Институт информационных и вычислительных технологий МОН РК

Казахский Национальный Университет имени аль-Фараби

Университет Туран

Люблинский технический университет, Польша

«Ғылым ордасы»

МАТЕРИАЛЫ

IV международной научно-практической конференции "Информатика и прикладная математика", посвященной 70-летнему юбилею профессоров Биярова Т.Н., Вальдемара Вуйцика и 60-летию профессора Амиргалиева Е.Н.

25-29 сентябрь 2019, Алматы, Казахстан

Часть 1

Алматы 2019

УДК 378 (063)
ББК 74.58
И74

Главный редактор:
Калимолдаев М.Н. - генеральный директор ИИВТ, академик НАН РК, доктор физико-математических наук, профессор

Ответственные редакторы:
Мамырбаев О.Ж. - заместитель генерального директора ИИВТ, доктор PhD
Калижанова А.У. - заместитель генерального директора ИИВТ, кандидат физико-математических наук
Юничева Н.Р. - ученый секретарь ИИВТ МОН РК, кандидат технических наук, доцент

И 74 **Информатика и прикладная математика:** Мат. IV Межд. науч. конф. (25-29 сентября 2018 г.). Часть 1. – Алматы, 2019. – с. 617

ISBN 978-601-332-384-8

В сборнике опубликованы доклады, представленные по 4 секциям от Республики Казахстан, Российской Федерации, США, Латвии, Польши, Республики Беларусь, Украины, Азербайджана, Узбекистана, Японии, Кореи, Ирана, Португалии, Испании, Великобритании, Греции, Кыргызской Республики и других.

Рассмотрены актуальные вопросы в области математики, информатики и управления: математического моделирования сложных систем и бизнес-процессов, исследования и разработки защищенных и интеллектуальных информационных и телекоммуникационных технологий, математической теории управления, технологий искусственного интеллекта.

Материалы сборника предназначены для научных работников, докторантов и магистрантов, а также студентов старших курсов.

УДК 378 (063)
ББК 74.58

ISBN 978-601-332-384-8

8. Y. Fu, M. Shahidehpour, and Z. Li, “Long-term security constrained unit commitment: Hybrid Dantzig-Wolfe decomposition and subgradient approach,” IEEE Trans. Power Syst., vol. 20, pp. 2093–2106, Nov. 2005

9. R. BILLINTON AND M. FOTUHI-FIRUZABAD, A reliability framework for generating unit commitment, Electric Power System Research, 56 (2000), pp. 81–88. (Cited on p. 5)

10. D. BERTSIMAS, E. LITVINOV, X. A. SUN, J. ZHAO, AND T. ZHENG, Adaptive robust optimization for the security constrained unit commitment problem, IEEE Transactions on Power Systems, 28 (2013), pp. 52–63. (Cited on pp. 2, 3, 5, 7, 8, 9)

11. R. JIANG, M. ZHANG, G. LI, AND Y. GUAN, Two-stage network constrained robust unit commitment problem, European Journal of Operational Research, 234 (2014), pp. 751–762. (Cited on pp. 3, 5, 7, 8)

12. Y. Zhang, J. Wang, B. Zeng, and Z. Hu. Chance-constrained two-stage unit commitment under uncertain load and wind power output using bilinear benders decomposition. IEEE Transactions on Power Systems, 32(5):3637–3647, 2017.

13. B. ZENG AND L. ZHAO, Solving two-stage robust optimization problems using a column-and-constraint generation method, Operations Research Letters, 41 (2013), pp. 457–461. (Cited on p. 8)

14. Q. WANG, J. P. WATSON, AND Y. GUAN, Two-stage robust optimization for n - k contingency-constrained unit commitment, IEEE Transactions on Power Systems, 28 (2013), pp. 2366–2375. (Cited on pp. 8, 9)

КОЛЛЕКТИВНЫЕ РАССТОЯНИЯ ДЛЯ КЛАСТЕРИЗАЦИИ МНОЖЕСТВ ФОРМУЛ *N*-ЗНАЧНОЙ ЛОГИКИ

[1,2]Викентьев А.А., [1]Серов М.С., [1,2]Бериков В.Б., [3,4]Черикбаева Л.Ш., [5]Тулегенова Б.А.

e-mail: berikov@math.nsc.ru, vikent@math.nsc.ru, mikhailserov3@gmail.com, lyailya_sh@mail.ru, tulegenova_bakhit@mail.ru

[1]Новосибирский государственный университет, Новосибирск, Россия

[2]Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия

[3]Институт информационных и вычислительных технологий КН МОН РК, Казахстан

[4]Казахский национальный университет им. аль-Фараби, Казахстан

[5]Казахский национальный исследовательский технический университет им. К.И. Сатпаева, Казахстан

Аннотация. *Рассматривается задача анализа логических экспертных высказываний, возникающая на стыке искусственного интеллекта, машинного обучения и математической логики. В рамках логической модели представления знаний, содержащиеся в базе знаний интеллектуальной системы предложения описываются с помощью формул некоторой логики. В работе проводится* *исследование свойств метрик и мер нетривиальности на формулах n-значной логики, предлагается понятие коллективной метрики. Введённые понятия используются для кластерного анализа логических высказываний; описываются примеры полученных решений.*

Ключевые слова: *расстояние на формулах, n-значная логика, коллективные метрики, меры нетривиальности, кластерный анализ*

Введение. В настоящее время достаточно хорошо развиты теория и методы машинного обучения на основе эмпирических данных [1]. Наряду с анализом данных, существует потребность в анализе знаний экспертов, накопленных в базах знаний интеллектуальных систем. Это необходимо для структурирования базы знаний, ускорения поиска знаний, наиболее подходящих конкретной решаемой задаче. Упорядочение знаний по степени близости дает возможность ранжировать их источники (экспертов, интернет-пользователей и т.д.) по похожести. В качестве примеров приложений разрабатываемых методов можно привести медицинские экспертные системы, рекомендательные системы по поиску телефильмов, выбору туристических маршрутов.

