УДК 681.3

О.Л. Вовк

ПРИМЕНЕНИЕ МУЛЬТИМНОЖЕСТВ ДЛЯ РАНЖИРОВАНИЯ КРИТЕРИЕВ ОЦЕНКИ ДИСЦИПЛИН В ВУЗЕ

Введение. Ранжирование и классификация объектов, заданных векторами противоречивых характеристик, находит свое применение в различных областях научных исследований. Примерами могут служить задачи теории кодирования, классификации и обработки разнородной информации [1].

Мультимножество или множество с повторяющимися элементами [1] 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 рамках введения Болонского процесса в Украине [2] одной из актуальных задач при составлении учебных планов для подготовки бакалавров и магистров является выделение дисциплин, которые будут входить в циклы по выбору студентов. В данной работе предлагается подход, благодаря которому, на основании оценки студентами (бакалаврами и магистрами) набора дисциплин, будут упорядочены критерии, по которым студенты выбирают предметы для обучения.

Цель работы – построение рейтинга критериев оценок дисциплин студентами на базе аппарата мультимножеств.

В соответствии с поставленной целью в работе решены следующие задачи:

- рассмотрены мультимножества как аппарат для многокритериального выбора;
- сформулированы критерии оценки дисциплин студентами, которые их уже изучили;
- построен рейтинг критериев оценки дисциплин.

Основные понятия и определения теории мультимножеств. В работах [1,3] введены базовые понятия теории мультимножеств. Рассмотрим основные из них.

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 $\mathbb{A} = \{A_1,...,A_k\}$ 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 $\mathbb{A}$ 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 $A_i=(q_{i1}^{e_1}, q_{i2}^{e_2}, \ldots, q_{im}^{e_m})$.

Мультимножеством $\boldsymbol{A}$, порожденным обычным множеством $U=\{x_1,x_2,\ldots\}$, все элементы которого различны, называется совокупность групп элементов вида $A=\{k_A(x)\bullet x|x\in U,\ k_A(x)\in Z_+\}$. Здесь $k_A:U\to Z_+=\{0,1,2,\ldots\}$ называется функцией числа экземпляров мультимножества, определяющей кратность вхождения элемента $x_i\in U$ в мультимножество A, что обозначено символом •.

Если $k_A(x)=\chi_A(x)$, где $\chi_A(x)=1$ при $x\in A$ и $\chi_A(x)=0$ при $x\notin A$, то мультимножество A становится обычным множеством.

Если все мультимножества семейства $A=\{A_1,A_2,\ldots\}$ образуются из элементов множества G, то G называется доменом для семейства A, а множество $\mathrm{Supp}A=\{x|x\in G,\ \chi_{\mathrm{Supp}A}(x)=\chi_A(x)\}$ – опорным множеством или носителем мультимножества A.

Мощность мультимножества $|A|=\Sigma_x k_A(x)$ определяется как общее число экземпляров всех его элементов; размерность мультимножества $/A/=\Sigma_x\chi_A(x)=|\mathrm{Supp}A|$ – как общее число различных элементов. Максимальное значение функции кратности $\mathrm{hgt}A=\max\limits_{x\in G} k_A(x)$ называется высотой, а элемент

$x_{A*}=\arg\max_{x\in G} k_A(x)$ – пиком мультимножества A. Мультимножество называется пустым $\varnothing$, если $k_\varnothing(x)=0$, и максимальным Z, если $k_Z(x)=\max_{A\in \mathrm{A}} k_A(x)$, $\forall x\in U$.

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

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

Мультимножества A и B будем называть одноименно или S-эквивалентными ($A\cong B$), если их носители совпадают ($\mathrm{Supp}A=\mathrm{Supp}B$) и существует взаимно однозначное соответствие f между одноименными компонентами: $k_B(x)=f(k_A(x))$, $\forall x \in G$; разноименно или D-эквива-лентными ($A\approx B$), если их носители эквивалентны ($\mathrm{Supp}A\sim\mathrm{Supp}B$) 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 Z_+.

Введем следующие основные операции над мультимножествами:

- объединение

$A\bigcup B = \{k_{A\cup B}(x)\bullet x \mid k_{A\cup B}(x)=\max(k_A(x), k_B(x))\}$;

- пересечение

$A\bigcap B = \{k_{A\cap B}(x)\bullet x \mid k_{A\cap B}(x)=\min(k_A(x), k_B(x))\}$;

- арифметическое сложение

$A+B = \{k_{A+B}(x)\bullet x \mid k_{A+B}(x)=k_A(x)+k_B(x)\}$;

- арифметическое вычитание

$A-B = \{k_{A-B}(x)\bullet x \mid k_{A-B}(x)=k_A(x)-k_{A\cap B}(x)\}$;

- симметрическая разность

$A\Delta B = \{k_{A\Delta B}(x)\bullet x \mid k_{A\Delta B}(x)=|k_A(x)-k_B(x)|\}$;

- дополнение

$\overline{\boldsymbol{A}} = Z-A = \{k_{\overline{A}}(x)\bullet x \mid k_{\overline{A}}(x)=k_Z(x)-k_A(x)\}$.

