УДК 510.254:517.977.5

Г.А. Кравцов, канд. техн. наук,
Ин-т проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины
(Украина, 03164, Киев-164, ул.Генерала Наумова, 15,
e-mail: hryhoriy.kravtsov@gmail.com)

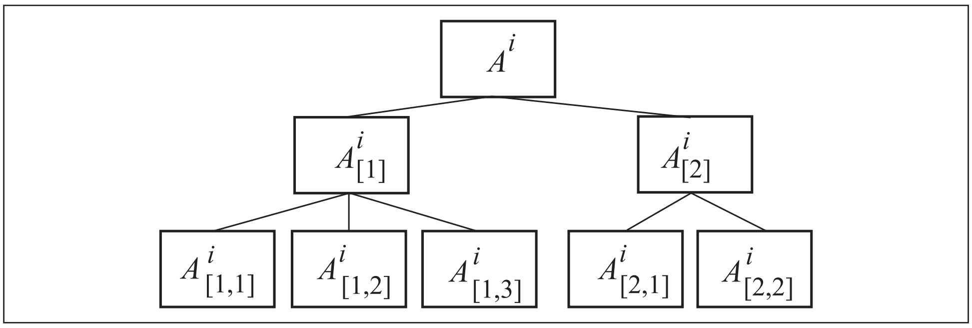
Модель вычислений на классификациях

Представлены результаты теоретических исследований модели вычислений на классификациях. На основе математической структуры «дерево» введен набор операций, позволяющих определить меру на плоских классификациях. Классификации, имеющие несколько плоскостей деления, представляют собой метрическое пространство и рассматриваются как логичное развитие плоских классификаций, применяемых, в частности, в задачах подбора экспертов. Показано соответствие полученной модели принципам математического моделирования.

Подано результати теоретичних досліджень моделі обчислень на класифікаціях. На основі математичної структури «дерево» введено набір операцій, які дозволяють визначити міру на плоских класифікаціях. Класифікації з декількома площинами поділу є метричним простором і розглядаються як логічний розвиток плоских класифікацій, що застосовуються, зокрема, в задачах підбору експертів. Показано, що отримана модель відповідає принципам математичного моделювання.

Ключевые слова: плоская классификация, пространственная классификация, мера, относительное расстояние, абсолютное расстояние, модель вычислений.

В настоящее время весьма актуальными являются исследования, связанные с применением классификаций в информационных технологиях [1]. Например, известная проблема подбора экспертов [2] имеет решение, если знания экспертов классифицированы в соответствии с некоторой произвольной классификацией. Тогда решение представляется как нахождение меры подобия (сходства) или меры отличия классифицированных знаний эксперта и классифицированных знаний, необходимых для решения некоторой проблемы (задачи). Одной из интерпретаций проблемы подбора экспертов является проблема подбора рецензента научной статьи. В этом случае классификацией знаний может быть Универсальная десятичная классификация (УДК) [3]. Данными о знаниях эксперта являются его публикации с кодами УДК.



Проблема поиска рецензента имеет простое решение, если среди публикаций рецензентов найдена работа, код УДК которой полностью совпадает с кодом УДК рецензируемой статьи. В противном случае требуется найти меру сходства или меру различия между двумя кодами, что является весьма нетривиальным решением при условии специальных определителей УДК, образующих дополнительные плоскости деления классификации. Можно принять УДК за иерархию и воспользоваться методом анализа иерархий (МАИ) [4]. Однако МАИ на практике малоприменим для таких задач, так как метод парных сравнений, лежащий в основе МАИ, требует значительных временных ресурсов экспертов для парного сравнения более чем 70 тыс. классов УДК.

В работе [5] решение проблемы подбора эксперта получено на основании использования ключевых слов, взаимооценок, самооценок и др. Для публикаций, не имеющих ключевых слов, можно было бы воспользоваться методами машинного обучения извлечению ключевых слов из текста, однако проблема автоматического извлечения ключевых слов остается нерешенной [6]. В то же время, можно рассматривать ключевые слова как элементы классификации. При этом вопрос равнозначности ключевых слов остается открытым.

Следовательно, актуальной задачей является разработка модели вычислений на классификациях, которая позволит представить классификацию как метрическое пространство и определить меру схожести или меру отличия на классификации. Кроме того, в известных работах, посвященных исследованию или применению классификаций, рассматривается одна плоскость деления классификации, что не позволяет эти результаты применить к исследованию классификаций с несколькими плоскостями деления.