Следуя логической модели представления знаний, имеющиеся знания отображаются совокупностью формул – логических высказываний в некоторой логике.

Имеется большое число работ, в которых решается задача формирования логических продукционных правил, описывающих закономерности, полученные из даных. Однако обработка информации, имеющей вид набора логических высказываний – сравнительно недавно начавшее развиваться направление (см., например, [2]).

Задача определения меры близости между логическими высказываниями была поставлена в работах [3,4]. Теоретико-модельный подход к определению расстояний между логическими формулами n-значной логики Лукасевича был впервые предложен и развит в работах [5-7]. Привлечение многозначной логики позволило учитывать возможную неопределенность знаний, неполную уверенность эксперта. Использование логических значений истинности на моделях при введении расстояний было предложено в работе [8] и в дальнейшем обобщено на общий конечный случай логики Лукасевича [9].

В работе [10] были введены меры недостоверности и расстояния, а также сформулированы принципы кластеризации формул, применяемые для их компьютерного анализа. Остался открытым вопрос о других возможностях задания таких расстояний. Этому вопросу, введению мер нетривиальности и их применению для кластеризации множеств формул и посвящена данная работа.

В работе рассматриваются экспертные знания, представленные в виде логических высказываний - формул логики Лукасевича. Очевидно, что высказывания могут различаться по их информативности, которая отражает важность сообщенной экспертом информации. Возникает вопрос о ранжировании высказываний экспертов и сравнении их по информативности. Понятно, что информативность высказывания должна зависить от информативности его компонент и степени различия содержащейся в них информации. Необходимо ввести такую метрику на логических высказываниях, которая будет служить мерой различия содержащейся в высказываниях информации. Тем самым задача введения меры информативности на логических формулах и их кластеризации предполагает рассмотрение вопроса определения расстояния на этом множестве [4].

При решении указанных задач используется коллективный (ансамблевый) подход в кластерном анализе, развиваемый в работах [11-13]. Привлечение ансамбля позволяет повысить устойчивость результатов анализа в случае неопределенности в структуре информации. Такая неопределенность возникает, например, когда неизвестно истинное число кластеров или могут существовать малоинформативные экспертные высказывания, сложные кластерные структуры.

Статья имеет следующий план. В разделе 2 вводятся основные определения и обозначения. В разделе 3 описываются вводимые расстояния и меры нетривиальности логических формул. Раздел 4 посвящен введению коллективных расстояний между формулами. В разлеле 5 описаны примеры кластеризации логических формул. В заключении делаются основные выводы работы обсуждаются дальнейшие планы.

Основные понятия и обозначения. В этом разделе мы рассмотрим n-значную логику L_n, а также некоторые теоретико-модельные свойства, необходимые в дальнейшем.

Определение 1. Пропозициональный язык L состоит из следующих пропозициональных символов:

1. x, y , z, ... - пропозициональные переменные;

2. $\neg$, $\rightarrow$ - пропозициональные связки (отрицание и импликация);

3. (,) - вспомогательные символы.

Определение 2. Будем понимать под формулами следующие последовательности пропозициональных символов:

1. x, y , z, ... - формулы;

2. если φ и ψ - формулы, то $\neg\varphi$, $\varphi\rightarrow\psi$ - тоже формулы;

3. никакие другие последовательности из исходных символов, кроме построенных в силу пунктов 1-2, формулами не являются.

Теперь мы можем построить n-значную логику L_n.

Определение 3. n-значная матричная логика L_n определяется через следующие понятия:

$M_n = \langle V_n, \neg, \rightarrow, \{1\} \rangle$ - логическая матрица;

$V_n = \left\{0, \frac{1}{n-1}, ..., \frac{n-2}{n-1}, 1\right\}$ - множество значений истинности;

$\neg : V_n \rightarrow V_n$, - операция отрицания;

$\rightarrow : V_n \times V_n \rightarrow V_n$ - операция импликации;

$\{1\}$ - выделенное значение истины.

Рассмотрим логические операции в n-значной логике L_n.

Определение 4. Логические операции на V_n задаются следующим образом:

$\neg x = 1 - x$ - отрицание;

$x \to y = min\{1, 1 - x + y\}$ - импликация;

$x \vee y = (x \to y) \to y = max\{x, y\}$ - дизъюнкция;

$x \wedge y = \neg(\neg x \vee \neg y) = min\{x, y\}$ - конъюнкция.

Сформулируем теоретико-модельные обозначения, которые будут использоваться в дальнейшем.

Определение 5. $S(\varphi)$ - носитель формулы φ из L_n (множество переменных, используемых при ее написании);

Через $S(\Sigma) = \cup_{\varphi \in \Sigma} S(\varphi)$ обозначим носитель множества Σ.

Определение 6. Модель M - кортеж из означенных переменных с определением истинности формул на нем согласно логике Лукасевича.

Пусть $P(S(\Sigma))$ - множество всех моделей. Будем использовать обозначение $\varphi_{\frac{k}{n-1}}$, когда φ принимает на модели значение $\frac{k}{n-1}$, $k = 0, \dots, n-1$. Пусть $M\left(\frac{k}{n-1}\right)$ - количество моделей, на которых φ принимает значение $\frac{k}{n-1}$, $M(\frac{k}{n-1}, \frac{l}{n-1})$ - количество моделей, на которых φ принимает значение $\frac{k}{n-1}$, а ψ принимает значение $\frac{l}{n-1}$.

