УДК 510.22

УПОРЯДОЧЕНИЕ И КЛАССИФИКАЦИЯ ОБЪЕКТОВ С ПРОТИВОРЕЧИВЫМИ ПРИЗНАКАМИ*

А.Б.Петровский
доктор технических наук, академик РАЕН,
заведующий отделом Института системного анализа РАН, pab@isa.ru

Мультимножество или множество с повторяющимися элементами служит удобной математической моделью для представления объектов, которые характеризуются многими разнородными (количественными и качественными) признаками и могут существовать в нескольких экземплярах с отличающимися, в частности, противоречивыми значениями признаков. В работе рассматриваются новые методы упорядочения и классификации таких многопризнаковых объектов, основанные на теории метрических пространств мультимножеств. Методы применены для решения практических задач: построения рейтинга компаний и конкурсного отбора проектов, оцененных несколькими экспертами по многим критериям.

1. Введение

В проблемах многокритериального принятия решений, распознавания образов, классификации, обработки разнородной информации, теории кодирования других предметных областях часто возникает необходимость сгруппировать или упорядочить анализируемые объекты, основываясь на их свойствах, выраженных признаками (атрибутами) объектов. Вместе с тем имеется достаточно широкий круг задач, где изучаемые объекты характеризуются многими разнородными признаками, которые могут быть и количественными, и качественными, и, кроме того, одни и те же объекты могут существовать в нескольких экземплярах с отличающимися значениями признаков, свертка которых или невозможна, или математически некорректна. В качестве примеров таких задач укажем классификацию и ранжирование объектов, оцененных несколькими экспертами по многим качественным критериям, распознавание графических символов, обработку текстовых документов. Множественность и повторяемость факторов, описывающих объекты, усложняет и затрудняет решение таких задач. Главные трудности обусловлены необходимостью одновременно учитывать большое количество вербальных и числовых данных и обрабатывать эти данные, не прибегая к дополнительным преобразованиям типа усреднения, смешивания, взвешивания, которые могут привести к необоснованным и необратимым искажениям исходных данных.

Удобной математической моделью для представления многопризнаковых объектов является мультимножество или множество с повторяющимися элементами. Кратность элементов – существенная особенность мультимножества, позволяющая отличать его от множества и рассматривать мультимножество как качественно новое математическое понятие. В работе предложены методы упорядочения и классификации совокупности многопризнаковых объектов, которые базируются на теории метрических пространств мультимножеств. Метод упорядочения объектов основан на оценке их близости по отношению к некоторому «идеальному» объекту в многопризнаковом пространстве. Метод классификации объектов позволяет строить обобщенное решающее правило для их отбора, которое аппроксимирует различные, в том числе и противоречивые, правила экспертной сортировки объектов.

* Работа частично поддержана грантом Президента Российской Федерации по поддержке ведущих научных школ НШ1964.2003.1, Российским фондом фундаментальных исследований (проекты 01-01-00514, 02-01-01077), Российской академией наук (программы фундаментальных исследований РАН «Математическое моделирование и интеллектуальные системы» и ОИТВС РАН «Фундаментальные основы информационных технологий и систем»).

2. Способы представления многопризнаковых объектов

Выбор той или иной модели для представления рассматриваемых объектов и исследования структуры их связей определяется свойствами этих объектов, которые выражаются признаками (атрибутами) объектов. Признаки, характеризующие свойства объектов, могут быть непрерывными и дискретными, количественными и качественными, или смешанными.

Обычно совокупность объектов представляется множеством точек в некотором многомерном (как правило, метрическом) пространстве, оси которого соотносятся с соответствующими признаками. В прикладных задачах в качестве такого пространства достаточно часто (но, заметим, не всегда обоснованно) выбирается пространство типа евклидового. Задание расстояния между объектами позволяет оценивать близость или удаленность этих объектов относительно друг друга вне зависимости от их природы, исследовать структурные особенности совокупности объектов и всего пространства в целом.

В различных предметных областях рассматриваются совокупности $A=\{A_1,...,A_k\}$ объектов, которые описываются m дискретными признаками $Q_1,\ldots,Q_m$, имеющими конечное число $q_s^{e_s}$, $e_s=1,\ldots,h_s$, $s=1,\ldots,m$ количественных (числовых) или качественных (номинальных, либо порядковых) значений. Каждый объект A_i, $i=1,\ldots,k$ из совокупности A можно представить как точку q_i в m-мерном векторном пространстве $Q=Q_1\times Q_2\times\ldots\times Q_m$, являющемся прямым произведением шкал значений признаков Q_s, и поставить объекту A_i в соответствие m-мерный вектор $A_i=(q_{i1}^{e_1}, q_{i2}^{e_2},\ldots, q_{im}^{e_m})$ [1], [2], [3], [4].

Ситуация существенным образом усложняется, если одному и тому же объекту A_i может соответствовать не один, а несколько m-мерных векторов с различающимися значениями признаков. Подобная ситуация возникает, например, когда необходимо одновременно учесть m параметров объекта A_i, измеренных n различными способами, либо когда объект A_i оценивается n независимыми экспертами по m критериям. В таком случае объект A_i представляется в m-мерном пространстве Q уже не одной точкой q_i, а группой ("облаком"), состоящей из n точек $\{q_i^{(1)},\ldots,q_i^{(n)}\}$ вида $A_i=\{(q_{i1}^{e_1(1)}, q_{i2}^{e_2(1)},\ldots, q_{im}^{e_m(1)}),\ldots,(q_{i1}^{e_1(n)}, q_{i2}^{e_2(n)},\ldots, q_{im}^{e_m(n)})\}$, которая должна рассматриваться и анализироваться как единое целое. При этом, очевидно, измеренные разными способами значения параметров, как и индивидуальные оценки экспертов, могут быть похожими, различающимися и даже противоречивыми, что в свою очередь может приводить к несравнимости m-мерных векторов $q_i^{(j)}=(q_{i1}^{e_1(j)}, q_{i2}^{e_2(j)},\ldots, q_{im}^{e_m(j)})$, характеризующих один и тот же объект A_i.

Совокупность таких многомерных объектов может иметь в пространстве Q сложную структуру, достаточно трудную для анализа. Непросто ввести в этом пространстве и метрику для измерения расстояний между объектами. Указанные трудности можно преодолеть, воспользовавшись иным способом представления многопризнаковых объектов, основанным на формализме мультимножеств [5], [6], который позволяет одновременно учесть все комбинации значений количественных и качественных признаков, а также число значений каждого из этих признаков. Вместо прямого произведения m шкал значений признаков $Q=Q_1\times Q_2\times\ldots\times Q_m$ введем обобщенную шкалу признаков – множество $G=\{Q_1,Q_2,\ldots,Q_m\}$, состоящее из m групп признаков, и представим объект $A_i\in A$ в таком символическом виде:

$$A_i=\{k_{Ai}(q_1^{\,1})\bullet q_1^{\,1},\ldots,k_{Ai}(q_1^{h_1})\bullet q_1^{h_1},\ldots,k_{Ai}(q_m^{\,1})\bullet q_m^{\,1},\ldots,k_{Ai}(q_m^{h_m})\bullet q_m^{h_m}\}, \qquad (1)$$

где число $k_{Ai}(q_s^{e_s})$ указывает, сколько раз признак $q_s^{e_s}\in Q_s$ встречается в описании объекта A_i, знак $\bullet$ обозначает кратность вхождения признака $q_s^{e_s}$. Например, при многокритериальной оценке объекта A_i несколькими экспертами число $k_{Ai}(q_s^{e_s})$ равно числу экспертов, давших

объекту A_i оценку $q_s^{e_s}$ по критерию Q_s. Объект A_i можно записать и более единообразно как $A_i=\{k_{Ai}(x_1)\bullet x_1,\ldots,k_{Ai}(x_h)\bullet x_h\}$, определив элементы множества $G=\{x_1,...,x_h\}$ следующим образом:

$$x_1=q_1^{\,1},x_2=q_1^{\,2},\ldots,x_{h_1}=q_1^{h_1},\ x_{h_1+1}=q_2^{\,1},\ldots,x_{h_1+h_2}=q_2^{h_2},\ldots,x_{h_1+\ldots+h_{m-1}+1}=q_m^{\,1},\ldots,x_{h_1+\ldots+h_m}=q_m^{h_m},$$

где $h=h_1+...+h_m$. Множество G определяет свойства совокупности объектов $\mathsf{A}=\{A_1,...,A_k\}$. Такие объекты A_i суть множества с повторяющимися элементами $x_j\in G$ или мультимножества, и их можно представлять точками в метрических пространствах мультимножеств.

3. Мультимножества и операции над ними

Дадим краткий обзор теории мультимножеств и метрических пространств мультимножеств [5], [6]. Мультимножеством $\boldsymbol{A}$, порожденным обычным множеством $U=\{x_1,x_2,\ldots\}$, все элементы которого различны, называется совокупность групп элементов вида $\boldsymbol{A}=\{k_A(x)\bullet x|x\in U,\ k_A(x)\in\mathbb{Z}_+\}$. Здесь $k_A:U\to\mathbb{Z}_+=\{0,1,2,\ldots\}$ называется функцией числа экземпляров мультимножества, определяющей кратность вхождения элемента $x_i\in U$ в мультимножество $\boldsymbol{A}$, что обозначено символом $\bullet$. Если $k_A(x)=\chi_A(x)$, где $\chi_A(x)=1$ при $x\in\boldsymbol{A}$ и $\chi_A(x)=0$ при $x\notin A$, то мультимножество $\boldsymbol{A}$ становится обычным множеством. Если все мультимножества семейства $\mathsf{A}=\{\boldsymbol{A}_1,\boldsymbol{A}_2,\ldots\}$ образуются из элементов множества G, то G называется доменом для семейства A, а множество $\mathrm{Supp}\boldsymbol{A}=\{x|x\in G,\ \chi_{\mathrm{Supp}A}(x)=\chi_A(x)\}$ – опорным множеством или носителем мультимножества $\boldsymbol{A}$. Мощность мультимножества $|\boldsymbol{A}|=\Sigma_x k_A(x)$ определяется как общее число экземпляров всех его элементов; размерность мультимножества $/\boldsymbol{A}/=\Sigma_x\chi_A(x)=|\mathrm{Supp}\boldsymbol{A}|$ – как общее число различных элементов. Максимальное значение функции кратности $\mathrm{hgt}\boldsymbol{A}=\max\limits_{x\in G}k_A(x)$ называется высотой, а элемент $x_{A*}=\arg\max\limits_{x\in G}k_A(x)$ – пиком мультимножества $\boldsymbol{A}$. Мультимножество называется пустым $\varnothing$, если $k_\varnothing(x)=0$, и максимальным $\boldsymbol{Z}$, если $k_Z(x)=\max\limits_{A\in\mathsf{A}}k_A(x),\ \forall x\in U$.