Понятие классификации и ее формальное представление. Согласно [7] классификация — это особого вида деление или система мериологических или таксономических делений. Формальное представление классификации: это математическая структура ориентированного дерева, в котором вершинами являются математические классы [8, 9].

Класс в классификации — это тройка $< F, i, Y >$, обозначаемая как F_Y^i, где F — классификация; i — плоскость деления классификации F; Y — путь уточнения плоской классификации F^i. Путь уточнения Y — это последовательность вида $[a, b, c, \ldots]$, однозначно определяющая место класса в некоторой плоскости деления классификации. Путь уточнения однозначно определяет ранг [10] класса в классификации, который равен длине пути уточнения. Так, ранг $A_{[a]}$ равен единице, а ранг $A_{[a,\, b]}$ равен двум.

Рассмотрим использование обозначений на некоторой условной классификации A^i (см. рисунок). Семантически класс в классификации сам является классификацией. Например, как видно из рисунка, класс $A_{[1]}^i$ — это классификация из двух классов: $A_{[1,\,1]}^i$, $A_{[1,\,2]}^i$ и $A_{[1,\,3]}^i$. Очевидно, что $A_{[2,\,1]}^i \neq A_{[1,\,2]}^i$, и тогда можно записать

$$A_I^i \neq A_J^i, \; I \neq J. \tag{1}$$

Поскольку A^i представляет собой одну из плоскостей деления классификации A с единым корнем ориентированного дерева, семантически A^i и A суть одно и то же.

Операции над классами плоской классификации. Пусть даны классы $A_{[a]}^i$, $A_{[b]}^i$, $A_{[a,\,b]}^i$ классификации A. Представим ассоциативную бинарную операцию обобщения классов классификации в виде

$$A_{[a]}^i \cdot A_{[b]}^i = A,$$

$$A_{[a]}^i \cdot A_{[a,\,b]}^i = A_{[a]}^i,$$

$$A_{[b]}^i \cdot A_{[a,\,b]}^i = A,$$

$$A_{[a]}^i \cdot A = A.$$

Здесь A является идемпотентом или нулем операции обобщения классов относительно самой себя, $A \cdot A = A$, и всех уточняющих классов классификации $A_I^i \cdot A = A$.

Ассоциативную бинарную операцию уточнения класса «+» можно записать в виде

$$A_{[a]}^i + A = A_{[a]}^i,$$

$$A_{[a]}^i + A_{[b]}^i = A_{[a]}^i + A_{[b]}^i,$$

$$A_{[a]}^i + A_{[a,\,b]}^i = A_{[a,\,b]}^i,$$

$$A_{[b]}^i + A_{[a,\,b]}^i = A_{[b]}^i + A_{[a,\,b]}^i.$$

Семантически операция уточнения класса эквивалентна дуге дерева [9] в случаях $A^i_{[a]} + A = A^i_{[a]}$ и $A^i_{[a]} + A^i_{[a,\, b]} = A^i_{[a,\, b]}$, что отражает семантику уточнения одного класса другим при условии, что оба класса находятся на разных ветках дерева.

Операции обобщения и уточнения класса необходимы при построении и доказательстве существования метрики на классификациях. Наличие бинарной ассоциативной операции и идемпотенты (1) позволяет сделать вывод о том, что классификация представляет собой конечную полугруппу классов [8].

Метрическое пространство плоской классификации. Относительным расстоянием R между двумя классами, $A^i_{[a]}$ и $A^i_{[b]}$, классификации A назовем неотрицательное целое число, равное числу уникальных операций уточнения от ближайшего общего обобщающего класса. Так, $R(A^i_{[a]}, A^i_{[b]}) = 2$ потому, что потребуется выполнить две операции уточнения, $A + A^i_{[a]} = A^i_{[a]}$ и $A + A^i_{[b]} = A^i_{[b]}$, к каждому классу из ближайшего обобщающего класса A.

Согласно данному определению относительного расстояния выполняются равенства

$$R(A^i_I, A^i_Y) = R(A^i_Y, A^i_I),$$
$$R(A^i_I, A^i_I) = 0, \tag{2}$$

где I, Y — произвольные пути уточнения (деления) классификации A в плоскости деления i. Расстояние $R(A^i_I, A^i_Y)$ названо относительным, так как оно измеряется относительно выбранных классов и никак не учитывает положение этих классов в классификации.

