УДК 51

А.А.Викентьев[1,2], М.С.Авилов[2]

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

Новые полные метрики и меры нетривиальности для формул многозначных логик в автоматической кластеризации формул из логической базы знаний. I

В статье рассмотрены следующие задачи: обобщены ранее введённые расстояния и меры нетривиальности на n-значный случай любой логики, сняв ограничения на параметры; используя новые расстояния, разработаны алгоритмы кластеризации множеств формул в n-значной логике; проведены кластеризации подмножеств формул из различных баз знаний и предложены способы сравнения результатов различных адаптированных алгоритмов кластеризации.

Ключевые слова: расстояния и меры, кластеризация множеств, логика, алгоритмы.

Введение

Задачи анализа экспертной информации и извлечения знаний из баз данных приобретают всё большую важность [1-11]. Одним из направлений этой проблематики являются информативность высказываний экспертов [8] и работа с базами знаний [9]. Определение информативности экспертного высказывания зависит от определения информативности его компонент, а также от степени различия этих компонент. Если мы сможем ввести метрику на множестве экспертных высказываний, то их, возможно, стоит подвергнуть кластерному анализу, что даст мощный инструмент для анализа имеющейся информации и изучения новых знаний.

Задача сравнения экспертных высказываний по информативности, их ранжированию и введению метрики на пространстве высказываний была поставлена Г.С. Лбовым и Н.Г. Загоруйко в середине 1990-х гг. [10], [12].

Экспертные высказывания было предложено записывать в виде логических формул исчисления. Таким образом, задача сравнения и ранжирования высказываний экспертов превращается в задачу сравнения и ранжирования логических формул между собой. Для этого надо ввести расстояние между формулами и меру нетривиальности каждой формулы.

Для 2-значной логики решением этой задачи занимались Г.С. Лбов и А. А. Викентьев [7]. Впоследствии А. А. Викентьевым был предложен более общий теоретико-модельный подход к анализу логических высказываний [13]. В качестве исчисления высказываний берется n-значная логика L_n. С помощью развитой теории моделей для формул многозначной логики вводились разные варианты расстояний и мер нетривиальности (опровержимости, недостоверности, информативности), исследовались свойства этих величин [2]. Затем на основе введённых расстояний и мер нетривиальности стали проводить кластеризацию множеств формул n-значной логики Лукасевича [14] и решалась задача ранжирования экспертных высказываний.

Кластеризация проводилась для малых n и небольшого числа переменных в формулах, а расстояние (более полное) вводилось с логическими (жёсткими) весами [6]. Чтобы автоматизировать вычислительную работу, с необходимостью создавали программы для вычисления расстояний, мер нетривиальности и кластеризации множеств формул многозначной логики.

1 Определения и обозначения

Определим n-значную логику L_n, формулами которой представляются высказывания, а также рассмотрим применение теории моделей к формулам L_n.

1.1 Конструкция n-значной логики L_n

Определение 1.1.1. 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 1.1.2. Под формулами будем понимать следующие последовательности слов из 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, не являются формулами.*

Теперь мы можем построить типичную n-значную логику L_n.

Определение 1.1.3. n-значная матричная логика L_n определяется через $M_n = \langle V_n, \neg, \rightarrow, \{1\}\rangle$ – *логическую матрицу*;

$V_n = \left\{0, \dfrac{1}{n-1}, ..., \dfrac{n-2}{n-1}, 1\right\}$ – *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.

Определение 1.1.4. Логические операции на V_n задаются следующим образом:

$\neg x = 1 - x$ – *отрицание*;

$x \rightarrow y = min\{1, 1 - x + y\}$ – *импликация*;

$x \vee y = (x \rightarrow y) \rightarrow y = max\{x, y\}$ – *дизъюнкция*;

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

1.2 Основные понятия и модели для логик

Пусть Σ – конечное множество формул L_n.

Определение 1.2.1. $S(\varphi)$ – носитель формулы φ из L_n (множество переменных, используемых при ее написании).