Рассмотрим возможные способы сопоставления мультимножеств, обусловленные особенностями их различных характеристик. Мультимножества $\boldsymbol{A}$ и $\boldsymbol{B}$ называются равными ($\boldsymbol{A}=\boldsymbol{B}$), если $k_A(x)=k_B(x)$ для всех элементов $x\in G$, и неравными ($\boldsymbol{A}\neq\boldsymbol{B}$), если $k_A(x)\neq k_B(x)$ хотя бы для одного $x\in G$. Для равных мультимножеств имеем $|\boldsymbol{A}|=|\boldsymbol{B}|$, $/\boldsymbol{A}/=/\boldsymbol{B}/$, $\mathrm{hgt}\boldsymbol{A}=\mathrm{hgt}\boldsymbol{B}$, $x_{A*}=x_{B*}$, $\mathrm{Supp}\boldsymbol{A}=\mathrm{Supp}\boldsymbol{B}$. Мультимножества $\boldsymbol{A}$ и $\boldsymbol{B}$ будем называть равномощными, если $|\boldsymbol{A}|=|\boldsymbol{B}|$; равноразмерными, если $/\boldsymbol{A}/=/\boldsymbol{B}/$; равновеликими, если они равномощны и равноразмерны. Равные мультимножества равновелики, обратное утверждение, вообще говоря, неверно.

Будем говорить, что мультимножество $\boldsymbol{B}$ содержится или включено в мультимножество $\boldsymbol{A}$ ($\boldsymbol{B}\subseteq\boldsymbol{A}$), если $k_B(x)\le k_A(x)$, для каждого элемента $x\in G$. Мультимножество $\boldsymbol{B}$ называется тогда подмультимножеством мультимножества $\boldsymbol{A}$, а мультимножество $\boldsymbol{A}$ – надмультимножеством мультимножества $\boldsymbol{B}$. В этом случае $|\boldsymbol{B}|\le|\boldsymbol{A}|$, $/\boldsymbol{B}/\le/\boldsymbol{A}/$, $\mathrm{hgt}\boldsymbol{B}\le\mathrm{hgt}\boldsymbol{A}$, $\mathrm{Supp}\boldsymbol{B}\subseteq\mathrm{Supp}\boldsymbol{A}$, а $x_{A*}=x_{B*}$, либо $x_{A*}\neq x_{B*}$. Как и в случае обычных множеств, одновременное выполнение условий $\boldsymbol{B}\subseteq\boldsymbol{A}$ и $\boldsymbol{A}\subseteq\boldsymbol{B}$ влечет равенство мультимножеств $\boldsymbol{A}=\boldsymbol{B}$. Включение мультимножества обладает свойствами рефлексивности ($\boldsymbol{A}\subseteq\boldsymbol{A}$) и транзитивности ($\boldsymbol{A}\subseteq\boldsymbol{B}$, $\boldsymbol{B}\subseteq\boldsymbol{C}\Rightarrow\boldsymbol{A}\subseteq\boldsymbol{C}$), а значит, является отношением предпорядка.

Мультимножества $\boldsymbol{A}$ и $\boldsymbol{B}$ будем называть одноименно или S-эквивалентными ($\boldsymbol{A}\cong\boldsymbol{B}$), если их носители совпадают ($\mathrm{Supp}\boldsymbol{A}=\mathrm{Supp}\boldsymbol{B}$) и существует взаимно однозначное соответствие f между одноименными компонентами: $k_B(x)=f(k_A(x)),\ \forall x\in G$; разноименно или D-эквивалентными ($\boldsymbol{A}\underline{\cong}\boldsymbol{B}$), если их носители эквивалентны ($\mathrm{Supp}\boldsymbol{A}\sim\mathrm{Supp}\boldsymbol{B}$) и существует взаимно однозначное соответствие f между разноименными компонентами: $k_B(x_i)=f(k_A(x_j))$, $x_i,x_j\in G$, где f – целочисленная функция с областью значений $\mathbb{Z}_+$. S- и D-эквивалентные мультимножества равноразмерны $/\boldsymbol{B}/=/\boldsymbol{A}/$, их высоты связаны равенством $\mathrm{hgt}\boldsymbol{B}=f(\mathrm{hgt}\boldsymbol{A})$. Одно из S-

эквивалентных мультимножеств всегда является подмультимножеством другого, а для *D*-эквивалентных мультимножеств это утверждение не выполняется. *D*-эквивалентные мультимножества становятся *S*-эквивалентными, если в одном из мультимножеств переобозначить элементы $x_i \to x_j$. Частными случаями *S*-эквивалентности будут равные мультимножества; сдвинутые мультимножества, для которых $k_B(x)=k_A(x)+p$, $p\geq 0$ – целое; растянутые или пропорциональные мультимножества, для которых $k_B(x)=qk_A(x)$, $q\geq 1$ – целое. Важным частным случаем *D*-эквивалентности являются равносоставленные мультимножества, чьи разноименные компоненты равны $k_A(x_i)=k_B(x_j)$, $x_i,x_j\in G$. Равные мультимножества равносоставлены, обратное утверждение неверно.

Введем следующие основные операции над мультимножествами:

объединение	$\boldsymbol{A}\cup\boldsymbol{B} = \{k_{A\cup B}(x)\bullet x \mid k_{A\cup B}(x)=\max(k_A(x), k_B(x))\};$
пересечение	$\boldsymbol{A}\cap\boldsymbol{B} = \{k_{A\cap B}(x)\bullet x \mid k_{A\cap B}(x)=\min(k_A(x), k_B(x))\};$
арифметическое сложение	$\boldsymbol{A}+\boldsymbol{B} = \{k_{A+B}(x)\bullet x \mid k_{A+B}(x)=k_A(x)+k_B(x)\};$
арифметическое вычитание	$\boldsymbol{A}-\boldsymbol{B} = \{k_{A-B}(x)\bullet x \mid k_{A-B}(x)=k_A(x)-k_{A\cap B}(x)\};$
симметрическая разность	$\boldsymbol{A}\Delta\boldsymbol{B} = \{k_{A\Delta B}(x)\bullet x \mid k_{A\Delta B}(x)=\lvert k_A(x)-k_B(x)\rvert\};$
дополнение	$\overline{\boldsymbol{A}} = \boldsymbol{Z}-\boldsymbol{A} = \{k_{\overline{A}}(x)\bullet x \mid k_{\overline{A}}(x)=k_Z(x)-k_A(x)\};$
умножение на число (репродукция)	$h\bullet\boldsymbol{A} = \{k_{h\bullet A}(x)\bullet x \mid k_{h\bullet A}(x)=h\cdot k_A(x),\ h\in\mathbb{Z}_+\};$
арифметическое умножение	$\boldsymbol{A}\bullet\boldsymbol{B} = \{k_{A\bullet B}(x)\bullet x \mid k_{A\bullet B}(x) = k_A(x)\cdot k_B(x)\};$
арифметическая *n*-ая степень	$\boldsymbol{A}^n = \{k_{A^n}(x)\bullet x \mid k_{A^n}(x) = (k_A(x))^n\};$
прямое произведение	$\boldsymbol{A}\times\boldsymbol{B} = \{k_{A\times B}\bullet\langle x_i, x_j\rangle \mid k_{A\times B}=k_A(x_i)\cdot k_B(x_j),\ x_i\in\boldsymbol{A},\ x_j\in\boldsymbol{B}\};$
прямая *n*-ая степень	$(\times\boldsymbol{A})^n = \{k_{(\times A)^n}\bullet\langle x_1,\ldots,x_n\rangle \mid k_{(\times A)^n}=\prod_{i=1}^{n} k_A(x_i),\ x_i\in\boldsymbol{A}\}.$

Носители операций над мультимножествами определяются следующими выражениями:

$\mathrm{Supp}(\boldsymbol{A}\cup\boldsymbol{B}) = \mathrm{Supp}(\boldsymbol{A}+\boldsymbol{B}) = (\mathrm{Supp}\boldsymbol{A})\cup(\mathrm{Supp}\boldsymbol{B});$

$\mathrm{Supp}(\boldsymbol{A}\cap\boldsymbol{B}) = \mathrm{Supp}\,(\boldsymbol{A}\bullet\boldsymbol{B}) = (\mathrm{Supp}\boldsymbol{A})\cap(\mathrm{Supp}\boldsymbol{B});$

$\mathrm{Supp}(\boldsymbol{A}\Delta\boldsymbol{B}) = (\mathrm{Supp}(\boldsymbol{A}-\boldsymbol{B}))\cup(\mathrm{Supp}(\boldsymbol{B}-\boldsymbol{A}));$

$(\mathrm{Supp}\boldsymbol{A})\Delta(\mathrm{Supp}\boldsymbol{B}) = (\mathrm{Supp}\boldsymbol{A}\backslash\mathrm{Supp}\boldsymbol{B})\cup(\mathrm{Supp}\boldsymbol{B}\backslash\mathrm{Supp}\boldsymbol{A});$