Рассмотрим пример. Найдем относительное расстояние $R(A^i_{[a,b,c,d]}$, $A^i_{[a,b,c,k]})$, воспользовавшись заменой переменной $B^i = A^i_{[a,b,c]}$, и представим $R(A^i_{[a,b,c,d]}, A^i_{[a,b,c,k]})$ в виде $R(B^i_{[d]}, B^i_{[k]})$. По аналогии с $R(A^i_{[a]}, A^i_{[b]}) = 2$ получим $R(B^i_{[d]}, B^i_{[k]}) = 2$ или $R(A^i_{[a,b,c,d]}, A^i_{[a,b,c,k]}) = 2$.

Следует заметить, что относительное расстояние $R(A^i_I, A^i_Y)$ всегда единственно и конечно в силу того, что классификация A есть ориентированное дерево [9]. Относительное расстояние $R(A^i_I, A^i_Y)$ инвариантно, если класс A, уточняемый классами A^i_I и A^i_Y, становится уточняющей классификацией (классом) классификации B. Поскольку структура деления класса A неизменна, то и путь между двумя классами, A^i_I и A^i_Y, в A неизменен.

Относительное расстояние $R(A_I^i, A)$ есть ранг [10]. Для относительного расстояния между двумя классами, A_I^i и A_Y^i, в A справедливо равенство

$$R(A_I^i, A_Y^i) = R(A, A_I^i) + R(A, A_Y^i) - 2R(A, A_I^i \cdot A_Y^i). \tag{3}$$

Поэтому для $A_{[a]}^i$ и $A_{[a,\, b]}^i$ исоответственно $R(A_{[a]}^i, A) = 1$ и $R(A_{[a,\, b]}^i, A) = 2$.

Выражение (3) описывает связь относительного расстояния между классами A_I^i и A_Y^i и абсолютными расстояниями до этих классов от вершины классификации A и относительного расстояния от вершины классификации до общего обобщающего класса $A_I^i \cdot A_Y^i$. Если $R(A, A_I^i \cdot A_Y^i) = 0$, то (3) принимает вид

$$R(A_I^i, A_Y^i) = R(A, A_I^i) + R(A, A_Y^i). \tag{4}$$

Если $R(A, A_I^i \cdot A_Y^i) > 0$, то для (3) справедливо неравенство

$$R(A_I^i, A_Y^i) < R(A, A_I^i) + R(A, A_Y^i). \tag{5}$$

На основании (2) равенство (4) и неравенство (5) могут быть обобщены к виду

$$R(A_I^i, A_Y^i) \le R(A_I^i, A) + R(A, A_Y^i), \tag{6}$$

что соответствует условию неравенства треугольника.

Равенства (2), в которых определены свойства симметричности и рефлексивности, и неравенство треугольника (6) позволяют сделать вывод о том, что относительное расстояние $R(A_I^i, A_Y^i)$ есть мера на классификации A. Однако относительное расстояние в качестве меры оказывается малополезным, так как не отвечает на вопрос о том, насколько похожи или различны два класса классификации.

Мера сходства [11] для двух классов одной плоскости деления классификации может быть определена как мера Жоккара [12], если положить, что для множественной меры Жоккара

$$K_{-1,\,1} = \frac{n(A \cap B)}{n(A \cdot B)} \tag{7}$$

определено следующее соответствие:

$$O(A_I^i, A_Y^i) = \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i + A_Y^i) + 1} =$$

$$= \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i \cdot A_Y^i) + R(A_I^i, A_I^i \cdot A_Y^i) + R(A_Y^i, A_I^i \cdot A_Y^i) + 1}, \tag{8}$$

где A_I^i и A_Y^i — классы в A. Добавление единицы объясняется приведением к множеству вершин графа дерева, в котором для одной ветки число вершин на единицу больше числа связей (согласно свойствам дерева), чтобы мера (8) семантически была эквивалентна мере Жоккара (7). В выражении (8) преобразование знаменателя корректно в силу определения относительного расстояния. Для удобства меру (8) обозначим $O(A^i)$ и назовем мерой сходства на плоскости деления классификации A^i. Мера сходства (8) является конкурентной [11].

Покажем, что мера (8) всегда равна единице. Согласно (2) $R(A_I^i, A_I^i \cdot A_Y^i)=0$ и $R(A_Y^i, A_I^i \cdot A_Y^i)=0$ при $A_I^i = A_Y^i$, а следовательно,

$$O(A_I^i, A_Y^i)=\frac{R(A, A_I^i \cdot A_Y^i)+1}{R(A, A_I^i \cdot A_Y^i)+1}=1,$$

что и требовалось доказать.

