

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin
(editors)

Natural and Artificial Intelligence

ITHEA

SOFIA

2010

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin (ed.)

Natural and Artificial Intelligence

ITHEA®

Sofia, Bulgaria, 2010

ISBN 978-954-16-0043-9

First edition

Recommended for publication by The Scientific Council of the Institute of Information Theories and Applications FOI ITHEA

This book is engraved in prof. Zinovy Lvovich Rabinovich memory. He was a great Ukrainian scientist, co-founder of ITHEA International Scientific Society (ITHEA ISS). To do homage to the remarkable world-known scientific leader and teacher this book is published in Russian language and is concerned to some of the main areas of interest of Prof. Rabinovich.

The book is opened by the last paper of Prof. Rabinovich specially written for ITHEA ISS. Further the book maintains articles on actual problems of natural and artificial intelligence, information interaction and corresponded intelligent technologies, expert systems, robotics, classification, business intelligence; etc. In more details, the papers are concerned in: conceptual problems of the natural and artificial intelligent systems: structures and functions of the human memory, ontological models of knowledge representation, knowledge extraction from the natural language texts; network technologies; evolution and perspectives of development of the mechatronics and robotics; visual communication by gestures and movements, psychology of vision and information technologies of computer vision, image processing; object classification using qualitative characteristics; methods for comparing of alternatives and their ranging in the procedures of expert knowledge processing; ecology of programming – a new trend in the software engineering; decision support systems for economics and banking; systems for automated support of disaster risk management; and etc.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

General Sponsor: Consortium FOI Bulgaria (www.foibg.com).

Printed in Bulgaria

Copyright © 2010 All rights reserved

© 2010 ITHEA® – Publisher; Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org; e-mail: info@foibg.com

© 2010 Krassimir Markov, Vitalii Velychko, Oleksy Voloshin – Editors

© 2010 Ina Markova – Technical editor

© 2010 For all authors in the book.

© ITHEA is a registered trade mark of FOI-COMMERCE Co.

ISBN 978-954-16-0043-9

C/o Jusautor, Sofia, 2010

ПРИМЕНЕНИЕ МЕТОДА «МАСКА» ДЛЯ ГРУППОВОЙ ЭКСПЕРТНОЙ КЛАССИФИКАЦИИ НАУЧНЫХ ПРОЕКТОВ ПО МНОГИМ КРИТЕРИЯМ

Петровский Алексей Борисович, Ройзензон Григорий Владимирович,
Тихонов Игорь Петрович, Балышев Александр Владимирович

Аннотация: В работе излагается новый метод «МАСКА» групповой классификации многопризнаковых объектов, описываемых качественными признаками. Метод основан на теории мультимножеств. Разработаны алгоритмы для построения согласованных правил отнесения объектов к классам решений и выделения противоречиво классифицированных объектов. Рассмотрен пример применения метода для конкурсного отбора научных проектов в Российском фонде фундаментальных исследований.

Ключевые слова: групповой вербальный анализ решений, многокритериальная порядковая классификация, агрегирование признаков, интегральные показатели оценки, результативность научных проектов

ACM Classification Keywords: A.0 General Literature - Conference proceedings

Введение

Многокритериальная классификация рассматриваемых объектов по их свойствам, выраженных признаками (атрибутами), в индивидуальном принятии решений основывается на субъективных предпочтениях единственного лица, принимающего решение (ЛПР). При коллективном принятии решений необходимо учитывать одновременно различные интересы многих участников, разнообразие и несовпадение их целей и способов выражения их предпочтений. Примером такой задачи служит конкурсный отбор научных проектов, оцененных несколькими экспертами по многим качественным критериям, в котором необходимо разделить проекты на группы, в той или иной степени отвечающие целям конкурса и им не соответствующие.

Известно достаточно много подходов к индивидуальной классификации объектов, характеризующихся многими количественными и/или качественными признаками [Ларичев, 2006], [Doumpos, Zorounidis, 2002], [Greco et al., 2002], [Köksalan et al., 2003], [Roy, Bouyssou, 1993]. Методы групповой многокритериальной классификации в настоящее время менее развиты [Петровский, 2003], [Petrovsky, 2001]. Главные трудности при коллективном принятии решений связаны с необходимостью учета большого числа вербальных и количественных данных и обработкой этих данных, не прибегая к дополнительным преобразованиям типа усреднения, смешивания, взвешивания, которые могут привести к необоснованным и необратимым искажениям исходной информации. Воспользовавшись аппаратом теории мультимножеств для представления многопризнаковых объектов, можно преодолеть указанные выше трудности.

В работе рассматривается модификация метода МАСКА (МногоАспектная Согласованная Классификация Альтернатив) для групповой классификации многопризнаковых объектов, которые характеризуются многими разнородными (количественными и качественными) признаками и могут существовать в нескольких экземплярах с отличающимися, в частности, противоречивыми значениями признаков [Петровский, 2003], [Petrovsky, 2001].

Метод базируется на теории метрических пространств мультимножеств и позволяет строить согласованное групповое правило для классификации объектов, которое аппроксимирует

индивидуальные правила экспертной сортировки объектов. Этот метод допускает использование различных, в том числе и противоречивых, данных для описания объектов и позволяет находить несогласованные индивидуальные правила классификации.

Предложенный подход к построению согласованного решающего правила для классификации многопризнаковых объектов был апробирован на базе данных конкурсов целевых фундаментальных исследований, выполняемым по грантам Российского фонда фундаментальных исследований (РФФИ) в интересах Федеральных агентств и ведомств России [Петровский, Тихонов, 2009]. Для различных научных направлений были построены групповые решающие правила классификации, основанные на многокритериальных оценках проектов многими экспертами, которые позволили выделить наиболее важные для отбора проектов критерии и выявить расхождения в индивидуальных правилах сортировки проектов, применявшихся экспертами.