$\mathrm{Supp}(h\bullet\boldsymbol{A}) = \mathrm{Supp}\boldsymbol{A} = \mathrm{Supp}(\boldsymbol{A}^n);$ $\quad\mathrm{Supp}(\boldsymbol{A}\times\boldsymbol{B}) = (\mathrm{Supp}\boldsymbol{A})\times(\mathrm{Supp}\boldsymbol{B}).$

В теории множеств операции арифметического сложения, умножения на число, арифметического умножения и возведения в степень множеств в общем случае не определяются. Аналогами этих операций могут служить соответственно покомпонентное сложение и умножение на скаляр векторов $\mathbf{a}+\mathbf{b}=(a_1+b_1,\ldots,a_n+b_n)$, $h\cdot\mathbf{a}=(ha_1,\ldots,ha_n)$ и матриц $\mathrm{A}+\mathrm{B}=\|a_{ij}+b_{ij}\|_{m\times n}$, $h\cdot\mathrm{A}=\|h\,a_{ij}\|_{m\times n}$, поэлементное умножение матриц $\mathrm{A}\cdot\mathrm{B}=\|a_{ij}\cdot b_{ij}\|_{m\times n}$. Последняя операция, введенная в алгебраической теории распознавания образов [7], отличается от традиционной операции умножения матриц. При переходе к множествам арифметическое умножение и возведение в степень мультимножеств вырождаются в пересечение множеств, а арифметическое сложение множеств и умножение множества на число будут неосуществимы.

Семейство мультимножеств, замкнутое относительно операций объединения, пересечения, сложения и дополнения, называется алгеброй мультимножеств $L\ (\boldsymbol{Z})$, где максимальное мультимножество $\boldsymbol{Z}$ является единицей алгебры, а пустое мультимножество $\varnothing$ – нулем. Действительная неотрицательная функция $m(\boldsymbol{A})$, определенная на алгебре $L\ (\boldsymbol{Z})$ и удовлетворяющая условию коаддитивности (сильной аддитивности): $m(\boldsymbol{A})+m(\boldsymbol{B})=m(\boldsymbol{A}+\boldsymbol{B})$, называется мерой мультимножества. Мера мультимножества $m(\boldsymbol{A})$ обладает следующим свойствами: $m(\varnothing)=0$; монотонность $m(\boldsymbol{A})\leq m(\boldsymbol{B})\Leftrightarrow\boldsymbol{A}\subseteq\boldsymbol{B}$; непрерывность $\lim\limits_{i\to\infty} m(A_i)=m(\lim\limits_{i\to\infty} A_i)$; симметричность $m(\boldsymbol{A})+m(\overline{\boldsymbol{A}})=m(\boldsymbol{Z})$; эластичность $m(h\bullet\boldsymbol{A})=hm(\boldsymbol{A})$. Меру мультимножества можно определить различными способами, например, как линейную комбинацию функций кратности:

$m(\boldsymbol{A})=\sum_j w_j k_A(x_j)$, $w_j>0$. Заметим, что мощность мультимножества $|\boldsymbol{A}|$ также будет мерой мультимножества

Метрические пространства мультимножеств $(\mathcal{A},d)$ введены в [5], где определены следующие виды расстояний между мультимножествами:

$$d_1(\boldsymbol{A},\boldsymbol{B}) = m(\boldsymbol{A}\Delta\boldsymbol{B}); \quad d_2(\boldsymbol{A},\boldsymbol{B}) = m(\boldsymbol{A}\Delta\boldsymbol{B})/m(\boldsymbol{Z}); \quad d_3(\boldsymbol{A},\boldsymbol{B}) = m(\boldsymbol{A}\Delta\boldsymbol{B})/m(\boldsymbol{A}\cup\boldsymbol{B}). \qquad (2)$$

Функции $d_2(\boldsymbol{A},\boldsymbol{B})$ и $d_3(\boldsymbol{A},\boldsymbol{B})$ удовлетворяют условию нормировки $0\le d(\boldsymbol{A},\boldsymbol{B})\le 1$. По определению принимается $d_3(\varnothing,\varnothing)=0$. Основное расстояние $d_1(\boldsymbol{A},\boldsymbol{B})$ является метрикой типа Хемминга, традиционно используемым во многих приложениях. Полностью усредненное расстояние $d_2(\boldsymbol{A},\boldsymbol{B})$ характеризует различие между двумя мультимножествами $\boldsymbol{A}$ и $\boldsymbol{B}$, отнесенное к расстоянию, максимально возможному в исходном пространстве. Локально усредненное расстояние $d_3(\boldsymbol{A},\boldsymbol{B})$ задает различие, отнесенное к максимально возможной «общей части» только этих двух мультимножеств в исходном пространстве.

4. Построение рейтинга компаний

Одним из весьма распространенных подходов к структуризации совокупности объектов $\mathcal{A}=\{A_1,...,A_k\}$ является их строгое или нестрогое упорядочение, которое представляет собой введение между объектами бинарных отношений строгого или нестрогого порядка, эквивалентности или несравнимости, заданных на множестве свойств объектов. Сравнение объектов по их свойствам производится на основе признаков, характеризующих объекты.

Рассмотрим достаточно часто встречающуюся практическую задачу нахождения рейтинга компаний, занимающихся бизнесом в некоторой области. Решить такую задачу, можно, например, голосованием – рейтинг компаний определяется тогда по количеству поданных за нее голосов. Но в этом случае получается оценка компании «в целом» без каких-либо деталей.

Более сложной является задача построения рейтинга компаний, основываясь на фактических показателях их деятельности и/или экспертных оценках по многим критериям. Перечень таких критериев формируется заранее, он зависит от целей анализа. Например, компании, действующие в некотором секторе рынка, можно оценивать по следующим критериям: Q_1. Уровень деловой активности; Q_2. Объем прибыли от реализации продукции; Q_3. Объем продаж; Q_4. Число выполненных проектов; Q_5. Квалификация персонала; Q_6. Численность сотрудников компании; и тому подобное. Шкалы критериев оценки могут быть как количественными, так и качественными. Для удобства оценки и сравнения компаний количественные критерии можно трансформировать в качественные с небольшим числом упорядоченных градаций шкал. Шкалы критериев Q_4. «Число выполненных проектов» и Q_6. «Численность сотрудников компании» могут иметь, например, такой вид:

q_4^1 – очень высокое (больше ста);
q_4^2 – высокое (от пятидесяти до ста);
q_4^3 – среднее (от десяти до пятидесяти);
q_4^4 – низкое (меньше десяти).

Пусть каждая компания из некоторой совокупности оценивается несколькими экспертами по всем критериям. В частности, возможна ситуация, когда представитель каждой компании является экспертом, ставящим свои оценки всем рассматриваемым компаниям, в том числе и своей собственной. При этом оценки разных экспертов могут отличаться друг от друга и даже быть противоречивыми. В таком случае каждую компанию можно рассматривать как многопризнаковый объект, а определение рейтинга компаний представляет собой тогда задачу упорядочивания многопризнаковых объектов. Основной трудностью при решении таких задач является необходимость учета всех описаний объекта – различающихся оценок, сделанных разными экспертами.

К числу наиболее популярных методов упорядочения объектов относятся непосредственная порядковая классификация, ранжирование, парные сравнения.

Наименее трудоемким для эксперта методом упорядочения объектов является метод непосредственной классификации с именованными и упорядоченными классами – метод сортировки [8]. В этом методе эксперт непосредственно относит объект A_i к одному из выделенных классов, назначая объекту одну из оценок по порядковой или номинальной шкале критериев. При коллективной экспертизе сортировка объектов проводится обычно на основе распределений экспертных оценок. Если согласованность оценок оказывается приемлемой, то в качестве коллективной средней оценки используется медиана Кемени-Снелла [9], [10], которая практически достаточно часто совпадает с модой распределения. Итоговое упорядочение объектов строится на основе средних оценок.

При упорядочении объектов с помощью метода ранжирования для каждого объекта A_i тем или иным образом, например, на основе предпочтений лица, принимающего решение (ЛПР), или оценок эксперта вычисляется натуральное число r_i, называемое рангом. Упорядочению объектов соответствует упорядочение рангов $r_1<r_2<\ldots<r_i<\ldots<r_c$. Возможны различные способы ранжирования объектов. Например, объекты могут предъявляться эксперту все сразу или поочередно. При небольшом числе объектов и одном признаке (критерии) оценке объектов ранжирование не представляет больших трудностей для экспертов. При увеличении числа объектов, критериев и экспертов, оценивающих объекты, количество связей между оценками резко возрастает. Поэтому эксперты могут допускать в таких случаях существенные ошибки при определении рангов объектов. В силу ограниченных возможностей человека при обработке информации методы ранжирования объектов являются для экспертов более трудоемкими по сравнению с методами непосредственной классификации.

В методах парных сравнений итоговое упорядочение объектов строится на основе сравнения всех пар объектов. ЛПР или эксперту предъявляется пара объектов и предлагается указать, какой из объектов более предпочтителен. В случае сравнения всех пар объектов и транзитивности предпочтений эксперта, получается полное упорядочение объектов. Если эксперт считает некоторые из объектов несравнимыми, то упорядочение будет частичным. Для каждого эксперта и признака (критерия) составляется своя матрица парных сравнений «объект-объект». Таким образом, появляется набор матриц, обработка которых для получения итогового упорядочения требует создания специальных вычислительных алгоритмов.

