

А.А. Викентьев

*Институт математики им. С.Л.Соболева СО РАН, Новосибирск, Россия;
Новосибирский государственный университет, Россия;
Новосибирский государственный технический университет, Россия
(E-mail: vikent@math.nsc.ru)*

О машинных реализациях вычислений новых модельных расстояний, структуризаций и распознавания в знаниях

Рассмотрены примеры по обработке структуры конечного множества логических формул (многозначных суждений экспертов) и задача по распознаванию образов. Используя наработанное программное обеспечение, даны иллюстрации их применения. Предложены подходы по обработке множеств суждений или экспертной информации, которые применимы при обучении студентов, например, для оценочного тестирования знаний по конкретному разделу, в коллективном управлении качеством образования с учетом пожеланий сторон, для обработки экспертных оценок и предложений по улучшению окружающей среды. Использование этих подходов позволит повысить учет достоверности знаний, качество управления образованием, получаемой информации и пожеланий различных платформ.

Ключевые слова: логические формулы, обработка множеств, распознавание образов, экспертная оценка, множество суждений, учет, достоверная информация.

1 Введение в проблему

В настоящее время возрос интерес к построению решающих функций на основе анализа экспертной информации, заданной в виде вероятностных логических высказываний нескольких экспертов, реализации процессов адаптации и согласования логических формул [1–12]. Предлагаемые ниже подходы по обработке множеств суждений экспертов найдут применение в обучении студентов, например, математике (оценочное тестирование по разделу), в коллективном управлении качеством образования (учет пожеланий сторон), а также при обработке множеств формализованных суждений по улучшению окружающей среды. При применении данной технологии пользователь в процессе работы формирует базы знаний, которые впоследствии можно включать в процесс алгоритмической обработки для принятия решений. В этом случае используются различные модельные расстояния для формул логики Лукасевича, которые отражают многозначность суждений (высказываемых экспертом); определяются коллективные расстояния, которые служат некоторым согласованием мер близости, предлагаемых для кластеризации множеств высказываний и нахождения по ним новых кластеризаций, дающих более высокие индексы результатов. Предполагается знакомство с работой [10, 13–18].

Проблема распознавания образов уже давно привлекает внимание психологов, физиологов, инженеров и математиков. Методы распознавания образов находят применение в различных сферах деятельности человека: диагностика заболеваний, сельское хозяйство, добыча полезных ископаемых и многое другое.

Для решения проблемы распознавания образов необходимо проанализировать информацию, поступающую в виде «данных», «знаний» и других структур. Такой анализ включает в себя две процедуры: процедуру обнаружения закономерностей, содержащихся в предоставленной информации; процедуру структурирования знаний и использования обнаруженных закономерностей для предсказания значения одной части информации по известным значениям другой её части. Напомним, что в работе [10] отмечено, что при увеличении числа знаний возникает потребность в анализе этих знаний. В частности, допустим, что задана некоторая структурированная база знаний (например, кластерами), на вход которой подаётся некоторое новое знание q . Требуется определить, к какому из имеющихся k таксонов (именованных областей, содержащих элементы, похожие друг на друга по каким-то характеристикам) следует отнести это новое знание, т.е. получаем задачу распознавания образов:

Постановка задачи

Пусть в пространстве знаний заданы:

1. Набор характеристик X .
2. Список наименований фиксированных областей (таксонов, называемых также образами), на которые разделено выборочное пространство $i = 1 \dots I$.
3. Обучающая выборка в виде знаний экспертов Doi (в пространстве X) для каждого Si .
4. Контрольное знание q .

Требуется определить номер iSi : $q \in Si$, используя алгоритм k -ближайших соседей по прецедентам (типичным представителям каждого образа):

$$i = \arg \min_{i \in I} \sum_{k=1}^K Rik \mid S, Do, X, k, R,$$

где Rik — k -минимальные расстояния от q до M знаний для каждого таксона; R — ошибка распознавания. То есть находятся расстояния от контрольного знания до реализации каждого образа, выбираются k -минимальные расстояния, определяются средние (для каждого образа), среди которых находится минимальное, и таким образом восстанавливается номер таксона, которому принадлежит контрольное знание.

Для решения поставленной задачи была написана компьютерная программа. Кроме того, в ней рассмотрен алгоритм, реализованный ранее, отличие которого от рассмотренного в [10] заключается в использовании для определения i эталонных знаний, создаваемых для каждого образа:

$$i = \arg \min_{i \in I} Ri,$$

где Ri — расстояние от q до Eti (эталонного знания i -го образа). Далее все описано аналогично статье [10].