Строго говоря, мера (8) не является математической мерой в силу невыполнения условия неравенства треугольника [11]. Однако следует заметить, что равенство $\overline{O}(A_I^i, A_Y^i)=1-O(A_I^i, A_Y^i)$ обладает свойствами симметричности и рефлексивности:

$$\begin{aligned}\overline{O}(A_I^i, A_Y^i)&=\overline{O}(A_Y^i, A_I^i),\\ \overline{O}(A_I^i, A_I^i)&=0.\end{aligned} \tag{9}$$

Пусть даны три меры: $\overline{O}(A_I^i, A_Y^i)$, $\overline{O}(A_I^i, A)$ и $\overline{O}(A_Y^i, A)$. Воспользовавшись (8), получим

$$\overline{O}(A_I^i, A_Y^i)=1-\frac{R(A, A_I^i \cdot A_Y^i)+1}{R(A, A_I^i \cdot A_Y^i)+R(A_I^i, A_I^i \cdot A_Y^i)+R(A_Y^i, A_I^i \cdot A_Y^i)+1}, \tag{10}$$

$$\overline{O}(A_I^i, A)=1-\frac{1}{R(A_I^i, A)+1}, \tag{11}$$

$$\overline{O}(A_Y^i, A)=1-\frac{1}{R(A_Y^i, A)+1}. \tag{12}$$

Относительное расстояние $R(A_I^i, A)=R(A, A_I^i \cdot A_Y^i)+R(A_I^i, A_I^i \cdot A_Y^i)$ по определению аналогично расстоянию $R(A_Y^i, A)=R(A, A_I^i \cdot A_Y^i)+R(A_Y^i, A_I^i \cdot A_Y^i)$. Тогда (11) и (12) могут быть представлены в виде

$$\overline{O}(A_I^i, A)=\frac{R(A, A_I^i \cdot A_Y^i)+R(A_I^i, A_I^i \cdot A_Y^i)}{R(A, A_I^i \cdot A_Y^i)+R(A_I^i, A_I^i \cdot A_Y^i)+1}, \tag{13}$$

$$\overline{O}\,(A_Y^i\,, A) = \frac{R\,(A,\, A_I^i \cdot A_Y^i\,) + R\,(A_Y^i\,,\, A_I^i \cdot A_Y^i\,)}{R\,(A,\, A_I^i \cdot A_Y^i\,) + R\,(A_Y^i\,,\, A_I^i \cdot A_Y^i\,) + 1}\,, \tag{14}$$

а (10) — в виде

$$\overline{O}\,(A_I^i\,, A_Y^i\,) = \frac{R\,(A_I^i\,,\, A_I^i \cdot A_Y^i\,) + R\,(A_Y^i\,,\, A_I^i \cdot A_Y^i\,)}{R\,(A,\, A_I^i \cdot A_Y^i\,) + R\,(A_I^i\,,\, A_I^i \cdot A_Y^i\,) + R\,(A_Y^i\,,\, A_I^i \cdot A_Y^i\,) + 1}\,.$$

Выполним замену переменных $a = R\,(A,\, A_I^i \cdot A_Y^i)$, $b = R\,(A_I^i\,,\, A_I^i \cdot A_Y^i\,)$ и $c = R\,(A_Y^i\,,\, A_I^i \cdot A_Y^i)$. Выразим $\overline{O}\,(A_I^i\,, A) + \overline{O}\,(A_Y^i\,, A) - \overline{O}\,(A_I^i\,, A_Y^i\,)$ через переменные a, b и c:

$$\overline{O}(A_I^i\,, A) + \overline{O}(A_Y^i\,, A) - \overline{O}(A_I^i\,, A_Y^i\,) = \frac{a+b}{a+b+1} + \frac{a+c}{a+c+1} - \frac{b+c}{a+b+c+1}\,. \tag{15}$$

Выполним замену переменной $\overline{O}\,(A_I^i\,, A) + \overline{O}\,(A_Y^i\,, A) - \overline{O}\,(A_I^i\,, A_Y^i\,) = f\,(a,\, b,\, c)$ и приведем правую часть (15) к общему знаменателю $W = (a+b+1)(a+c+1)(a+b+c+1)$. Тогда получим $f\,(a,\, b,\, c) = V\,/\,W$, где $V = 2a^3 + 3a^2 b + 3a^2 c + 4a^2 + ab^2 + 4abc + 3ab + ac^2 + 3ac + 2a + b^2 c + bc^2 + 2bc$. Поскольку $a \ge 0$, $b \ge 0$ и $c \ge 0$ (по определению относительного расстояния), то $W \ge 1$ и $V \ge 0$, откуда следует $\overline{O}\,(A_I^i\,, A) + \overline{O}\,(A_Y^i\,, A) - \overline{O}\,(A_I^i\,, A_Y^i\,) \ge 0$ или

$$\overline{O}\,(A_I^i\,, A) + \overline{O}\,(A_Y^i\,, A) \ge \overline{O}\,(A_I^i\,, A_Y^i\,)\,. \tag{16}$$