ЛПР и эксперты могут быть не всегда последовательными в своих ответах, могут допускать неточности, особенно в трудных случаях, предпочтения ЛПР могут быть противоречивыми. Для преодоления таких трудностей при построении итоговых упорядочений разрабатываются специальные процедуры. Так, например, в группе методов ЗАПРОС (Замкнутые Процедуры у Опорных Ситуаций) [1] для упорядочения многокритериальных объектов используется процедура выявления цепочек сравнений, образующих нетранзитивные триады. Выявленные нарушения предъявляются ЛПР для изменения его оценок с тем, чтобы устранить противоречия и построить единую порядковую шкалу оценок. В группе методов ELECTRE (Elimination et Choix Traduisant la Realite) [11] упорядочение многокритериальных объектов осуществляется путем их попарного сравнения с использованием специальных индексов согласия и несогласия, рассчитываемых на основе предпочтений ЛПР.

Когда объекты имеют многопараметрическое описание, а сами объекты должны рассматриваться и анализироваться как единое целое, например, когда объекты оцениваются несколькими экспертами по многим качественным критериям $Q_1,\ldots,Q_m$, построение итогового упорядочения k объектов на основании m отдельных ранжировок, полученных по каждому из параметров вызывает значительные трудности. Исторически сложились два подхода к их преодолению, которые можно условно назвать статистическим и алгебраическим [8]. При статистическом подходе каждое из индивидуальных упорядочений, к примеру, заданное экспертом, рассматривается как одна из возможных реализаций одного и того же наиболее ве-

роятного упорядочения объектов. Известны различные модели для построения такого вероятностного упорядочения, например, модели Льюса, Терстоуна и другие.

В алгебраическом подходе итоговое упорядочение ищется как наиболее близкое ко всем индивидуальным упорядочениям. Близость ранжировок оценивается по некоторому расстоянию, обычно вводимому аксиоматически. Одним из широко используемых видов таких компромиссных решений является медиана Кемени-Снелла. Другим часто употребляемым методом построения итогового упорядочения служит упорядочение объектов по средним рангам, то есть по среднему арифметическому значению рангов, присвоенных каждому объекту разными экспертами. Как отмечается в работе [10], со статистической точки зрения и медиана Кемени-Снелла, и упорядочение по средним рангам представляют собой упорядочения, наиболее коррелированные в среднем с индивидуальными экспертными предпочтениями. В первом случае корреляции ищутся с использованием в качестве коэффициента ранговой корреляции коэффициента Кендалла, а во втором – коэффициента Спирмена. При упорядочении несравнимых объектов необходимо учитывать дополнительную информацию, например, предпочтения ЛПР [1] или относительную важность критериев [4].

В перечисленных выше подходах построение итогового упорядочения объектов производится либо на основе информации, полученной от одного источника, либо путем согласования или усреднения различных оценок. Однако, если имеются различные источники информации, например, объекты оцениваются несколькими экспертами, которые работают независимо и не знают оценок друг от друга, то получить согласованное мнение экспертов крайне сложно или вообще невозможно. Поэтому необходимы методы упорядочения многопризнаковых объектов, которые позволяли бы одновременно учитывать оценки, в том числе и противоречивые, всех экспертов без поиска компромисса между мнениями отдельных экспертов.

5. Упорядочение многопризнаковых объектов

Дадим формальную постановку задачи упорядочения многопризнаковых объектов. Пусть $A = \{A_1,...,A_k\}$ – совокупность объектов, которые оцениваются n экспертами по m критериям $Q_1,\ldots,Q_m$. Каждый критерий Q_s имеет порядковую шкалу количественных или качественных оценок $\{q_s^{e_s}\}$, $e_s=1,\ldots,h_s$, $s=1,\ldots,m$, которые упорядочены от лучшего значения к худшему $q_s^1 \succ q_s^2 \succ \ldots \succ q_s^{h_s}$. Предполагается, что разные критерии могут иметь различную относительную важность, но значения оценок, относящихся к одному и тому же критерию, равноценны. Будем также считать, что каждый объект оценивается всеми n экспертами по всем m критериям, не существует «главного» эксперта и мнения всех экспертов одинаково важны, экспертные оценки независимы. Можно выделить два объекта (возможно, гипотетических) – абсолютно лучший и абсолютно худший, которым все эксперты дали соответственно наивысшие и наинизшие оценки по всем критериям. Требуется, исходя из многокритериальных оценок объектов, упорядочить объекты от лучшего к худшему.

Представим объект A_i как мультимножество вида (1) над доменом $G=\{Q_1,\ldots,Q_m\}$, являющимся множеством критериальных оценок, где функция кратности $k_{Ai}(q_s^{e_s})$ мультимножества характеризует количество экспертов, давших объекту A_i оценку $q_s^{e_s}$. Наилучшему и наихудшему объектам соответствуют мультимножества

$$\boldsymbol{A}_{\max} = \{n\bullet q_1^{\,1},0,\ldots,0,\ n\bullet q_2^{\,1},0,\ldots,0,\ldots,\ n\bullet q_m^{\,1},0,\ldots,0\}, \quad (3)$$

$$\boldsymbol{A}_{\min} = \{0,\ldots,0,n\bullet q_1^{h_1},\ 0,\ldots,0,n\bullet q_2^{h_2},\ldots,\ 0,\ldots,0,n\bullet q_m^{h_m}\}, \quad (4)$$

и их принято называть идеальным и антиидеальным решениями. В дальнейшем мы не будем делать различия между объектом A_i и представляющим его мультимножеством $\boldsymbol{A}_i$. Задача

упорядочения многопризнаковых объектов сводится, таким образом, к упорядочению мультимножеств. Рассмотрим возможные подходы к ее решению.

Простейший способ сравнения и упорядочения объектов состоит в упорядочении мультимножеств по включению. В этом случае i-ый объект $\boldsymbol{A}_i$ будет лучше j-ого объекта $\boldsymbol{A}_j$ ($\boldsymbol{A}_i \succ \boldsymbol{A}_j$), если для мультимножеств выполняется включение $\boldsymbol{A}_i \supset \boldsymbol{A}_j$, что равносильно условию $k_{Ai}(q_s^{e_s}) > k_{Aj}(q_s^{e_s})$ для всех $q_s^{e_s} \in G$. Однако такая возможность на практике встречается достаточно редко.

Мультимножество $\boldsymbol{A}$ в определенном смысле эквивалентно целочисленному вектору $\mathbf{k}_A=(k_{A1},\ldots,k_{Ah1},\ldots,\ k_{Am1},\ldots,k_{Ahm})$, различные компоненты k_{As} которого являются значениями функции кратности $k_A(q_s^{e_s})$ мультимножества $\boldsymbol{A}$. Используя представление объекта $\boldsymbol{A}$ с помощью вектора $\mathbf{k}_A$, мы возвращаемся к методам группового сравнения и упорядочения многопризнаковых объектов, рассмотренным выше. Важнейшим недостатком этих методов является их малая пригодность для противоречиво описанных объектов, а также трудоемкость процедур сбора и обработки информации об объектах.

Будем теперь считать многопризнаковые объекты точками метрического пространства мультимножеств (A , d), например, с основной метрикой (типа Хемминга), которая задается формулой (2), принимающей вид

$$d_1(\boldsymbol{A},\boldsymbol{B}) = m(\boldsymbol{A}\Delta\boldsymbol{B}) = \sum_{s=1}^{m} w_s \sum_{e_s=1}^{h_s} | k_A(q_s^{e_s}) - k_B(q_s^{e_s}) |, \qquad (5)$$

где $w_s > 0$ – коэффициенты относительной важности критериев Q_s. Будем сравнивать объекты по их близости к идеальному решению $\boldsymbol{A}_{\max}$ и говорить, что объект $\boldsymbol{A}_i$ лучше объекта $\boldsymbol{A}_j$ ($\boldsymbol{A}_i \succ \boldsymbol{A}_j$), если он находится ближе к идеальному решению $\boldsymbol{A}_{\max}$, то есть выполняется условие

$$d_1(\boldsymbol{A}_{\max}, \boldsymbol{A}_i) < d_1(\boldsymbol{A}_{\max}, \boldsymbol{A}_j). \qquad (6)$$

Упорядочим все объекты по величине их расстояния от идеального решения. Если для некоторых объектов $d_1(\boldsymbol{A}_{\max},\boldsymbol{A}_i)=d_1(\boldsymbol{A}_{\max},\boldsymbol{A}_j)$, то объекты $\boldsymbol{A}_i$ и $\boldsymbol{A}_j$ будут или эквивалентными, или несравнимыми. Тем самым полученное ранжирование объектов окажется нестрогим.

Так как каждый объект $\boldsymbol{A}_i$ оценивается n экспертами по всем m критериям, то нетрудно убедиться, что выполняются равенства

$$\sum_{e_s=1}^{h_s} k_{Ai}(q_s^{e_s}) = n, \qquad \sum_{s=1}^{m}\sum_{e_s=1}^{h_s} k_{Ai}(q_s^{e_s}) = m{\cdot}n,$$

Отсюда, в частности, для любого критерия Q_s следует соотношение

$$\sum_{e_s=2}^{h_s} k_{Ai}(q_s^{e_s}) = n - k_{Ai}(q_s{}^1). \qquad (7)$$

Воспользовавшись формулами (3), (5), условием равноценности оценок по каждому критерию и учитывая равенство (7), запишем выражение для расстояния от идеального решения $\boldsymbol{A}_{\max}$ до объекта $\boldsymbol{A}_i$ в виде:

$$d_1(\boldsymbol{A}_{\max}, \boldsymbol{A}_i) = \sum_{s=1}^{m} w_s \sum_{e_s=1}^{h_s} | k_{A\max}(q_s^{e_s}) - k_{Ai}(q_s^{e_s}) | = 2\sum_{s=1}^{m} w_s\,[n - k_{Ai}(q_s{}^1)].$$

Условие (6) сравнения многопризнаковых объектов приобретает тогда следующую форму: объект $\boldsymbol{A}_i$ лучше объекта $\boldsymbol{A}_j$ ($\boldsymbol{A}_i \succ \boldsymbol{A}_j$), если

$$\sum_{s=1}^{m} w_s\, k_{Ai}(q_s{}^1) > \sum_{s=1}^{m} w_s\, k_{Aj}(q_s{}^1). \qquad (8)$$

Таким образом, правило упорядочения многопризнаковых объектов сводится к сравнению взвешенных сумм $S_{Ai}{}^1 = \sum_s w_s\, k_{Ai}(q_s{}^1)$ первых (наилучших) оценок объектов по всем критериям Q_s. Лучшим будет тот объект $\boldsymbol{A}_i$, у которого эта сумма $S_{Ai}{}^1$ будет больше.