2 О новых (коллективных) модельных расстояниях

Известно [16–18], что устойчивость решений в задачах кластеризации может быть повышена благодаря формированию ансамбля алгоритмов и построению на его основе коллективного решения. При этом используются результаты, полученные разными алгоритмами либо одним алгоритмом с различными значениями параметров. Кроме того, для формирования ансамбля могут быть применены разные подсистемы переменных. Ансамблевый подход является одним из наиболее перспективных направлений в кластерном анализе [1, 4, 5; 18–27; 11–18].

В данной работе для построения коллективного решения используются результаты работы иерархического алгоритма.

Пусть имеется множество из l формул. Допустим, мы рассмотрели k наборов α . Для каждого набора провели кластеризацию. Выбираем и получаем некоторые кластеры. Для каждого полученного набора кластеров рассчитаем индекс качества кластеризации. Тогда *коллективным расстоянием* для φ_i и φ_j назовем следующую величину:

$$\tilde{\rho}(\varphi_i \varphi_j) = s_1 \rho_1(\varphi_i \varphi_j) + \dots + s_k \rho_k(\varphi_i \varphi_j),$$

где s_m — индекс качества кластеризации для m -го набора α ; $m = 1, \dots, k$; $\rho_m(\varphi_i \varphi_j)$ — расстояние между формулами φ_i и φ_j для m -го набора α , $m = 1, \dots, k$.

Далее введем коллективную меру достоверности. *Коллективной мерой достоверности* для формулы φ_i назовем следующую величину:

$$\tilde{I}(\varphi_i) = s_1 I_1(\varphi_i) + \dots + s_k I_k(\varphi_i),$$

где s_m — индекс качества кластеризации для m -го набора α , $m = 1, \dots, k$; $I_m(\varphi_i)$ — мера достоверности для формулы φ_i для m -го набора α , $m = 1, \dots, k$ (табл. 1–5).

Например, для предыдущего примера при $n = 5$:

$$\tilde{\rho}(\varphi_1 \varphi_2) = 0,056 * 0,760 + 0,016 * 0,760 = 0,055.$$

$$\tilde{I}(\varphi_1) = 0,056 * 0,200 + 0,016 * 0,190 = 0,014. \text{ Далее по аналогии.}$$

Матрица коллективных расстояний для примера ($n = 5$) множества состоит из восьми формул:

$$\begin{aligned} \varphi_1 &= x \rightarrow y; \\ \varphi_2 &= \neg(x \rightarrow y); \\ \varphi_3 &= (x \vee z) \rightarrow y; \\ \varphi_4 &= \neg((x \wedge y) \vee z) \rightarrow w; \\ \varphi_5 &= y \rightarrow (x \wedge z); \\ \varphi_6 &= (\neg y \vee (x \rightarrow z)) \rightarrow w; \\ \varphi_7 &= ((x \rightarrow y) \rightarrow z) \rightarrow w; \\ \varphi_8 &= (w \rightarrow z) \wedge (y \rightarrow x). \end{aligned}$$

Таблица 1

Матрица коллективных мер недоверности для примера 5.1 (n = 5)

	1	2	3	4	5	6	7	8
1	0	0.055	0.007	0.025	0.033	0.028	0.018	0.030
2		0	0.049	0.039	0.036	0.036	0.046	0.036
3			0	0.023	0.037	0.026	0.017	0.034
4				0	0.029	0.004	0.009	0.032
5					0	0.030	0.031	0.010
6						0	0.012	0.033
7							0	0.034
8								0

Таблица 2

Матрица коллективных расстояний для того же примера (n = 7)

<i>i</i>	1	2	3	4	5	6	7	8
$\tilde{I}(\varphi_i)$	0.014	0.058	0.021	0.026	0.021	0.029	0.019	0.024

Таблица 3

Матрица коллективных мер недоверности для примера 5.1 (n = 7)

	1	2	3	4	5	6	7	8
1	0	0.024	0.003	0.011	0.014	0.012	0.007	0.012
2		0	0.022	0.017	0.017	0.016	0.021	0.017
3			0	0.011	0.015	0.011	0.007	0.014
4				0	0.013	0.001	0.004	0.015
5					0	0.013	0.013	0.004
6						0	0.005	0.014
7							0	0.014
8								0

Таблица 4

Используем иерархический алгоритм. Получим следующие кластеры.