Неравенство (16) — это неравенство треугольника. Следовательно, с учетом (9) и (16) $\overline{O}\,(A_I^i\,, A_Y^i\,)$ — мера отличия на классификации. Отсюда можно сделать вывод о том, что классификации с несколькими плоскостями деления являются метрическими пространствами.

Мера отличия (16) обладает следующим свойством: если выбрать два произвольных класса, относительное расстояние между которыми постоянно и конечно, то мера отличия (16) убывает с увеличением относительного расстояния от самой общей классификации до ближайшего обобщающего класса выбранных классов. Так, если $C = R\,(A_I^i\,,\, A_I^i \cdot A_Y^i\,) + R\,(A_Y^i\,,\, A_I^i \cdot A_Y^i\,)$, то

$$\overline{O}\,(A_I^i\,, A_Y^i\,) = 1 - \frac{R\,(A,\, A_I^i \cdot A_Y^i\,) + 1}{R\,(A,\, A_I^i \cdot A_Y^i\,) + C + 1}\,,$$

где $C \ge 0$ по определению относительного расстояния. Следовательно,

$$\overline{O}\,(A_I^i\,, A_Y^i\,) = 1 - \lim_{R(A,\, A_I^i \cdot A_Y^i) \to \infty} \frac{R\,(A,\, A_I^i \cdot A_Y^i\,) + 1}{R\,(A,\, A_I^i \cdot A_Y^i\,) + C + 1} = 0\,.$$

В то же время,

$$\overline{O}\,(A_I^i,\ A_Y^i)=1-\lim_{R\,(A,\ A_I^i\cdot A_Y^i)\to 0}\ \lim_{C\to 1}\ \frac{R\,(A,\ A_I^i\cdot A_Y^i)+1}{R\,(A,\ A_I^i\cdot A_Y^i)+C+1}=\frac{1}{2}. \qquad (17)$$

Выражение (17) свидетельствует о том, что если известен только один уточняющий класс классификации, то мера различия между ним и его обобщающей классификацией равна 1/2 или эквивалентна вероятности, что классифицируемый объект либо попадает в уточняющий класс, либо не попадает с вероятностью 1/2. Уточнение классификации по одному классифицируемому признаку на каждом уровне приводит к следующей мере различия:

$$\overline{O}\,(A_I^i,\ A_Y^i)=1-\lim_{R\,(A,\ A_I^i\cdot A_Y^i)\to 0}\ \frac{R\,(A,\ A_I^i\cdot A_Y^i)+1}{R\,(A,\ A_I^i\cdot A_Y^i)+C+1}=\frac{C}{C+1}. \qquad (18)$$

Если в выражение (11) вместо константы C подставить удвоенную максимальную длину пути уточнения (максимальный ранг), то получим максимальное значение меры отличия на плоскости деления классификации A^i. Из (18) следует

$$O\,(A_I^i,A_Y^i)=\lim_{R\,(A,\ A_I^i\cdot A_Y^i)\to 0}\ \frac{R\,(A,\ A_I^i\cdot A_Y^i)+1}{R\,(A,\ A_I^i\cdot A_Y^i)+C+1}=\frac{C}{C+1}.$$

Формализация модели. Если определены плоские классификации $F^1, ..., F^i, ..., F^N$, $i=\overline{1,\ N}$, совокупность которых представляет собой классификацию $F=F^1\times F^2\times ... \times F^N$ с мерой отличия, то

$$\overline{O}\,(F)=\frac{1}{N}\sum_{i=1}^{N}\overline{O}\,(F^i).$$

Мера (11) есть среднее арифметическое мер отличия плоских классификаций. Несложно показать, что если $O\,(F^i)=1$, $i=\overline{1,N}$, то $\overline{O}\,(F)=1$. Если существуют такие $p\,(F^i)$, $i=\overline{1,\ N}$, что $K=\sum_{i=1}^{N} p\,(F^i)$, то

$$\overline{O}_K\,(F)=\frac{1}{NK}\sum_{i=1}^{N}\overline{O}\,(F^i)\,p\,(F^i) \qquad (19)$$