Для некоторых объектов $\boldsymbol{A}_{ir}$ вместо неравенств (6) или (8) выполняются равенства

$d_1(\boldsymbol{A}_{\max}, \boldsymbol{A}_{i1})=\ldots=d_1(\boldsymbol{A}_{\max}, \boldsymbol{A}_{it})$, $r=1,\ldots t$. В таком случае получим частичное упорядочение объектов, в котором объекты $\boldsymbol{A}_{i1},\ldots,\boldsymbol{A}_{it}$ «делят» одно и то же место. Чтобы упорядочить эти объекты внутри группы воспользуемся следующим приемом. Подсчитаем для объектов взвешенные суммы $S_{Air}{}^2=\sum_s w_s\, k_{Air}(q_s{}^2)$ вторых оценок по всем критериям, и будем считать, что объект $\boldsymbol{A}_{iu}$ лучше объекта $\boldsymbol{A}_{iv}$, если выполняется условие

$$\sum_{s=1}^{m} w_s\, k_{Aiu}(q_s{}^2) > \sum_{s=1}^{m} w_s\, k_{Aiv}(q_s{}^2). \qquad (9)$$

Если для каких-то объектов $\boldsymbol{A}_{irp}$ и эти суммы окажутся одинаковыми, то упорядочим объекты из этой подгруппы по суммам $S_{Airp}{}^3=\sum_s w_s\, k_{Airp}(q_s{}^3)$ третьих оценок по всем критериям. И так далее, пока не расставим по своим местам все объекты $\boldsymbol{A}_{i1},\ldots,\boldsymbol{A}_{it}$ данной группы и всей совокупности $\boldsymbol{A}=\{\boldsymbol{A}_1,\ldots,\boldsymbol{A}_k\}$ в целом.

Представим рассмотренную процедуру упорядочения совокупности многопризнаковых объектов в виде следующего алгоритма [12].

Шаг 1. Вычислить для каждого объекта $\boldsymbol{A}_i$ из совокупности $\boldsymbol{A}=\{\boldsymbol{A}_1,\ldots,\boldsymbol{A}_k\}$ взвешенную сумму $S_{Ai}{}^1=\sum_s w_s\, k_{Ai}(q_s{}^1)$ всех первых (наилучших) оценок по всем критериям Q_s и упорядочить объекты от лучшего к худшему по величинам $S_{Ai}{}^1$ сумм первых оценок. Если найдутся группы эквивалентных или несравнимых объектов $\boldsymbol{A}_{i1},\ldots,\boldsymbol{A}_{it}$, имеющих одинаковые суммы $S_{Ai}{}^1$, перейти к шагу 2.

Шаг 2. Вычислить для каждого объекта $\boldsymbol{A}_{ir}$, $r=1,\ldots t$ в соответствующей группе взвешенную сумму $S_{Air}{}^2=\sum_s w_s\, k_{Air}(q_s{}^2)$ всех вторых оценок по всем критериям Q_s и упорядочить объекты внутри каждой группы от лучшего к худшему по величинам $S_{Air}{}^2$ сумм вторых оценок. Если останутся подгруппы эквивалентных или несравнимых объектов $\boldsymbol{A}_{iru},\ldots,\boldsymbol{A}_{irv}$, имеющих одинаковые суммы $S_{Air}{}^2$, перейти к шагу 3.

Шаг 3. Вычислить для каждого объекта $\boldsymbol{A}_{irp}$ в соответствующей подгруппе взвешенную сумму $S_{Airp}{}^3=\sum_s w_s\, k_{Airp}(q_s{}^3)$ всех третьих оценок по всем критериям Q_s и упорядочить объекты внутри каждой подгруппы от лучшего к худшему по величинам сумм $S_{Airp}{}^3$ третьих оценок. Продолжить процедуру до полного упорядочения всех объектов из совокупности $\boldsymbol{A}=\{\boldsymbol{A}_1,\ldots,\boldsymbol{A}_k\}$. Если число h_s значений оценок $q_s^{e_s}$ у некоторых критериев Q_s окажется меньше требуемого на данном b-ом шаге алгоритма, то следует считать $k_{Air\ldots p}(q_s{}^b)=0$. ◙

В приведенном выше алгоритме предполагалась различная относительная важность критериев Q_s, выражаемая коэффициентами $w_s>0$, на которые могут накладываться некоторые условия, например, $\sum_s w_s=1$. Проблема определения важности критериев имеет самостоятельное значение и в контексте данной работы не рассматривается. В случае, когда все критерии одинаково важны, все коэффициенты w_s считаются равными 1.

Аналогичным образом можно построить процедуру упорядочения многопризнаковых объектов $\boldsymbol{A}_i$ по отношению к антиидеальному решению $\boldsymbol{A}_{\min}$, заданному выражением (4), считая, что объект $\boldsymbol{A}_i$ лучше объекта $\boldsymbol{A}_j$ ($\boldsymbol{A}_i \succ \boldsymbol{A}_j$), если он находится дальше от антиидеального решения $\boldsymbol{A}_{\min}$, то есть $d_1(\boldsymbol{A}_{\min}, \boldsymbol{A}_i)>d_1(\boldsymbol{A}_{\min}, \boldsymbol{A}_j)$. Как и выше, объекты $\boldsymbol{A}_i$ и $\boldsymbol{A}_j$ будут эквивалентными или несравнимыми, если $d_1(\boldsymbol{A}_{\min},\boldsymbol{A}_i)=d_1(\boldsymbol{A}_{\min},\boldsymbol{A}_j)$. Подчеркнем, что упорядочение совокупности многопризнаковых объектов $\boldsymbol{A}=\{\boldsymbol{A}_1,\ldots,\boldsymbol{A}_k\}$ по отношению к антиидеальному решению может не совпадать с упорядочением по отношению к идеальному решению.

Данный метод упорядочения многопризнаковых объектов был применен для построения рейтинга российских компаний, работающих в секторе информационно-коммуникационных технологий [13]. Экспертная оценка деятельности компаний давалась по специально разработанным критериям с качественными оценками, аналогичным указанным выше, а результаты обрабатывались по описанной процедуре. Всего было оценено около 50 компаний, из которых были выделены 30 ведущих высокотехнологичных компаний, а также составлены рейтинги 10 ведущих разработчиков программного обеспечения и 10 наиболее динамично развивающихся компаний.

6. Классификация многопризнаковых объектов

Рассмотрим еще одну практическую задачу, в которой, исходя из некоторой предварительной сортировки совокупности многопризнаковых объектов, требуется распределить эти объекты по нескольким классам. Допустим, что для решения какой-либо важной проблемы (научно-технической, экономической, производственной, экологической) необходимо сформировать программу, которая будет состоять из отдельных работ, проектов, заданий и тому подобное, отобранных на конкурсной основе. Каждая представленная на конкурс заявка оценивается несколькими экспертами по специально разработанным качественным критериям. Основываясь на заключениях экспертов, орган, ответственный за формирование программы, принимает решение о включении того или иного проекта в программу.

Например, при формировании государственной научно-технической программы по высокотемпературной сверхпроводимости [14] экспертная оценка и конкурсный отбор проектов проводился по следующим качественным критериям: Q_1. Важность проекта для программы; Q_2. Перспективность проекта; Q_3. Новизна подхода к решению поставленных задач; Q_4. Квалификация исполнителей проекта; Q_5. Ресурсное обеспечение работ; Q_6. Возможность быстрого выхода результатов в практику.

Каждый критерий имел порядковую или номинальную шкалу оценок с развернутыми словесными формулировками градаций качества. Так, шкала критерия Q_4. «Квалификация исполнителей проекта» имела вид:

q_4^1 – по опыту и квалификации исполнители проекта являются одним из лучших научных коллективов;

q_4^2 – опыт и квалификация исполнители находятся на уровне, достаточном для проведения работ;

q_4^3 – исполнители не обладают необходимыми опытом и квалификацией;

q_4^4 – опыт и квалификация исполнителей неизвестны.

Шкала оценок по критерию Q_6. «Возможность быстрого выхода результатов в практику» выглядела следующим образом:

q_6^1 – результаты будут обладать достаточной степенью технологичности, обеспечивающей их быстрое использования в практике;

q_6^2 – для использования запланированных результатов на практике потребуются дополнительные исследования и разработки;

q_6^3 – результаты будут носить в основном теоретический характер.

Экспертиза заявок осуществлялась экспертами независимо друг от друга без согласования их мнений. Каждый эксперт, наряду с оценкой заявки по всем критериям, давал одну из следующих рекомендаций:

r_1 – включить проект в программу;

r_2 – отклонить проект;

r_3 – отложить рассмотрение заявки и отправить проект на доработку.

Указанные рекомендации экспертов являются, по существу, правилами предварительной классификации (сортировки) рассматриваемых заявок. В других задачах критерии оценки объектов и правила их сортировки могут быть и иными.