<i>i</i>	1	2	3	4	5	6	7	8
$\tilde{I}(\varphi_i)$	0.006	0.026	0.009	0.012	0.009	0.013	0.008	0.010

Они и задают структуру множества рассматриваемых формул, позволяющую решать вопрос о принадлежности новой формулы к одному из кластеров.

Номер итерации	$n = 5$	Δ	$n = 7$	Δ
1	{4, 6}, 1, 2, 3, 5, 7, 8	0.003	{4, 6}, 1, 2, 3, 5, 7, 8	0.001
2	{1, 3}, {4, 6}, 2, 5, 7, 8	0.007	{1, 3}, {4, 6}, 2, 5, 7, 8	0.003
3	{1, 3}, {4, 6, 7}, 2, 5, 8	0.01	{1, 3}, {4, 6, 7}, 2, 5, 8	0.005
4	{1, 3}, {4, 6, 7}, {5, 8}, 2	0.01	{1, 3, 4, 6, 7}, 2, 5, 8	0.007
5	{1, 3, 4, 6, 7}, {5, 8}, 2	0.015	{1, 3, 4, 6, 7}, {5, 8}, 2	0.007
6	{1, 3, 5, 8, 6, 7, 4}, 2	0.015	{1, 3, 5, 8, 6, 7, 4}, 2	0.007
7	{1, 3, 5, 8, 6, 7, 4, 2}	0.044	{1, 3, 5, 8, 6, 7, 4, 2}	0.02

3 Наблюдения и выводы для различных n , $n \geq 2$ логики Лукасевича

Была создана база из более 250 различных логических формул, откуда случайным образом выбирались подмножества формул. С помощью адаптированных алгоритмов, описанных в [8, 9], было кластеризовано более 100 таких подмножеств при различных n и α .

Исходя из рассмотренных примеров для новых метрик, были сделаны следующие выводы:

1. Для $n = 2, \dots, 7$ наблюдается разница в составе кластеров, при $n > 7$ кластеры и последовательность итераций не меняются. Таким образом, возникает гипотеза о нецелесообразности использования логики большой значности в реальных задачах от малого числа переменных.

2. Для алгоритма k -средних при вычислении центров масс наблюдаются одни и те же результаты, как при замене среднего арифметического ближайшим сверху значением из V_n , так и ближайшим снизу.

3. В большинстве случаев коллективные расстояния дают наилучшую кластеризацию.

В данном разделе нами использованы новые меры, обобщающие формулы расстояния и меры недоверности, полученные впервые для шестизначной логики Лукасевича совместно с В.В. Фефеловой. Также использована формула для расстояния, когда некоторые значения переменных заранее известны. Ранее [6–10] доказаны свойства этих величин, схожие со свойствами расстояния, и меры недоверности для случая классической логики. Полученные величины могут быть использованы при построении логических решающих функций в распознавании, при анализе баз знаний и их кластеризации (например, таких, как медицинская экспертная база или база рекомендательных систем).

Были адаптированы два алгоритма кластеризации – иерархический и k -средних (k -means). Проведены численные эксперименты для различных n и α , показаны их различия. По введенному индексу качества кластеризации из множества примеров была выбрана наилучшая кластеризация, построены коллективные расстояния и по ним – новая кластеризация.

Для автоматической кластеризации впервые совместно с В.В. Фефеловой [8, 9] разработано и написано рабочее Java-приложение, расширенное в данной работе. Сложность алгоритма вычисления расстояния между формулами – экспоненциальная. Далее описаны содержание программы для вычисления расстояний и меры недоверности, в которой представлена реализация иерархического алгоритма и алгоритма k -средних. Программа реализована на языке *Java* и состоит из семи классов. Первый класс *Fraction* определяет класс рациональных дробей и операции над ними (рациональные дроби являются истинностными значениями многозначной логики). Следующий класс *Formula* описывает формулы (высказывания экспертов или часть базы знаний). Класс *Distance* реализует расчет мер недоверности и расстояний между формулами. Класс *HierarchicAlg* содержит иерархический алгоритм. Класс *K-Means* содержит алгоритм k -средних. В классе *Helper* задаются множества истинностных значений и весов, также в нем содержатся вспомогательные методы. Класс *N-valued Logic* является основным, в нем производится вызов объектов других классов для необходимых расчетов.