Расстояние и мера нетривиальности логических формул

Для определения расстояния мы учитываем разницу между значениями двух формул на каждой модели. Чем меньше модуль разности между значениями формул φ и ψ, тем они более близки в данной модели. Объединим модели с одинаковыми модулями разности между значениями φ и ψ и возьмем их с некоторым весом, учитывающим близость. Получим следующую формулу:

$$\begin{aligned}\rho_0(\varphi,\psi) &= \lambda_0\left(M(0,0) + M(\frac{1}{n-1},\frac{1}{n-1}) + \dots + M(\frac{n-2}{n-1},\frac{n-2}{n-1}) + M(1,1)\right) + \\ &+\lambda_1\left(M(0,\frac{1}{n-1}) + M(\frac{1}{n-1},0) + \dots + M(\frac{n-2}{n-1},1) + M(1,\frac{n-2}{n-1})\right) + \\ &\dots \\ &+\lambda_{n-2}\left(M(0,\frac{n-2}{n-1}) + M(\frac{n-2}{n-1},0) + M(\frac{1}{n-1},1) + M(1,\frac{1}{n-1})\right) +\end{aligned}$$

$$+\lambda_{n-1}(M(0,1)+M(1,0)).$$

Очевидно, что модели, на которых значения формул совпадают, рассматривать не нужно, поэтому полагаем $\lambda_0=0$. Модели, на которых формула φ принимает значение 0, а формула ψ принимает значение 1 (и наоборот) берем с весом $\lambda_{n-1}=1$. Полагаем, что чем меньше модуль разности между значениями формул φ и ψ на модели, тем они ближе на данной модели, значит, их нужно брать с меньшим весом, поэтому $\lambda_i \le \lambda_{i+1}$ для $i = 0,..,n-1$.

Остается нормировать величину ρ_0.

Определение 7. Растоянием между формулами φ и ψ n-значной логики L_n при $S(\varphi)\cup S(\psi)\subseteq S(\Sigma)$ на множестве $P(S(\Sigma))$ назовём величину

$$\rho(\varphi,\psi)=\frac{1}{n^{|S(\Sigma)|}}\cdot\sum_{k=0}^{n-1}\sum_{l=0}^{n-1}\lambda_{|k-l|}\cdot M(\frac{k}{n-1},\frac{l}{n-1}),$$

где λ_i удовлетворяют условиям:

$$0=\lambda_0 \leqslant \lambda_i \leqslant \lambda_{n-1}=1;$$

$$\lambda_i \le \lambda_{i+1},\ i=0,...,n-1;$$

$$\lambda_i+\lambda_{n-1-i}=1\ \ \forall i=0,...,\frac{n-1}{2}.$$

Установим свойства модельного расстояния ρ .

Теорема 1. Любые формулы φ,ψ,χ из Σ удовлетворяют следующим свойствам:

1. $0\leqslant\rho(\varphi,\psi)\leqslant 1$;
2. $\rho(\varphi,\psi)=0 \Leftrightarrow \varphi\equiv\psi$;
3. $\rho(\varphi,\psi)=\rho(\psi,\varphi)$;
4. $\rho(\varphi,\psi)\leqslant\rho(\varphi,\chi)+\rho(\chi,\psi)$;
5. $\varphi\equiv\varphi_1,\psi\equiv\psi_1\Rightarrow\rho(\varphi,\psi)=\rho(\varphi_1,\psi_1)$;
6. $\rho(\varphi,\psi)=\rho(\neg\varphi,\neg\psi)$;
7. $\rho((\varphi\wedge\psi),(\varphi\vee\psi))=\rho(\varphi,\psi)$.

Доказательство пунктов Теоремы 1 изложено в работах [7,8,9].

В классической логике под информативностью высказывания понимают относительное число моделей, на которых высказывание эксперта ложно. Другими словами, это расстояние от высказывания до тождественно истинной формулы. Чем меньше моделей, на которых высказывание истинно, тем оно информативней, то есть менее тривиально. Поэтому вместо термина "мера информативности" будем использовать термин "мера нетривиальности".

Зададим меру нетривиальности для случая конечнозначной логики L_n. Так как в L_n имеется $(n-2)$ истинностных значений, отличных от 1, то нужно учитывать модели, на которых формула принимает значение $\frac{k}{n-1}$, $k=0,...,n-2$. Также нужно учесть, насколько близко значение формулы к 1. Естественно положить, что коэффициент при модели, на которой формула принимает значение $\frac{k}{n-1}$ должен быть больше, чем коэффициент при модели, на которой формула принимает значение $\frac{l}{n-1}$ при $l>k$, так как $\frac{l}{n-1}$ ближе к 1.

Определение 8. Мерой нетривиальности I формулы φ n-значной логики L_n, $S(\varphi)\cup S(\psi)\subseteq S(\Sigma)$ из $P(S(\Sigma))$ называется величина

$$I(\varphi)=\frac{1}{n^{|S(\Sigma)|}}\sum_{k=0}^{n-1}\alpha_{|(n-1)-k|}\cdot M(\frac{k}{n-1}), \tag{1}$$

где величины α_i удовлетворяют условиям:

$$0=\alpha_0\leqslant\alpha_i\leqslant\alpha_{n-1}=1;$$

$$\alpha_i\leqslant\alpha_{i+1},\ i=0,...,n-1;$$

$$\alpha_i+\alpha_{n-1-i}=1\ \ \forall\, i=0,...,\frac{n-1}{2}.$$

В выражении (1) набор коэффициентов $\vec{\alpha}=(\alpha_0,...,\alpha_{n-1})$ для меры нетривиальности I зависит от набора коэффициентов $\Lambda=(\lambda_0,...,\lambda_{n-1})$ модельного расстояния ρ, так как мера нетривиальности сама по себе является расстоянием. Кроме того, при совпадении набора $\vec{\alpha}$ и Λ выполняется:

$$I(\varphi)=\rho(\varphi,1).$$

В частном случае двузначной логики, данное свойство было положено в основу определения меры информативности высказываний в работе [14].