Если бы заявка оценивалась только одним экспертом, то найти на множестве многокритериальных оценок обобщенное решающее правило для отбора предложений не составило бы особого труда. Известно большое число разных подходов к решению подобного рода задач классификации, например, [1], [2], [10], [15]. Однако когда заявка рассматривается несколькими экспертами, то появляется несколько различных вариантов («экземпляров») одной и той же заявки, поскольку и многокритериальные экспертные оценки, и заключения экспертов могут быть как схожими, так и противоречивыми. В силу качественного характера

экспертных данных их агрегирование тем или иным способом представляет самостоятельную, достаточно сложную проблему. Помимо этого, вырабатывая решение о включении заявки в программу, необходимо учесть все, даже и не совпадающие заключения экспертов по принятию или отклонению заявки. Желательно поэтому иметь некое единое решающее правило для отнесения заявки к какому-либо классу, которое, во-первых, базировалось бы на характеристиках заявок, выраженных их многокритериальными оценками, а во-вторых, в наибольшей степени соответствовало бы индивидуальным экспертным правилам сортировки. Прежде, чем переходить к изложению путей решения этой задачи, напомним некоторые общие положения.

Наиболее общим определением класса является следующее: класс – это совокупность (семейство) объектов, обладающих общими свойствами. Информация о свойствах объекта может быть получена путем наблюдений, измерений, оценок и тому подобное и представлена совокупностью признаков, значения которых выражаются в числовых и/или вербальных шкалах. Входящие в один и тот же класс объекты считаются неразличимыми (эквивалентными), а каждый класс объектов характеризуется некоторым качеством, отличающим его от других классов. Все классы вместе должны составлять исходную совокупность объектов.

Свойство сходства и различимости объектов, относящихся к одному и тому же классу, широко используется при построении различных методов классификации. Так, например, в ряде методов сортировки объектов, основанных на теориях нечетких [16], [17] и грубых [18] множеств, допускается неоднозначность классификации объектов, связанная с разной степенью принадлежности объекта к классу, то есть объекты, которые «несомненно» и «возможно» принадлежат к некоторому классу, считаются различающимися.

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

$$\text{ЕСЛИ } \langle\text{условия}\rangle, \text{ ТО } \langle\text{решение}\rangle. \qquad (10)$$

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

При достаточно небольшом числе классифицируемых объектов и признаков, их описывающих, семейство решающих правил легко обозримо и доступно для анализа. Чем больше количество рассматриваемых объектов и разнообразнее решающее правила их классификации, тем труднее становится анализ этих правил. Могут существовать различные причины, обусловливающие неоднозначность классификации, к примеру, если объекты сортируются разными экспертами. Эксперты могут относить сильно различающиеся объекты в один и тот же класс, а объекты со сходными значениями признаков – в разные классы. Несогласованность индивидуальных решающих правил может быть вызвана неоднозначностью понимания экспертами решаемой задачи, ошибками или неточностями, допущенными экспертами при первоначальной классификации объектов, субъективным различием решающих правил, используемых разными экспертами, специфичностью знаний самих экспертов, нетранзитивностью отдельных экспертных суждений и многими другими причинами. В итоге может появиться семейство решающих правил, среди которых будут одинаковые, сходные, различающиеся и противоречивые правила.

В этом случае возникает проблема: построить такое обобщенное решающее правило или небольшую группу правил, которые наилучшим (в некотором смысле) образом аппроксимируют совокупность всех индивидуальных правил сортировки объектов, включают минимальный набор признаков и относят объекты в заданные классы с допустимой точностью.

7. Аппроксимация индивидуальных правил сортировки

Перейдем к формальной постановке задачи аппроксимации большого числа правил сортировки многопризнаковых объектов компактным набором простых решающих правил. Пусть $\mathbf{A} = \{A_1,...,A_k\}$ – совокупность объектов, которые описываются m дискретными признаками $Q_1,\ldots,Q_m$, имеющими качественные значения. Каждая группа признаков $Q_s=\{q_s^{e_s}\}$, $e_s=1,\ldots,h_s$, $s=1,\ldots,m$ отражает содержательное качество объектов, например, $q_s^{e_s}$ может быть значением показателя, характеризующего какое-либо свойство объекта, или оценкой объекта по критерию, и тому подобное. Объекты A_i, $i=1,\ldots,k$ предварительно рассортированы по нескольким классам X_t, $t=1,\ldots,f$ путем прямой классификации. Принадлежность объекта A_i к некоторому классу X_t выражается правилом сортировки R, которое может считаться еще одним качественным признаком со шкалой значений $R=\{r_t\}$. Любой объект A_i может существовать в n экземплярах, которые отличаются наборами признаков, его характеризующих. Однако в описании каждого экземпляра объекта присутствует только одно какое-то значение признака из каждой группы $Q_1,\ldots,Q_m,R$. Других дополнительных предположений об особенностях классов, признаков объектов и их значений (важности, предпочтительности, характерности, упорядоченности и прочее) не делается. Требуется построить одно или несколько решающих правил, составленных из небольшого числа значений признаков, которые относили бы объекты к заданным классам наилучшим (в смысле близости к предварительной сортировке) образом. Само понятие близости также должно быть определено.

Сопоставим каждый многопризнаковый объект с мультимножеством вида, аналогичного выражению (1)

$$\boldsymbol{A}_i = \{(k_{Ai}(q_s^{e_s})\bullet q_s^{e_s}), (k_{Ai}(r_t)\bullet r_t)\} \qquad (11)$$

над доменом $G=\{Q_1,\ldots,Q_m,R\}$. Запись объекта $\boldsymbol{A}_i$ в таком виде может трактоваться как еще один способ выражения индивидуальных правил сортировки (10). А именно: терм ⟨условия⟩ ассоциируется тогда с различными комбинациями значений признаков $q_s^{e_s}$, описывающими свойства объекта $\boldsymbol{A}_i$, а терм ⟨решение⟩ – с принадлежностью объекта $\boldsymbol{A}_i$ к классу $\boldsymbol{X}_t$. В терм ⟨решение⟩ входит также некоторое правило, позволяющее говорить о принадлежности объекта $\boldsymbol{A}_i$ к какому-то определенному классу $\boldsymbol{X}_t$. Это может быть, например, правило простого большинства голосов, в соответствии с которым объект $\boldsymbol{A}_i$ будет считаться принадлежащим к классу $\boldsymbol{X}_t$, если $k_{Ai}(r_t)>k_{Ai}(r_p)$ для всех $p\neq t$, или правило квалифицированного большинства голосов, по которому должно выполняться условие $k_{Ai}(r_t)>\sum_{p\neq t} k_{Ai}(r_p)$, или любое другое правило. При этом предполагается, что каждый объект оценивается всеми n экспертами.

Вся совокупность объектов $\mathbf{A} = \{\boldsymbol{A}_1,...,\boldsymbol{A}_k\}$, представленных мультимножествами (11), порождает семейство первичных решающих правил сортировки. Правила совпадают или являются похожими, когда различные объекты с идентичными или схожими (близкими) значениями признаков включаются в один класс. Противоречивые правила относят слабо различимые объекты в разные классы.

Для простоты будем считать, что результатом классификации должно быть разложение совокупности объектов $\mathbf{A}$ только на два класса $\boldsymbol{X}_a$ и $\boldsymbol{X}_b$. Требование бинарной декомпозиции $\mathbf{A} = \{\boldsymbol{X}_a, \boldsymbol{X}_b\}$ не является принципиальным ограничением. Если необходимо 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. Например, заявки можно разделить на принятые и отклоненные, отклоненные заявки – на отложенные для дальнейшей доработки и окончательно не принятые, и так далее.

Рассмотрим наиболее простой и типичный случай, когда все группы объектов формируются как суммы соответствующих им мультимножеств. Тогда каждое из мультимножеств

$\boldsymbol{X}_t$, $t=a,b$, представляющее свой класс объектов, можно записать в виде следующего разложения на мультимножества по группам признаков:

$$\boldsymbol{X}_t = \sum_{s=1}^{m} \boldsymbol{Q}_{st} + \boldsymbol{R}_t \,, \tag{12}$$

где каждое слагаемое есть в свою очередь разложение

$$\boldsymbol{Q}_{st} = \sum_{e_s=1}^{h_s} \boldsymbol{Q}_{st}^{e_s}, \qquad \boldsymbol{Q}_{st}^{e_s} = \sum_{i \in I_{st}^{e_s}} \boldsymbol{A}_i, \qquad \boldsymbol{R}_t = \sum_{i \in I_{rt}} \boldsymbol{A}_i,$$

подмножества индексов $I_{st}^{e_s} = I_s^{e_s} \bigcap I_t$ и $I_{rt}=I_r \bigcap I_t$; I_t – подмножество индексов i для объектов $\boldsymbol{A}_i$, имеющих функции кратности $k_{Ai}(r_t) > \sum_{p \neq t} k_{Ai}(r_p)$ или $k_{Ai}(r_t) > k_{Ai}(r_p), p \neq t$; $I_s^{e_s}$ – подмножество индексов i для объектов $\boldsymbol{A}_i$, имеющих $k_{Ai}(q_s^{e_s}) \neq 0$, $k_{Ai}(q_v^{e_v}) = 0$, $v \neq s$, $k_{Ai}(r_t) = 0$; I_r – подмножество индексов i для объектов $\boldsymbol{A}_i$, имеющих $k_{Ai}(r_t) \neq 0$, $k_{Ai}(q_s^{e_s}) = 0$. Так как каждый экземпляр объекта $\boldsymbol{A}_i$ может обладать только единственными значениями $k_{Ai}(q_s^{e_s})$ и $k_{Ai}(r_t)$ из каждой группы признаков Q_s и R, то выполняются следующие условия для мощностей мультимножеств:

$$|\boldsymbol{Q}_{sa}| + |\boldsymbol{Q}_{sb}| = k, \quad |\boldsymbol{R}_a| + |\boldsymbol{R}_b| = k, \quad |\boldsymbol{X}_a| + |\boldsymbol{X}_b| = k \cdot (m+1),$$

где, напомним, k равно числу объектов, а m – числу групп признаков.