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

Определение 1.2.2. Модель M – кортеж из означенных переменных, на котором вычислимо значение истинности любой формулы.

$P(S(\Sigma))$ – множество всех моделей, $\mid P(S(\Sigma)) \mid= n^{|S(\Sigma)|}$ [13]. Будем использовать обозначение $\varphi_{\frac{k}{n-1}}$, когда φ принимает на модели значение $\frac{k}{n-1}, k = 0, ..., n-1$. $M(\varphi_{\frac{k}{n-1}})$ – 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 $\frac{l}{n-1}$.

2 Расстояние и мера нетривиальности формул в L_n

В определении расстояния мы учитываем разницу между любыми различными значениями двух формул на каждой модели. Таким образом, более полно используем многозначность логических формул. Мера нетривиальности служит для ответа на вопрос: насколько добавление формулы в базу знаний повышает ее информативность. Чем более формула недостоверна (а значит, информативна), тем выше мера её нетривиальности.

Покажем, как строится расстояние для формул n-значной логики L_n. Естественно предположить, что чем меньше модуль разности между значениями φ и ψ, тем ближе формулы в данной модели. Будем умножать количество моделей с одинаковыми модулями разности на коэффициент, учитывающий близость значений формул. Ранее в качестве таких коэффициентов брались n истинностных значений (для логики Лукасевича) L_n:

$$\tilde{\rho_0}(\varphi,\psi) = 0\cdot\left(M(0,0)+M(\frac{1}{n-1},\frac{1}{n-1})+M(\frac{2}{n-1},\frac{2}{n-1})+...+M(\frac{n-2}{n-1},\frac{n-2}{n-1})+M(1,1)\right)+$$

$$+\frac{1}{n-1}\cdot\left(M(0,\frac{1}{n-1})+M(\frac{1}{n-1},0)+M(\frac{1}{n-1},\frac{2}{n-1})+...+M(\frac{n-2}{n-1},1)+M(1,\frac{n-2}{n-1})\right)+...$$

$$...+\frac{n-2}{n-1}\cdot\left(M(0,\frac{n-2}{n-1})+M(\frac{n-2}{n-1},0)+M(\frac{1}{n-1},1)+M(1,\frac{1}{n-1})\right)+$$

$$+1\cdot(M(0,1)+M(1,0)) = \sum_{k=0}^{n-1}\sum_{l=0}^{n-1}\frac{\mid k-l\mid}{n-1}\cdot M(\frac{k}{n-1},\frac{l}{n-1}).$$

Таким образом, в качестве расстояния между формулами φ и ψ n-значной логики L_n при $S(\varphi)\cup S(\psi)\subseteq S(\Sigma)$ на множестве $P(S(\Sigma))$ ранее принималась нормированная величина $\tilde{\rho_0}$:

$$\rho_0(\varphi,\psi) = \frac{1}{n^{|S(\Sigma)|}}\cdot\sum_{k=0}^{n-1}\sum_{l=0}^{n-1}\frac{\mid k-l\mid}{n-1}\cdot M(\frac{k}{n-1},\frac{l}{n-1}). \qquad (1)$$

Пример. Для случая $n=5$ $\rho_0(\varphi\wedge\psi,\varphi\vee\psi)=0.4$, $\rho_0(\varphi\wedge\psi\wedge\chi\wedge\omega,\varphi\to\omega)=0.2576$ [6]. Недостатком величины является взятая конструкция весов $\frac{|k-l|}{n-1}$. Она не позволяет выбирать, с каким именно коэффициентом величина $M(\frac{k}{n-1},\frac{l}{n-1})$ (число моделей с одинаковым модулем разности) будет входить в итоговое расстояние ρ_0. Такая же проблема возникает для ранее использованной меры нетривиальности формул n-значной логики L_n:

$$I_0(\varphi) = \rho_0(\varphi,1) = \sum_{i=0}^{n-2}\frac{n-1-i}{n-1}\cdot\frac{M(\varphi_{\frac{i}{n-1}})}{n^{|S(\Sigma)|}}. \qquad (2)$$

