи с этим весьма актуально применение технологий интеллектуального анализа данных (ИАД). К настоящему времени известно большое количество работ в области создания алгоритмов ИАД, к которым относятся и алгоритмы, представленные в публикациях [1–6]. В отличие от применяемой в публикациях [1–3] модели представления данных и знаний с одним механизмом классификации, нами используется матричная модель данных и знаний [4] в пространстве k-значных признаков с произвольным количеством механизмов классификации (классификационных признаков), на основе которой одновременно с выявлением различного рода закономерностей осуществляется построение диагностических тестов (ДТ) . Задача построения ДТ в данном исследовании, как и в публикациях [1, 7] сводится к нахождению безызбыточных [4, 5] (тупиковых [1]) столбцовых покрытий k-значных матриц. В связи с полиномиальной сложностью задачи нахождения столбцовых покрытий [7] весьма актуальна задача сокращения перебора.

Основная идея данной работы заключается в преобразовании исходного матричного описания данных и знаний к матричному описанию меньшей размерности, что приводит к сокращению перебора при выявлении различного рода закономерностей.

Основные понятия и определения. Закономерности в данных и знаниях

Воспользуемся определениями и понятиями, введёнными в [1, 4, 5, 7].

Работа выполнена при финансовой поддержке РФФИ, проект № 10-01-00462-а и проект № 09-01-99014-р_офи.

Данные и знания, используемые для выявления закономерностей, представляются в виде матричной модели, включающей целочисленную матрицу описаний ($\mathbf{Q}$), задающую описание объектов в пространстве характеристических признаков $z_1, \ldots, z_m$, и целочисленную матрицу различений ($\mathbf{R}$), задающую разбиение объектов на классы эквивалентности по каждому механизму классификации. Если значение характеристического признака несущественно для объекта, то данный факт отмечается прочерком («–») в соответствующем элементе матрицы $\mathbf{Q}$. Множество всех неповторяющихся строк матрицы $\mathbf{R}$ сопоставлено множеству выделенных образов, представленных одностолбцовой матрицей $\mathbf{R}'$, элементами которой являются номера образов [4, 5].

Приведём пример матриц $\mathbf{Q}$, $\mathbf{R}$ и $\mathbf{R}'$ (рис. 1).

$$\begin{array}{c} z_1\ z_2\ z_3\ z_4\ z_5\ z_6\ z_7\ z_8 \\ \mathbf{Q}=\begin{bmatrix} 3&4&8&9&5&1&3&4\\ 2&5&7&9&4&1&5&4\\ 3&5&7&8&3&0&8&4\\ 4&2&6&8&4&1&5&4\\ 3&3&8&8&4&1&5&4\\ 2&3&8&8&4&1&3&4\\ 1&2&8&8&3&-&3&4\\ 1&-&7&8&4&0&6&4\\ 2&6&7&9&5&1&3&4\\ 1&4&8&8&3&0&4&4 \end{bmatrix} \end{array} \begin{array}{c} 1\\2\\3\\4\\5\\6\\7\\8\\9\\10 \end{array} \begin{array}{c} k_1\ k_2 \\ \mathbf{R}=\begin{bmatrix} 1&1\\1&1\\1&2\\1&2\\1&1\\2&1\\1&2\\2&2\\2&1\\2&2 \end{bmatrix} \end{array} \quad \mathbf{R}'=\begin{bmatrix} 1\\1\\2\\2\\1\\3\\2\\4\\3\\4 \end{bmatrix}$$

Рис. 1. Матричное представление данных и знаний.

Данная модель позволяет представлять не только данные, но и знания экспертов, поскольку одной строкой матрицы $\mathbf{Q}$ можно задавать в интервальной форме подмножество объектов, для которых характерны одни и те же итоговые решения, задаваемые соответствующими строками матрицы $\mathbf{R}$.

Отметим, что в рассматриваемой модели недопустимо пересечение описаний объектов из разных образов.

Тестом [1] называется совокупность признаков, различающих любые пары объектов, принадлежащих разным образам (классам по каждому из механизмов классификации [5]). Тест называется минимальным, если содержит минимальное количество признаков. Тест называется безызбыточным, если содержит безызбыточное количество признаков.

