А.Б. Петровский

Методы групповой классификации многопризнаковых объектов (часть 2)[1]

Аннотация. Рассматривается новый подход к групповой классификации объектов, которые описываются многими качественными и/или количественными признаками и могут присутствовать в нескольких версиях. Обсуждаются новые способы агрегирования многопризнаковых объектов и свойства построенных классов. Методы кластеризации и сортировки объектов основаны на теории метрических пространств мультимножеств.

Ключевые слова: групповая классификация, кластеризация, сортировка, многопризнаковые объекты, метрические пространства мультимножеств.

Введение

Объединение объектов (альтернатив, вариантов, действий, ситуаций, персон и т.п.) в классы (кластеры, категории, группы) является одним из наиболее популярных средств для обнаружения, извлечения, формализации и фиксации знаний. Свойства объектов обычно задаются множеством характеристик или признаков, значения которых могут быть количественными (числовыми) и/или качественными (символьными, вербальными). Класс включает объекты, обладающие общими особенностями. Классы могут быть определены заранее или появляться и конструироваться в процессе классификации. Группирование объектов является полезным приемом структуризации их совокупности. Классифицируя объекты, мы относим каждый объект к некоторому классу и используем информацию о принадлежности объектов к классам для выработки и коррекции правил классификации.

В первой части статьи [30] были рассмотрены различные способы представления и агрегирования объектов, описываемых многими признаками; введены основные понятия теории мультимножеств и метрических пространств мультимножеств, которые использованы для задания групп объектов, существующих в нескольких версиях (экземплярах). Описаны обобщенные схемы алгоритмов иерархической кластеризации, в которой число генерируемых кластеров заранее неизвестно, и неиерархической кластеризации, где число кластеров считается фиксированным и установлено заранее.

Для удобства читателя в обеих частях статьи принята единая нумерация разделов, формул и ссылок на литературу.

4. Метод групповой сортировки многопризнаковых объектов

Рассмотрим теперь задачу групповой классификации многопризнаковых объектов при наличии нескольких учителей (задачу групповой многокритериальной сортировки). Пусть имеется совокупность объектов $A=\{A_1,...,A_n\}$, каждый из которых оценивается k экспертами по m критериям $Q_1,...,Q_m$, характеризующим особенности объекта. Каждый критерий (со-

[1] 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).

держательный признак) Q_s, s=1,…,m имеет собственную числовую, символьную и/или вербальную шкалу $X_s=\{x_s^1,\ldots,x_s^{hs}\}$, которая упорядочена, например, от худшей оценки к лучшей. Кроме того, каждый эксперт относит каждый объект A_i, i=1,…,n к одному из классов $C_1,\ldots C_g$, которые различаются по своим свойствам и могут быть либо упорядочены, либо не упорядочены по предпочтительности. Принадлежность объекта A_i к классу C_l, l=1,…,g описывается сортирующим атрибутом R со шкалой значений $R=\{r_1,\ldots,r_g\}$, который можно считать еще одним качественным признаком. Таким образом, присутствует k различных версий каждого объекта A_i и k индивидуальных экспертных правил их сортировки, обычно не согласованных между собой.

В самом общем смысле правило классификации есть логическое утверждение вида:

$$\text{IF } \langle\text{условия}\rangle\text{, THEN } \langle\text{решение}\rangle. \qquad (4)$$

Здесь терм 〈условия〉 определяет требования, которым должен удовлетворять выбираемый объект, терм 〈решение〉 обозначает имя формируемого класса и/или принадлежность объекта к заранее определенному классу при выполнении требуемых условий.

В задаче групповой многокритериальной сортировки требуется найти одно или несколько достаточно простых обобщенных правил групповой классификации вида (4), которые в наибольшей степени совпадают с несогласованными индивидуальными правилами экспертной классификации объектов, позволяет отнести объекты к заданным классам, не отвергая возможную противоречивость оценок объектов, и выявить противоречиво классифицированные объекты. Задача групповой сортировки совокупности объектов, описываемых многими качественными признаками, относится к числу наиболее трудных задач классификации. Трудности связаны, главным образом, с необходимостью обработки больших массивов символьных и/или вербальных данных, свертка которых или невозможна, или математически некорректна.

Для групповой сортировки многопризнаковых объектов разработан метод МАСКА (МногоАспектная Согласованная Классификация Альтернатив), основанный на теории мультимножеств [3, 7, 21, 22]. Представим каждый многопризнаковый объект A_i как мультимножество

$$\begin{aligned}\boldsymbol{A}_i=\{&k_{\boldsymbol{A}i}(x_1^1)\circ x_1^1,\ldots,k_{\boldsymbol{A}i}(x_1^{h1})\circ x_1^{h1},\ldots,\\ &k_{\boldsymbol{A}i}(x_m^1)\circ x_m^1,\ldots,k_{\boldsymbol{A}i}(x_m^{hm})\circ x_m^{hm},\\ &k_{\boldsymbol{A}i}(r_1)\circ r_1,\ldots,k_{\boldsymbol{A}i}(r_g)\circ r_g\},\end{aligned} \qquad (11)$$

порожденное множеством $X'=X_1\cup\ldots\cup X_m\cup R=X\cup R$.

Расширенное множество признаков X' объединяет подмножества критериальных оценок $X_s=\{x_s^1,\ldots,x_s^{hs}\}$, s=1,…,m и подмножество сортирующих признаков $R=\{r_1,\ldots,r_g\}$, где r_l – экспертное заключение, относящее объект к одному из классов C_l, l=1,…,g. Значения $k_{\boldsymbol{A}i}(x_s^{es})$ и $k_{\boldsymbol{A}i}(r_l)$ равны соответственно числам экспертов, оценивших объект A_i значением признака x_s^{es} и давших заключение r_l. Очевидно, что оценки и заключения многих разных экспертов могут быть схожими, различающимися и противоречивыми. Эта несогласованность выражает субъективность экспертных суждений, которые нельзя рассматривать как случайные ошибки экспертов.

Взаимосвязь между совокупностью многопризнаковых объектов $A=\{A_1,...,A_n\}$ и множеством признаков X' описывается расширенной матрицей данных $\mathrm{K}'=\|k_{ij}\|_{n\times(h+g)}$, $h=h_1+...+h_m$. Каждая строка матрицы K' представляет объект A_i, каждый столбец соответствует некоторому значению $x_j'\in X'$ критерия Q_s или сортирующего признака R, а элемент k_{ij} есть кратность $k_{\boldsymbol{A}i}(x_j')$ значения признака x_j'. Матрица K' аналогична по смыслу так называемой таблице решений или информационной таблице, часто используемой в анализе данных, принятии решений, распознавании образов и других областях.

Представление многопризнакового объекта A_i в виде мультимножества (11) можно рассматривать также как коллективное решающее правило (4), объединяющее индивидуальные правила сортировки нескольких экспертов. Это правило следующим образом ассоциируется с аргументами мультимножества (11). Терм 〈условия〉 включает различные комбинации оценок по критериям x_s^{es}, которые описывают свойства объекта, и признакам r_l. Терм 〈решение〉 обозначает, что объект A_i входит в класс C_l, если выполнены необходимые условия. Например, объект A_i относится к классу C_l согласно одному из следующих правил «большинства голосов»: если $k_{\boldsymbol{A}i}(r_l)>k_{\boldsymbol{A}i}(r_p)$ для всех $p\neq l$, или если $k_{\boldsymbol{A}i}(r_l)>\sum_{p\neq l}k_{\boldsymbol{A}i}(r_p)$. Нужно найти одно или несколько обобщенных правил для классификации всей совокупности объектов вида (4), которые наилучшим образом были бы согласованы с индивидуальными правилами сортировки и позволяли выявлять противоречивые правила.

Для упрощения задачи групповой классификации предположим, что совокупность объектов $A=\{A_1,...,A_n\}$ нужно разделить только на два класса C_a (более предпочтительные объекты) и C_b (менее предпочтительные объекты). Разбиение объектов именно на два класса 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, если требуется отобрать некоторые группы конкурсных проектов, то сначала эти проекты можно разделить на принятые и отложенные, затем принятые проекты – на безусловно и условно поддержанные, а отложенные проекты – на возможно поддержанные и неподдержанные, и так далее.