есть нормированная мера отличия на классификации F, в которой $p\,(F^i)$ — вес или степень влияния F^i на нормированную меру отличия $\overline{O}_K\,(F)$. Очевидно, что если $p\,(F^i)=1$, то $\overline{O}_K\,(F)=\overline{O}(F)$. Также очевидно, что если существуют такие весовые коэффициенты $p(F^i)$, из которых один, например $p(F^y)$, равен единице при $0< y< N$, а остальные равны нулю ($i=\overline{1, N}$, $i\neq y$), то $\overline{O}_K(F)=\overline{O}(F^y)$. Несложно заметить, что в выражении (19) отношение $\dfrac{p(F^i)}{K}$ есть вероятность, так как $K=\sum_{i=1}^{N} p(F^i)$. Следовательно, $\overline{O}_K(F)$ есть математическое ожидание меры отличия классификации, если $\overline{O}(F^i)$ считать случайными величинами.

Если для любых $p(F^k)$ и $p(F^{k+1})$ выполняется неравенство $p(F^{k+1})>$ $> p(F^k)$ так, что справедливо рекурсивное неравенство

$$\min(O(F^{k+1}))p(F^{k+1})>\sum_{i=1}^{k}\max(O(F^i))p(F^i), \tag{20}$$

то мера (19) обладает определенной степенью доверия.

Максимальное значение сходства по определению $\max(O(F^{k+1}))=1$. Полагая, что L_{F^k} есть максимальный ранг плоской классификации F^k, неравенство (20) можно представить в виде

$$p(F^{k+1})>(2L_{F^{k+1}}+1)\sum_{i=1}^{k} p(F^i). \tag{21}$$

Полученная система

$$\begin{gathered}\overline{O}_K(F)=\frac{1}{NK}\sum_{i=1}^{N}\overline{O}(F^i)\,p(F^i),\\ p(F^{k+1})>(2\,L_{F^{k+1}}+1)\sum_{i=1}^{k} p(F^i)\end{gathered} \tag{22}$$

представляет собой модель вычисления на классификациях с определенной степенью доверия. Свойствами такой модели являются:

независимость от принципа выбора плоскостей деления классификации;

масштабируемость, ограниченная лишь целесообразностью;

определенность влияния неучтенных плоскостей деления классификации F.

Поясним последнее свойство. Поскольку каждый последующий весовой коэффициент больше предыдущего, $p(F^{k+1})>p(F^k)$, и выполняется условие (21), то

$$\dot{K}=\frac{p(F^1)}{2\,L_{F^1}+1}+K$$

есть теоретическая сумма весовых коэффициентов, где $\frac{p(F^1)}{2L_{F^1}+1}$ — максимальная абсолютная погрешность вычисления $\dot{K}$ при наличии неучтенных плоскостей деления классификации F.

Теоретическая мера отличия имеет вид

$$\overline{O}_{\dot{K}}(F)=\frac{p(F^1)+\sum_{i=1}^{N}\overline{O}(F^i)p(F^i)}{(N+1)((2L_{F^1}+1)K+p(F^1))}.$$

Относительная погрешность ρ модели равна $\frac{\overline{O}_K(F)-\overline{O}_{\dot{K}}(F)}{\overline{O}_K(F)}$ и отражает степень доверия к модели. Если положить $\overline{O}(F^i)=1$, то получим $\overline{O}_K(F)=1/N$ и

$$\overline{O}_{\dot{K}}(F)=\frac{p(F^1)+K}{(N+1)((2L_{F^1}+1)K+p(F^1))}.$$

Следовательно,

$$\rho=1-\frac{N(p(F^1)+K)}{(N+1)(2L_{F^1}K+K+p(F^1))}.$$

Если для для любых $p(F^k)$ и $p(F^{k+1})$ выполняется неравенство $p(F^{k+1})\geq p(F^k)$ или выполняется неравенство $p(F^{k+1})>p(F^k)$, но не выполняется условие (20), то возникают следующие трудности.