Под закономерностями в знаниях будем понимать подмножества признаков с определенными легко интерпретируемыми свойствами, влияющими на различимость объектов из разных образов, устойчиво наблюдаемыми для объектов из обучающей выборки и проявляющимися на других объектах той же природы, а также весовые коэффициенты признаков, задающие различимость объектов из разных образов.

К упомянутым подмножествам будем относить константные (принимающие одно и тоже значение для всех образов), устойчивые (константные внутри образа, но не являющиеся константными), неинформативные (не различающие ни одной пары объектов), альтернативные (в смысле включения в ДТ), зависимые (в смысле включения подмножеств различимых пар объектов), несущественные (не входящие ни в один безызбыточный ДТ), обязательные (входящие во все безызбыточные ДТ) признаки, а также все минимальные и все (либо часть — при большом признаковом пространстве) безызбыточные различающие подмножества признаков, являющиеся, по сути, соответственно минимальными и безызбыточными ДТ [5].

Далее устойчивые признаки будем обозначать через r_i^j, где i — номер признака, j — номер образа.

Также к закономерностям относятся весовые коэффициенты признаков, формула для вычисления которых приведена в [5].

Безызбыточный безусловный диагностический тест (ББДТ) [1] характеризуется одновременным предъявлением всех входящих в него признаков исследуемого объекта при принятии решений.

Для поиска ББДТ применяется процедура построения матрицы импликаций [4, 5] — целочисленной матрицы $\mathbf{U}$, столбцы которой сопоставлены характеристическим признакам $z_1, z_2, \ldots, z_m$, а строки — всевозможным парам объектов v, l, соответственно из разных образов a, b (классов); $v \in \{1, 2, \ldots, \sigma(\mathbf{Q^a})\}$, $l \in \{1, 2, \ldots, \sigma(\mathbf{Q^b})\}$, где $\sigma(\mathbf{Q^a})$ ($\sigma(\mathbf{Q^b})$) — количество строк в подматрице $\mathbf{Q^a}$ ($\mathbf{Q^b}$) матрицы $\mathbf{Q}$. Строка $\mathbf{U}_i$ матрицы $\mathbf{U}$ представляет собой значение целочисленной вектор-функции различения, j-я ($j \in \{1, 2, \ldots, m\}$) компонента u_{ij} которой вычисляется по формуле:

$$u_{i,j} = |q^a_{v,j} - q^b_{l,j}|, \qquad (1)$$

где $q^a_{v,j}$ ($q^b_{l,j}$) — значение признака z_j для объекта v (l); $i = \prod\limits_{r=1}^{a} \sigma(\mathbf{Q}^r) \sum\limits_{s=r+1}^{b-1} \sigma(\mathbf{Q}^s) + \sum\limits_{r=1}^{v-1} \sigma(\mathbf{Q}^b) + vl$, а $u_{i,j}$ вычисляется для каждой пары образов.

Будем говорить, что строка $\mathbf{U}_d$ поглощает строку $\mathbf{U}_l$ ($\mathbf{U}_d \succ \mathbf{U}_l$), если

$$(\mathbf{U}_d \succ \mathbf{U}_l) \leftrightarrow \forall i \in I(u_{di} \geqslant u_{li}),$$

где $i \in \{1, \ldots, s\}$ — множество строк матрицы $\mathbf{U}$, и строки $\mathbf{U}_d$ и $\mathbf{U}_l$ несравнимы ($\mathbf{U}_d \neq \mathbf{U}_l$)

$$(\mathbf{U}_d \neq \mathbf{U}_l) \leftrightarrow (\mathbf{U}_d \not\succ \mathbf{U}_l) \;\&\; (\mathbf{U}_l \not\succ \mathbf{U}_d).$$

Безызбыточной матрицей импликаций (БМИ) назовем такую матрицу $\mathbf{U}'$, в которой отсутствуют поглощающие строки.

Отметим, что понятие БМИ для булевых [5] и k-значных признаков различаются, что приводит к необходимости модификации алгоритма выявления закономерностей, приведенного в статье [5].

Безызбыточным столбцовым покрытием (БСП) БМИ называется подмножество столбцов, покомпонентная сумма элементов которых равна столбцу, состоящему из ненулевых значений элементов и при исключении любого столбца из данного подмножества указанное свойство нарушается.

Каждое столбцовое покрытие БМИ задает полностью различающее подмножество признаков, состоящее из признаков, соответствующих столбцам, входящим в данное покрытие.