Конкурсный отбор научных проектов

По установленному в РФФИ порядку экспертиза научных проектов, в том числе целевых фундаментальных исследований, производится при конкурсном отборе заявок. Оценку заявок выполняют эксперты, работающие в различных научно-исследовательских институтах, высших учебных заведениях, научно-производственных организациях. Каждый проект независимо оценивается несколькими экспертами, как правило, тремя без согласования их мнений. Для оценки содержания заявки используются специальные анкеты, которые содержат критерии, имеющие словесные шкалы оценок с развернутыми формулировками градаций качества. По каждому критерию эксперт выбирает только одну из оценок (такая экспертиза называется закрытой). Кроме того, эксперт дает свое заключение о целесообразности поддержки проекта. На основе заключений экспертов Экспертный совет РФФИ принимает решение о принятии проекта и объемах его финансирования.

Экспертная оценка содержания целевых фундаментальных исследований проводится по 12 качественным критериям, представленным в экспертной анкете и объединенным в группы: Научная характеристика проекта; Оценка возможностей практической реализации проекта.

В группу «Научная характеристика проекта» входят 9 критериев:

Q_{11} . Уровень фундаментальности проекта; Q_{12} . Ориентированность результата; Q_{13} . Цели исследования; Q_{14} . Методы достижения цели проекта; Q_{15} . Характер исследований; Q_{16} . Научная значимость проекта; Q_{17} . Степень новизны предлагаемых решений; Q_{18} . Потенциал исполнителей; Q_{19} . Техническая оснащенность.

Группа «Оценка возможностей практической реализации проекта» состоит из 2 критериев:

Q_{21} . Завершающая стадия фундаментальных исследований, предлагаемых в проекте;
 Q_{22} . Масштабы применимости результатов заявленных исследований.

Каждый критерий имеет порядковую или номинальную шкалу оценок с развернутыми словесными формулировками градаций качества. Например, шкала критерия Q_{17} . «Степень новизны предлагаемых решений» выглядит следующим образом:

q_{17}^1 - решения сформулированы впервые и существенно превосходят уровень имеющихся решений;
 q_{17}^2 - решение находится на уровне имеющихся решений;
 q_{17}^3 - решение уступает имеющимся решениям.

По каждому проекту каждый эксперт давал также свое заключение о поддержке проекта:

r_1 - безусловно поддержать (оценка «5»),

r_2 - целесообразно поддержать (оценка «4»),

r_3 - поддержка возможна (оценка «3»),

r_4 - не поддерживать (оценка «2»).

Совокупность экспертных заключений можно считать еще одним критерием R , характеризующим проекты. Таким образом, каждый проект представляет собой объект, описываемый многими качественными (вербальными) признаками.

Принятие того или иного заключения из множества R может рассматриваться как решение задачи классификации рассматриваемых объектов по их свойствам, выраженных признаками (атрибутами) Q_{11} – Q_{22} , которая состоит в разбиении множества объектов на несколько заранее заданных упорядоченных групп (классов решений) D_1 – D_4 . Рекомендации экспертов являются индивидуальными правилами предварительной классификации (сортировки) рассматриваемых проектов, которые могут совпадать или различаться.

Для отбора проектов нужно построить агрегированное групповое правило, которое обобщает индивидуальные решения всех экспертов и позволяет отнести проект в один из заданных классов. Это решающее правило должно формулироваться на естественном языке и учитывать экспертные оценки заявок по многим критериям. Построение такого решающего правила в при наличии многих несоответствующих индивидуальных заключений экспертов относится к числу сложных задач многокритериального принятия решений. Вместе с тем наличие такого решающего правила позволяет формировать разные стратегии конкурсного отбора проектов, проводить анализ качества работы экспертов, выявлять различные точки зрения.

Для построения решающих правил конкурсного отбора целевых фундаментальных исследований использована оригинальная методология группового вербального анализа решений, которая успешно применялась при решении различных практических задач [Ларичев, 2006], [Петровский, 2003], [Петровский, 2009]. В вербальном анализе решений рассматриваемые варианты и классы решений описываются с помощью качественных критериев, имеющих словесные формулировки градаций на шкалах оценок. Обычно вводится небольшое число таких градаций (3-5), чтобы обеспечить ясную различимость соседних оценок. Именно такой способ формализации экспертных оценок проектов принят в РФФИ.

В вербальном анализе решений численные коэффициенты важности критериев и ценности вариантов не рассчитываются и не применяются, а качественные оценки признаков не преобразуются в какие-либо числовые показатели. Даже при небольшом числе градаций оценок на шкалах критериев оказывается возможным описать достаточно сложные свойства объектов. Тем самым, используя только качественные измерения, на множестве кортежей многокритериальных оценок можно задать отношения превосходства и эквивалентности вариантов решения, с помощью которых осуществляется их порядковая классификация, частичное упорядочение или выделение лучшего варианта.

Вербальный анализ решений отличается активным участием ЛПР в анализе и решении стоящей проблемы, позволяя разносторонне и достаточно подробно выражать предпочтения ЛПР, уточнять и корректировать их в ходе решения задачи, генерировать и обосновывать новые варианты решения. Методы группового вербального анализа решений дают возможность при решении задач коллективного выбора одновременно учитывать различные интересы многих участников, разнообразие и несоответствие их целей и способов выражения их предпочтений. В целом вербальные методы более «прозрачны», мало чувствительны к ошибкам измерения и менее трудоемки для человека.

Представление многопризнаковых объектов

Многопризнаковые объекты $A_i, i=1, \dots, n$ обычно принято представлять как векторы или кортежи $q_i=(q_{i1}^{e_1}, \dots, q_{im}^{e_m})$ в пространстве $Q=Q_1 \times \dots \times Q_m$, где $Q_s=\{q_s^{e_s}\}$ – непрерывная или дискретная шкала s -го признака, $e_s=1, \dots, h_s, s=1, \dots, m$. Ситуация существенно усложняется, если одному и тому же объекту A_i может соответствовать не один, а несколько m -мерных векторов. В таких случаях объект A_i представляется в пространстве Q группой, состоящей из k векторов $\{q_i^{(1)}, \dots, q_i^{(k)}\}$ вида $q_i^{(j)}=(q_{i1}^{e_1(j)}, \dots, q_{im}^{e_m(j)})$, $j=1, \dots, k$, которая должна рассматриваться как единое целое. Измеренные разными способами значения параметров или индивидуальные оценки экспертов, могут быть похожими, различающимися и даже противоречивыми, что может приводить к несравнимости m -мерных векторов $q_i^{(j)}=(q_{i1}^{e_1(j)}, \dots, q_{im}^{e_m(j)})$, характеризующих один и тот же объект A_i . Совокупность таких многомерных объектов может иметь в пространстве Q сложную структуру, достаточно трудную для анализа.