Сформируем каждый класс объектов как сумму мультимножеств вида (11), которые отражают принадлежность многопризнаковых объектов к этому классу. Тогда классам C_a и C_b соответствуют следующие мультимножества:

$$\begin{aligned}\boldsymbol{Y}_t=\{&k_{\boldsymbol{Y}t}(x_1^{\,1})\circ x_1^{\,1},\ldots,k_{\boldsymbol{Y}t}(x_1^{\,h1})\circ x_1^{\,h1},\ldots,\\ &k_{\boldsymbol{Y}t}(x_m^{\,1})\circ x_m^{\,1},\ldots,k_{\boldsymbol{Y}t}(x_m^{\,hm})\circ x_m^{\,hm},\\ &k_{\boldsymbol{Y}t}(r_1)\circ r_1,\ldots,k_{\boldsymbol{Y}t}(r_g)\circ r_g\},\end{aligned} \qquad (12)$$

где $k_{\boldsymbol{Y}t}(x_s^{es})=\sum_{i\in It}k_{\boldsymbol{A}i}(x_s^{es})$, $k_{\boldsymbol{Y}t}(r_l)=\sum_{i\in It}k_{\boldsymbol{A}i}(r_l)$, $t=a,b$, $I_a\bigcup I_b=\{1,\ldots,n\}$, $I_a\bigcap I_b=\varnothing$. Взаимосвязь между классами C_a, C_b и множеством признаков X' характеризуется агрегированной таблицей решений – матрицей данных $\mathrm{M}=\|k_{ij}'\|_{2\times(h+g)}$, которая состоит из 2 строк и $h+g$ столбцов. Каждая строка матрицы M представляет класс C_t, каждый столбец соответствует некоторому значению $x_j'\in X'$ критерия Q_s или сортирующего признака R, а элемент k_{ij}' есть кратность $k_{\boldsymbol{Y}t}(x_j')$ значения признака x_j'. Выражение (12) итерпретировать как коллективное решающее правило вида (4), которое агрегирует индивидуальные правила всех экспертов по сортировке совокупности многопризнаковых объектов $A=\{A_1,...,A_n\}$ в класс C_t.

Рассмотрим инвертированную таблицу решений – обратную матрицу данных $\mathrm{M}^{-1}=\|k_{ji}'\|_{(h+g)\times 2}$, которая состоит из $h+g$ строк и 2 столбцов. Каждая строка матрицы M^{-1} представляет одно из значений $x_j'\in X'$ критерия Q_s или сортирующего признака R, каждый столбец соответствует некоторому классу C_t, а элемент k_{ji}' есть кратность $k_{\boldsymbol{Y}t}(x_j')$ значения признака x_j'. Введем множество новых признаков $Y'=\{y_a,y_b\}$, элементы которого характеризуют принадлежность объектов к классам C_a и C_b. Тогда строки матрицы M^{-1} формируют совокупность B новых объектов, представленных следующими новыми мультимножествами:

$$\begin{aligned}\boldsymbol{R}_l&=\{k_{\boldsymbol{R}l}(y_a)\circ y_a,\ k_{\boldsymbol{R}l}(y_b)\circ y_b\},\\ \boldsymbol{Q}_j&=\{k_{\boldsymbol{Q}j}(y_a)\circ y_a,\ k_{\boldsymbol{Q}j}(y_b)\circ y_b\}\end{aligned} \qquad (13)$$

над множеством Y'. Здесь $k_{\boldsymbol{R}l}(y_t)=k_{\boldsymbol{Y}t}(r_l)$, $k_{\boldsymbol{Q}j}(y_t)=k_{\boldsymbol{Y}t}(x_j)$, $j=1,\ldots,h$. Назовем мультимножества $\boldsymbol{R}_l$ «категориальными», а мультимножества $\boldsymbol{Q}_j$ – «содержательными».

Нетрудно убедиться, что согласно правилу большинства голосов категориальные мультимножества образуют два мультимножества $\boldsymbol{R}_a=\sum_{l\in La}\boldsymbol{R}_l$ и $\boldsymbol{R}_b=\sum_{l\in Lb}\boldsymbol{R}_l$, где подмножества индексов $L_a\bigcup L_b=\{1,\ldots,g\}$, $L_a\bigcap L_b=\varnothing$. При этом в метрическом пространстве $(\boldsymbol{B},d)$ новых мультимножеств, где $\boldsymbol{B}=\{\boldsymbol{R}_1,\ldots,\boldsymbol{R}_g,\ \boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_h\}$, d – одна из метрик вида (2), расстояние $d^*=d(\boldsymbol{R}_a,\boldsymbol{R}_b)$ будет наибольшим расстоянием между объектами, входящими в разные классы.

Иными словами, категориальные мультимножества $\boldsymbol{R}_a$ и $\boldsymbol{R}_b$ определяют декомпозицию рассматриваемой совокупности объектов A на классы C_a и C_b, лучшую из всех возможных для заданных индивидуальных экспертных правил сортировки:

$$\begin{aligned}&\text{IF }\langle(\textstyle\sum_{l\in La}k_{\boldsymbol{A}i}(r_l)>\sum_{l\in Lb}k_{\boldsymbol{A}i}(r_l))\rangle,\text{ THEN }\langle\text{Объект } A_i\in C_a\rangle;\\ &\text{IF }\langle(\textstyle\sum_{l\in La}k_{\boldsymbol{A}i}(r_l)<\sum_{l\in Lb}k_{\boldsymbol{A}i}(r_l))\rangle,\text{ THEN }\langle\text{Объект } A_i\in C_b\rangle.\end{aligned} \qquad (14)$$

При идеальной предварительной сортировке объектов, когда отсутствуют противоречия между индивидуальными правилами экспертов, расстояние d^* для выражений (5) будет равно соответственно $d_{11}{}^*=kn$, $d_{12}{}^*=1/(h+g)$, $d_{13}{}^*=1$.

Основная идея нахождения простого обобщенного правила групповой многокритериальной классификации объектов, наилучшим образом агрегирующего несогласованные индивидуальные правила сортировки многих экспертов, формулируется так. Для каждой s-ой группы признаков Q_s, $s=1,\ldots,m$ нужно построить содержательные мультимножества $\boldsymbol{Q}_{sa}{}^*$ и $\boldsymbol{Q}_{sb}{}^*$, которые будут определять s-ую лучшую декомпозицию совокупности объектов A на классы C_a и C_b.

По аналогии с категориальными мультимножествами $\boldsymbol{R}_a$ и $\boldsymbol{R}_b$ можно сформировать содержательные мультимножества $\boldsymbol{Q}_{sa}$ и $\boldsymbol{Q}_{sb}$, которые образуются как суммы мультимножеств: $\boldsymbol{Q}_{sa}=\sum_{j\in Jsa}\boldsymbol{Q}_j$ и $\boldsymbol{Q}_{sb}=\sum_{j\in Jsb}\boldsymbol{Q}_j$, где подмножества индексов $J_{sa}\bigcup J_{sb}=\{1,\ldots,h_s\}$, $J_{sa}\bigcap J_{sb}=\varnothing$. Каждое мультимножество $\boldsymbol{Q}_{st}$ соответствует некоторой подгруппе значений признаков $Q_{st}=\bigcup_{j\in Jst}Q_j$, $t=a,b$. Как и выше, содержательные мультимножества $\boldsymbol{Q}_{sa}$ и $\boldsymbol{Q}_{sb}$ задают некоторую декомпозицию совокупности объектов A на классы C_a и C_b.

Рассматривая содержательные мультимножества как точки в пространстве мультимножеств ($\boldsymbol{B}$,d), требуется построить для всех признаков такие пары мультимножеств $\boldsymbol{Q}_{sa}$* и $\boldsymbol{Q}_{sb}$*, которые находились бы на максимально возможном расстоянии. Содержательные мультимножества $\boldsymbol{Q}_{sa}$* и $\boldsymbol{Q}_{sb}$*, соответствующие наилучшей декомпозиции объектов на классы C_a и C_b, являются решением следующей оптимизационной задачи:

$$d(\boldsymbol{Q}_{sa},\boldsymbol{Q}_{sb}) \rightarrow \max d(\boldsymbol{Q}_{sa},\boldsymbol{Q}_{sb}) = d(\boldsymbol{Q}_{sa}^*,\boldsymbol{Q}_{sb}^*). \quad (15)$$

Значения признаков $x_j\in\boldsymbol{Q}_{st}$*, получаемые из решения m оптимизационных задач (15) и характеризующие класс C_t, будем называть классифицирующими признаками для этого класса. Классифицирующие признаки можно упорядочить по величине расстояния $d(\boldsymbol{Q}_{sa}^*,\boldsymbol{Q}_{sb}^*)$ или точности аппроксимации $V_s=d(\boldsymbol{Q}_{sa}^*,\boldsymbol{Q}_{sb}^*)/d(\boldsymbol{R}_a,\boldsymbol{R}_b)$. Классифицирующий признак $x_j\in\boldsymbol{Q}_{st}$*, обеспечивающий приемлемую точность аппроксимации $V_s\geq V_0$, включается в обобщенное решающее правило для групповой многокритериальной сортировки объектов. Значение точности аппроксимации V_s в некотором смысле отражает относительную значимость s-ой группы признаков Q_s в обобщенном правиле.