Очевидно, что объекты $\boldsymbol{A}_i$, которые попали в разложение $\{\boldsymbol{R}_a, \boldsymbol{R}_b\}$, сделанное только по правилам сортировки, образуют наилучшую из всех возможных декомпозиций рассматриваемой совокупности объектов $\mathbf{A} = \{\boldsymbol{A}_1, ..., \boldsymbol{A}_k\}$ на два класса для имеющегося набора первичных правил сортировки. Обозначим через $d^* = d(\boldsymbol{R}_a, \boldsymbol{R}_b)$ расстояние между мультимножествами $\boldsymbol{R}_a$ и $\boldsymbol{R}_b$ в метрическом пространстве мультимножеств $(\mathbf{A}, d)$ с метрикой d, определяемой одним из выражений (2) или (5). В каждой конкретной задаче классификации расстояние d^* является предельно возможным расстоянием между объектами, входящими в разные классы. При идеальной предварительной сортировке объектов противоречия в индивидуальных правилах отсутствуют. В этом случае максимально возможное расстояние в метрическом пространстве мультимножеств $(\mathbf{A}, d)$, на котором могут находиться объекты, принадлежащие разным классам, будет равно соответственно $d_1^* = kn$, $d_2^* = 1/h$, $d_3^* = 1$. Здесь n есть число индивидуальных правил сортировки, приходящихся на один объект, совпадающее, в частности, с числом экспертов, h – общее число значений всех признаков, описывающих объекты, равное для задачи классификации $h = h_1 + ... + h_m + f$.

Сформулируем теперь основную идею нахождения обобщенного решающего правила, аппроксимирующего большое семейство противоречивых правил сортировки многопризнаковых объектов. Для каждой группы признаков Q_s нужно сгенерировать пары новых мультимножеств таким образом, чтобы мультимножества внутри каждой пары были удалены друг от друга в метрическом пространстве $(\mathbf{A}, d)$ как можно больше и с достаточной точностью совпадали с первоначальной сортировкой объектов по классам $\boldsymbol{X}_a$ и $\boldsymbol{X}_b$, заданной разложением $\{\boldsymbol{R}_a, \boldsymbol{R}_b\}$. Разные комбинации признаков, определяющих границы между сгенерированными мультимножествами внутри каждой пары, дадут желаемые обобщенные решающие правила для классификации объектов.

Решение задачи аппроксимации решающих правил для классификации многопризнаковых объектов сводится, таким образом, к решению m оптимизационных задач вида

$$d(\boldsymbol{Q}_{sa}, \boldsymbol{Q}_{sb}) \to \max \; d(\boldsymbol{Q}_{sa}, \boldsymbol{Q}_{sb}) = d(\boldsymbol{Q}_{sa}^*, \boldsymbol{Q}_{sb}^*), \tag{13}$$

где мультимножества $\boldsymbol{Q}_{sa}^*$ и $\boldsymbol{Q}_{sb}^*$ принадлежат к разным классам и находятся на максимально возможном расстоянии в метрическом пространстве мультимножеств $(\mathbf{A}, d)$. Решение каждой из задач (13) является наилучшей бинарной декомпозиций $\{\boldsymbol{Q}_{sa}^*, \boldsymbol{Q}_{sb}^*\}$ имеющейся совокупности многопризнаковых объектов $\mathbf{A} = \{\boldsymbol{A}_1, ..., \boldsymbol{A}_k\}$ по s-ой группе признаков. Когда

число h_s значений $q_s^{e_s}$ каждого из признаков невелико (h_s=2÷5), решение задачи (13) не вызывает существенных трудностей и может быть получено даже путем простого перебора.

Каждое мультимножество $\boldsymbol{Q}_{st}^*$ (t=a,b), относящееся к одному и тому же классу, представляет собой сумму двух подмультимножеств $\boldsymbol{Q}_{st}^* = \boldsymbol{Q}_{st}^{*1} + \boldsymbol{Q}_{st}^{*2}$. Значение признака q_s^*, которое определяет границу между слагаемыми $\boldsymbol{Q}_{st}^{*1}$ и $\boldsymbol{Q}_{st}^{*2}$, назовем аппроксимирующим признаком. Комбинации аппроксимирующих признаков $\{q_s^*\}$ для разных номеров s групп признаков Q_s задают условия отнесения объекта $\boldsymbol{A}_i$ к соответствующему классу $\boldsymbol{X}_t$ и образуют в совокупности искомые обобщенные правила классификации объектов вида (11).

Аппроксимирующие признаки q_s^* для различных групп признаков можно упорядочить по величине расстояния $d(\boldsymbol{Q}_{sa}^*, \boldsymbol{Q}_{sb}^*)$. Для построения обобщенных правил классификации следует использовать признаки q_s^*, занимающие первые места в такой ранжировке. Чем ближе значения расстояний $d(\boldsymbol{Q}_{sa}^*, \boldsymbol{Q}_{sb}^*)$ к расстоянию $d^*=d(\boldsymbol{R}_a, \boldsymbol{R}_b)$, тем более точной будет аппроксимация первоначальной индивидуальной сортировки объектов. Оценить точность аппроксимации по s-ой группе признаков можно, например, выражением

$$\rho_s = d(\boldsymbol{Q}_{sa}^*, \boldsymbol{Q}_{sb}^*)/d(\boldsymbol{R}_a, \boldsymbol{R}_b), \qquad (14)$$

В обобщенное решающее правило должны тогда включаться аппроксимирующие признаки q_s^*, имеющие показатель точности ρ_s, превышающей некоторый желаемый пороговый уровень ρ_0. Заметим, что величина ρ_s показателя точности аппроксимации характеризует в определенном смысле относительную важность s-ой группы признаков Q_s в обобщенном правиле классификации.

8. Конкурсный отбор проектов

Соотношения между совокупностью объектов $\boldsymbol{A} = \{A_1,...,A_k\}$ и множеством их признаков $G=\{x_1,...,x_h\}$ удобно выражать с помощью матрицы $C=\|c_{ij}\|_{k\times h}$, которая часто используется в анализе данных, теории принятия решений, распознавании образов, других приложениях и называется таблицей «объекты-признаки», информационной таблицей или таблицей решений [2], [18]. Строки этой матрицы соответствуют объектам, столбцы – признакам, а элементы матрицы являются значениями признаков. Таким образом, каждая строка матрицы C характеризует свойства рассматриваемого объекта, а каждый столбец дает информацию об объектах, обладающих данным свойством. Свойства совокупности $\boldsymbol{A} = \{\boldsymbol{A}_1,...,\boldsymbol{A}_k\}$ многопризнаковых объектов $\boldsymbol{A}_i$, представленных мультимножествами, и их принадлежность к некоторому классу решений $\boldsymbol{X}_t$ также можно описать с помощью таблиц решений. В исходной таблице решений $C=\|c_{ij}\|_{k\times h}$, элементы которой задаются как $c_{ij} = k_{Ai}(x_j)$, $x_j = q_s^{e_s}, r_t$, каждая строка является аргументом выражения (11). Разложению совокупности объектов $\boldsymbol{A}$ на два класса $\boldsymbol{X}_a$ и $\boldsymbol{X}_b$, которые задаются формулой (12), соответствует преобразованная таблица решений $C'=\|k_{Xt}'(x_j)\|_{2\times h}$, состоящая из двух строк $k_{Xa}'(x_j)$ и $k_{Xb}'(x_j)$. Матрицы C и C' состоят из $2(m+1)$ блоков, которые соответствуют мультимножествам признаков $\boldsymbol{Q}_{sa}$, $\boldsymbol{Q}_{sb}$ и решений $\boldsymbol{R}_a$, $\boldsymbol{R}_b$.

Процедура построения обобщенного решающего правила для классификации многопризнаковых объектов включает следующие основные этапы.

Шаг 1. Построить таблицу решений $C=\|k_{Ai}(x_j)\|_{k\times h}$ для рассматриваемой совокупности многопризнаковых объектов $\boldsymbol{A} = \{\boldsymbol{A}_1,...,\boldsymbol{A}_k\}$, строки которой соответствуют мультимножествам $\boldsymbol{A}_i$ вида (11).

Шаг 2. Объединить объекты $\boldsymbol{A}_i$, относящиеся к заданным классам $\boldsymbol{X}_a$ и $\boldsymbol{X}_b$, воспользовавшись формулами (12). Получить преобразованную таблицу решений $C'=\|k_{Xt}'(x_j)\|_{2\times h}$, строки которой соответствуют мультимножествам $\boldsymbol{X}_a$ и $\boldsymbol{X}_b$.

Шаг 3. Решить задачу оптимизации (13) для каждого бинарного разложения $\{\boldsymbol{Q}_{sa}^*, \boldsymbol{Q}_{sb}^*\}$ по s-ой группе признаков Q_s и найти аппроксимирующий признак q_s^* в каждом s-ом блоке преобразованной матрицы C'.

Шаг 4. Проранжировать аппроксимирующие признаки q_s^* по убыванию величины расстояния $d^*=d(\boldsymbol{R}_a, \boldsymbol{R}_b)$ или показателя точности ρ_s (14).

Шаг 5. Выбрать аппроксимирующие признаки q_s^*, которые обеспечивают необходимую точность аппроксимации $\rho_s \geq \rho_0$, и сформировать из них обобщенное решающее правило для классификации многопризнаковых объектов. ◙

Проиллюстрируем предложенный подход к построению обобщенного решающего правила для классификации многопризнаковых объектов, которое аппроксимирует большое число противоречивых правил сортировки, на примере конкурсного отбора проектов для формирования государственной научно-технической программы по высокотемпературной сверхпроводимости [14]. Каждая представленная на конкурс заявка независимо оценивалась 3 экспертами по 6 качественным критериям, которые давали также свое заключение по принятию или отклонению заявки. Всего было подано более 250 заявок и около 170 из них было отобрано для включения в программу.