Установим свойства меры нетривиальности I.

Теорема 2. Любые формулы φ,ψ из Σ удовлетворяют следующим свойствам:

1. $0\leqslant I(\varphi)\leqslant 1$;
2. $I(\varphi)+I(\neg\varphi)=1$;
3. $I(\varphi\wedge\psi)\geqslant max\{I(\varphi),I(\psi)\}$;
4. $I(\varphi\vee\psi)\leqslant min\{I(\varphi),I(\psi)\}$;

5. $I(\varphi \wedge \psi) + I(\varphi \vee \psi) \geqslant I(\varphi) + I(\psi)$.

Доказательство пунктов Теоремы 2 изложено в работах [7,8,9].

Коллективное расстояние и мера нетривиальности

Известно, что устойчивость решений в задачах машинного обучения, в частности, в кластерном анализе [11-13], 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 машинном обучении [1]. Идея, лежащая в основе этого подхода, может быть применена и при анализе логических высказываний.

Пусть имеется множество, состоящее из l формул $\varphi_1,...,\varphi_l$. Допустим, рассмотрено k произвольных наборов $\Lambda_1,...,\Lambda_k$ и для каждого набора проведено разбиение множества формул на некоторое заданное число кластеров. Для этого может применяться алгоритм кластерного анализа, который в качестве входной информации использует матрицу попарных расстояний между объектами (в нашем случае под объектами понимаются формулы). Поскольку каждый набор Λ_m определяет, вообще говоря, свой тип расстояния, то и соответствующие разбиения на кластеры могут не совпадать. Для каждого *m*-го полученного набора кластеров вычислим индекс качества группировки $s_m,\ m=1,...,k$, который показывает общую степень компактности кластеров и их удаленности друг от друга [15] (например, можно использовать индекс Дунна или индекс кофенетической корреляции). Пусть индексы нормированы так, что $s_m \in [0,1],\ \sum s_m = 1$ и будем считать, что при улучшении качества значение индекса увеличивается.

Для каждой пары формул φ_i , φ_j имеется несколько вариантов расстояний.

Определение 9. Коллективным расстоянием для φ_i и φ_j назовем следующую величину:

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

где $\rho_m(\varphi_i,\varphi_j)$ - расстояние между формулами φ_i и φ_j для набора $\vec{\alpha}_m,\ m=1,...,k.$ Таким образом, коллективное расстояние между парой формул основано на усреднении по всем имеющимся вариантам расстояний с весами, равными показателям качества группировки, проведенной на основе каждого варианта. При таком способе больший вес имеют те варианты задания расстояния, которые позволяют получить более четко выраженные кластерные структуры на множестве формул.

Аналогично введем коллективную меру нетривиальности.

Определение 10. Коллективной мерой нетривиальности для формулы φ_i назовем следующую величину:

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

где $I_m(\varphi_i)$ - мера нетривиальности формулы φ_i для m-го набора $\vec{\alpha}_m,\ m=1,...,k.$

Численные эксперименты по кластеризации формул *n*-значной логики

Для множеств высказываний известны только попарные расстояния между формулами и меры нетривиальности. Поэтому для кластеризации были выбраны два широкоизвестных алгоритма, которые предназначены для анализа данных, информация о которых имеет форму матрицы попарных расстояний - иерархический агломеративный алгоритм [16] и алгоритм k-медоидов [17], который можно рассматривать как модификацию алгоритма k-средних.

Разработан программный комплекс, в котором эти алгоритмы кластеризации адаптированы для работы с конечными множествами формул в L_n [10]. Сложность вычисления расстояния между высказываниями экспоненциально увеличивается с ростом их числа, поэтому эксперименты проведены для относительно небольших значений. Ниже приведены результаты численных экспериментов, проведенных с помощью этого программного комплекса.

Использование иерархического алгоритма, $n = 9$

Пусть $n = 9$. Рассмотрим множество $test$ 1, состоящее из 10 формул девятизначной логики L_9: $\varphi_1 = x \to y$, $\varphi_2 = \neg(x \to y)$, $\varphi_3 = (x \vee z) \to y$, $\varphi_4 = \neg((x \wedge y) \vee z) \to w$, $\varphi_5 = y \to (x \wedge z)$, $\varphi_6 = (\neg y \vee (x \to z)) \to w$, $\varphi_7 = ((x \to y) \to z) \to w$, $\varphi_8 = (w \to z) \wedge (y \to x)$, $\varphi_9 = (\neg x \to z) \to (w \wedge y)$, $\varphi_{10} = \neg((x \wedge (y \vee z)) \to (w \wedge \neg z))$.

Проведём кластерный анализ множества $test$ 1, используя иерархический алгоритм. Выберем следующие веса Λ_1:

$\lambda_0 = 0, \lambda_1 = \frac{1}{8}, \lambda_2 = \frac{2}{8}, \lambda_3 = \frac{3}{8}, \lambda_4 = \frac{4}{8}, \lambda_5 = \frac{5}{8}, \lambda_6 = \frac{6}{8}, \lambda_7 = \frac{7}{8}, \lambda_8 = 1$ (стандартный случай). После 8 итераций результаты кластеризации представлены в Таблице 1. Здесь через Δ обозначено пороговое расстояние, испозуемое для отсечения кластеров на очередной итерации объединения, а через $\varphi_{ij...r}$ множество формул $\varphi_i, \varphi_j, ..., \varphi_r$, объединенных в один кластер.