Разные комбинации классифицирующих признаков дадут следующие обобщенные решающие правила для групповой многокритериальной сортировки объектов на классы C_a и C_b:

IF $\langle(x_j\in Q_{ua}^*)$ AND $(x_j\in Q_{va}^*)$ AND $(x_j\in Q_{wa}^*)$ AND…$\rangle$, THEN $\langle$Объект $A_i\in C_a\rangle$; (16)

IF $\langle(x_j\in Q_{ub}^*)$ AND $(x_j\in Q_{vb}^*)$ AND $(x_j\in Q_{wb}^*)$ AND…$\rangle$, THEN $\langle$Объект $A_i\in C_b\rangle$. (17)

Заметим, что эти правила, вообще говоря, различны.

При одновременном учете экспертных правил сортировки (14) и обобщенного решающего правила (16) или (17) среди отобранных в данный класс C_a или C_b окажутся как правильно, так и неправильно классифицированные объекты. Желательно поэтому найти такие значения классифицирующих признаков, которые обеспечат максимальные числа N_a и N_b правильно классифицированных объектов и минимальные числа N_{ac} и N_{bc} неправильно классифицированных объектов.

Начнем с обобщенного решающего правила (16). Найдем сначала критерий Q_u*, обеспечивающий максимальную разность $N_a–N_{ac}$ между числом правильно и числом неправильно классифицированных объектов. Затем рассмотрим все комбинации критерия Q_u* с остальными критериями и выберем пару критериев Q_u*, Q_v*, обеспечивающих максимальную разность $N_a–N_{ac}$. Далее, рассмотрев все комбинации пары критериев Q_u*, Q_v* с остальными критериями, выберем тройку Q_u*,Q_v*,Q_w*, затем четверку и т.д. критериев, обеспечивающих максимальную разность $N_a–N_{ac}$. Последовательно шаг за шагом уменьшая разность $N_a–N_{ac}$ между числами правильно и неправильно классифицированных объектов, найдем совокупность критериев, которые включаются в согласованное решающее правило для построения уточненного класса $C_a\backslash C_{ac}$ безусловно предпочтительных объектов:

IF $\langle(\sum_{x\in Qua^*}k_{Ai}(x_j)>\sum_{x\in Qub^*}k_{Ai}(x_j))$ AND $(\sum_{x\in Qva^*}k_{Ai}(x_j)>\sum_{x\in Qvb^*}k_{Ai}(x_j))$ AND… AND $(\sum_{r\in Ra}k_{Ai}(r_l)>\sum_{r\in Rb}k_{Ai}(r_l))\rangle$, THEN $\langle$Объект $A_i\in C_a\backslash C_{ac}\rangle$. (18)

Аналогичным образом строится согласованное решающее правило для формирования уточненного класса $C_b\backslash C_{bc}$ безусловно непредпочтительных объектов:

IF $\langle(\sum_{x\in Qua^*}k_{Ai}(x_j)<\sum_{x\in Qub^*}k_{Ai}(x_j))$ AND $(\sum_{x\in Qva^*}k_{Ai}(x_j)<\sum_{x\in Qvb^*}k_{Ai}(x_j))$ AND… AND $(\sum_{r\in Ra}k_{Ai}(r_l)<\sum_{r\in Rb}k_{Ai}(r_l))\rangle$, THEN $\langle$Объект $A_i\in C_b\backslash C_{bc}\rangle$, (19)

которое минимизирует разность $N_b–N_{bc}$ между числом правильно и числом неправильно классифицированных объектов.

Согласованные групповые правила (18) и (19) наилучшим образом аппроксимируют семейство индивидуальных правил сортировки многих экспертов и, вообще говоря, различны.

Одновременно формируется класс $C_c=C_{ac}\bigcup C_{bc}$ противоречиво классифицированных объектов, которые удовлетворяют следующему решающему правилу:

$$\text{IF } \langle [(\sum_{x\in Qua^*} k_{Ai}(x_j) > \sum_{x\in Qub^*} k_{Ai}(x_j)) \text{ AND } (\sum_{x\in Qva^*} k_{Ai}(x_j) > \sum_{x\in Qvb^*} k_{Ai}(x_j)) \text{ AND}\ldots \text{ AND } (\sum_{r\in Ra} k_{Ai}(r_l) < \sum_{r\in Rb} k_{Ai}(r_l))] \text{ OR } [(\sum_{x\in Qua^*} k_{Ai}(x_j) < \sum_{x\in Qub^*} k_{Ai}(x_j)) \text{ AND } (\sum_{x\in Qva^*} k_{Ai}(x_j) < \sum_{x\in Qvb^*} k_{Ai}(x_j)) \text{ AND}\ldots \text{ AND } (\sum_{r\in Ra} k_{Ai}(r_l) > \sum_{r\in Rb} k_{Ai}(r_l))]\rangle, \text{ THEN } \langle\text{Объект } A_i \in C_c\rangle. \quad (20)$$

Эти объекты нуждаются в дополнительном анализе. Правило (20) помогает обнаружить возможные противоречия в индивидуальных экспертных правилах сортировки.

Алгоритм для построения обобщенных решающих правил, которые обеспечивают согласованную групповую классификацию многопризнаковых объектов, включает следующие шаги.

Шаг 1. Построить таблицу решений $K'=\|k_{ij}\|_{n\times(h+g)}$, которая представляет совокупность многопризнаковых объектов $A=\{A_1,...,A_n\}$ и расширенное множество признаков $X'=\{x_1^{\ 1},\ldots,x_1^{\ h1};\ \ldots;\ x_m^{\ 1},\ldots,x_m^{\ hm},\ r_1,\ldots,r_g\}$.

Шаг 2. Объединить объекты $A_1,...,A_n$ в заданные классы C_a, C_b путем сложения соответствующих мультимножеств согласно прнятому коллективному решающему правилу (4), которое интегрирует индивидуальные правила сортировки нескольких экспертов. Построить агрегированную таблицу решений $M=\|k_{ij}'\|_{2\times(h+g)}$ и инвертированную таблицу решений $M^{-1}=\|k_{ji}'\|_{(h+g)\times 2}$, которые представляют классы C_a, C_b объектов и значения критериев и сортирующих признаков.

Шаг 3. Решить задачу оптимизации (15) для каждой группы признаков Q_s, $s=1,\ldots,m$ и найти классифицирующие признаки $x_j\in \boldsymbol{Q}_{sa}^*$ и $x_j\in \boldsymbol{Q}_{sb}^*$. Проранжировать классифицирующие признаки по величине расстояния $d(\boldsymbol{Q}_{sa}^*,\boldsymbol{Q}_{sb}^*)$ или точности аппроксимации $V_s=d(\boldsymbol{Q}_{sa}^*,\boldsymbol{Q}_{sb}^*)/d(\boldsymbol{R}_a,\boldsymbol{R}_b)$.

Шаг 4. Выбрать классифицирующие признаки $x_j\in \boldsymbol{Q}_{st}^*$, $t=a,b$, обеспечивающие приемлемую точность аппроксимации $V_s\geq V_0$. Сформировать обобщенные решающие правила (16) и (17) для групповой многокритериальной сортировки объектов.

Шаг 5. По каждому критерию Q_s^*, входящему в обобщенное решающее правило (16) отнесения объектов к классу C_a, найти правильно и неправильно классифицированные объекты. Выбрать критерий Q_u^*, обеспечивающий максимальную разность между числами правильно и неправильно классифицированных объектов, который включить в согласованное решающее правило (18) для отнесения объектов к уточненному классу $C_a\backslash C_{ac}$ безусловно предпочтительных объектов.

Шаг 6. Рассмотреть все комбинации критерия Q_u^* с остальными критериями, входящими в обобщенное решающее правило (16), найти правильно и неправильно классифицированные объекты. Выбрать пару критериев Q_u^*, Q_v^*, обеспечивающих максимальную разность между числами правильно и неправильно классифицированных объектов, которые включить в согласованное решающее правило (18) для отнесения объектов к уточненному классу $C_a\backslash C_{ac}$.

Шаг 7. Рассмотреть все комбинации пары критериев Q_u^*, Q_v^* с остальными критериями, входящими в обобщенное решающее правило (16), найти правильно и неправильно классифицированные объекты. Выбрать тройку, затем четверку и т.д. критериев $Q_u^*,Q_v^*,Q_w^*,\ldots$ обеспечивающих максимальную разность между числами правильно и неправильно классифицированных объектов, которые включить в согласованное решающее правило (18) для отнесения объектов к уточненному классу $C_a\backslash C_{ac}$ безусловно предпочтительных объектов.