Эти трудности можно преодолеть, если воспользоваться иным способом представления многопризнаковых объектов, основанным на формализме мультимножеств, который позволяет одновременно учесть все комбинации значений количественных и качественных признаков, а также различное число значений каждого из этих признаков. Введем вместо прямого произведения m шкал признаков $Q=Q_1 \times \dots \times Q_m$ множество $G=Q_1 \cup \dots \cup Q_m$ – обобщенную шкалу признаков, состоящую из m групп признаков, и представим объект в таком символическом виде:

$$A_i = \{k_{A_i}(q_1^{e_1}) \circ q_1^{e_1}, \dots, k_{A_i}(q_1^{h_1}) \circ q_1^{h_1}, \dots, k_{A_i}(q_m^1) \circ q_m^1, \dots, k_{A_i}(q_m^{h_m}) \circ q_m^{h_m}\}, \quad (1)$$

где число $k_{A_i}(q_s^{e_s})$ указывает, сколько раз признак $q_s^{e_s} \in Q_s$ встречается в описании объекта A_i , а знак \circ обозначает кратность признака $q_s^{e_s}$. Множество G характеризует свойства совокупности объектов $A=\{A_1, \dots, A_n\}$. Такая запись объекта A_i представляет его как мультимножество или множество с повторяющимися элементами [Петровский, 2003], [Petrovsky, 2001].

Определяются следующие операции над мультимножествами: объединение $A \cup B$, пересечение $A \cap B$, сложение $A + B$, вычитание $A - B$, симметрическая разность $A \Delta B$, дополнение $\bar{A} = Z - A$, умножение на целое число $c \cdot A, c > 0$, арифметическое умножение $A \cdot B$, арифметическая n -ая степень A^n , прямое произведение $A \times B$, прямая n -ая степень $(\times A)^n, Z$ – максимальное мультимножество с $k_Z(x) = \max_{A \in A} k_A(x)$.

Новые типы операций над мультимножествами открывают новые возможности для агрегирования многопризнаковых объектов. Например, группа X_a объектов может быть получена как сумма $X_a = \sum_i A_i$, объединение $X_a = \cup_i A_i$ или пересечение $X_a = \cap_i A_i$ мультимножеств A_i , описывающих объекты A_i , либо как линейная комбинация различных мультимножеств вида $X_a = \sum_i c_i \cdot A_i, X_a = \cup_i c_i \cdot A_i$ или $X_a = \cap_i c_i \cdot A_i$. Заметим кстати, что и само мультимножество A_i вида (2) формально можно представить как сумму $A_i = \sum \{q_i^{(j)}\}$ k различных множеств (m -мерных векторов или кортежей) $q_i^{(j)}$, характеризующих один и тот же объект A_i .

На семействе мультимножеств $A=\{A_1, \dots, A_n\}$ можно ввести новые классы метрических пространств мультимножеств (A, d) , которые задаются следующими видами метрик:

$$d_{1p}(A, B) = \left(\sum_{x_i \in G} w_i |k_A(x_i) - k_B(x_i)| \right)^{1/p}, \quad p > 0;$$

$$d_{2p}(A, B) = \left(\sum_{x_i \in G} w'_i |k_A(x_i) - k_B(x_i)| \right)^{1/p}, \quad w'_i = w_i / \sum_{j=1}^h w_j k_Z(x_j);$$

$$d_{3p}(\mathbf{A}, \mathbf{B}) = \left(\frac{\sum_{x_i \in G} w_i |k_A(x_i) - k_B(x_i)|}{\sum_{x_i \in G} w_i \max[k_A(x_i), k_B(x_i)]} \right)^{1/p}$$

Основная метрика $d_{1p}(\mathbf{A}, \mathbf{B})$ является метрикой типа Хемминга при $p=1$, используемой во многих приложениях. Полностью усредненная метрика $d_{2p}(\mathbf{A}, \mathbf{B})$ характеризует различие между двумя мультимножествами A и B , отнесенное к расстоянию, максимально возможному в исходном пространстве. Локально усредненная метрика $d_{3p}(\mathbf{A}, \mathbf{B})$ задает различие, отнесенное к максимально возможной «общей части» $A \cup B$ только этих двух мультимножеств в исходном пространстве.

Метод согласованной групповой классификации объектов

Процедура классификации объектов в рамках формальной логики может быть описана как совокупность решающих правил, которые представляются выражениями вида:

$$\text{ЕСЛИ } \langle \text{условия} \rangle, \text{ ТО } \langle \text{решение} \rangle \quad (2)$$

При прямой классификации терм $\langle \text{условия} \rangle$ включает названия объектов или перечень значений признаков, описывающих объекты класса. При непрямой классификации один или несколько термов $\langle \text{условия} \rangle$ конструируются как отношения между различными признаками и/или их значениями. Терм $\langle \text{решение} \rangle$ в обоих случаях означает, что объект принадлежит к определенному классу. Заметим, что подобным образом формируются и базы знаний экспертных систем продукционного типа.

Рассмотрим задачу групповой классификации объектов в следующей постановке. Задано множество объектов (вариантов, альтернатив) A_1, \dots, A_n , которые описываются m дискретными признаками Q_1, \dots, Q_m , имеющими количественные и/или качественные шкалы оценок. Каждая группа признаков $Q_s = \{q_s^1, \dots, q_s^{h_s}\}$, $e_s = 1, \dots, h_s$, $s = 1, \dots, m$ отражает содержательное качество объектов. Например, $q_s^{e_s}$ может быть значением показателя, определяющего какое-либо свойство объекта, или оценкой объекта по критерию, и тому подобное. Объекты A_1, \dots, A_n предварительно рассортированы по нескольким классам D_1, \dots, D_f путем прямой классификации. Принадлежность объекта A_i , $i = 1, \dots, n$ к некоторому классу D_t , $t = 1, \dots, f$ выражается атрибутом R , которое можно считать еще одним качественным признаком со шкалой значений $R = \{r_{it}\}$. Совокупность признаков $\{q_i^{e_1}, \dots, q_m^{e_m}; r_{it}\}$ представляет собой индивидуальное правило классификации вида (2), относящее объект A_i к классу D_t .