Реализация на высокоуровневом ООП языке позволяет добиться универсальности применения программы. Программа представлена в виде консольного приложения, все необходимые параметры задаются в тексте программы. В консоль выводятся результаты вычислений: матрица расстояний, меры недоверности и результат поэтапного выполнения иерархического алгоритма и алгоритма k -средних (индекс качества кластеризации также выдается на каждом этапе).

В дальнейшем планируется адаптировать и другие алгоритмы кластеризации, а также провести кластеризацию большего числа формул и проиллюстрировать применение полученных величин на практике.

4 Машинная реализация алгоритмов распознавания

До недавнего времени существовал только один реализованный алгоритм распознавания образов в пространстве знаний, основанный на гипотезе унимодальной компактности. В процессе работы с В.В. Ивановым [10] была создана компьютерная программа «Knowledge» в среде DELPHI 5.0., реализующая новый алгоритм из [10], алгоритм k -ближайших соседей, основанный уже на гипотезе полимодальной компактности, которая является более слабой и потому позволяет расширить возможности распознавания в пространстве знаний.

Работа с программой начинается с Главного окна (рис. 1), где задаются число характеристик (по которым производится обучение и распознавание); количество экспертов, задающих параметры; количество образов, на которое разбивается выборочное пространство; k -ближайших соседей; максимальное значение критерия значимости для характеристик.

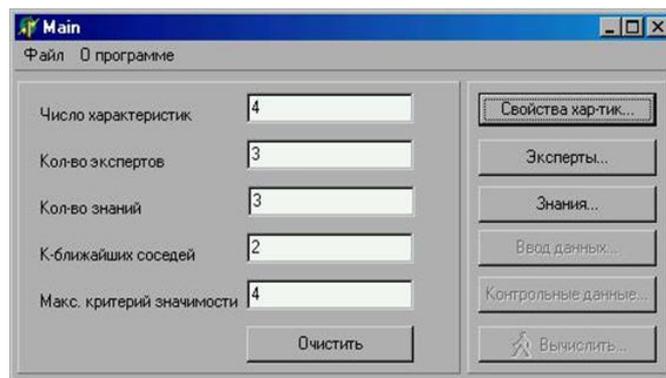


Рисунок 1. Главное окно

Переходя в режим «Свойства характеристик» (рис. 2), можем задать максимальное значение характеристики, минимальное значение характеристики, т.е. крайнее левое значение (от которого отсчитывается первый интервал) и разбиение — количество интервалов.



Рисунок 2. Режим «Свойства характеристик»

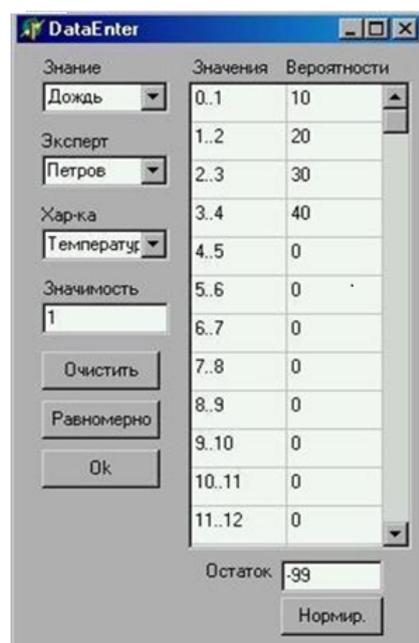


Рисунок 3. Обучающая выборка

Нажав на кнопку «Ввод данных» (рис. 3), мы получаем возможность ввести обучающую выборку — заполняется для каждого образа, каждого эксперта, каждой характеристики, каждого интервала значения вероятностей (если эксперт не может предоставить распределение для какой-либо характеристики, то для неё задаём равномерное распределение). После ввода данных (в процентах или в долях) в поле «вероятности» (рис. 3, 4), их необходимо нормировать (кнопка «Нормир»), затем задаётся значимость характеристик, т.е. значение, характеризующее, насколько эксперт считает важной данную характеристику (при расчётах используются средние значения значимостей по всем экспертам для данной характеристики, т.е. проводится своего рода «голосование»).

После ввода обучающей выборки переходим в режим «Контрольные данные», где задаётся контрольное знание (рис. 4).