Допустим, что теоретически

$$\dot{K}=\sum_{i=1}^{N}p(F^i)+\sum_{j=1}^{M}p(F^j),$$

где N — число учтенных плоскостей деления классификации F; M — число неучтенных плоскостей деления, которое неизвестно. При этом $\min(p(F^i))\geq\max(p(F^j))$. Очевидно, что при таких условиях вычислить $\dot{K}$ не представляется возможным. Следовательно, определить степень доверия к модели

$$\overline{O}_K(F)=\frac{1}{NK}\sum_{i=1}^{N}\overline{O}(F^i)\,p(F^i),$$
$$p(F^{k+1})\geq p(F^i), \qquad (23)$$

Свойство математической модели	Свойство полученной математической модели
Множественность	Возможность исследовать различные классификации как метрическое пространство
Единство	Возможность исследовать различные классификации независимо от способа выбора плоскостей деления классификации
Конечность	Конечность (определяется значением N)
Адекватность	Представление любой плоской классификации ориентированным деревом
Эффективность	Экспериментальное подтверждение
Простота	Простота (линейно-рекурсивная модель)
Устойчивость	Соответствует принципу моноцентризма Богданова [13], так как обладает единым центром $p(F^N)$

также не представляется возможным. Следовательно, к свойствам модели с неопределенной степенью доверия можно отнести лишь независимость от принципа выбора плоскостей деления классификации и масштабируемость.

В таблице представлены обоснования соответствия предложенной модели (23) принципам математического моделирования [14].

Следует заметить, что модель (22) соответствует теореме Бира [15] в общей теории систем. С.Т. Бир доказал, что «если жизнеспособная система содержит в себе жизнеспособную систему, то их организационные структуры должны быть рекурсивны». Этому соответствует выражение (21).

Выводы

Введение набора операций на плоскости деления классификации позволило показать, что относительное расстояние между классами на плоскости деления классификации является малополезной мерой. Используя относительное расстояние, удалось построить математически доказанную меру отличия двух классов классификации. Классификации с несколькими плоскостями деления рассмотрены как логичное развитие плоских классификаций. Доказано, что классификации с несколькими плоскостями деления представляют собой метрическое пространство.

СПИСОК ЛИТЕРАТУРЫ

1. *Association* for Information Science and Technology (2015). Conceptual Crowbars and Classification at the Crossroads: The Impact and Future of Classification Research (SIG/CR). — [On-line] Available from: https://www.asist.org/events/annual-meeting/annual-meeting-2015/seminars-and-workshops/conceptual-crowbars-and-classification-at-the-crossroads-the-impact-and-future-of-classification-research-sigcr/. [Accessed: December, 2015].
2. *Орлов А.И.* Организационно-экономическое моделирование: теория принятия решений —М.: КНОРУС, 2010. — 568 с.
3. *Universal* Decimal Classification [Электронный ресурс]. — Режим доступа: http://www.udcsummary.info/. — Ноябрь 2015.
4. *Саати Т.* Принятие решений. Метод анализа иерархий. — М. : Радио и связь, 1993. — 278 с.
5. *Тоценко В.Г.* Методы и системы поддержки принятия решений. Алгоритмический аспект. — Киев: Наук. думка, 2002. — 382 с.
6. *Шереметьева С.О., Осминин П.Г.* Киберленинка. Методы и модели автоматического извлечения ключевых слов. [Электронный ресурс]. — Режим доступа: http://cyberleninka.ru/article/n/metody-i-modeli-avtomaticheskogo-izvlecheniya-klyuchevyh-slov. — Ноябрь 2015.
7. *Ивлев Ю.В.* ЛОГИКА. Учебник. Изд.четвертое, пер. и доп. — М. : Изд-во «Проспект», 2008. — 304 с.
8. *Фейс К.* Алгебра: кольца, модули и категории. Т. 1. — М. : «Мир», 1977. — 688 с.
9. *Берзтисс А.* Структуры данных. — М. : Статистика, 1974. — 408 с.
10. *Шаталкин А.И.* Таксономия. Основания, принципы и правила. —М. : Товарищество научных изданий КМК, 2012. — 600 с.
11. *Загоруйко Н.Г., Борисова И.А., Дюбанов В.В., Кунтенко О.А.* Меры сходства, компактности, информативности и однородности обучающей выборки // Тр. Всеросийской конф. «Знания — Онтологии — Теории» (ЗОНТ-09). Том 1. — Новосибирск, 2009. — С. 93—102.
12. *Елисеева И.И., Рукавишников В.О.* Группировка, корреляция, распознавание образов. Статистические методы классификации и измерения связей. — М. : Статистика, 1977. — 144 с.
13. *Богданов А.А.* Тектология: Всеобщая организационная наука. —М. : «Финансы», 2003. — 287 с.
14. *Самарский А.А., Михайлов А.П.* Математическое моделирование: Идеи. Методы. Примеры. — М. : Физматлит, 2001. — 320 с.
15. *Бир С.Т.* Мозг фирмы. — М. : «Едиториал УРСС», 2005. — 416 с.