Шаг 8. Сформировать аналогичным образом согласованное решающее правило (19) для отнесения объектов к уточненному классу $C_b\backslash C_{bc}$ безусловно непредпочтительных объектов. Найти правильно и неправильно классифицированные объекты.

Шаг 9. Сформировать решающее правило (20) для построения класса $C_c=C_{ac}\bigcup C_{bc}$ объектов, которые имеют противоречивые индивидуальные правила сортировки. Провести дополнительный анализ таких объектов для устранения противоречий и принятия окончательного решения.

Обобщенные решающие правила для групповой многокритериальной сортировки объектов можно легко записать на естественном языке, используя формулировки вербальных значений классифицирующих признаков.

Проиллюстрируем, как работает изложенный алгоритм, на примере задачи классификации многопризнаковых объектов с несколькими учителями. Требуется разделить совокупность объектов $A=\{A_1,...,A_{10}\}$ на два упорядоченных класса C_a (более предпочтительные) и C_b (менее предпочтительные). Объекты характеризуются признаками $Q_1,\ldots,Q_8$, каждый из которых принимает одну из оценок по упорядоченной пятибалльной шкале $X=\{x_1, x_2, x_3, x_4, x_5\}$. Напомним, что эти оценки являются числовыми, символьными или вербальными переменными. Описания объектов приведены в таблице данных 2 «Объекты-Признаки» $L=\|k_{ij}\|_{10\times 5}$.

Построим таблицу решений $L'=\|k_{ij}\|_{10\times(5+2)}$ – расширенную таблицу данных с двумя дополнительными столбцами, соответствующими сортирующим признакам r_a и r_b. Эти признаки характеризуют заключения экспертов («голоса» дисциплин или респондентов) о принадлежности объекта к одному из классов C_a или C_b, например, следующим образом: признаку r_a соответствуют лучшие оценки объекта (x_5 и x_4), а признаку r_b – худшие оценки (x_3, x_2, x_1). Таблица решений L', строки которой суть мультимножества вида

$$\boldsymbol{A}_i=\{k_{\boldsymbol{A}i}(x_1)\circ x_1, k_{\boldsymbol{A}i}(x_2)\circ x_2, k_{\boldsymbol{A}i}(x_3)\circ x_3, k_{\boldsymbol{A}i}(x_4)\circ x_4, k_{\boldsymbol{A}i}(x_5)\circ x_5, k_{\boldsymbol{A}i}(r_a)\circ r_a, k_{\boldsymbol{A}i}(r_b)\circ r_b\},$$

представлена Табл. 6.

Табл. 6. Таблица решений L'

A\X'	x_1	x_2	x_3	x_4	x_5	r_a	r_b
A_1	0	0	0	4	4	8	0
A_2	2	4	1	1	0	1	7
A_3	5	0	1	2	0	2	6
A_4	0	1	1	3	3	6	2
A_5	0	0	0	7	1	8	0
A_6	0	0	0	4	4	8	0
A_7	2	2	3	1	0	1	7
A_8	0	1	2	3	2	5	3
A_9	1	3	4	0	0	0	8
A_{10}	0	0	1	2	5	7	1

Табл. 7. Агрегированная таблица решений M

C\X'	x_1	x_2	x_3	x_4	x_5	r_a	r_b
$\boldsymbol{Y}_a$	0	2	4	23	19	42	6
$\boldsymbol{Y}_b$	10	9	9	4	0	4	28

Объединим объекты в два класса C_a и C_b, например, по такому коллективному правилу «большинства голосов»: если число лучших оценок объекта больше числа худших оценок (x_3, x_2, x_1), то объект включается в класс C_a, в противоположном случае объект включается в класс C_b. Таким образом, в класс C_a войдут объекты A_1, A_4, A_5, A_6, A_8, A_{10}, а в класс C_b – объекты A_2, A_3, A_7, A_9. Классам C_a и C_b соответствуют следующие мультимножества:

$$\boldsymbol{Y}_a=\{0\circ x_1, 2\circ x_2, 4\circ x_3, 23\circ x_4, 19\circ x_5, 42\circ r_a, 6\circ r_b\},$$
$$\boldsymbol{Y}_b=\{10\circ x_1, 9\circ x_2, 9\circ x_3, 4\circ x_4, 0\circ x_5, 4\circ r_a, 28\circ r_b\},$$

которые являются строками агрегированной таблицы решений $M=\|k_{ij}'\|_{2\times(5+2)}$ (Табл. 7).

Рассмотрим инвертированную таблицу решений $M^{-1}=\|k_{ji}'\|_{(5+2)\times 2}$ и построим категориальные и содержательные мультимножества

$$\boldsymbol{Q}_1=\{0\circ y_a, 10\circ y_b\}, \boldsymbol{Q}_2=\{2\circ y_a, 9\circ y_b\},$$
$$\boldsymbol{Q}_3=\{4\circ y_a, 9\circ y_b\}, \boldsymbol{Q}_4=\{23\circ y_a, 4\circ y_b\},$$
$$\boldsymbol{Q}_5=\{19\circ y_a, 0\circ y_b\},$$
$$\boldsymbol{R}_a=\{42\circ y_a, 4\circ y_b\}, \boldsymbol{R}_b=\{6\circ y_a, 28\circ y_b\}.$$

Выберем в качестве меры близости мультимножеств в метрическом пространстве $(\boldsymbol{B},d)$ одну из метрик (5) вида $d_{11}(\boldsymbol{B}_p,\boldsymbol{B}_q)=D_{pq}=\sum_t w_t|k_{\boldsymbol{B}p}(y_t)-k_{\boldsymbol{B}q}(y_t)|$, $t=a,b$, считая все признаки y_t одинаково важными ($w_t=1$). Расстояние между категориальными мультимножествами равно

$$d(\boldsymbol{R}_a,\boldsymbol{R}_b)=|42-6|+|4-28|=60.$$

Решением единственной оптимизационной задачи (15) являются следующие содержательные мультимножества: $\boldsymbol{Q}_a^*=\boldsymbol{Q}_4+\boldsymbol{Q}_5$ и $\boldsymbol{Q}_b^*=\boldsymbol{Q}_1+\boldsymbol{Q}_2+\boldsymbol{Q}_3$, которые среди всех возможных комбинаций пар мультимножеств $\boldsymbol{Q}_a$ и $\boldsymbol{Q}_b$ находятся на максимальном расстоянии

$$d(\boldsymbol{Q}_a^*,\boldsymbol{Q}_b^*)=|19+23-4-2|+|4-9-9-10|=36+24=60.$$

Мультимножество $\boldsymbol{Q}_a^*$ определяет группу классифицирующих признаков $Q_a^*=Q_4\cup Q_5$, куда входят признаки x_4, x_5, характеризующие класс C_a, Мультимножество $\boldsymbol{Q}_b^*$ определяет группу классифицирующих признаков $Q_b^*=Q_1\cup Q_2\cup Q_3$, куда входят признаки x_1, x_2, x_3, характеризующие класс C_b,

Согласованные обобщенные решающие правила для групповой классификации многопризнаковых объектов, выраженные на естественном языке, выглядят так:

«В более предпочтительный класс C_a входят объекты, имеющие оценки «4/хорошо» или «5/отлично» (x_4 или x_5)»;

«В менее предпочтительный класс C_b входят объекты, имеющие оценки «1/очень плохо», или «2/плохо», или «3/удовлетворительно» (x_1, или x_2, или x_3)».

Заметим, что при выводе этих решающих правил нигде не использовалась информация о характере признаков, описывающих объекты, а именно являются ли признаки числовыми, символьными или вербальными переменными.

Обратим еще внимание на следующие моменты. Во-первых, построенные обобщенные решающие правила для групповой классификации объектов не отличаются от коллективного правила «большинства голосов», по которому объекты относились к тому или иному классу, а величины расстояний $d(\boldsymbol{R}_a,\boldsymbol{R}_b)$ и $d(\boldsymbol{Q}_a^*,\boldsymbol{Q}_b^*)$ равны между собой. Эти обстоятельства объясняются тем, что все индивидуальные решающие правила для первоначальной сортировки объектов не содержали противоречий. В общем случае это бывает не так. Во-вторых, сформированная классификация многопризнаковых объектов полностью совпадает с результатом, полученным с помощью алгоритма иерархической кластеризации в первой части статьи [30].

Так же достаточно легко и просто можно решить задачу групповой порядковой классификации объектов, описываемых многими количественными и/или качественными признаками, в случаях, когда объекты присутствуют в нескольких версиях (экземплярах), различающихся значениями признаков, задано несколько классов решений, а индивидуальные решающие правила для первоначальной сортировки объектов несогласованы или противоречивы.