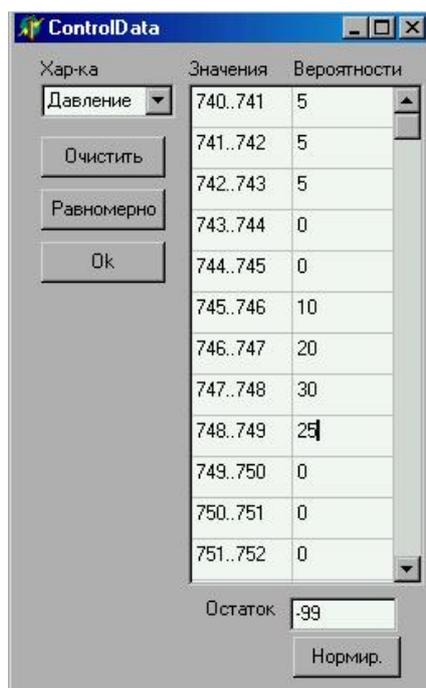


Рисунок 4. Режим «Контрольные данные»

По мере завершения ввода данных программа готова к распознаванию. Для этого нажимаем на кнопку «Вычислить», после чего появляется окно результатов (рис. 5), обозначение – * в каждом образе (имя образа выводится слева, например: *дождь* и *снег*) показывает каждого из k -ближайших знаний экспертов (в примере $= 2 \Rightarrow$ выбираем два знания в каждом образе, расстояние от контрольного до которых минимально, для образа *дождь* — это знания *Петрова* и *Иванова*, для образа *снег* — знания *Иванова* и *Сидорова*); ** — показывает образ, к которому относится контрольное знание, распознанное с использованием эталона и алгоритма k -ближайших соседей.

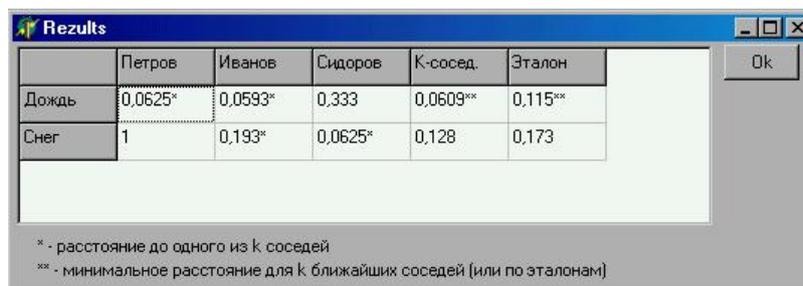


Рисунок 5. Окно результатов

Работа данной программы была проверена на тестовых выборках, одна из которых представляет собой набор знаний, высказанных студентами-пятикурсниками, выступающими в роли экспертов. Студентам было предложено назвать характеристики, по которым можно определить, что студент получит на экзамене: 5, 4 или 3. После опроса выбраны следующие характеристики: а) посещаемость лекций; б) посещаемость семинаров; в) средний балл; г) время подготовки к экзамену; д) курс, на котором учится студент. Для них были введены значения и вероятности, с которыми эти значения принимаются. Контрольная выборка представляла собой знания экспертов, не включенных в обучающую выборку, без имени образа, которому принадлежат эти знания. Проводилось распознавание, а затем сравнение результатов, полученных программой, и значений, заданных экспертами. Совпадение данных показало, что алгоритмы, реализованные в программе, позволяют проводить распознавание образов в пространстве знаний. Возникали случаи, когда распознавание с использованием алгоритма k -ближайших соседей и эталонов давали разные результаты. Это происходило при полимодальных распределениях характеристик, в данном случае распознавание по эталону давало ошибочный результат, что подтверждает целесообразность использования алгоритма k -ближайших соседей, а не эталонных знаний соседей в случае полимодальных распределений. Очевидно, что можно применять в этой задаче новые модельные расстояния [7–11].

Перечислим результаты данного раздела работы: в новой постановке рассмотрена задача распознавания образов в пространстве знаний, в виде программы реализованы алгоритм k -ближайших соседей, позволяющий решить данную задачу, и ранее рассмотренный алгоритм сравнения по эталонам. Заданы обучающие выборки, проведено распознавание знаний и подтверждена связь между характером распределений и правильностью работы алгоритмов, в случае унимодальных распределений оба алгоритма распознают практически одинаково, а в случае полимодальных — сравнение по эталонам даёт больше ошибок. Проведённые эксперименты показывают возможность дальнейшего использования программ и различных модельных расстояний для структуризации знаний в логических базах знаний.

Планируется дальнейшее развитие предложенных подходов для решения конкретных прикладных задач.

Автор выражает благодарность своим коллегам и особо сподвижникам по работе В.В. Фефеловой и В.В. Иванову за помощь в реализации алгоритмов и проверке конкретных вычислений, изложенных выше.