Любой объект A_i существует в k экземплярах, которые отличаются наборами признаков, его характеризующих. Например, когда объект независимо оценивается и сортируется по классам k экспертами или его параметры измеряются k различными способами. Тем самым имеется k различных и, возможно, противоречивых индивидуальных правил классификации объектов. Требуется построить одно или несколько агрегированных групповых правил вида (2), которые наилучшим (в некотором смысле) образом аппроксимируют совокупность всех индивидуальных правил классификации объектов и относят объекты в заданные классы с допустимой точностью.

Для решения задачи групповой классификации многопризнаковых объектов использован метод МАСКА (МногоАспектная Согласованная Классификация Альтернатив), основанный на теории мультимножеств [Петровский, 2003], [Petrovsky, 2001]. Метод предназначен для построения согласованных групповых правил классификации реально имеющих объектов, оцененных несколькими экспертами по многим критериям, которые агрегируют индивидуальные правила сортировки объектов, и выявления противоречивых индивидуальных правил классификации.

Сопоставим каждому многопризнаковому объекту мультимножество

$$\mathbf{A}_i = \{k_{A_i}(q_1^1) \circ q_1^1, \dots, k_{A_i}(q_1^{h_1}) \circ q_1^{h_1}, \dots, k_{A_i}(q_m^1) \circ q_m^1, \dots, k_{A_i}(q_m^{h_m}) \circ q_m^{h_m}, k_{A_i}(r_1) \circ r_1, \dots, k_{A_i}(r_f) \circ r_f\} \quad (3)$$

над доменом $G = Q_1 \cup \dots \cup Q_m \cup R$, где $k_{A_i}(q_s^{es})$ и $k_{A_i}(r_t)$ равны числу экспертов, давших объекту A_i оценку q_s^{es} и рекомендацию r_t . Принадлежность объекта A_i к некоторому классу D_t определяется, например, правилом большинства голосов, в соответствии с которым объект A_i считается принадлежащим к классу D_t , если $k_{A_i}(r_t) > k_{A_i}(r_p)$ для всех $p \neq t$. Представление каждого объекта как мультимножества (3) есть, по сути, групповое правило классификации вида (2), объединяющее k индивидуальных правил сортировки. Нужно найти одно или несколько агрегированных правил классификации всей совокупности объектов вида (2), которые наилучшим образом были бы согласованы с групповыми правилами классификации.

Для простоты положим, что результатом классификации должно быть разложение совокупности объектов только на два класса: D_a («хорошие») и D_b («плохие»). При необходимости рассортировать объекты на большее число классов, можно сначала разбить совокупность объектов на две группы, затем одну из них или обе группы – на подгруппы, и так далее. Сопоставим каждому классу решений D_a и D_b мультимножества X_a и X_b , которые представляют собой соответственно совокупности хороших и плохих объектов, сформированных путем сложения соответствующих им мультимножеств

$$\begin{aligned} \mathbf{X}_t = \sum_{i \in I_t} \mathbf{A}_i &= \{k_{X_t}(q_1^1) \circ q_1^1, \dots, k_{X_t}(q_1^{h_1}) \circ q_1^{h_1}, \dots, k_{X_t}(q_m^1) \circ q_m^1, \dots, k_{X_t}(q_m^{h_m}) \circ q_m^{h_m}, \\ k_{X_t}(r_a) \circ r_a, k_{X_t}(r_b) \circ r_b\}, \quad k_{X_t}(q_s^{es}) &= \sum_{i \in I_t} k_{A_i}(q_s^{es}), \quad k_{X_t}(r_t) = \sum_{i \in I_t} k_{A_i}(r_t), \quad t = a, b, \end{aligned} \quad (4)$$

где подмножества индексов $I_a \cup I_b = \{1, \dots, n\}$, $I_a \cap I_b = \emptyset$. При суммировании мультимножеств учитываются значения всех признаков, характеризующих все объекты, входящие в группу.

Процедура нахождения агрегированного правила, аппроксимирующего большое число противоречивых правил сортировки многопризнаковых объектов заключается в следующем. Образует новые мультимножества

$$\begin{aligned} \mathbf{R}_a &= \{k_{R_a}(X_a) \circ X_a, k_{R_a}(X_b) \circ X_b\}, \quad \mathbf{R}_b = \{k_{R_b}(X_a) \circ X_a, k_{R_b}(X_b) \circ X_b\}, \\ \mathbf{Q}_{as} &= \sum_{j \in J_{as}} \mathbf{Q}_j, \quad \mathbf{Q}_{bs} = \sum_{j \in J_{bs}} \mathbf{Q}_j, \quad \mathbf{Q}_j = \{k_j(X_a) \circ X_a, k_j(X_b) \circ X_b\}, \end{aligned}$$

которые назовем «категориальными» и «содержательными», где подмножества индексов $J_{as} \cup J_{bs} = \{1, \dots, h_s\}$, $J_{as} \cap J_{bs} = \emptyset$, $s = 1, \dots, m$. Здесь аргументы X_a и X_b соответствуют классам \mathbf{X}_a и \mathbf{X}_b .

«Категориальные» мультимножества \mathbf{R}_a и \mathbf{R}_b задают наилучшую из всех возможных декомпозиций рассматриваемой совокупности объектов на классы D_a и D_b для имеющегося набора первичных правил сортировки. При идеальных предварительных сортировках объектов, в которых отсутствуют противоречия между индивидуальными правилами экспертов, расстояние $d^* = d(\mathbf{R}_a, \mathbf{R}_b)$ между мультимножествами \mathbf{R}_a и \mathbf{R}_b будет максимально возможным расстоянием между объектами, входящими в разные классы. При идеальных предварительных сортировках объектов, в которых отсутствуют противоречия между индивидуальными правилами экспертов, расстояние d^* будет равно соответственно $d_{1p}^* = [kn]^{1/p}$, $d_{2p}^* = [1/h]^{1/p}$, $d_{3p}^* = 1$, где $h = h_1 + \dots + h_m + f$ – общее число значений всех признаков, описывающих объекты.