Анализ k-значных данных

Анализ k-значных данных сводится к выявлению закономерностей.

В целях сокращения перебора при выявлении закономерностей для больших размеров матриц $\mathbf{Q}$ и $\mathbf{R}'$ предлагается ряд оптимизирующих преобразований, таких как построение сокращённых целочисленных матриц описания ($\mathbf{Q}'$), различений ($\mathbf{R}''$) и по матрицам $\mathbf{Q}'$ и $\mathbf{R}''$ сокращённой БМИ ($\mathbf{U}''$).

Отметим, что в практических целях целесообразно строить БМИ, секционированную по механизмам классификации [5].

Постановка задачи. Даны матрицы $\mathbf{Q}$ и $\mathbf{R}$ в пространстве k-значных признаков.

Необходимо построить матрицу $\mathbf{U}''$, найти все её безызбыточные столбцовые покрытия и соответствующие множества k-значных ББДТ одновременно с выявлением других закономерностей.

По исходным матрицам $\mathbf{Q}$ и $\mathbf{R}'$ построим матрицу $\mathbf{U}'$ одновременно с вычислением весовых коэффициентов признаков [5].

Построим матрицы $\mathbf{Q}'$ и $\mathbf{R}''$ по матрицам $\mathbf{Q}$ и $\mathbf{R}'$. Каждый образ представим двумя строками матрицы $\mathbf{Q}'$. Каждую чётную (нечётную) строку $\mathbf{Q}'_i$ ($\mathbf{Q}'_f$) матрицы $\mathbf{Q}'$ представим целочисленным вектором ($i = 2c$, $f = 2c-1$, $c \in \{1, \ldots, v\}$, v — количество образов), j-я ($j \in \{1, \ldots, m\}$) компонента которого соответствует максимальному (минимальному) значению j-го характеристического признака внутри каждого образа, и вычисляется по формуле:

$$q'_{i,j} = \max_{l} q^{l}{}_{j}(c), \tag{2}$$

$$q'_{f,j} = \min_{l} q^{l}{}_{j}(c), \tag{3}$$

где $q^{l}{}_{j}(c)$ — значение признака z_j для объекта $l \in \{1, \ldots, r_c\}$ из образа c, а r_c — количество объектов, принадлежащих образу c.

При этом сохраняется взаимнооднозначное соответствие между строками матриц $\mathbf{Q}'$ и $\mathbf{R}''$ такое же, как между строками матриц $\mathbf{Q}$ и $\mathbf{R}'$.

Как указывалось ранее, в рамках рассматриваемой матричной модели не допускается пересечение описания образов, поэтому если при построении матрицы $\mathbf{Q}'$ строки из разных образов совпали, то для обеспечения непересечения образов предлагается использовать правило 1.

Правило 1: Если строка в матрице $\mathbf{Q}'$ из образа c совпадает со строкой из образа e ($c \neq e$), то заменяются все строки матриц $\mathbf{Q}'$ и $\mathbf{R}''$ из образов c, e строками матриц $\mathbf{Q}$ и $\mathbf{R}'$ из образов c, e соответственно.

Отметим, что процесс применения правила 1 является итеративным и повторяется до тех пор, пока в матрице $\mathbf{Q}'$ не останется совпадающих строк из разных образов. При этом правило целесообразно применять только в случаях, когда количество совпадающих строк не превышает заданного числа g', зависимого от размерности и самих знаний, представленных матрицами $\mathbf{Q}$ и $\mathbf{R}$. В противном случае мы не достигнем сокращения перебора.

Обозначим скорректированные матрицы через $\mathbf{Q}''$ и $\mathbf{R}'''$ соответственно и по ним построим матрицу $\mathbf{U}''$, строка $\mathbf{U}''_i$ которой представляет собой значение целочисленной вектор-функции различения, j-ю ($j \in \{1, \ldots, m\}$) компоненту $u''_{i,j}$ которой будем вычислять аналогично формуле 1.