Работа выполнена при поддержке грантов РФФИ, проекты 14-07-00851а, 14-07-00249а.

Список литературы

- 1 Загоруйко Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. — Новосибирск: Изд-во Ин-та математики, 1999.
- 2 Загоруйко Н.Г. Меры расстояния в пространстве знаний / Н.Г. Загоруйко, М.В. Бушуев // Анализ данных в экспертных системах. — 1986. — Вып. 117. Вычислительные системы. — С. 24–35.
- 3 Загоруйко Н.Г. Алгоритмы обнаружения эмпирических закономерностей / Н.Г. Загоруйко, В.Н. Ёлкина, Г.С. Лбов. — Новосибирск: Наука, 1985.
- 4 Загоруйко Н.Г. Методы распознавания и их применение / Н.Г. Загоруйко. — М.: Сов. радио, 1972.
- 5 Викентьев А.А. Расстояние между формулами пятизначной логики Лукасевича и мера недоверности высказываний экспертов / А.А. Викентьев, Е.С. Кабанова // Вестн. Караганд. ун-та. Сер. Математика. — 2013. — № 1(69). — С. 18–27.
- 6 Викентьев А.А. О возможных расстояниях и степенях недоверности в многозначных высказываниях экспертов и приложении этих понятий в проблемах кластеризации и распознавания / А.А. Викентьев // Проблемы информатики. — 2011. — № 3(11). — С. 33–45.
- 7 Vikent'ev A.A. Concerning distances and degrees of uncertainty for many-valued expert statements and application of those concepts in pattern recognition and clustering // Pattern Recognition and Image Analysis. — 2014. — Vol. 24. — No. 4. — P. 489–501.
- 8 Викентьев А.А. Введение полных расстояний и мер недоверности для формул логик Лукасевича для автоматической кластеризации множеств логических высказываний из базы знаний / А.А. Викентьев, В.В. Фефелова // Вестн. Караганд. ун-та. — Сер. Математика. — 2015. — № 3(79). — С. 17–24.

- 9 Викентьев А.А. Новые расстояния и меры достоверности для формул логики Лукасевича в кластеризации логических высказываний базы знаний / А.А. Викентьев, В.В. Фефелова // Математические методы распознавания образов ММО-17: тезисы докл. 17-й Всерос. конф. с междунар. участием, г. Светлогорск, Калининградская обл. — М.: Торус пресс, 2015. — С. 68, 69.
- 10 Викентьев А.А. Методы распознавания образов в пространстве знаний / А.А. Викентьев, В.В. Иванов // Вестн. Караганд. ун-та. — Сер. Математика. — 2016. — № 1(81). — С. 26–34.
- 11 Викентьев А.А. Изучение модельных расстояний на логических высказываниях с учетом экспертных интерпретаций для формул многозначных логик Лукасевича и автоматической кластеризации в базах знаний. I / А.А. Викентьев // Вестн. Караганд. ун-та. Сер. Математика. — 2016. — № 2(82). — С. 23–31.
- 12 Викентьев А.А. Изучение модельных расстояний на логических высказываниях с учетом экспертных интерпретаций для формул многозначных логик Лукасевича и автоматической кластеризации в базах знаний. II / А.А. Викентьев // Вестн. Караганд. ун-та. Сер. Математика. — 2016. — № 2(82). — С. 32–39.
- 13 Ершов Ю.Л. Математическая логика / Ю.Л. Ершов, Е.А. Палютин. — 2-е изд. — М.: Наука, 1987. — 336 с.
- 14 Карпенко А.С. Логика Лукасевича и простые числа / А.С. Карпенко. — М.: Наука, 2000. — 319 с.
- 15 Лбов Г.С. Логические решающие функции и вопросы статистической устойчивости решений / Г.С. Лбов, Н.Г. Старцева. — Новосибирск: Изд-во Ин-та математики, 1999. — 212 с.
- 16 Vikent'ev A.A., Lbov G.S. Setting the metric and informativeness on statements of experts // Pattern Recognition and Image Analysis. — 1997. — Vol. 7. — No. 2. — P. 175–183.
- 17 Лбов Г.С. Устойчивость решающих функций в задачах распознавания образов и анализа разнотипной информации / Г.С. Лбов, В.Б. Бериков. — Новосибирск: Изд-во ИМ СО РАН, 2005. — 200 с.
- 18 Strehl A., Ghosh J. Clustering ensembles – a knowledge reuse framework for combining multiple partitions // Journal Machine Learning Research. — 2002. — No. 3. — P. 583–617.