Приведем некоторые данные, иллюстрирующие рассматриваемый пример: часть решающей таблицы $C=\|k_{Ai}(x_j)\|_{k\times h}$, характеризующей поданные на конкурс проекты $\boldsymbol{A}_i$; преобразованная решающая таблица $C'=\|k_{Xt}'(x_j)\|_{2\times h}$, соответствующая классам принятых $\boldsymbol{X}_a$ и отклоненных $\boldsymbol{X}_b$ проектов; значения расстояний между бинарными разложениями $d_1(\boldsymbol{Q}_{sa}^*, \boldsymbol{Q}_{sb}^*)$ и $d_1(\boldsymbol{R}_a, \boldsymbol{R}_b)$ в пространстве мультимножеств (A, d_1) с метрикой (5) при $w_s=1$; значения показателей точности ρ_s для аппроксимирующих признаков q_s^* по каждому s-ому блоку матрицы.

Объекты — Признаки

	$q_1^1\ q_1^2\ q_1^3$	$q_2^1\ q_2^2\ q_2^3$	$q_3^1\ q_3^2\ q_3^3$	$q_4^1\ q_4^2\ q_4^3\ q_4^4$	$q_5^1\ q_5^2\ q_5^3\ q_5^4$	$q_6^1\ q_6^2\ q_6^3$	$r_a\ r_b$
A_1	1 2 0	2 1 0	3 0 0	2 1 0 0	0 2 1 0	2 1 0	3 0
…							
A_i	1 1 1	0 2 1	1 2 0	0 2 1 0	0 1 2 0	0 0 3	2 1
A_{i+1}	1 1 1	0 2 1	1 2 0	0 2 1 0	0 1 2 0	0 0 3	1 2
…							
A_k	0 2 1	0 1 2	0 3 0	0 1 1 1	0 0 2 1	0 3 0	0 3

Классы — Признаки

	$q_1^1\ q_1^2\ q_1^3$	$q_2^1\ q_2^2\ q_2^3$	$q_3^1\ q_3^2\ q_3^3$	$q_4^1\ q_4^2\ q_4^3\ q_4^4$	$q_5^1\ q_5^2\ q_5^3\ q_5^4$	$q_6^1\ q_6^2\ q_6^3$	$r_a\ r_b$
X_a	144 360 21	81 324 120	99 336 90	219 297 9 0	72 435 18 0	126 300 99	510 15
X_b	45 156 51	27 93 132	36 111 105	51 132 63 6	60 147 30 15	45 135 72	78 174

d_1	333	297	303	393	327	273	591
ρ_s	0,563	0,503	0,517	0,665	0,553	0,462	

Принятые проекты $\boldsymbol{A}_1$–$\boldsymbol{A}_i$ входят в класс $\boldsymbol{X}_a$, отклоненные проекты $\boldsymbol{A}_{i+1}$–$\boldsymbol{A}_k$ относятся к классу $\boldsymbol{X}_b$. Обратим внимание читателя, что хотя проекты $\boldsymbol{A}_i$ и $\boldsymbol{A}_{i+1}$ имеют одинаковые значения оценок $\{q_s\}$ по всем признакам, но наборы их индивидуальных правил сортировки не совпадают, и поэтому $\boldsymbol{A}_i \in \boldsymbol{X}_a$, а $\boldsymbol{A}_{i+1} \in \boldsymbol{X}_b$. Множество аппроксимирующих признаков q_s^*, упорядоченное по величине расстояния $d_1(\boldsymbol{Q}_{sa}^*, \boldsymbol{Q}_{sb}^*)$, выглядит следующим образом:

$$\{q_s^*\} = \{q_4^1, q_4^2; q_1^1, q_1^2; q_5^1, q_5^2; q_3^1, q_3^2; q_2^1, q_2^2\}. \qquad (15)$$

Заметим, что задача (13) не имеет оптимального решения по критерию Q_6, то есть любое значение признака q_6 является неаппроксимирующим. Выбрав некоторое желаемое значение точности аппроксимации ρ_0, получим следующие обобщенные решающие правила для отбора проектов.

«Исполнители проекта должны быть одними из лучших или обладать опытом и квалификацией, достаточными для проведения работ» (оценки q_4^1 или q_4^2; точность аппроксимации $\rho_s \geq 0{,}66$).

«Проект должен быть крайне важным или важным для достижения одной из основных целей программы; исполнители проекта должны быть одними из лучших или обладать опытом, квалификацией и материально-техническими ресурсами, достаточными для проведения работ» (оценки q_4^1 или q_4^2; и q_1^1 или q_1^2; и q_5^1 или q_5^2; точность аппроксимации $\rho_s \geq 0{,}55$).

Отметим, что последнее правило полностью совпадает с решающим правилом для отбора проектов, приведенным ранее в работе [14]. Обобщенное решающее правило классификации объектов позволяет также выявить расхождения в индивидуальных правилах сортировки, применявшихся экспертами, и при необходимости скорректировать их. Ранжирование (15) аппроксимирующих признаков по величине расстояния d_1 показывает, что наиболее важным для отбора проектов оказывается критерий Q_4, характеризующий опыт и квалификацию исполнителей, а следующими по важности – критерии Q_1, оценивающий важность проекта для достижения целей программы, и Q_5, отражающий ресурсное обеспечение работ.

9. Заключение

Проблемы классификации и упорядочения объектов, которые описываются многими количественными и качественными признаками, причем каждый из объектов может существовать в нескольких различающихся, но равноправных «экземплярах», являются достаточно трудными. Эти трудности имеют и содержательные основания (например, некорректность применения процедур «усреднения» качественных признаков), и формальные причины (например, противоречивость данных, большая размерность задачи). Главные из перечисленных трудностей оказалось возможным преодолеть благодаря использованию нового теоретического инструментария, основанного на понятии мультимножества. Применение теории мультимножеств позволяет разрабатывать новые методы анализа данных и решения новых классов задач, которые не содержат необоснованных преобразований исходной информации и не приводят к потере или искажению данных.

Литература

[1]. О.И.Ларичев, Е.М.Мошкович. Качественные методы принятия решений. Вербальный анализ решений. – М.: Наука, Физматлит, 1996.

[2]. Б.Г.Миркин. Анализ качественных признаков и структур. – М.: Статистика, 1980.

[3] Л.Г.Евланов. Теория и практика принятия решений. – М.: Экономика, 1984.

[4]. В.Д.Ногин. Принятие решений в многокритериальной среде: количественный подход. – М.: Физматлит, 2002.

[5]. А.Б.Петровский. Метрические пространства мультимножеств.//Доклады Академии наук, 1995, Т.344, №2, 175-177.

[6]. А.Б.Петровский. Основные понятия теории мультимножеств. – М.: Едиториал УРСС, 2002.

[7]. Ю.И.Журавлев. Корректные алгебры над множествами некорректных (эвристических) алгоритмов.I,II,III.//Кибернетика, 1977, №4, 14-21; 1977, №6, 21-27; 1978, №2, 35-43.

[8]. Ю.Н.Тюрин. Экспертная классификация.//Экспертные методы в современных исследованиях. Сборник трудов. – М.: ВНИИСИ, 1979, 5-15.

[9]. J.G.Kemeni, J.L.Snell. Mathematical models in the social sciences. – Ginn, Boston, 1962. (Дж.Кемени, Дж.Снелл. Кибернетическое моделирование./Пер. с англ. – М.: Советское радио, 1972).

[10]. Б.Г.Литвак. Экспертная информация: методы получения и анализа. – М.: Радио и связь, 1982.
[11]. B.Roy. Multicriteria methodology for decision aiding. – Kluwer Academic Publishers, Dordrecht, 1996.
[12]. А.В.Литвинова. Упорядочивание многопризнаковых объектов на основе теории мультимножеств.//Дипломная работа на соискание степени магистра. Московский физико-технический институт (государственный университет), М., 2002.
[13]. Кто в России самый интеллектуальный? Рейтинг ведущих российских разработчиков высоких технологий.//Компания, 2000, №47(143), 38-39.
[14]. О.И.Ларичев, А.С.Прохоров, А.Б.Петровский, М.Ю.Стернин, Г.И.Шепелев. Опыт планирования фундаментальных исследований на конкурсной основе.//Вестник АН СССР, 1989, №7, 51-61.
[15]. А.А.Дорофеюк. Алгоритмы автоматической классификации.//Автоматика и телемеханика, 1971, №12, 78-113.
[16]. С.А.Орловский. Проблемы принятия решений при нечеткой исходной информации. – М.: Наука, 1981.
[17]. H.J.Zimmerman, L.A.Zadeh, B.R.Gaines. Fuzzy sets and decision analysis. – North-Holland, Amsterdam, 1984.
[18]. Z.Pawlak, R.Slowinsky. Rough set approach to multi-attribute decision analysis.//European Journal of Operational Research, 1994, №72, 443-459.

ODERING AND CLASSIFYING OBJECTS WITH CONTRADICTORY ATTRIBUTES

Alexey B.Petrovsky

Multiset or set with repeating elements is a convenient mathematical model for a representation of objects, which are characterized by many diverse (quantitative and qualitative) attributes, and can exist in several copies with different, in particular, contradictory values of attributes. New methods for ordering and classifying multi-attribute objects that is based on the theory of multiset metric spaces are considered in the paper. These methods are applied for solving case studies: ranking companies and competitive selection of projects, which are estimated by several experts by many qualitative criteria.

Петровский А. Б. Упорядочение и классификация объектов с противоречивыми признаками // *Новости искусственного интеллекта*.— 2003.— № 4. — С. 34–43.

```
@Article{Petrovsky_2003b,
  author =       "Петровский, А. Б.",
  title =        "Упорядочение и классификация объектов с
                  противоречивыми признаками",
  journal =      "Новости искусственного интеллекта",
  number =       "4",
  pages =        "34--43",
  year =         "2003",
  language =     "russian",
}
```