Пусть, например, существует по две версии каждого из объектов $A_1,...,A_{10}$, которые характеризуются признаками $Q_1,\dots,Q_8$, имеющими разные шкалы оценок $X_s=\{x_s^1,\dots,x_s^{hs}\}$, $s=1,\dots,8$. Описания этих объектов, представленных мультимножествами вида (3), указаны в таблице данных 5 «Объекты-Признаки» $\mathrm{K}=\|k_{ij}\|_{10\times 40}$. Допустим, что объекты нужно опять разбить на два класса: C_a (более предпочтительные) и C_b (менее предпочтительные).

Расширенная таблица решений $\mathrm{K'}=\|k_{ij}\|_{10\times(40+2)}$ приведена в Табл. 8. Строками таблицы решений служат мультимножества вида (11)

$$\boldsymbol{A}_i=\{k_{\boldsymbol{A}i}(x_1^1)\circ x_1^1,\dots,k_{\boldsymbol{A}i}(x_1^5)\circ x_1^5;\ \dots;$$
$$k_{\boldsymbol{A}i}(x_8^1)\circ x_8^1,\dots,k_{\boldsymbol{A}i}(x_8^5)\circ x_8^5;\ k_{\boldsymbol{A}i}(r_a)\circ r_a,\ k_{\boldsymbol{A}i}(r_b)\circ r_b\}$$

над множеством оценок $X'=\{x_1^1,\dots,x_1^5;\ \dots;\ x_8^1,\dots,x_8^5;\ r_a,\ r_b\}$. Сортирующие признаки r_a и r_b отражают индивидуальные заключения экспертов, относивших объекты к классам C_a или C_b.

Коллективное правило «большинства голосов» выглядит теперь так: объект включается в класс C_a, если за это «проголосовало» большинство экспертов, то есть $k_{\boldsymbol{A}i}(r_a)\geq k_{\boldsymbol{A}i}(r_b)$; в противоположном случае объект включается в класс C_b. Заметим, что по объекту A_{10} голоса экспертов разделились поровну, то есть индивидуальные решающие правила противоречивы.

Согласно правилу «большинства голосов» объекты A_1, A_4, A_5, A_6, A_8, A_{10} входят в класс C_a, а объекты A_2, A_3, A_7, A_9 – в класс C_b. Классам C_a и C_b соответствуют мультимножества:

$$\boldsymbol{Y}_a=\{0\circ x_1^1,0\circ x_1^2,1\circ x_1^3,5\circ x_1^4,6\circ x_1^5;$$
$$0\circ x_2^1,0\circ x_2^2,1\circ x_2^3,4\circ x_2^4,7\circ x_2^5;$$
$$0\circ x_3^1,1\circ x_3^2,3\circ x_3^3,6\circ x_3^4,2\circ x_3^5;$$
$$0\circ x_4^1,1\circ x_4^2,1\circ x_4^3,6\circ x_4^4,4\circ x_4^5;$$
$$0\circ x_5^1,1\circ x_5^2,2\circ x_5^3,9\circ x_5^4,0\circ x_5^5;$$
$$0\circ x_6^1,0\circ x_6^2,0\circ x_6^3,5\circ x_6^4,7\circ x_6^5;$$
$$0\circ x_7^1,1\circ x_7^2,1\circ x_7^3,5\circ x_7^4,5\circ x_7^5;$$
$$0\circ x_8^1,0\circ x_8^2,1\circ x_8^3,7\circ x_8^4,4\circ x_8^5,\ 11\circ r_a,\ 1\circ r_b\},$$
$$\boldsymbol{Y}_b=\{2\circ x_1^1,0\circ x_1^2,3\circ x_1^3,\circ x_1^4,0\circ x_1^5;$$
$$3\circ x_2^1,4\circ x_2^2,1\circ x_2^3,0\circ x_2^4,0\circ x_2^5;$$
$$2\circ x_3^1,3\circ x_3^2,3\circ x_3^3,0\circ x_3^4,0\circ x_3^5;$$
$$5\circ x_4^1,1\circ x_4^2,1\circ x_4^3,1\circ x_4^4,0\circ x_4^5;$$
$$0\circ x_5^1,2\circ x_5^2,3\circ x_5^3,2\circ x_5^4,1\circ x_5^5;$$
$$1\circ x_6^1,2\circ x_6^2,4\circ x_6^3,1\circ x_6^4,0\circ x_6^5;$$
$$3\circ x_7^1,3\circ x_7^2,\circ x_7^3,0\circ x_7^4,0\circ x_7^5;$$
$$0\circ x_8^1,3\circ x_8^2,4\circ x_8^3,1\circ x_8^4,0\circ x_8^5,\ 0\circ r_a,\ 8\circ r_b\},$$

которые образуют строки агрегированной таблицы решений $\mathrm{M}=\|k_{ij}'\|_{2\times(40+2)}$ (Табл. 9).

Воспользуемся инвертированной таблицей решений $\mathrm{M}^{-1}=\|k_{ji}'\|_{(40+2)\times 2}$ для вычисления расстояний между категориальными и содержательными мультимножествами. Расстояние между категориальными мультимножествами $\boldsymbol{R}_a=\{11\circ y_a,\ 0\circ y_b\}$ и $\boldsymbol{R}_b=\{1\circ y_a,\ 8\circ y_b\}$ равно

$$d(\boldsymbol{R}_a,\boldsymbol{R}_b)=|11-1|+|0-8|=18.$$

Классифицирующие признаки для каждого критерия Q_s, $s=1,\dots,8$ ищутся как решение соответствующей оптимизационной задачи (15). Среди всех возможных комбинаций пар содержательных мультимножеств $\boldsymbol{Q}_{sa}$ и $\boldsymbol{Q}_{sb}$ на максимальных расстояниях будут находиться следующие мультимножества:

$$\boldsymbol{Q}_{1a}^*=\boldsymbol{Q}_{13}+\boldsymbol{Q}_{14}+\boldsymbol{Q}_{15} \text{ и } \boldsymbol{Q}_{1b}^*=\boldsymbol{Q}_{11}+\boldsymbol{Q}_{12},$$
$$d(\boldsymbol{Q}_{1a}^*,\boldsymbol{Q}_{1b}^*)=|1+5+6-0-0|+|3+3+0-2-0|=12+4=16;$$
$$\boldsymbol{Q}_{2a}^*=\boldsymbol{Q}_{24}+\boldsymbol{Q}_{25} \text{ и } \boldsymbol{Q}_{2b}^*=\boldsymbol{Q}_{21}+\boldsymbol{Q}_{22}+\boldsymbol{Q}_{23},$$
$$d(\boldsymbol{Q}_{2a}^*,\boldsymbol{Q}_{2b}^*)=|4+7-0-0-1|+|0+0-3-4-1|=10+8=18;$$
$$\boldsymbol{Q}_{3a}^*=\boldsymbol{Q}_{33}+\boldsymbol{Q}_{34}+\boldsymbol{Q}_{35} \text{ и } \boldsymbol{Q}_{3b}^*=\boldsymbol{Q}_{31}+\boldsymbol{Q}_{32},$$
$$d(\boldsymbol{Q}_{3a}^*,\boldsymbol{Q}_{3b}^*)=|3+6+2-0-1|+|3+0+0-2-3|=10+2=12;$$
$$\boldsymbol{Q}_{4a}^*=\boldsymbol{Q}_{44}+\boldsymbol{Q}_{45} \text{ и } \boldsymbol{Q}_{4b}^*=\boldsymbol{Q}_{41}+\boldsymbol{Q}_{42}+\boldsymbol{Q}_{43},$$
$$d(\boldsymbol{Q}_{4a}^*,\boldsymbol{Q}_{4b}^*)=|6+4-0-1-1|+|1+0-5-1-1|=8+6=14;$$
$$\boldsymbol{Q}_{5a}^*=\boldsymbol{Q}_{53}+\boldsymbol{Q}_{54}+\boldsymbol{Q}_{55} \text{ и } \boldsymbol{Q}_{5b}^*=\boldsymbol{Q}_{51}+\boldsymbol{Q}_{52},$$
$$d(\boldsymbol{Q}_{5a}^*,\boldsymbol{Q}_{5b}^*)=|2+9+0-0-1|+|3+2+1-0-2|=10+4=14;$$

Табл. 8. Таблица решений K’