А.А. Викентьев

Білімдегі айырып тану, құрылым және жаңа үлгілі арақашықтық, есептеулердің машиналық жүзеге асуы

Бейнелерді тану есебі мен логикалық формулалардың (сарапшылардың көп мағыналы пайымдауы) ақырлы жиынның құрылымдарын өңдеу мысалдары қарастырылды. Сыналған бағдарламалық құралды пайдаланып, оларды жүзеге асыру мысалдары берілген. Пікірлер жиыны мен сараптамалық ақпаратты өңдеу тәсілдері ұсынылған, мысалы, ұсыныс беруші жақтарды ескере отырып, сапалы білім берудегі ұжымдық басқаруда нақты бөлім бойынша білімді бағалау үшін және қоршаған ортаны жақсарту мақсатында экспертті бағалаулар мен ұсыныстарды беруде. Бұл тәсілдерді қолдану шынайы білімді тіркеуде сапалы білім беруді басқаруда, алынған ақпараттың растығын, сондай-ақ әр түрлі платформалардың ескертулерін тіркеу жұмыстарында жоғары көрсеткішке ие болуға көмектеседі.

Кілт сөздер: логикалық формулалар, жиындарды өңдеу, бейнелерді танып білу, тіркеу, расталған ақпарат.

A.A. Vikent'ev

On machine implementations of calculations of new model distances, structurizations and recognition in knowledge

Examples are given of processing the structure of a finite set of logical formulas (multi-valued judgments of experts) and the task of pattern recognition. Using the software, illustrations of their application are given.

The approaches to processing sets of judgments or expert information that are applicable to the training of students, for example, for evaluating knowledge testing on a specific section, in collective management of the quality of education, taking into account the wishes of the parties, for processing expert assessments and proposals for improving the environment are proposed. The use of these approaches will allow to increase the account of the reliability of knowledge, the quality of education management, increase the reliability of the information received and take into account the wishes of various platforms.

Keywords: logical forums, processing of sets, recognition of images, expert evaluation, sets of judgments, accounting, reliability of information.