Аналогичным образом для каждой s -ой группы признаков Q_s нужно сформировать такие пары «содержательных» мультимножеств \mathbf{Q}_{sa}^* и \mathbf{Q}_{sb}^* , которые находились бы на максимально возможном расстоянии в метрическом пространстве мультимножеств и с достаточной точностью совпадали бы с разложением $\{\mathbf{R}_a, \mathbf{R}_b\}$. Задача нахождения классифицирующих признаков $q_s^j \in \mathbf{Q}_{as}^*$, $q_s^j \in \mathbf{Q}_{bs}^*$ сводится, таким образом, к решению m оптимизационных задач вида

$$d(\mathbf{Q}_{as}, \mathbf{Q}_{bs}) \rightarrow \max d(\mathbf{Q}_{as}, \mathbf{Q}_{bs}) = d(\mathbf{Q}_{as}^*, \mathbf{Q}_{bs}^*). \quad (5)$$

Решение каждой из указанных задач задает наилучшее разложение $\{\mathbf{Q}_{sa}^*, \mathbf{Q}_{sb}^*\}$ имеющейся совокупности объектов по s -ой группе признаков. Классифицирующие признаки для различных групп признаков можно упорядочить по величине $L_s = d(\mathbf{Q}_{as}^*, \mathbf{Q}_{bs}^*) / d(\mathbf{R}_a, \mathbf{R}_b)$, которая определяет точность аппроксимации и в некотором смысле характеризует относительную значимость s -ой группы признаков \mathbf{Q}_s в агрегированном групповом правиле классификации объектов.

Разные комбинации классифицирующих признаков, входящих в «содержательные» и «категориальные» мультимножества, дадут разные варианты агрегированных групповых правил для классификации совокупности многопризнаковых объектов:

$$\text{IF } \langle (q_{i1} \in Q_{au}^*) \text{ AND } (q_{i2} \in Q_{av}^*) \text{ AND } \dots \text{ AND } (q_{in} \in Q_{aw}^*) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_a \rangle, \quad (6)$$

$$\text{IF } \langle (q_{i1} \in Q_{bu}^*) \text{ AND } (q_{i2} \in Q_{bv}^*) \text{ AND } \dots \text{ AND } (q_{in} \in Q_{bw}^*) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_b \rangle. \quad (7)$$

Среди объектов, отобранных по агрегированному правилу в заданный класс D_a или D_b , будут как правильно, так и неправильно классифицированные объекты. Желательно найти такие согласованные групповые правила, которые обеспечат максимальную разность между числом правильно и числом неправильно классифицированных объектов. Комбинации соответствующих классифицирующих признаков включаются в согласованные групповые правила отнесения объектов к уточненным классам $D_a \setminus D_{ac}$ («безусловно хорошие») и $D_b \setminus D_{bc}$ («безусловно плохие»), которые записываются в виде:

$$\begin{aligned} & \text{IF } \langle (\sum_{q \in Q_{au}^*} k_{Ai}(q) > \sum_{q \in Q_{bu}^*} k_{Ai}(q)) \text{ AND } (\sum_{q \in Q_{av}^*} k_{Ai}(q) > \sum_{q \in Q_{bv}^*} k_{Ai}(q)) \\ & \text{AND } \dots \text{ AND } (k_{Ai}(r_a) > k_{Ai}(r_b)) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_a \setminus D_{ac} \rangle, \end{aligned} \quad (8)$$

$$\begin{aligned} & \text{IF } \langle (\sum_{q \in Q_{au}^*} k_{Ai}(q) < \sum_{q \in Q_{bu}^*} k_{Ai}(q)) \text{ AND } (\sum_{q \in Q_{av}^*} k_{Ai}(q) < \sum_{q \in Q_{bv}^*} k_{Ai}(q)) \\ & \text{AND } \dots \text{ AND } (k_{Ai}(r_a) < k_{Ai}(r_b)) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_b \setminus D_{bc} \rangle. \end{aligned} \quad (9)$$

Одновременно формируется класс противоречиво классифицированных объектов $D_c = D_{ac} \cup D_{bc}$, которые имеют противоречивые индивидуальные правила сортировки

$$\begin{aligned} & \text{IF } \langle [(\sum_{q \in Q_{au}^*} k_{Ai}(q) > \sum_{q \in Q_{bu}^*} k_{Ai}(q)) \text{ AND } \dots \text{ AND } (k_{Ai}(r_a) < k_{Ai}(r_b))] \\ & \text{OR } [(\sum_{q \in Q_{au}^*} k_{Ai}(q) < \sum_{q \in Q_{bu}^*} k_{Ai}(q)) \text{ AND } \dots \text{ AND } (k_{Ai}(r_a) > k_{Ai}(r_b))] \rangle, \text{ THEN } \langle \text{Object } A_i \in D_c \rangle. \end{aligned} \quad (10)$$

Эти объекты нуждаются в дополнительном анализе.

Алгоритмы построения агрегированных групповых правил

Процедура построения согласованных групповых правил для классификации многопризнаковых объектов, которые агрегируют большое число индивидуальных правил сортировки, включает в себя следующие алгоритмы:

алгоритм 1 – поиск классифицирующих признаков и построение агрегированных групповых правил отнесения объектов к классам D_a и D_b ;

алгоритм 2 – построение согласованных групповых правил отнесения объектов к уточненным классам $D_a \setminus D_{ac}$, $D_b \setminus D_{bc}$ и к классу противоречиво классифицированных объектов $D_c = D_{ac} \cup D_{bc}$.

Алгоритм 1 состоит из следующих шагов.

- Построить таблицу решений $K = \|k_{Ai}(q)\|_{k \times n}$, строки которой представляют собой мультимножества A_i вида (3).

• Сформировать путем сложения мультимножеств A_i группы объектов A_i , относящихся к заданным классам D_a («хорошие») и D_b («плохие»), которым соответствуют мультимножествам X_a и X_b (4). Отнесение объекта в один из двух классов производится по правилу большинства голосов на основе индивидуальных рекомендаций экспертов по признаку R без учета оценок по критериям Q_1, \dots, Q_{22} . Заметим, что объекты, получившие одинаковые оценки по каждому из критериев, но отличающиеся оценками по признаку R , попадают в разные классы. Построить агрегированную таблицу решений $K = \|k_{xi}(q)\|_{2 \times n}$, строки которой представляют собой мультимножества X_a и X_b .