Таблица 1. Результаты работы иерархического алгоритма, $test$ 1, $n = 9$, набор весов Λ_1

Итерация	Δ	Кластеры
0	0,0000	$\varphi_1,\varphi_2,\varphi_3,\varphi_4,\varphi_5,\varphi_6,\varphi_7,\varphi_8,\varphi_9,\varphi_{10}$
1	0,0073	$\varphi_{46},\varphi_1,\varphi_2,\varphi_3,\varphi_5,\varphi_7,\varphi_8,\varphi_9,\varphi_{10}$
2	0,0173	$\varphi_{13},\varphi_{46},\varphi_2,\varphi_5,\varphi_7,\varphi_8,\varphi_9,\varphi_{10}$
3	0,0952	$\varphi_{13},\varphi_{467},\varphi_2,\varphi_5,\varphi_8,\varphi_9,\varphi_{10}$
4	0,0952	$\varphi_{13},\varphi_{467},\varphi_{58},\varphi_2,\varphi_9,\varphi_{10}$
5	0,1907	$\varphi_{13467},\varphi_{589},\varphi_2,\varphi_{10}$
6	0,1907	$\varphi_{13467},\varphi_{58910},\varphi_2$
7	0,3433	$\varphi_{1345678910},\varphi_2$
8	0,5892	$\varphi_{12345678910}$

Теперь зададим набор весов Λ_2:

$\lambda_0 = 0, \lambda_1 = \frac{1}{9}, \lambda_2 = \frac{2}{8}, \lambda_3 = \frac{5}{16}, \lambda_4 = \frac{4}{8}, \lambda_5 = \frac{11}{16}, \lambda_6 = \frac{6}{8}, \lambda_7 = \frac{8}{9}, \lambda_8 = 1$. Результаты группировки показаны в Таблице 2.

Таблица 2 - Результаты работы иерархического алгоритма, $test$ 1, $n = 9$, набор весов Λ_2

Итерация	Δ	Кластеры
0	0,0000	$\varphi_1,\varphi_2,\varphi_3,\varphi_4,\varphi_5,\varphi_6,\varphi_7,\varphi_8,\varphi_9,\varphi_{10}$
1	0,0051	$\varphi_{46},\varphi_1,\varphi_2,\varphi_3,\varphi_5,\varphi_7,\varphi_8,\varphi_9,\varphi_{10}$
2	0,0218	$\varphi_{13},\varphi_{46},\varphi_2,\varphi_5,\varphi_7,\varphi_8,\varphi_9,\varphi_{10}$
3	0,0464	$\varphi_{13},\varphi_{46},\varphi_{58},\varphi_2,\varphi_7,\varphi_9,\varphi_{10}$
4	0,0859	$\varphi_{13},\varphi_{467},\varphi_{58},\varphi_2,\varphi_9,\varphi_{10}$
5	0,1933	$\varphi_{13467},\varphi_{589},\varphi_2,\varphi_{10}$
6	0,2496	$\varphi_{13467},\varphi_{58910},\varphi_2$
7	0,3572	$\varphi_{1345678910},\varphi_2$
8	0,5713	$\varphi_{12345678910}$

Использование иерархического алгоритма, $n = 10$

Пусть $n = 10$. Рассмотрим множество $test$ 1 и проведём кластерный анализ множества $test$ 1, используя иерархический алгоритм. Выберем следующие веса Λ_3:

$$\lambda_0 = 0, \lambda_1 = \frac{1}{9}, \lambda_2 = \frac{2}{9}, \lambda_3 = \frac{3}{9}, \lambda_4 = \frac{4}{9}, \lambda_5 = \frac{5}{9}, \lambda_6 = \frac{6}{9}, \lambda_7 = \frac{7}{9}, \lambda_8 = \frac{8}{9}, \lambda_9 = 1.$$

(стандартный случай). После 9 итераций результаты кластеризации представлены в Таблице 3.

Таблица 3. Результаты работы иерархического алгоритма, $test$ 1, $n = 10$, набор весов Λ_3.

Итерация	Δ	Кластеры
0	0,0000	$\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
1	0,0081	$\varphi_{46}, \varphi_1, \varphi_2, \varphi_3, \varphi_5, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
2	0,0081	$\varphi_{13}, \varphi_{46}, \varphi_2, \varphi_5, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
3	0,0173	$\varphi_{13}, \varphi_{467}, \varphi_2, \varphi_5, \varphi_8, \varphi_9, \varphi_{10}$
4	0,0952	$\varphi_{13467}, \varphi_2, \varphi_5, \varphi_8, \varphi_9, \varphi_{10}$
5	0,0952	$\varphi_{13467}, \varphi_{58}, \varphi_2, \varphi_9, \varphi_{10}$
6	0,1907	$\varphi_{13467}, \varphi_{589}, \varphi_2, \varphi_{10}$
7	0,1907	$\varphi_{13467}, \varphi_{58910}, \varphi_2$
8	0,3864	$\varphi_{1345678910}, \varphi_2$
9	0,5121	$\varphi_{12345678910}$

Теперь выберем набор весов Λ_4:

$$\lambda_0 = 0, \lambda_1 = \frac{1}{10}, \lambda_2 = \frac{5}{18}, \lambda_3 = \frac{3}{9}, \lambda_4 = \frac{7}{18}, \lambda_5 = \frac{11}{18}, \lambda_6 = \frac{6}{9}, \lambda_7 = \frac{13}{18}, \lambda_8 = \frac{9}{10}, \lambda_9 = 1.$$

Результаты группирования приведены в Таблице 4.