A\X’	$x_1^1\ x_1^2\ x_1^3\ x_1^4\ x_1^5$	$x_2^1\ x_2^2\ x_2^3\ x_2^4\ x_2^5$	$x_1^1\ x_3^2\ x_3^3\ x_3^4\ x_3^5$	$x_4^1\ x_4^2\ x_4^3\ x_4^4\ x_4^5$	$x_5^1\ x_5^2\ x_5^3\ x_5^4\ x_5^5$	$x_6^1\ x_6^2\ x_6^3\ x_6^4\ x_6^5$	$x_7^1\ x_7^2\ x_7^3\ x_7^4\ x_7^5$	$x_8^1\ x_8^2\ x_8^3\ x_8^4\ x_8^5$	$r_a\ \ r_b$
A_1	0 0 0 1 1	0 0 0 0 2	0 0 0 1 1	0 0 0 0 2	0 0 0 2 0	0 0 0 1 1	0 0 0 2 0	0 0 0 0 2	2 0
A_2	0 0 1 1 0	1 1 0 0 0	1 1 0 0 0	2 0 0 0 0	0 0 1 1 0	0 1 1 0 0	0 1 1 0 0	0 2 0 0 0	0 2
A_3	2 0 0 0 0	1 1 0 0 0	0 0 2 0 0	2 0 0 0 0	0 0 0 1 1	1 1 0 0 0	2 0 0 0 0	0 0 1 1 0	0 2
A_4	0 0 0 1 1	0 0 1 1 0	0 1 1 0 0	0 0 0 1 1	0 0 0 2 0	0 0 0 0 2	0 0 1 1 0	0 0 0 1 1	2 0
A_5	0 0 0 1 1	0 0 0 1 1	0 0 1 1 0	0 0 0 2 0	0 0 0 2 0	0 0 0 1 1	0 0 0 1 1	0 0 0 2 0	2 0
A_6	0 0 0 1 1	0 0 0 0 2	0 0 0 2 0	0 0 0 2 0	0 0 0 2 0	0 0 0 1 1	0 0 0 0 2	0 0 0 1 1	2 0
A_7	0 0 1 1 0	1 1 0 0 0	1 1 0 0 0	0 0 1 1 0	0 1 1 0 0	0 0 1 1 0	1 1 0 0 0	0 1 1 0 0	0 2
A_8	0 0 0 1 1	0 0 0 1 1	0 0 0 1 1	0 1 1 0 0	0 0 1 1 0	0 0 0 1 1	0 0 0 1 1	0 0 1 1 0	2 0
A_9	0 0 1 1 0	0 1 1 0 0	0 1 1 0 0	1 1 0 0 0	0 1 1 0 0	0 0 2 0 0	0 1 1 0 0	0 0 2 0 0	0 2
A_{10}	0 0 1 0 1	0 0 0 1 1	0 0 1 1 0	0 0 0 1 1	0 1 1 0 0	0 0 0 1 1	0 1 0 0 1	0 0 0 2 0	1 1

Табл. 9. Агрегированная таблица решений M

C\X’	$x_1^1\ x_1^2\ x_1^3\ x_1^4\ x_1^5$	$x_2^1\ x_2^2\ x_2^3\ x_2^4\ x_2^5$	$x_1^1\ x_3^2\ x_3^3\ x_3^4\ x_3^5$	$x_4^1\ x_4^2\ x_4^3\ x_4^4\ x_4^5$	$x_5^1\ x_5^2\ x_5^3\ x_5^4\ x_5^5$	$x_6^1\ x_6^2\ x_6^3\ x_6^4\ x_6^5$	$x_7^1\ x_7^2\ x_7^3\ x_7^4\ x_7^5$	$x_8^1\ x_8^2\ x_8^3\ x_8^4\ x_8^5$	$r_a\ \ r_b$
Y_a	0 0 1 5 6	0 0 1 4 7	0 1 3 6 2	0 1 1 6 4	0 1 2 9 0	0 0 0 5 7	0 1 1 5 5	0 0 1 7 4	11 1
Y_b	2 0 3 3 0	3 4 1 0 0	2 3 3 0 0	5 1 1 1 0	0 2 3 2 1	1 2 4 1 0	3 3 2 0 0	0 3 4 1 0	0 8

$$\boldsymbol{Q}_{6a}{*}=\boldsymbol{Q}_{64}+\boldsymbol{Q}_{65} \text{ и } \boldsymbol{Q}_{6b}{*}=\boldsymbol{Q}_{61}+\boldsymbol{Q}_{62}+\boldsymbol{Q}_{63},$$
$$d(\boldsymbol{Q}_{6a}{*},\boldsymbol{Q}_{6b}{*})=|5+7–0–0–0|+|1+0–1–2–4|=12+6=18;$$
$$\boldsymbol{Q}_{7a}{*}=\boldsymbol{Q}_{74}+\boldsymbol{Q}_{75} \text{ и } \boldsymbol{Q}_{7b}{*}=\boldsymbol{Q}_{71}+\boldsymbol{Q}_{72}+\boldsymbol{Q}_{73},$$
$$d(\boldsymbol{Q}_{7a}{*},\boldsymbol{Q}_{7b}{*})=|5+5–0–1–1|+|0+0–3–3–2|=8+8=16;$$
$$\boldsymbol{Q}_{8a}{*}=\boldsymbol{Q}_{84}+\boldsymbol{Q}_{85} \text{ и } \boldsymbol{Q}_{8b}{*}=\boldsymbol{Q}_{81}+\boldsymbol{Q}_{82}+\boldsymbol{Q}_{83},$$
$$d(\boldsymbol{Q}_{8a}{*},\boldsymbol{Q}_{8b}{*})=|7+4–0–0–1|+|1+0–0–3–4|=10+6=16.$$

Отсюда вытекает, что для класса C_a характерными являются классифицирующие признаки

$$Q_a{*}=\{x_2^4,x_2^5\}\cup\{x_6^4,x_6^5\}\cup\{x_1^3,x_1^4,x_1^5\}\cup\{x_7^4,x_7^5\}$$
$$\cup\{x_8^4,x_8^5\}\cup\{x_4^4,x_4^5\}\cup\{x_5^3,x_5^4,x_5^5\}\cup\{x_3^3,x_3^4,x_3^5\},$$

а для класса C_b – классифицирующие признаки

$$Q_b{*}=\{x_2^1,x_2^2,x_2^3\}\cup\{x_6^1,x_6^2,x_6^3\}\cup\{x_1^1,x_1^2\}\cup\{x_7^1,x_7^2,x_7^3\}$$
$$\cup\{x_8^1,x_8^2,x_8^3\}\cup\{x_4^1,x_4^2,x_4^3\}\cup\{x_5^1,x_5^2\}\cup\{x_3^1,x_3^2\}.$$

Классифицирующие признаки упорядочены здесь по их значимости, определяемой величиной расстояния $d(\boldsymbol{Q}_{sa}{*},\boldsymbol{Q}_{sb}{*})$.

Выбирая наиболее значимые классифицирующие признаки, получим, например, такие обобщенные решающие правила для групповой классификации многопризнаковых объектов, записанные на естественном языке:

«В более предпочтительный класс C_a входят объекты, имеющие по критериям Q_2 и Q_6 оценки «4/хорошо» или «5/отлично» (x_2^4 или x_2^5, и x_6^4 или x_6^5)»;

«В менее предпочтительный класс C_b входят объекты, имеющие по критериям Q_2 и Q_6 оценки «1/очень плохо», или «2/плохо», или «3/удовлетворительно» (x_2^1 или x_2^2, и x_2^3 или x_6^1, и x_6^2 или x_6^3)».

Указанные обобщенные решающие правила для групповой классификации объектов конкретизируют построенные выше решающие правила классификации, принимая во внимание не общее число оценок того или иного рода, а оценки по отдельным критериям (дисциплинам, респондентам). При необходимости можно расширить эти обобщенные решающие правила, включив в них следующие по значимости классифицирующие признаки.

5. Практические примеры

Приведем несколько примеров практического применения предложенных методов.

Конкурсный отбор предложений широко используется для различных целей во многих областях деятельности. Часто такой отбор осуществляется на основе многокритериальных экспертных оценок и заключений, данных несколькими экспертами. С помощью метода МАСКА были построены правила для групповой сортировки предложений, которые содержат полезную вспомогательную информацию для руководителя или органа, отвечающего за принятие решения. Эти обобщенные решающие правила базируются на многих противоречивых индивидуальных правилах сортировки, оперирующих нечисловыми данными. Подобные правила нельзя найти другими известными методами многокритериального принятия решений.