• Найти классифицирующие признаки $q_s^j \in Q_{as}^*$, $q_s^j \in Q_{bs}^*$ по каждой s -ой группе признаков Q_s , которые являются решением оптимизационной задачи (5) с метрикой типа Хемминга. Проранжировать классифицирующие признаки по убыванию уровня аппроксимации L_s .

• Построить агрегированные групповые правила классификации объектов (6) и (7):

Укажем основные шаги алгоритма 2 применительно к классу лучших объектов D_a .

• По каждому критерию Q_s^* , входящему в групповое правило отнесения объектов к классу D_a , найти правильно классифицированные объекты, которые удовлетворяют условию

$$(\sum_{q \in Q_{au}} k_{Ai}(q) > \sum_{q \in Q_{bu}} k_{Ai}(q)) \text{ AND } (k_{Ai}(r_a) > k_{Ai}(r_b)),$$

и неправильно классифицированные объекты, которые удовлетворяют условию

$$(\sum_{q \in Q_{au}} k_{Ai}(q) > \sum_{q \in Q_{bu}} k_{Ai}(q)) \text{ AND } (k_{Ai}(r_a) < k_{Ai}(r_b)).$$

• Выбрать критерий Q_u^* , обеспечивающий максимальную разность между числом правильно и числом неправильно классифицированных объектов, который включается в согласованное групповое правило отнесения объектов к уточненному классу $D_a \setminus D_{ac}$.

• Рассмотреть все комбинации критерия Q_u^* с остальными критериями, входящими в групповое правило отнесения объектов к классу D_a , и найти правильно и неправильно классифицированные объекты, которые удовлетворяют условиям

$$(\sum_{q \in Q_{au}} k_{Ai}(q) > \sum_{q \in Q_{bu}} k_{Ai}(q)) \text{ AND } (\sum_{q \in Q_{av}} k_{Ai}(q) > \sum_{q \in Q_{bv}} k_{Ai}(q)) \text{ AND } (k_{Ai}(r_a) > k_{Ai}(r_b)),$$

$$(\sum_{q \in Q_{au}} k_{Ai}(q) > \sum_{q \in Q_{bu}} k_{Ai}(q)) \text{ AND } (\sum_{q \in Q_{av}} k_{Ai}(q) > \sum_{q \in Q_{bv}} k_{Ai}(q)) \text{ AND } (k_{Ai}(r_a) < k_{Ai}(r_b)).$$

• Выбрать пару критериев Q_u^* , Q_v^* , обеспечивающих максимальную разность между числом правильно и числом неправильно классифицированных объектов, которые включаются в согласованное групповое правило отнесения объектов к уточненному классу $D_a \setminus D_{ac}$.

• Рассмотреть все комбинации пары критериев Q_u^* , Q_v^* с остальными критериями, входящими в групповое правило отнесения объектов к классу D_a , и найти правильно и неправильно классифицированные объекты.

• Выбрать тройку, затем четверку и т.д. критериев Q_u^* , Q_v^* , Q_w^* , ... обеспечивающих максимальную разность между числом правильно и числом неправильно классифицированных объектов, которые включаются в согласованное групповое правило отнесения объектов к уточненному классу $D_a \setminus D_{ac}$.

• Включить критерии Q_u^* , Q_v^* , Q_w^* , ..., обеспечивающие максимальную разность между числами правильно и неправильно классифицированных объектов, в согласованное групповое правило отнесения объектов к уточненному классу решений $D_a \setminus D_{ac}$ (8)

$$\begin{aligned} & \text{IF } \langle (\sum_{q \in Q_{a2}} k_{Ai}(q) > \sum_{q \in Q_{b2}} k_{Ai}(q)) \text{ AND } (\sum_{q \in Q_{a3}} k_{Ai}(q) > \sum_{q \in Q_{b3}} k_{Ai}(q)) \\ & \text{AND } (\sum_{q \in Q_{a1}} k_{Ai}(q) > \sum_{q \in Q_{b1}} k_{Ai}(q)) \text{ AND } (\sum_{q \in Q_{a5}} k_{Ai}(q) > \sum_{q \in Q_{b5}} k_{Ai}(q)) \\ & \text{AND } (k_{Ai}(r_a) > k_{Ai}(r_b)) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_a \setminus D_{ac} \rangle. \end{aligned}$$

Аналогичным образом с помощью алгоритма 2 строится согласованное групповое правило отнесения объектов к уточненному классу решений $D_b \setminus D_{bc}$ (9)

$$\begin{aligned} & \text{IF } \langle (\sum_{q \in Q_{a2}} k_{A_i}(q) < \sum_{q \in Q_{b2}} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{a3}} k_{A_i}(q) < \sum_{q \in Q_{b3}} k_{A_i}(q)) \\ & \text{AND} (\sum_{q \in Q_{a1}} k_{A_i}(q) < \sum_{q \in Q_{b1}} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{a5}} k_{A_i}(q) < \sum_{q \in Q_{b5}} k_{A_i}(q)) \\ & \text{AND } (k_{A_i}(r_a) < k_{A_i}(r_b)) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_b \setminus D_{bc} \rangle. \end{aligned}$$

После этого формируется класс $D_c = D_{ac} \cup D_{bc}$ объектов, которые имеют противоречивые индивидуальные правила сортировки (10)

$$\begin{aligned} & \text{IF } \langle [(\sum_{q \in Q_{a2}} k_{A_i}(q) > \sum_{q \in Q_{b2}} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{a3}} k_{A_i}(q) > \sum_{q \in Q_{b3}} k_{A_i}(q)) \\ & \text{AND} (\sum_{q \in Q_{a1}} k_{A_i}(q) > \sum_{q \in Q_{b1}} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{a5}} k_{A_i}(q) > \sum_{q \in Q_{b5}} k_{A_i}(q)) \\ & \text{AND } (k_{A_i}(r_a) < k_{A_i}(r_b))] \\ & \text{OR } [\sum_{q \in Q_{a2}} k_{A_i}(q) < \sum_{q \in Q_{b2}} k_{A_i}(q) \text{ AND } (\sum_{q \in Q_{a3}} k_{A_i}(q) < \sum_{q \in Q_{b3}} k_{A_i}(q)) \\ & \text{AND} (\sum_{q \in Q_{a1}} k_{A_i}(q) < \sum_{q \in Q_{b1}} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{a5}} k_{A_i}(q) < \sum_{q \in Q_{b5}} k_{A_i}(q)) \\ & \text{AND } (k_{A_i}(r_a) < k_{A_i}(r_b))] \rangle, \text{ THEN } \langle \text{Object } A_i \in D_c \rangle. \end{aligned}$$

Провести дополнительный анализ таких объектов для устранения противоречий и принятия окончательного решения.