Таблица 4. Результаты работы иерархического алгоритма, $test$ 1, $n = 10$, набор весов Λ_4

Итерация	Δ	Кластеры
0	0,0000	$\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
1	0,0074	$\varphi_{46}, \varphi_1, \varphi_2, \varphi_3, \varphi_5, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
2	0,0102	$\varphi_{13}, \varphi_{46}, \varphi_2, \varphi_5, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
3	0,0195	$\varphi_{13}, \varphi_{46}, \varphi_{58}, \varphi_2, \varphi_7, \varphi_9, \varphi_{10}$
4	0,0514	$\varphi_{13}, \varphi_{467}, \varphi_{58}, \varphi_2, \varphi_9, \varphi_{10}$
5	0,0701	$\varphi_{13467}, \varphi_{58}, \varphi_2, \varphi_9, \varphi_{10}$
6	0,1387	$\varphi_{13467}, \varphi_{589}, \varphi_2, \varphi_{10}$
7	0,2017	$\varphi_{13467}, \varphi_{58910}, \varphi_2$
8	0,4553	$\varphi_{1345678910}, \varphi_2$
9	0,5216	$\varphi_{12345678910}$

Использование алгоритма k-медоидов, $n = 7$

Пусть $n = 7$. Рассмотрим множество $test$ 2, состоящее из 15 формул семизначной логики L_7: $\varphi_1 = \neg(z \to w)$, $\varphi_2 = x \wedge (y \vee z)$, $\varphi_3 = x \to (z \wedge y)$, $\varphi_4 = (x \vee w) \to z$, $\varphi_5 = x \wedge (z \to \neg y)$, $\varphi_6 = \neg(z \to (x \wedge z))$, $\varphi_7 = y \vee (\neg x \to w)$, $\varphi_8 = (x \to z) \wedge \neg w$, $\varphi_9 = \neg((w \wedge z) \to (x \vee y))$, $\varphi_{10} = (x \to \neg y) \vee (\neg z \to w)$, $\varphi_{11} = x \vee \neg y$, $\varphi_{12} = z \wedge (y \to x)$, $\varphi_{13} = \neg z \vee (x \to (y \wedge v))$, $\varphi_{14} = (y \to z) \to \neg y$, $\varphi_{15} = \neg((x \to y) \wedge z)$.

Произведём кластерный анализ множества $test$ 2, используя алгоритм k-медоидов. Допустим, что нужно получить 4 кластера. Выберем следующие веса: $\lambda_0 = 0, \lambda_1 = \frac{1}{6}, \lambda_2 = \frac{2}{6}, \lambda_3 = \frac{3}{6}, \lambda_4 = \frac{4}{6}, \lambda_5 = \frac{5}{6}, \lambda_6 = 1$. Основываясь на матрице расстояний, выберем 4 центра: $\varphi_3, \varphi_4, \varphi_9, \varphi_{12}$ (центры примерно равноудалены друг от друга и сумма расстояний между ними максимальна). Оставшиеся формулы распределяются в исходные кластеры в соответствии с расстоянием до центров. Это даёт нам следующие кластеры: $\varphi_{3,7,8,11}, \varphi_{1,2,4,6,13,14}, \varphi_{5,9,15}, \varphi_{10,12}$. Затем алгоритм снова вычисляет центры масс и перераспределяет формулы по обновлённым центрам. Алгоритм продолжает работу, пока кластеры не прекратят изменяться. Результаты кластеризации представлены в Таблице 5.

Таблица 5. Результаты работы алгоритма k-медоидов, $test$ 2, $n=7$

Итерация	Центры	Кластеры
1	$\varphi_3,\varphi_4,\varphi_9,\varphi_{12}$	$\varphi_{3,7,8,11},\varphi_{1,2,4,6,13,14},\varphi_{5,9,15},\varphi_{10,12}$
2	$\varphi_3,\varphi_6,\varphi_9,\varphi_{12}$	$\varphi_{3,4,7,8,11,13},\varphi_{1,2,6,14},\varphi_{9,15},\varphi_{5,10,12}$
3	$\varphi_4,\varphi_6,\varphi_9,\varphi_{10}$	$\varphi_{3,4,7,8,11,13},\varphi_{1,2,6,14},\varphi_{9,15},\varphi_{5,10,12}$

Алгоритм останавливается после третьей итерации и даёт в результате 4 кластера: $\varphi_{3,4,7,8,11,13},\varphi_{1,2,6,14},\varphi_{9,15},\varphi_{5,10,12}$.

Использование алгоритма k-медоидов, $n=10$

Пусть $n=10$. Рассмотрим множество $test$ 3, состоящее из 15 формул десятизначной логики L_{10}: $\varphi_1 = x\wedge(\neg z\to y)$, $\varphi_2 = w\to(x\wedge y)$, $\varphi_3=(y\to z)\vee(\neg(y\wedge z))$, $\varphi_4=(\neg x\wedge w)\to w$, $\varphi_5=y\to(y\to w)$, $\varphi_6=x\to(w\vee y)$, $\varphi_7=x\wedge(x\to\neg z)$, $\varphi_8=\neg((y\vee z)\wedge w)$, $\varphi_9=x\vee(x\wedge(y\to w))$, $\varphi_{10}=(x\to\neg y)\vee(\neg z\to w)$, $\varphi_{11}=(\neg x\to z)\to(w\wedge y)$, $\varphi_{12}=x\to(w\vee(x\to\neg z))$, $\varphi_{13}=(x\to w)\to(z\vee w)$, $\varphi_{14}=(w\vee x)\wedge x$, $\varphi_{15}=(x\to y)\to(\neg z\wedge w))$.