Другие операции, а также способы определения носителей операций над мультимножествами рассмотрены подробно в [3].

Действительная неотрицательная функция $m(A)$, определенная на алгебре L (Z) и удовлетворяющая условию коаддитивности: $m(A)+m(B)=m(A+B)$, называется мерой мультимножества. Мера мультимножества $m(A)$ обладает следующим свойствами: $m(\varnothing)=0$; монотонность $m(A)\le m(B)\Leftrightarrow A\subseteq B$; непрерывность $\lim_{i\to\infty} m(A_i)=m(\lim_{i\to\infty} A_i)$; симметричность $m(A)+m(\overline{\boldsymbol{A}})=m(Z)$; эластичность $m(h\bullet A)=hm(A)$. Меру мультимножества можно определить различными способами, например, как линейную комбинацию функций кратности: $m(A)=\sum_j w_j k_A(x_j)$, $w_j>0$. Заметим, что мощность мультимножества $|A|$ также будет мерой мультимножества

Метрические пространства мультимножеств (A ,d) введены в [3], где определены следующие виды расстояний между мультимножествами:

$$d_1(A,B) = m(A\Delta B);\quad d_2(A,B) = m(A\Delta B)/m(Z);$$
$$d_3(A,B) = m(A\Delta B)/m(A\bigcup B).$$

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

Построение модели оценки дисциплин студентами. Предварительно отметим, что проводился опрос студентов специальности «Программное обеспечение автоматизированных систем» Донецкого национального технического университета.

В качестве дисциплин-объектов для анализа A предлагаются следующие дисциплины:

- A_1 – функциональное и логическое программирование;
- A_2 – автоматизированная разработка программного обеспечения;
- A_3 – программирование в Интернет;
- A_4 – основы автоматизированного проектирования систем;
- A_5 – системное программирование и операционные системы;
- A_6 – графическое и геометрическое моделирование;
- A_7 – системы искусственного интеллекта;
- A_8 – организация функционирования ЭВМ.

Введем множество Q – множество критериев оценки дисциплин студентами:

- Q_1 – получение новых знаний о языках программирования (возможные ответы: $q_1^1 =$ да, $q_1^2 =$нет);
- Q_2 - получение новых знаний о средах разработки программ (возможные ответы: $q_2^1 =$ да, $q_2^2 =$нет);
- Q_3 - получение знаний о проектировании и описании программ (возможные ответы: $q_3^1 =$ да, $q_3^2 =$нет);
- Q_4 - получение фундаментальных теоретических знаний по специальности (возможные ответы: $q_4^1 =$ да, $q_4^2 =$нет);
- Q_5 - уровень изложения материала (возможные ответы: $q_5^1 =$ высокий, $q_5^2 =$средний, $q_5^3 =$низкий);
- Q_6 - развитие логического и образного мышления (возможные ответы: $q_6^1 =$ да, $q_6^2 =$нет);
- Q_7 - использование материалов на практике (возможные ответы: $q_7^1 =$ да, $q_7^2 =$нет).

Кроме оценки каждой из предлагаемой дисциплины множества A по критериям множества Q каждому эксперту (студенту) предлагалось отнести каждую дисциплину к одному из множеств X_a ($r_a =$дисциплина, необходимая при подготовке специалиста) или X_b ($r_b =$дисциплина, которая не нужна при подготовке специалиста).

Результаты оценок дисциплин по критериям. В качестве экспертов для оценки дисциплин было выбрано 30 студентов четвертого и пятого курса специальности «Программное обеспечение автоматизированных систем» Донецкого национального технического университета, 12 из которых проходят подготовку в качестве магистров, а 18 – в качестве специалистов.

Результаты анкетирования экспертов приведены в таблице 1.

Объединим объекты A_i, относящиеся к заданным классам X_a и X_b. Получим преобразованную таблицу решений, строки которой соответствуют мультимножествам X_a и X_b (таблица 2). Считаем, что дисциплина A_i относится к классу X_a, если $r_a(A_i) > r_b(A_i)$, иначе дисциплина относится к классу X_b, для $i \in [1,8]$.

Рассчитаем расстояние d_1 для каждого критерия согласно формуле:

$$d_1(Q_{sa}*,Q_{sb}*) = \sum_{x\in Qs*}|k_{Xa}(x_j)–k_{Xb}(x_j)|, \quad d_1(R_a,R_b) = \sum_{x\in R}|k_{Xa}(x_j)–k_{Xb}(x_j)|.$$

Результаты расчета приведены в таблице 3.