References

- 1 Zagoruiko, N.G. (1999). *Prikladnye metody analiza dannykh i znanii [Applied methods of data and knowledge analysis]*. Novosibirsk: Izdatelstvo Instituta matematiki [in Russian].
- 2 Zagoruiko, N.G., Bushuev, M.V. (1986). Mery rasstoianiiia v prostranstve znanii [Distance measures in the space of knowledge]. *Analiz dannykh v ekspertnykh sistemakh – Data Analysis in Expert Systems*, Issue. 117: Computing systems, 24–35 [in Russian].
- 3 Zagoruiko, N.G., Yolkina, V.N., Lbov, G.S. (1985). *Algoritmy obnaruzheniia empiricheskikh zakonomernostei [Algorithms for detecting empirical regularities]*. Novosibirsk: Nauka [in Russian].
- 4 Zagoruiko, N.G. (1972). *Metody raspoznavaniia i ikh primenenie [Methods of recognition and their application]*. Moscow: Sovetskoe radio [in Russian].
- 5 Vikent'ev, A.A., Kabanova, E.S. (2013). Rasstoianie mezhdru formulami piatiznachnoi logiki Lukasevicha i mera nedostovernosti vyskazyvanii ekspertov [The distance between the formulas of the three-valued logic of Lukasiewicz and the measure of unreliability of the statements of experts]. *Vestnik Karagandinskogo universiteta. Ser. Matematika – Bulletin of the Karaganda University, Ser. Mathematics*, 1 (69), 18–27 [in Russian].
- 6 Vikent'ev, A.A. (2011). O vozmozhnykh rasstoianiiakh i stepeniakh nedostovernosti v mnogoznachnykh vyskazyvaniakh ekspertov i prilozhenie etikh poniatii v problemakh klasterizatsii i raspoznavaniia [On the possible distances and degrees of inauthenticity in polysemantic statements of experts and the application of these concepts in the problems of clustering and recognition]. *Problemy informatiki – Problems of Informatics*, 3 (11), 33–45 [in Russian].
- 7 Vikent'ev, A.A. (2014). Concerning distances and degrees of uncertainty for many-valued expert statements and application of those concepts in pattern recognition and clustering. *Pattern Recognition and Image Analysis, Vol. 24, 4*, 489–501 [in Russian].
- 8 Vikent'ev, A.A., Fefelova, V.V. (2015). Vvedenie polnykh rasstoianii i mer nedostovernosti dlia formul logik Lukasevicha dlia avtomaticheskoi klasterizatsii mnozhestv logicheskikh vyskazyvanii iz bazy znanii [The introduction of full distances and measures of inauthenticity for Lukasiewicz's formulas for automatic clustering of sets of logical statements from the knowledge base]. *Vestnik Karagandinskogo universiteta. Ser. Matematika – Bulletin of the Karaganda University, Ser. Mathematics*, 3 (79), 17–24 [in Russian].
- 9 Vikent'ev, A.A., Fefelova, V.V. (2015). Novye rasstoianiiia i mery dostovernosti dlia formul logiki Lukasevicha v klasterizatsii logicheskikh vyskazyvanii bazy znanii [New distances and confidence measures for formulas of Lukasiewicz's logic in the clustering of logical statements of the knowledge base]. *Mathematical methods of pattern recognition MMRO-17: 17-aia Vserossiiskaia konferantsaia s mezhdunarodnym uchastiem, g. Svetlogorsk, Kaliningradskaia obl. – 17th All-Russian Conference with international participation. City of Svetlogorsk, Kaliningrad region*, pp. 68, 69. Moscow: Torus press [in Russian].
- 10 Vikent'ev, A.A., Ivanov, V.V. (2016). Metody raspoznavaniia obrazov v prostranstve znanii [Methods for recognizing images in the space of knowledge]. *Vestnik Karagandinskogo universiteta. Ser. Matematika – Bulletin of the Karaganda University, Ser. Mathematics*, 1 (81), 26–34 [in Russian].
- 11 Vikent'ev, A.A. (2016). Izuchenie modelnykh rasstoianii na logicheskikh vyskazyvaniakh s uchetom ekspertnykh interpretatsii dlia formul mnogoznachnykh logik Lukasevicha i avtomaticheskoi klasterizatsii v bazakh znanii. I [Study of model distances on logical statements based on expert interpretation of the formulas of many-valued logics of Lukasiewicz and automatic clustering of knowledge bases. I]. *Vestnik Karagandinskogo universiteta. Ser. Matematika – Bulletin of the Karaganda University, Ser. Mathematics*, 2 (82), 23–31 [in Russian].

- 12 Vikent'ev, A.A. (2016). Izuchenie modelnykh rasstoianii na logicheskikh vyskazyvaniiah s uchetom ekspertnykh interpretatsii dlia formul mnogoznachnykh logik Lukasevicha i avtomaticheskoi klasterizatsii v bazakh znani. II [Study of model distances on logical statements based on expert interpretation of the formulas of many-valued logics of Lukasiewicz and automatic clustering of knowledge bases. II]. *Vestnik Karagandinskogo universiteta. Ser. Matematika – Bulletin of the Karaganda University, Ser. Mathematics, 2 (82)*, 32–39 [in Russian].
- 13 Yershov, Yu.L., Palyutin, E.A. (1987). *Matematicheskaiia logika [Mathematical Logic]*. (2d ed.). Moscow: Nauka [in Russian].
- 14 Karpenko, A.S. (2000). *Logiki Lukasevicha i prostye chisla [Lukasevich's logic and prime numbers]*. Moscow: Nauka [in Russian].
- 15 Lbov, G.S., Startseva, N.G. (1999). *Logicheskie reshaiushchie funktsii i voprosy statisticheskoi ustoiichivosti reshenii [Logical decision functions and questions of statistical stability of solutions]*. Novosibirsk: Izdatel'stvo Instituta matematiki [in Russian].
- 16 Vikent'ev, A.A., Lbov, G.S. (1997). Setting the metric and informativeness on statements of experts. *Pattern Recognition and Image Analysis, Vol. 7, 2*, 175–183.
- 17 Lbov, G.S., Berikov, V.B. (2005). *Ustoiichivost reshaiushchikh funktsii v zadachakh raspoznavaniia obrazov i analiza raznotipnoi informatsii [Stability of decision functions in problems of pattern recognition and analysis of heterogeneous information]*. Novosibirsk: Izdatel'stvo IM SO RAN [in Russian].
- 18 Strehl, A., Ghosh, J. (2002). Clustering ensembles – a knowledge reuse framework for combining multiple partitions. *Journal Machine Learning Research, 3*, 583–617.