По построенной матрице $\mathbf{U}''$ выявим закономерности с применением модифицированного алгоритма, основанного на алгоритмах [4, 6]. Модификация заключается в запоминании и удалении столбцов матрицы $\mathbf{U}''$, сопоставленных альтернативным (за исключением одного) и зависимым признакам при построении столбцовых покрытий матрицы $\mathbf{U}''$. В целях сокращения перебора при поиске БСП матрицы $\mathbf{U}''$, запоминаются и удаляются столбцы, сопоставленные признакам в разной степени различающие одни и те же объекты, за исключением одного, а также пара столбцов, покомпонентная сумма элементов которых не равна 0 при наличии столбца в матрице $\mathbf{U}''$, соответствующие элементы которого не равны 0, при равенстве 0 остальных элементов. Нахождение БСП матрицы $\mathbf{U}''$ будем продолжать до достижения количества найденых БСП величины g'', определяемой экспериментально.

При количестве найденных БСП, меньшем, чем g'', поглощаемые столбцы в матрице $\mathbf{U}''$, сопоставленные зависимым признакам и удалённые из матрицы $\mathbf{U}''$ при нахождении БСП, используются для поиска дополнительных БСП в целях принятия решения с заданной точностью.

Поскольку процедура оптимизации не является эквивалентным преобразованием, то возможно появление БСП матрицы $\mathbf{U}''$, не являющихся БСП матрицы $\mathbf{U}'$. В связи с этим необходимо проверить являются ли построенные БСП матрицы $\mathbf{U}''$ БСП для матрицы $\mathbf{U}'$.

Сформулируем правило 2 для нахождения и исключения тех БСП матрицы $\mathbf{U}''$, которые не являются БСП матрицы $\mathbf{U}'$.

Правило 2. Если покомпонентная сумма элементов столбцов матрицы $\mathbf{U}'$ j-го БСП, построенного по матрице $\mathbf{U}''$, будет содержать хотя бы один нулевой элемент, то j-й БСП для матрицы $\mathbf{U}''$ не является БСП для матрицы $\mathbf{U}'$.

По оставшимся после применения правила 2 БСП, построим ББДТ.

Далее приведём пример, иллюстрирующий поиск вышеупомянутых закономерностей.

Иллюстративный пример

Пусть заданы матрицы $\mathbf{Q}$ и $\mathbf{R}$ (см. рис. 1). Упорядочим строки матриц $\mathbf{Q}$ и $\mathbf{R}$ по принадлежности к образам (рис. 2).

$$\boldsymbol{Q}= \begin{array}{cc} \begin{array}{cccccccc} z_1 & z_2 & z_3 & z_4 & z_5 & z_6 & z_7 & z_8 \end{array} & \\ \left[\begin{array}{cccccccc} 3 & 4 & 8 & 9 & 5 & 1 & 3 & 4 \\ 2 & 5 & 7 & 9 & 4 & 1 & 5 & 4 \\ 3 & 3 & 8 & 8 & 4 & 1 & 5 & 4 \\ \hline 3 & 5 & 7 & 8 & 3 & 0 & 8 & 4 \\ 4 & 2 & 6 & 8 & 4 & 1 & 5 & 4 \\ 1 & 2 & 8 & 8 & 3 & - & 3 & 4 \\ \hline 2 & 3 & 8 & 8 & 4 & 1 & 3 & 4 \\ 2 & 6 & 7 & 9 & 5 & 1 & 3 & 4 \\ \hline 1 & - & 7 & 8 & 4 & 0 & 6 & 4 \\ 1 & 4 & 8 & 8 & 3 & 0 & 4 & 4 \end{array}\right] & \begin{array}{c} 1 \\ 2 \\ 5 \\ 3 \\ 4 \\ 7 \\ 6 \\ 9 \\ 8 \\ 10 \end{array} \end{array} \quad \boldsymbol{R}= \begin{array}{c} \begin{array}{cc} k_1 & k_2 \end{array} \\ \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ \hline 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ \hline 2 & 1 \\ 2 & 1 \\ \hline 2 & 2 \\ 2 & 2 \end{array}\right] \end{array} \quad \boldsymbol{R'}= \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ \hline 2 \\ 2 \\ 2 \\ \hline 3 \\ 3 \\ \hline 4 \\ 4 \end{array}\right]$$

Рис. 2. Матрицы описания и различений.

По матрицам $\mathbf{Q}$ и $\mathbf{R}'$ построим матрицу $\mathbf{U}'$, пример которой не приводится, поскольку рамки статьи ограничены.