Оценим точность аппроксимации по s-ой группе признаков ($s \in [1,7]$):

$$\rho_s = d(Q_{sa}*, Q_{sb}*)/d(R_a, R_b).$$

Результаты оценки точности аппроксимации даны в таблице 4.

Выберем аппроксимирующие признаки q_s* для каждого критерия (аппроксимирующим считаем тот признак, для которого выполняется условие для данных таблицы 2 $X_a(q_s^*) > X_b(q_s^*)$):

$$\{q_s^*\}=\{q_1^1,q_2^1,q_3^1,q_4^1,q_5^1,q_6^1,q_7^1\}.$$

Упорядочим аппроксимирующие признаки q_s* по убыванию точности аппроксимации ρ_s:

$$\{q_s^*\}=\{q_7^1,q_2^1,q_1^1,q_4^1,q_5^1,q_6^1,q_3^1\}.$$

Выбрав некоторое желаемое значение точности аппроксимации ρ_0, получим обобщенные решающие правила для отбора дисциплин. Предварительно отметим, что критерии q_6^1, q_3^1 не могут быть значимыми в связи с низкой точностью аппроксимации.

«Материалы дисциплины должны использоваться на практике» (оценка q_7^1, точность аппроксимации ρ_0=0,8).

«Материалы дисциплины должны использоваться на практике и дисциплина должна давать знания о средах разработки программ» (оценки q_7^1, q_2^1, точность аппроксимации ρ_0=0,52).

«Материалы дисциплины должны использоваться на практике, дисциплина должна давать знания о средах разработки программ и языках программирования» (оценки q_7^1, q_2^1, q_1^1, точность аппроксимации ρ_0=0,44).

Ранжирование аппроксимирующих признаков по величине расстояния d_1 показывает, что наиболее важным при выборе дисциплины студентом является критерий Q_7, характеризующий применение материалов дисциплины на практике.

Таблица 1. Результаты анкетирования экспертов

	q_1^1	q_1^2	q_2^1	q_2^2	q_3^1	q_3^2	q_4^1	q_4^2	q_5^1	q_5^2	q_5^3	q_6^1	q_6^2	q_7^1	q_7^2	r_a	r_b
A_1	29	1	20	10	15	15	18	12	16	13	1	26	4	10	20	14	16
A_2	15	15	21	9	27	3	19	11	10	12	8	21	9	18	12	28	2
A_3	25	5	15	15	8	22	19	11	4	14	12	8	22	18	12	26	4
A_4	4	26	8	22	11	19	5	25	3	17	10	11	19	3	27	9	21
A_5	28	2	26	4	21	9	22	8	18	11	1	19	11	28	2	30	0
A_6	12	18	20	10	7	23	16	14	11	16	3	25	5	15	15	27	3
A_7	11	19	14	16	16	14	13	17	3	15	12	11	19	4	26	13	17
A_8	9	21	7	23	18	12	21	9	8	11	11	19	11	12	18	14	16

Таблица 2. Результаты разделения дисциплин на принятые X_a и непринятые X_b

	q_1^1	q_1^2	q_2^1	q_2^2	q_3^1	q_3^2	q_4^1	q_4^2	q_5^1	q_5^2	q_5^3	q_6^1	q_6^2	q_7^1	q_7^2	r_a	r_b
X_a	80	40	82	38	63	57	76	44	43	53	24	73	47	79	41	111	9
X_b	53	67	49	71	60	60	57	63	30	56	34	67	53	29	91	50	70

Таблица 3. Результаты расчета расстояний d_1

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7
d_1	54	66	6	38	26	12	100

Таблица 4. Результаты оценки точности аппроксимации

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7
ρ_s	0,443	0,541	0,049	0,311	0,213	0,098	0,82

Заключение. Классификация и ранжирование объектов, заданных противоречивыми признаками является сложной и одновременно актуальной задачей. Аппарат мультимножеств, рассмотренный в работе, дает возможность решения задач многокритериального выбора. Построение обобщенного решающего правила для отбора дисциплин студентами является актуальной задачей в период внедрения положений Болонской декларации в систему высшего образования.

Результаты, полученные в работе, показывают, что наиболее интересными для студентов (для конкретной рассматриваемой специальности) являются дисциплины, материалы которых находят свое применение на практике.

В дальнейшем планируется построение обобщающих правил ранжирования критериев выбора дисциплин студентами ВУЗов вне зависимости от специальности, по которой они обучаются.

ЛИТЕРАТУРА:

1. Петровский А.Б. Упорядочение и классификация объектов с противоречивыми признаками // Новости искусственного интеллекта. – 2003. – №4. – 17 с.
2. Наказ № 49 від 23.01.2004р. "Про затвердження Програми дій щодо реалізації положень Болонської декларації в системі вищої освіти і науки України на 2004-2005 роки.", http://www.mon.gov.ua/
3. Петровский А.Б. Пространства множеств и мультимножесв. – Москва: Едиториал УРСС, 2003. – 248 с.