Государственная научно-техническая программа по высокотемпературной сверхпроводимости [6] формировалась путем конкурсного отбора проектов. Всего было подано более 250 заявок и около 170 из них было отобрано для включения в программу. Каждая представленная на конкурс заявка независимо оценивалась 3 экспертами по 6 качественным критериям. Эксперты давали также свое заключение по принятию или отклонению заявки. Оценка заявок 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, которые были упорядочены от лучших оценок к худшим. Так, шкала критерия Q_4. «Квалификация исполнителей проекта» имела вид:

x_4^1 – по опыту и квалификации исполнители проекта являются одним из лучших научных коллективов;

x_4^2 – опыт и квалификация исполнители достаточны для проведения работ;

x_4^3 – исполнители не обладают необходимыми опытом и квалификацией;

x_4^4 – опыт и квалификация исполнителей неизвестны.

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 по ее предварительной сортировке:

r_1 – включить проект в программу;

r_2 – отклонить проект;

r_3 – отложить рассмотрение заявки и отправить проект на доработку.

Приведем некоторые данные, иллюстрирующие рассматриваемый пример. Табл. 10 содержит (i): часть таблицы решений K', характеризующей поданные на конкурс проекты; (ii) агрегированную таблицу решений M, которая представляет классы проектов C_a, C_b; (iii) расстояния $d_{11}(\boldsymbol{Q}_{sa}{}^*,\boldsymbol{Q}_{sb}{}^*)$ между содержательными и $d_{11}(\boldsymbol{R}_a,\boldsymbol{R}_b)$ между категориальными мультимножествами в пространстве мультимножеств, точность аппроксимации V_s. Столбец r_a соответствует экспертной рекомендации r_1, колонка r_b объединяет рекомендации r_2 и r_3.

Как явствует из таблицы 10(iii), наиболее значимым для отбора проектов является критерий Q_4, характеризующий опыт и квалификацию исполнителей. Следующими по значимости будут критерий Q_1, отражающий важность проекта для программы, и критерий Q_5, оценивающий ресурсное обеспечение работ.

Если бы заявка оценивалась только одним экспертом, то найти обобщенное решающее правило для отбора предложений на множестве их многокритериальных оценок можно с помощью разных методов [5, 12-14, 17, 18, 23-27]. Однако когда заявка рассматривается несколькими экспертами, то появляется несколько различных версий (экземпляров) одной и той же заявки, поскольку и многокритериальные экспертные оценки, и заключения экспертов могут быть как схожими, так и противоречивыми. Методов решения задачи групповой классификации объектов существенно меньше [16]. Кроме того, все они ориентированы на количественные данные.

При данных качественного характера их агрегирование тем или иным способом представляет самостоятельную, достаточно сложную

Табл. 10. Таблицы решений K' и M для конкурсных заявок

(i)

A/X'	x_1^1 x_1^2 x_1^3	x_2^1 x_2^2 x_2^3	x_3^1 x_3^2 x_3^3	x_4^1 x_4^2 x_4^3 x_4^4	x_5^1 x_5^2 x_5^3 x_5^4	x_6^1 x_6^2 x_6^3	r_a r_b
$\boldsymbol{A}_1$	1 2 0	2 1 0	3 0 0	2 1 0 0	0 2 1 0	2 1 0	3 0
…							
$\boldsymbol{A}_i$	1 1 1	0 2 1	1 2 0	0 2 1 0	0 1 2 0	0 0 3	2 1
$\boldsymbol{A}_{i+1}$	1 1 1	0 2 1	1 2 0	0 2 1 0	0 1 2 0	0 0 3	1 2
…							
$\boldsymbol{A}_n$	0 2 1	0 1 2	0 3 0	0 1 1 1	0 0 2 1	0 3 0	0 3

(ii)

C/X'	x_1^1 x_1^2 x_1^3	x_2^1 x_2^2 x_2^3	x_3^1 x_3^2 x_3^3	x_4^1 x_4^2 x_4^3 x_4^4	x_5^1 x_5^2 x_5^3 x_5^4	x_6^1 x_6^2 x_6^3	r_a r_b
$\boldsymbol{Y}_a$	144 360 21	81 324 120	99 336 90	219 297 9 0	72 435 18 0	126 300 99	510 15
$\boldsymbol{Y}_b$	45 156 51	27 93 132	36 111 105	51 132 63 6	60 147 30 15	45 135 72	78 174

(iii)

d_{11}	333	297	303	393	327	273	591
V_s	0,563	0,503	0,517	0,665	0,553	0,462	

проблему. Помимо этого, вырабатывая решение о включении заявки в программу, необходимо учесть все, в том числе не совпадающие заключения экспертов по принятию или отклонению заявки. Нужно 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.

Использование метода групповой многокритериальной сортировки объектов, представленных в виде мультимножеств, позволило построить несколько обобщенных решающих правил классификации, которые формулируются так:

«Проект включается в программу, если исполнители проекта имеют наибольшие или достаточные опыт и квалификацию для проведения работ» (оценки по критериям x_4^1 или x_4^2; точность аппроксимации $V_s \geq 0{,}65$);

«Проект включается в программу, если исполнители проекта имеют опыт и квалификацию, наибольшие или достаточные для проведения работ; проект обеспечивает достижение всех или одной из основных целей программы; исполнители имеют необходимые для проведения работ материально-технические ресурсы в полном или достаточном объеме» (оценки по критериям x_4^1 или x_4^2; и x_1^1 или x_1^2; и x_5^1 или x_5^2; точность аппроксимации $V_s \geq 0{,}55$).

Последнее решающее правило полностью совпало с полученным на практике [6]. Заметим также, что ни одно значение признака x_6 не являлось классифицирующим, поскольку задача (15) не имела оптимального решения для критерия Q_6.

Согласованное групповое правило для отбора лучших проектов, агрегирующее индивидуальные экспертные правила сортировки, записывается следующим образом:

«Проект включается в программу, если он обеспечивает достижение всех или одной из основных целей программы; имеет высокую или достаточную перспективность; предлагает совершенно новые или модернизированные подходы к решению поставленных задач; исполнители имеют необходимые для проведения работ материально-технические ресурсы в полном или достаточном объеме» (оценки по критериям x_1^1 или x_1^2; и x_2^1 или x_2^2; и x_3^1 или x_3^2; и x_5^1 или x_5^2).

Предложенный подход к построению согласованного решающего правила для классификации многопризнаковых объектов был апробирован также при анализе итогов конкурсов целевых фундаментальных исследований, выполняемых по грантам Российского фонда фундаментальных исследований (РФФИ) в интересах Федеральных агентств и ведомств России [9]. 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.

По установленному в РФФИ порядку экспертиза научных проектов осуществляется независимо несколькими экспертами, как правило, тремя. Для оценки содержания заявки используются специальные анкеты, которые содержат критерии, имеющие словесные шкалы оценок с развернутыми формулировками градаций качества. Экспертная оценка содержания целевых фундаментальных исследований проводится по 11 качественным критериям, представленным в экспертной анкете и объединенным в две группы: научная характеристика проекта; оценка возможностей практической реализации проекта. Например, шкала критерия Q_7. «Степень новизны предлагаемых решений» выглядит следующим образом:

q_7^1 - решения сформулированы впервые и существенно превосходят уровень имеющихся;

q_7^2 - решение находится на уровне имеющихся решений;

q_7^3 - решение уступает имеющимся решениям.

Кроме того, эксперт дает свое заключение о целесообразности поддержки проекта, проставляя следующие оценки:

r_1 - безусловно поддержать (оценка «5»),

r_2 - целесообразно поддержать (оценка «4»),

r_3 - поддержка возможна (оценка «3»),

r_4 - не поддерживать (оценка «2»).

Окончательное решение о поддержке проекта и объемах его финансирования принимает Экспертный совет РФФИ, опираясь на оценки экспертов.

Экспериментальные расчеты проводились на модельной базе данных одного из конкурсов РФФИ целевых фундаментальных исследований, которая включала в себя экспертные оценки поддержанных и отклоненных проектов по следующим областям 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 включался в число безусловно поддержанных, когда все эксперты проставляли оценку «5» (то есть $k_{Ai}(r_2)=k_{Ai}(r_3)=k_{Ai}(r_4)=0$) или число оценок «5» было больше или равно числу оценок «4» (то есть $k_{Ai}(r_1)\geq k_{Ai}(r_2)$, $k_{Ai}(r_3)=k_{Ai}(r_4)=0$). Проект включался в число безусловно отклоненных, когда ни один из экспертов не давал оценку «5» (то есть $k_{Ai}(r_1)=0$).