Определим константные и устойчивые признаки; признак z_8, равный 4-м, является константным; признаки $z_1^3 = 2$, $z_1^4 = 1$, $z_4^2 = 8$, $z_4^4 = 8$, $z_6^1 = 1$, $z_6^3 = 1$ и $z_6^4 = 0$ являются устойчивыми.

Удалив константный признак z_8 построим сокращенные матрицы $\mathbf{Q}'$ и $\mathbf{R}''$ (рис. 3). Заметим, что в матрице $\mathbf{Q}'$ совпали две строки: 1-я и 5-я, выделенные на рис. 3 полужирным шрифтом.

$$
\mathbf{Q'} = \begin{array}{c}
\begin{array}{ccccccc} z_1 & z_2 & z_3 & z_4 & z_5 & z_6 & z_7 \end{array} \\
\left[\begin{array}{ccccccc}
\mathbf{2} & \mathbf{3} & \mathbf{7} & \mathbf{8} & \mathbf{4} & \mathbf{1} & \mathbf{3} \\
3 & 5 & 8 & 9 & 5 & 1 & 5 \\ \hdashline
1 & 2 & 6 & 8 & 3 & 0 & 3 \\
4 & 5 & 8 & 8 & 4 & 1 & 8 \\ \hdashline
\mathbf{2} & \mathbf{3} & \mathbf{7} & \mathbf{8} & \mathbf{4} & \mathbf{1} & \mathbf{3} \\
2 & 6 & 8 & 9 & 5 & 1 & 3 \\ \hdashline
1 & 2 & 7 & 8 & 3 & 0 & 4 \\
1 & 6 & 8 & 8 & 4 & 0 & 6
\end{array}\right]
\end{array}
\begin{array}{ll}
1 & \mathrm{min} \\ 2 & \mathrm{max} \\ 3 & \mathrm{min} \\ 4 & \mathrm{max} \\ 5 & \mathrm{min} \\ 6 & \mathrm{max} \\ 7 & \mathrm{min} \\ 8 & \mathrm{max}
\end{array}
\qquad
\mathbf{R''} = \left[\begin{array}{c} 1 \\ 1 \\ \hdashline 2 \\ 2 \\ \hdashline 3 \\ 3 \\ \hdashline 4 \\ 4 \end{array}\right]
$$

Рис. 3. Матрицы $\mathbf{Q}'$ и $\mathbf{R}''$.

Скорректируем матрицы $\mathbf{Q}'$ и $\mathbf{R}''$, применив правило 1 и обозначим их через $\mathbf{Q}''$ и $\mathbf{R}'''$ соответственно (рис. 4). Заменённые строки выделены полужирным шрифтом на рис. 4.

$$
\mathbf{Q''} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} z_1 & z_2 & z_3 & z_4 & z_5 & z_6 & z_7 \end{array} \\
\left[\begin{array}{ccccccc}
\mathbf{3} & \mathbf{4} & \mathbf{8} & \mathbf{9} & \mathbf{5} & \mathbf{1} & \mathbf{3} \\
\mathbf{2} & \mathbf{5} & \mathbf{7} & \mathbf{9} & \mathbf{4} & \mathbf{1} & \mathbf{5} \\
\mathbf{3} & \mathbf{3} & \mathbf{8} & \mathbf{8} & \mathbf{4} & \mathbf{1} & \mathbf{5} \\ \hdashline
1 & 2 & 6 & 8 & 3 & 0 & 3 \\
4 & 5 & 8 & 8 & 4 & 1 & 8 \\ \hdashline
\mathbf{2} & \mathbf{3} & \mathbf{8} & \mathbf{8} & \mathbf{4} & \mathbf{1} & \mathbf{3} \\
\mathbf{2} & \mathbf{6} & \mathbf{7} & \mathbf{9} & \mathbf{5} & \mathbf{1} & \mathbf{3} \\ \hdashline
1 & 2 & 7 & 8 & 3 & 0 & 4 \\
1 & 6 & 8 & 8 & 4 & 0 & 6
\end{array}\right]
\end{array}
\begin{array}{l} \\ \\ \\ \\ \mathrm{min} \\ \mathrm{max} \\ \\ \\ \mathrm{min} \\ \mathrm{max} \end{array}
\qquad
\mathbf{R'''} = \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ \hdashline 2 \\ 2 \\ \hdashline 3 \\ 3 \\ \hdashline 4 \\ 4 \end{array}\right]
$$

Рис. 4. Скорректированные матрицы $\mathbf{Q}''$ и $\mathbf{R}'''$.