Апробация метода

Предложенный подход к построению согласованного решающего правила для классификации многопризнаковых объектов был апробирован для анализа итогов конкурса РФФИ целевых фундаментальных исследований 2006 года. Экспериментальные расчеты проводились на модельной базе данных, которая включала в себя экспертные оценки поддержанных и отклоненных проектов по областям следующим знаний:

02. Физика и астрономия – всего 127 проектов, в том числе 39 поддержанных и 88 отклоненных проектов,
04. Биология и медицинская наука – всего 252 проекта, в том числе 68 поддержанных и 184 отклоненных проектов.

Были приняты следующие правила агрегирования индивидуальных правил сортировки проектов. Проект безусловно включался в число поддержанных, когда все эксперты давали оценку $r_1 \ll 5$ (то есть $k_{A_i}(r_2) = k_{A_i}(r_3) = k_{A_i}(r_4) = 0$) или число оценок $r_1 \ll 5$ было большее или равно числу оценок $r_2 \ll 4$ (то есть $k_{A_i}(r_1) \geq k_{A_i}(r_2)$, $k_{A_i}(r_3) = k_{A_i}(r_4) = 0$). Проект безусловно включался в число отклоненных, когда ни один из экспертов не давал оценку $r_1 \ll 5$ (то есть $k_{A_i}(r_1) = 0$). Проекты, не удовлетворявшие этим правилам, были исключены из таблиц решений и отнесены к проектам, требующих дополнительного анализа. В эту же группу были включены проекты, которые имели одинаковые заключения всех экспертов и были отнесены к классам либо поддержанных D_a , либо отклоненных D_b .

Совокупности классифицирующих признаков $Q_a^* = \cup_s Q_{as}^*$ и $Q_b^* = \cup_s Q_{bs}^*$, построенных с помощью алгоритма 1, которые определяют принадлежность проекта к классам поддержанных D_a и отклоненных D_b проектов, оказались одинаковыми по областям 02 и 04 и выглядят следующим образом (признаки расположены в порядке убывания их относительной значимости):

$$Q_a^* = Q_{a16}^* \cup Q_{a22}^* \cup Q_{a21}^* \cup Q_{a15}^* = \{q_{16}^1, q_{16}^2; q_{22}^1, q_{22}^2; q_{21}^1, q_{21}^2; q_{15}^3\},$$

$$Q_b^* = Q_{b16}^* \cup Q_{b22}^* \cup Q_{b21}^* \cup Q_{b15}^* = \{q_{16}^3; q_{22}^3; q_{21}^3, q_{21}^4; q_{15}^1, q_{15}^2\}.$$

Множество классифицирующих признаков Q_b^* дополняет множество Q_a^* .

Последовательное рассмотрение классифицирующих признаков с помощью алгоритма 2 показало, что построенная комбинация признаков дает максимальное количество правильно классифицированных проектов и минимальное количество неправильно классифицированных. Согласованное агрегированное групповое правило для отнесения проектов к уточненному классу $D_a \setminus D_{ac}$ (безусловно поддержанные) имеет вид:

для критериев $Q_{16}, Q_{22} - k_{A_i}(q_s^1) + k_{A_i}(q_s^2) > k_{A_i}(q_s^3)$,

для критерия $Q_{15} - k_{A_i}(q_s^1) + k_{A_i}(q_s^2) < k_{A_i}(q_s^3)$,

для критерия $Q_{21} - k_{A_i}(q_{21}^1) + k_{A_i}(q_{21}^2) > k_{A_i}(q_{21}^3) + k_{A_i}(q_{21}^4)$.

При применении этих правил следует также учитывать выполнение приведенных выше правил агрегирования индивидуальных правил сортировки проектов по заключениям экспертов. Согласованное агрегированное групповое правило для отнесения проектов к уточненному классу $D_b \setminus D_{bc}$ (безусловно отклоненные) получается заменой знаков на противоположные.

В рассмотренных областях знаний конечный результат классификации проектов достигается уже при комбинации только нескольких первых критериев из полученных перечней, а именно:

по области 02 – критериев Q_{16}^*, Q_{22}^* ;

по области 04 – критериев $Q_{16}^*, Q_{22}^*, Q_{21}^*$.

Таким образом, наиболее важными для отбора проектов в рассмотренных областях знаний являются критерии Q_{16} . Научная значимость проекта, Q_{22} . Масштабы применимости результатов заявленных исследований, Q_{21} . Завершающая стадия фундаментальных исследований, предлагаемых в проекте. Фактически, экспертные оценки только по этим критериям и следует принимать во внимание при разбиении проектов по уточненным классам решений и выявлении проектов, требующих дополнительного рассмотрения из-за расхождения индивидуальных заключений экспертов. Учитывая эти факты, можно значительно сократить время, используемое для классификации проектов.

Каждое из правил классификации можно записать на естественном языке, что является отличительной чертой вербального анализа решений. Согласованное агрегированное групповое правило для отбора безусловно поддержанных проектов выглядит следующим образом:

«Проект имеет исключительно высокую или значительную научную значимость (оценки q_{16}^1, q_{16}^2); массовый или междисциплинарный масштабы применимости результатов заявленных исследований (оценки q_{22}^1, q_{22}^2); предлагаемые в проекте исследования завершаются лабораторным образцом или ключевыми элементами разработки (оценки q_{21}^1, q_{21}^2)».