Для построения согласованных групповых правил по безусловной поддержке проекта в указанных областях знаний оказалось достаточно комбинаций только нескольких критериев, а именно: Q_6. Научная значимость проекта; Q_{10}. Завершающая стадия фундаментальных исследований, предлагаемых в проекте; Q_{11}. Масштабы применимости результатов заявленных исследований. Фактически, экспертные оценки только по этим критерии влияли на разбиение проектов по уточненным классам решений и выявление проектов, требующих дополнительного рассмотрения из-за расхождения индивидуальных заключений экспертов. Учет эти факторов обеспечил значительное сокращение времени, требуемого для нахождения классифицирующих признаков и классификации проектов.

Согласованное агрегированное групповое правило для поддержки проекта, записанное на естественном языке, выглядит следующим образом:

«Проект безусловно поддерживается, если он имеет исключительно высокую или значительную научную значимость (оценки x_6^1 или x_6^2); массовый или междисциплинарный масштабы применимости результатов заявленных исследований (оценки x_{11}^1 или x_{11}^2); предлагаемые в проекте исследования завершаются лабораторным образцом или ключевыми элементами разработки (оценки x_{10}^1 или x_{10}^2)».

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

Заключение

Анализ рассматриваемой проблемы является важным этапом процесса выработки решения. Изучение классов объектов и возможных отношений между ними может помочь лицу, принимающему решение, сформулировать стратегии выбора и решающие правила, более адекватные реальности, сделать его решения более обоснованными и разумными. Многоаспектный анализ и структуризация альтернативных вариантов позволяет глубже понять природу проблемы, а значит, и найти ее лучшее решение. Однако есть ситуации, когда известные методы непосредственно неприменимы для анализа и решения проблемы. Наиболее важные особенности этих проблем – множественность и повторяемость данных, которые характеризуют объекты, варианты, альтернативы и их свойства.

В статье предложены средства для обработки и структурирования совокупности объектов, которые описываются многими качественными и/или количественными признаками и могут существовать в нескольких версиях. Предлагаемые методы основаны на теории метрических пространств мультимножеств и более пригодны для работы с подобными объектами, чем другие хорошо известные подходы, использующие функции полезности разного вида [12, 16, 17, 23, 26, 27], ограниченную пороговую предпочтительность (outranking) [24, 25], нечеткие и огрубленные (rough) множества [14, 29]. Применение мультимножеств дает возможность решать традиционные задачи классификации более простым и конструктивным образом и, кроме того, решать новые виды задач, никогда ранее не решавшихся из-за их специфических особенностей, в частности, задачу групповой сортировки объектов по многим качественным критериям.

Методы, основанные на мультимножествах, обеспечивают эффективное решение широкого спектра практических задач принятия многокритериальных решений, например, отбор научно-исследовательских проектов [6, 9, 21], оценка кредитоспособности держателей кредитных карт [22]. Эти средства полезны также для всестороннего анализа многопризнаковых объектов в распознавании образов [1], обработке текстов [20], анализе данных, извлечении знаний и других областях. Еще одним многообещающим приложением теории мультимножеств может оказаться контентный анализ, поиск, упорядочение и классификация мультимедийной информации. Подчеркнем еще раз, что доступная информация, которая содержится в описаниях объектов, используется исключительно в первоначальной форме без каких-либо преобразований качественных данных (символьных, вербальных) в количественные (числовые) данные.

Литература

1. Арлазаров В.Л., Логинов А.С., Славин О.А.. Характеристики программ оптического распознавания текста. //Программирование, 2002, №3, с.45-63.
2. Деза М.М., Лоран М.. Геометрия разрезов и метрик./Пер. с англ. – М.: МЦНМО, 2001.
3. Комарова Н.А., Петровский А.Б. Метод согласованной групповой классификации многопризнаковых объектов. // Поддержка принятия решенений. Труды Института системного анализа Российской академии наук. Т. 35, с.19-32. – М: Издательство ЛКИ, 2008.
4. Кнут Д.Э. Искусство программирования. Т.2. Получисленные алгоритмы./Пер. с англ. – М.: Мир, 1977.
5. Ларичев О.И. Вербальный анализ решений. – М.: Наука, 2006.
6. Ларичев О.И., Прохоров А.С., Петровский А.Б., Стернин М.Ю., Шепелев Г.И. Опыт планирования фундаментальных исследований на конкурсной основе. // Вестник АН СССР, 1989, № 7, с.51-61.
7. Петровский А.Б. Пространства множеств и мультимножеств. – М: Едиториал УРСС, 2003.
8. Петровский А.Б. Кластерный анализ объектов с противоречивыми свойствами. // Труды Десятой национальной конференции по искусственному интеллекту с международным участием (КИИ-2006). В 3-х томах. Т.3, с.755-764. – М.: Физматлит, 2006.
9. А.Б.Петровский, И.П.Тихонов, А.В.Балышев, Н.А.Комарова. Построение согласованных групповых правил для конкурсного отбора научных проектов. // Третья международная конференция «Системный анализ и информационные технологии» (САИТ-2009). Труды конференции, с.337-348. – М.: ПолиПринтСервис, 2009.
10. Anderberg M.R. Cluster Analysis for Applications. – New York: Academic Press, 1973.
11. Blizard W. Multiset theory. // Notre Dame Journal of Formal Logic, 1989, Vol. 30, N 1, p. 36-66
12. Doumpos M., Zopounidis C. Multicriteria Decision Aid Classification Methods. – Dordrecht: Kluwer Academic Publishers, 2002.
13. Furems E. Knowledge-based multi-attribute classification problems structuring. // Ruan D., Montero J., Lu J., Martinez L., D'hondt P., Kerre E. (eds.) Computational Intelligence in Decision And Control, p. 465-470. – Singapore: World Scientific Publisher, 2008.
14. 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, N 2, p. 247-259.
15. Hartigan J.A. Clustering Algoritms. – New York: Wiley, 1975.
16. Hwang C.L., Lin M.J. Group Decision Making under Multiple Criteria. – Berlin: Springer-Verlag, 1987.
17. Köksalan M., Ulu C. An interactive approach for placing alternatives in preference classes. // European Journal of Operational Research, 2003, Vol. 144, N. 2, p. 429-439.
18. Larichev O.I., Olson D.L. Multiple Criteria Analysis in Strategic Siting Problems. – Boston: Kluwer Academic Publishers, 2001.
19. Miyamoto S. Cluster analysis as a tool of interpretation of complex systems. // Working Paper WP-87-41. – Laxenburg, Austria: IIASA, 1987.
20. Petrovsky A.B. Structuring techniques in multiset spaces. // Fandel G., Gal T. (eds.) Multiple Criteria Decision Making, p. 174-184. – Berlin: Springer-Verlag, 1997.
21. Petrovsky A.B. Multi-attribute sorting of qualitative objects in multiset spaces. // Koksalan M., Zionts S. (eds.) Multiple Criteria Decision Making in the New Millenium, p. 124-131. – Berlin: Springer-Verlag, 2001.
22. Petrovsky A.B. Multi-attribute classification of credit cardholders: multiset approach. // International Journal of Management and Decision Making, 2006, Vol. 7, N 2/3, p.166-179.
23. Roubens M. Ordinal multiattribute sorting and ordering in the presence of interacting points of view. // Bouyssou D., Jacquet-Lagrèze E., Perny P., Slowinski R., Vanderpooten D., Vincke P. (eds.) Aiding Decisions with Multiple Criteria: Essays in Honor of Bernard Roy, p. 229-246. – Dordrecht: Kluwer Academic Publishers, 2001.
24. Roy B. Multicriteria Methodology for Decision Aiding. – Dordrecht: Kluwer Academic Publishers, 1996.
25. Roy B., Bouyssou D. Aide Multicritère à la Décision: Méthodes et Cas. – Paris: Economica, 1993.
26. Saaty T. Multicriteria Decision Making: The Analytic Hierarchy Process. – Pittsburgh: RWS Publications, 1990.
27. Vincke P. Multicriteria Decision Aid. – Chichester: Wiley, 1992.
28. Yager R.R. On the theory of bags. // International Journal of General Systems, 1986, Vol. 13, p. 23-37.
29. Zadeh L.A. From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. // IEEE Transactions on Circuits and Systems, 1999, Vol. 45 N 1, p. 105-119.
30. Петровский А.Б. Методы групповой классификации многопризнаковых объектов (часть 1). // Искусственный интеллект и принятие решений, 2009, № 3, с.3-14.

Петровский Алексей Борисович. Заведующий лабораторией методов и систем поддержки принятия решений Института системного анализа РАН. Окончил Московский государственный университет им. М.В.Ломоносова в 1967 году. Доктор технических наук. Автор более 130 научных работ, в том числе 3 монографий и 1 учебника. Области научных интересов: дискретная математика, теория мультимножеств, многокритериальный анализ решений, системы поддержки принятия решений, информационные технологии, системный анализ, научно-техническая политика, прогнозирование, планирование и организация научных исследований.