Далее построим матрицу $\mathbf{U}''$ (рис. 5).

$$
\mathbf{U''} = \begin{array}{c}
\begin{array}{ccccccc} z_1 & z_2 & z_3 & z_4 & z_5 & z_6 & z_7 \end{array} \\
\left[\begin{array}{ccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 2 \\
2 & 3 & 0 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 2 \\
1 & 2 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 2 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1
\end{array}\right]
\end{array}
\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
$$

Рис. 5. Матрица $\mathbf{U}''$.

В результате выявления закономерностей получим: 1) признак z_4 — зависит от признаков z_1, z_2, z_5; 2) признак z_6 — зависит от признаков z_1, z_2; 3) признак z_6 — несущественный. Альтернативные и обязательные признаки отсутствуют.

Построим БСП матрицы $\mathbf{U}''$: $\{1;7\}$, $\{2;7\}$, $\{1;2;3\}$, $\{1;3;5\}$, $\{3;4;7\}$, $\{3;5;7\}$.

Проверим являются ли они БСП матрицы $\mathbf{U}'$ (правило 2). БСП матрицы $\mathbf{U}''$ $\{1;\,3;\,5\}$, $\{3;\,4;\,7\}$ не являются БСП матрицы $\mathbf{U}'$. Следовательно, при построении ББДТ они исключаются. По оставшимся БСП матрицы $\mathbf{U}''$ построим ББДТ. В результате получились следующие ББДТ: $T_1 = \{z_1, z_7\}$, $T_2 = \{z_2, z_7\}$, $T_3 = \{z_1, z_2, z_3\}$, $T_4 = \{z_3, z_5, z_7\}$.

Программная реализация

Алгоритм интеллектуального анализа k-значных данных реализован в виде динамически подключаемого модуля к интеллектуальному инструментальному средству ИМСЛОГ [6], на базе которого строятся прикладные интеллектуальные системы.

Модуль разработан на основе программной среды Borland C++ Builder. Входными данными модуля служат матрицы $\mathbf{Q}$ и $\mathbf{R}$. На выходе получаем выявленные закономерности, включая все ББДТ.

Библиотека классов модуля содержит математические объекты (k-значные векторы и матрицы), используемые в алгоритме, которые динамически создаются, обрабатываются и удаляются.

Заключение

Предложен оригинальный алгоритм интеллектуального анализа k-значных данных. Алгоритм позволяет производить построение ББДТ одновременно с выявлением других закономерностей, тем самым уменьшая временные и стоимостные затраты на их построение. Приведён иллюстративный пример, наглядно показывающий процесс выявления закономерностей и описана программная реализация предложенного алгоритма.

Исследование алгоритма на ряде псевдослучайных и реальных данных и знаний позволит выявить область его целесообразного использования.

Литература

[1] *Журавлёв Ю. И.* Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики: Вып. 33. Москва: Наука— 1978. — С. 5–68.

[2] *Загоруйко Н. Г.* Прикладные методы анализа данных и знаний— Н-ск: ИМ СО РАН, 1999. — 270 с.

[3] *Закревский А. Д.* Логика распознавания. Изд. 2-е, доп.— Москва: Едиториал УРСС, 2003. — 144 с.

[4] *Янковская А. Е.* Построение k-значных диагностических тестов в интеллектуальной системе с матричным представлением знаний // Сб. научн. труд. VI Национальной конф. по искусственному интеллекту с межд. уч. Т. I, Пущино, 1998. — С. 264–269.

[5] *Янковская А. Е.* Логические тесты и средства когнитивной графики в интеллектуальной системе // Новые информационные технологии в исследовании дискретных структур. Доклады 3-ей Всерос. конф. с междунар. участием, Томск: СО РАН, 2000. — С. 163–168.

[6] *Янковская А. Е., Гедике А. И.* Интеллектуальный анализ информации на базе инструментального средства ИСМЛОГ // Новости искусственного интеллекта. —2005.— № 1. — С. 36–47.

[7] *Дюкова Е. В.* О числе тупиковых покрытий целочисленной матрицы // Журнал вычислительной математики и математической физики. Том 45. – 2005.— № 5. — С. 935–940.