Произведём кластерный анализ множества $test$ 3, используя алгоритм k-медоидов. Допустим, что нужно получить 3 кластера. Выберем следующие веса: $\lambda_0=0,\lambda_1=\frac{1}{9},\lambda_2=\frac{2}{9},\lambda_3=\frac{3}{9},\lambda_4=\frac{4}{9},\lambda_5=\frac{5}{9},\lambda_6=\frac{6}{9},\lambda_7=\frac{7}{9},\lambda_8=\frac{8}{9},\lambda_9=1$. Основываясь на матрице расстояний, выберем 3 центра: $\varphi_1,\varphi_8,\varphi_{15}$. Оставшиеся формулы распределяются в исходные кластеры в соответствии с расстоянием до центров. Затем алгоритм снова вычисляет центры масс и перераспределяет формулы по обновлённым центрам. Алгоритм продолжает работу, пока кластеры не прекратят изменяться. Результаты кластеризации представлены в Таблице 6.

Таблица 6. Результаты работы алгоритма k-медоидов, $test$ 3, $n=10$

Итерация	Центры	Кластеры
1	$\varphi_1,\varphi_8,\varphi_{15}$	$\varphi_{1,3,4,10,11},\varphi_{6,8,9,12,13},\varphi_{2,5,7,14,15}$
2	$\varphi_{10},\varphi_6,\varphi_7$	$\varphi_{1,3,4,7,10,11},\varphi_{5,6,8,9,12,13},\varphi_{2,14,15}$
3	$\varphi_{10},\varphi_8,\varphi_7$	$\varphi_{1,3,4,7,10,11},\varphi_{5,6,8,9,12,13},\varphi_{2,14,15}$

Алгоритм останавливается после третьей итерации и даёт в результате 3 кластера: $\varphi_{1,3,4,7,10,11}, \varphi_{5,6,8,9,12,13}, \varphi_{2,14,15}$.

Использование коллективных расстояний и иерархического алгоритма, $n = 9,10$

Рассмотрим множество $test$ 1 и проведём кластерный анализ, используя иерархический алгоритм на основе коллективных расстояний, вычисляемых на основе наборов весов Λ_1, Λ_2. Качество кластеризации определялось с помощью индекса Дунна. После 9 итераций результаты кластеризации представлены в Таблице 7.

Таблица 7. Результаты работы иерархического алгоритма, $test$ 1, $n = 9$, коллективное расстояние

Итерация	Δ	Кластеры
0	0,0000	$\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
1	0,0076	$\varphi_{46}, \varphi_1, \varphi_2, \varphi_3, \varphi_5, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
2	0,0165	$\varphi_{13}, \varphi_{46}, \varphi_2, \varphi_5, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}$
3	0,0252	$\varphi_{13}, \varphi_{467}, \varphi_2, \varphi_5, \varphi_8, \varphi_9, \varphi_{10}$
4	0,0300	$\varphi_{13}, \varphi_{467}, \varphi_{58}, \varphi_2, \varphi_9, \varphi_{10}$
5	0,0383	$\varphi_{13467}, \varphi_{58}, \varphi_2, \varphi_9, \varphi_{10}$
6	0,0498	$\varphi_{13467}, \varphi_{589}, \varphi_2, \varphi_{10}$
7	0,0577	$\varphi_{13467}, \varphi_{58910}, \varphi_2$
8	0,0654	$\varphi_{1345678910}, \varphi_2$
9	0,0761	$\varphi_{12345678910}$

Теперь повысим размерность логики до 10. Проведём кластерный анализ множества $test$ 1, используя иерархический алгоритм на основе коллективных расстояний, вычисляемых на основе наборов весов Λ_3, Λ_4 После 9 итераций результаты кластеризации представлены в Таблице 8.

Таблица 8. Результаты работы иерархического алгоритма, $test$ 1, $n = 10$, коллективное расстояние

Итерация	Δ	Кластеры
0	0,0000	$\varphi_1,\varphi_2,\varphi_3,\varphi_4,\varphi_5,\varphi_6,\varphi_7,\varphi_8,\varphi_9,\varphi_{10}$
1	0,0066	$\varphi_{46},\varphi_1,\varphi_2,\varphi_3,\varphi_5,\varphi_7,\varphi_8,\varphi_9,\varphi_{10}$
2	0,0178	$\varphi_{13},\varphi_{46},\varphi_2,\varphi_5,\varphi_7,\varphi_8,\varphi_9,\varphi_{10}$
3	0,0245	$\varphi_{13},\varphi_{467},\varphi_2,\varphi_5,\varphi_8,\varphi_9,\varphi_{10}$
4	0,0324	$\varphi_{13467},\varphi_2,\varphi_5,\varphi_8,\varphi_9,\varphi_{10}$
5	0,0453	$\varphi_{13467},\varphi_{58},\varphi_2,\varphi_9,\varphi_{10}$
6	0,0518	$\varphi_{13467},\varphi_{589},\varphi_2,\varphi_{10}$
7	0,0679	$\varphi_{13467},\varphi_{58910},\varphi_2$
8	0,0822	$\varphi_{1345678910},\varphi_2$
9	0,0937	$\varphi_{12345678910}$

Наблюдения и выводы для различных значений n

Была создана база из 300 различных логических формул, откуда случайным образом выбирались подмножества формул. С помощью адаптированных алгоритмов, применяемых в предыдущей главе, было кластеризовано более 50 таких подмножеств при различных значениях n и λ .

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

1. Для $n = 2,...,9$ наблюдается разница в составе кластеров, а при $n > 9$ кластеры и последовательность итераций не меняется. Таким образом, возникает гипотеза о нецелесообразности использования логики большой значности в реальных задачах от малого числа переменных.

2. В большинстве случаев применение коллективных расстояний дает лучшее значение индексов качества кластеризации.