Построенные правила классификации проектов позволяют провести анализ работы экспертов, определить критерии, которые оказывают решающее влияние на отбор проектов, оценить качество и согласованность экспертных оценок, выявить существенные расхождения в их мнениях.

Уровень согласованности индивидуальных экспертных оценок проектов по критериям и заключений экспертов по поддержке проектов оказался невысоким. Выявлено следующее число проектов, имеющих расхождения между оценкой их содержания и заключениями экспертов и требующих дополнительного анализа:

по области 02 – 33 проекта или 26% от общего числа конкурсных заявок, включенных в таблицу решений (в том числе 15 проектов на предварительной стадии и 18 проектов по результатам работы алгоритмов);

по области 04 – 84 проекта или 33% от общего числа конкурсных заявок, включенных в таблицу решений (в том числе 44 проекта на предварительной стадии и 40 проектов по результатам работы алгоритмов).

Эти факты свидетельствуют о недостаточном внимании и аккуратности некоторых экспертов при проведении экспертизы проектов.

Важное значение для работы алгоритмов построения правил классификации проектов имеет число экспертов, назначаемых для проведения экспертизы. Если заявка оценивается только двумя экспертами, то не исключена возможность возникновения ситуации, когда заключения экспертов будут противоположными: один эксперт поддерживает проект, а другой отклоняет. Для большей достоверности выводов при классификации проектов целесообразно, чтобы минимальное число экспертов, оценивающих заявку, было не менее трех, а желательно – чтобы заявка оценивалась четырьмя экспертами. Построение решающих правил упрощается, если каждый проект оценивается одним и тем же одинаковым числом экспертов.

Заключение

Проблемы классификации объектов, которые описываются многими количественными и качественными признаками, причем каждый из объектов может существовать в нескольких «экземплярах», являются достаточно трудными для решения и до настоящего момента плохо разработанными. Преодолеть трудности оказалось возможным, используя методы вербального анализа решений и аппарат теории множеств, которые не содержат необоснованных преобразований исходной информации и не приводят к потере или искажению данных.

Изложен новый метод согласованной групповой классификации многопризнаковых объектов, основанный на их представлении с помощью множеств. Метод обеспечивает построение согласованного группового правила для классификации многопризнаковых объектов, которое аппроксимирует большое число противоречивых индивидуальных правил сортировки.

Подход апробирован на модельной базе данных, содержащей результаты экспертной оценки и конкурсного отбора проектов целевых фундаментальных исследований в РФФИ. Сформированы решающие правила для отнесения объектов в заданные классы. Решающие правила позволили выделить наиболее важные для отбора проектов критерии по каждой области знания и выявить расхождения в индивидуальных правилах сортировки проектов, применявшихся экспертами.

Предлагаемый подход обеспечивает систематизацию имеющейся информации, анализ причин принятия окончательного решения, получение его обоснования. Полученные результаты могут быть использованы при проведении экспертного анализа итогов конкурсов целевых фундаментальных исследований, а также могут найти применение при анализе конкурсов инициативных и иных проектов (с учетом их специфики), выполняемых в РФФИ.

Благодарности

Работа поддержана программами фундаментальных исследований президиума РАН «Интеллектуальные информационные технологии, математическое моделирование, системный анализ и автоматизация» и ОНИТ РАН «Информационные технологии и методы анализа сложных систем», Российским фондом фундаментальных исследований (проекты 08-01-00247, 08-07-13532, 09-07-00009, 09-07-12111).

Список литературы

- [Ларичев, 2006] Ларичев О.И. Вербальный анализ решений / Под ред. А.Б.Петровского. — М.: Наука, 2006.
- [Петровский, 2003] Петровский А.Б. Пространства множеств и мультимножеств. — М: Едиториал УРСС, 2003.
- [Петровский, 2009] Петровский А. Б. Теория принятия решений. — М.: Издательский центр «Академия», 2009.
- [Петровский, Тихонов, 2009] Петровский А. Б., Тихонов И. П. Фундаментальные исследования, ориентированные на практический результат: подходы к оценке эффективности // Вестник РАН. — 2009.— Т. 79, № 11.— С. 1006–1011.
- [Doumpos, Zopounidis, 2002] Doumpos M., Zopounidis C. Multicriteria Decision Aid Classification Methods. — Dordrecht: Kluwer Academic Publishers, 2002.
- [Greco et al., 2002] Greco S., Matarazzo B., Slowinski R. Rough sets methodology for sorting problems in presence of multiple attributes and criteria // European Journal of Operational Research. — 2002. — Vol. 138, no. 2. — Pp. 247-259.
- [Köksalan et al., 2003] Köksalan M., Ulu C. An interactive approach for placing alternatives in preference classes // European Journal of Operational Research. — 2003. — Vol. 144, no. 2. — Pp. 429–439.
- [Petrovsky, 2001] Petrovsky A.B. Multi-Attribute Sorting of Qualitative Objects in Multiset Spaces. // Multiple Criteria Decision Making in the New Millenium. — Berlin: Springer-Verlag, 2001. — Pp.124-131.
- [Roy, Bouyssou, 1993] Roy B., Bouyssou D. Aide Multicritere a la decision: Methodes et cas — Paris: Economica, 1993.

Сведения об авторах

Петровский Алексей Борисович – д.т.н., заведующий лабораторией Института системного анализа РАН, Россия, Москва 117312, пр-т 60-летия Октября, 9, тел. (495)135-8503, e-mail: pab@isa.ru

Ройзензон Григорий Владимирович – к.т.н., старший научный сотрудник Института системного анализа РАН, Россия, Москва 117312, пр-т 60-летия Октября, 9, тел. (495)135-8503, e-mail: rgv@isa.ru

Тихонов Игорь Петрович – к.т.н., начальник отдела целевых фундаментальных исследований Российского фонда фундаментальных исследований, Россия, Москва 117334, Ленинский пр-кт, 32А, e-mail: tikhonov@rfbr.ru

Балышев Александр Владимирович – к.биол.н., заместитель начальника отдела целевых фундаментальных исследований Российского фонда фундаментальных исследований, Россия, Москва 117334, Ленинский пр-кт, 32А, e-mail: bav@rfbr.ru