H.A. Kravtsov

MODEL OF COMPUTATIONS OVER CLASSIFICATIONS

The article presents results of theoretical research of a model of calculations over classifications. Based on the mathematical structure tree a set of operations is proposed, which permits one to determine a measure for plane classifications. Classifications with several division planes are a metric space and are considered as logical development of plane classification used in the problems of experts' selection. It has been shown that the obtained model satisfies principles of mathematical modeling.

K e y w o r d s: flat classification, spatial classification, measure, relative distance, absolute distance, computing model.

REFERENCES

1. Association for Information Science and Technology (2015), Conceptual Crowbars and Classification at the Crossroads: The Impact and Future of Classification Research (SIG/CR), available at: https://www.asist.org/events/annual-meeting/annual-meeting-2015/seminars-and-workshops/conceptual-crowbars-and-classification-at-the-crossroads-the-impact-and-future-of-classification-research-sigcr/ (accessed December, 2015).
2. Orlov, A.I. (2011), *Organizatsionno-economicheskoe modelirovanie: teoriya prinyatiya resheniy* [Organizational-economic modeling:decision making theory], KNORUS, Moscow, Russia.
3. Universal Decimal Classification, available at: http://www.udcsummary.info/ [accessed December, 2015]
4. Saaty, T. (1993), *Prinyatie resheniy. Metod analiza ierarhiy* [Decission making. Hierarchy analysis method], Radio i svyaz, Moscow, Russia.
5. Totsenko, V. (2002), *Metody i sistemy podderzhki prinyatiya resheniy. Algoritmicheskiy aspect* [Methods and systems of decision making support. Algorithmic aspect], Naukova dumka, Kiev, Ukraine.
6. Sheremetyeva, S.O. and Osminin, P.G. Kiberleninka. The methods and models of automatic keyword extracting, available at: http://cyberleninka.ru/article/n/metody-i-modeli-avtomaticheskogo-izvlecheniya-klyuchevyh-slov (accessed December, 2015)
7. Ivlev, Yu.V. (2008), *Logika* [Logic], Manual, 4th ed., rev. and ad., Prospekt, Moscow, Russia.
8. Feys, K. (1977), *Algebra: koltsa, moduli i kategorii. Tom I* [Algebra: rings, modules and categories. Vol. 1], Mir, Moscow, Russia.
9. Berztiss, A. (1974), *Struktury dannykh* [Data structures], Statistika, Moscow, Russia.
10. Shatalkin, A.I. (2012), *Taksonomiya. Osnovy, printsipy i pravila* [Taxonomy. Fundamentals, principles and rules], Tovarishchestvo nauchnykh publikatsiy «KMK», Moscow, Russia.
11. Zagoruyko, N.G., Borisova, I.A., Dyubanov, V.V. and Kuntenko, O.A. (2009), "Extents of similarity, compactness, informativeness and uniformity of a training sample", *Trudy vserossiyskoi kohferentsii "Znaniya — Ontologii — Teorii" (ZONT-09)* [Proceedings of All-Russian Conf. Knowledge-Ontology-Theories (ZONT-09)], Novosibirsk, 2009, Vol. 1., pp. 93-102.
12. Eliseeva, I.I. and Rukavishnikov, V.O. (1977), *Gruppirovka, korrelyatsiya, raspoznavaniye obrazov. Statisticheskiye metody klassifikatsii i izmereniya svyazei* [Grouping, correlation and image recognition. Statistical methods of classification and relation measurement], Statistika, Moscow, Russia.
13. Bogdanov, A.A. (2003), *Tektologiya: Vseobshchaya organizatsionnaya nauka* [Tectology: Universal organizational science], Finansy, Moscow, Russia.
14. Samarsky, A.A. and Mikhailov, A.P. (2001), *Matematicheskoye modelirovaniye: Idei, metody, primery* [Mathematical modeling: Ideas, methods, examples], Fizmatlit, Moscow, Russia.
15. Beer S.T. (2005), *Mozg firmy* [Brain of the firm], Editorial URSS, Moscow, Russia.

Поступила 23.11.15;
после доработки 14.01.16

КРАВЦОВ Григорий Алексеевич, канд.техн.наук, докторант Ин-та проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины. В 2000 г. окончил Севастопольский военно-морской ин-т им. П.С. Нахимова. Область научных исследований — математическое моделирование, кибербезопасность смарт-грид, криптография, разработка распределенных гетерогенных вычислительных систем.