Заключение

В данной работе решалась задача анализа экспертных знаний, представленных в виде логических высказываний - формул *n*-значной логики Лукасевича. Основываясь на коллективном подходе в машинном обучении, предложены новые метрики на формулах, позволяющие проводить их кластеризацию и оценивать информативность высказываний. Для этого были адаптированы два алгоритма кластеризации – иерархический агломеративный и k-медоидов (k-medoids). Проведены численные эксперименты для различных примеров значений n и расстояний, показаны различия в полученных результатах группировки.

Для автоматической кластеризации разработано рабочее java-приложение. Сложность алгоритма вычисления расстояния между формулами - экспоненциальная.

В дальнейшем планируется адаптировать для решения рассматриваемой задачи и другие алгоритмы кластеризации (например спектральный алгоритм). Актуальной является задача уменьшения сложности алгоритма вычисления расстояний на основе приближенных полиномиальных алгоритмов, решение которой необходимо для проведения анализа большого числа формул. Представляется возможным обобщить развиваемый подход на случай задачи классификации логических высказываний (обучения «с учителем»). В этой задаче требуется отнести некоторое новое высказывание к одному из возможных классов, используя информацию об имеющихся формулах и их принадлежности классам. Планируется также применить полученные результаты для решения практических задач, связанных с разработкой интеллектуальных систем в различных прикладных областях.

Благодарности

Работа поддержана Российским фондом фундаментальных исследований, проект 18-07-00600а.

СПИСОК ЛИТЕРАТУРЫ

1. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / пер. с англ. А.А. Слинкина. - М.: ДМК Пресс, 2015.
2. Tiwari M., Mishra B. Application of cluster analysis in expert system – a brief survey // International Journal of Computer Science Issues. 2011. Vol. 8, Issue 5, No 1. P. 342-346.
3. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Ин-та математики, 1999. – С 270.
4. Vikent'ev A.A., Lbov G.S. Setting the metrics and informativeness on statements of experts // Pattern Recognition And Image Analysis, 1997, Vol. 7, No. 2, P. 175-183.
5. Викентьев А.А., Викентьев Р. А. Расстояния и меры недостоверности на высказываниях n-значной логики // Вестник НГУ. Сер. матем., мех., информ., 2011, Т. 11:2, – С. 51-64.
6. Викентьев А.А. О возможных расстояниях и степенях недостоверности в многозначных высказываниях экспертов и приложение этих понятий в проблемах кластеризации и распознавания // Проблемы информатики, Новосибирск: СО РАН, 2011, Т. 3, – С. 33-45.
7. Vikent'ev A.A. Distances and Degrees of Uncertainty in Many-Valued Propositions of Experts and Application of These Concepts in Problems of Pattern Recognition and Clustering // Pattern Recognition and Image Analysis, 2014, Vol. 24, No. 4, – P. 489-501.
8. Vikent'ev, A.A., Avilov M.S. New Model Distances and Uncertainty Measures for Multivalued Logic. // C. Dichev, G. Agre (Eds.): Artificial Intelligence: Methodology,

Systems, and Applications, Lecture Notes on Computer Science, 2016, LNCS 9883, – P. 89-98.

9. Викентьев А.А., Кабанова Е.С. Расстояния между формулами пятизначной логики Лукасевича и мера недостоверности высказываний экспертов в кластеризации баз знаний // Вестник Томского государственного университета, Томск, 2013, 2(23), – С 121-129.

10. Викентьев А.А., Фефелова В.В. Новые модельные расстояния и меры достоверности формул логики Лукасевича в автоматической кластеризации высказываний баз данных // Algebra and Model Theory 10. Collection of papers / Eds.: A.G.Pinus, K.N.Ponomaryov, S.V.Sudoplatov, and E.I.Timoshenko, Novosibirsk, 2015. – P. 197-209.

11. Авилов М.С. Программный комплекс вычислений расстояний, мер нетривиальности и кластеризации множеств высказываний в n-значной логике // Материалы 53-й международной научной студенческой конференции МНСК-2015, Новосибирск, 2015, Математика, – С. 6.

12. Berikov V.B. Weighted ensemble of algorithms for complex data clustering // Pattern Recognition Letters. 2014. Vol. 38. – P. 99-106.

13. Berikov, V., Pestunov, I.: Ensemble clustering based on weighted co-association matrices: Error bound and convergence properties // Pattern Recognition. 2017. Vol. 63. – P. 427-436.

14. Amirgaliyev Y., Berikov V., Latuta K., Bekturgan K., Cherikbayeva L. Group Approach to Solving the Tasks of Recognition // Yugoslav Journal of Operations Research. 2019. Vol. 29, N 2. – P. 177-192.

15. Блощицын В.Я., Лбов Г.С. О мерах информативности логических высказываний // Доклады Республиканской школы-Семинара "Технология разработки экспертных систем". Кишинев, 1978. – С.12-14.

16. Arbelaitz O., Gurrutxaga I., Muguerza J., Perez J., Perona I. An extensive comparative study of cluster validity indices // Pattern Recognition. 2013. – P. 243-256.

17. Дуда Р., Харт П. Распознавание образов и анализ сцен. - М.: Мир, 1976.

КРИВИЗНА ФОРМАНА-РИЧЧИ В ТЕКСТУРНОМ АНАЛИЗЕ ДДЗ АРИДНЫХ ТЕРРИТОРИЙ

Волобуев Д.М., Макаренко Н.Г., Терехов А.Г., Джаксылыкова А.Б.

e-mail: ng-makar@mail.ru

Институт информационных и вычислительных технологий КН МОН РК, Казахстан

Аннотация. *В работе обсуждается применение дискретной кривизны Формана-Риччи для выделения сезонной компоненты аридных территорий по цифровым изображениям космического мониторинга. Такие территории обычно имеют текстуру с высокой вегетативной вариативностью. Дискретный аналог*