Пример. Для случая $n=5$ $I_0(\varphi\to\psi)=0.2$, $I_0(\varphi\vee\psi\vee\chi\vee\omega)=0.1416$ [6]. Однозначно определенным весом величины **??** является коэффициент $\frac{n-1-i}{n-1}$. Естественно, необходимо попытаться пошевелить взятые когнитивно найденные и обоснованные величины, указанные выше, и избавиться от жестких ограничений, накладываемых ранее в поиске подходящих коэффициентов.

Список литературы

1 *Ершов Ю.Л., Палютин Е.А.* Математическая логика. — М.: Наука, 1987.

2 *Викентьев А.А., Викентьев Р.А.* Расстояния и меры недостоверности на высказываниях n-значной логики//Вестн. Новосиб. гос. ун-та. Сер. матем., мех., информ. — 2011. — Т. 11. — С. 51–64.

3 *Викентьев А.А.* О возможных расстояниях и степенях недостоверности в многозначных высказываниях экспертов и приложении этих понятий в проблемах кластеризации и распознавания // Проблемы информатики - (Новосибирск: СО РАН) 2011. — Т. 3. — С. 33–45.

4 *Бериков В.Б.* Коллектив алгоритмов с весами в кластерном анализе разнородных данных // Вестн. Томск. гос. ун-та. — 2013. — № 2(23). — С. 22–31.

5 *Авилов М.С.* Программный комплекс вычислений расстояний, мер нетривиальности и кластеризации множеств высказываний в n-значной логике// Студент и научно-технический прогресс: материалы 52-й Междунар. науч. студ. конф.-Новосибирск, 2015.

6 *Викентьев А.А., Кабанова Е.С.* Расстояния между формулами пятизначной логики Лукасевича и мера недостоверности высказываний экспертов в кластеризации баз знаний // Вестн. Томск. гос. ун-та. — 2013. — № 2(23). — С. 121–129.

7 *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.

8 *Cooke R., Goossens L.* Delft expert judgment data base//Reliability Engineering and System Safety. — 2008. — No. 93(5). — P. 657–674.

9 *Baskerville R., Dulipovici A.* The theoretical foundations of knowledge management // Knowledge Management Research & Practice. — 2006. — Vol. 4. — P. 83-105.

10 *Загоруйко Н.Г.* Прикладные методы анализа данных и знаний. — Новосибирск: Изд-во Ин-та математики, 1999.

11 *Лбов Г.С.,Бериков В.Б.* Устойчивость решающих функций в задачах распознавания образов и анализа разнотипной информации. — Новосибирск: Изд-во Ин-та математики, 2005.

12 *Лбов Г.С., Старцева Н.Г.* Логические решающие функции и вопросы статистической устойчивости решений. — Новосибирск: Изд-во Ин-та математики, 1999.

13 *Викентьев А.А.* Мера опровержимости высказываний экспертов, расстояния в многозначной логике и процессы адаптации // Knowledge-Dialogue-Solution: Proceedings of the XIII-th International Conference (June, 18-24, 2007, Varna (Bulgaria)). — 2008. — P. 179–188.

14 *Карпенко А.С.* Логики Лукасевича и простые числа. — М.: Наука, 2000.

А.А.Викентьев, М.С.Авилов

Логикалық білім қорынан автоматтандырылған формулалар кластеризациясында көпмәнді логика формулалары үшін нөльдік емес өлшемдер мен жаңа толық метрикалар. I

Мақалада келесі есептер қарастырылған: параметрлерге шектеу алынып, кез келген n-мәнді логиканың жағдайына бұрын енгізілген нөльдік емес өлшемдер мен аралықтар жалпыланған; жаңа аралықтарды қолдана отырып, n-мәнді логикадағы формулалар жиындарын кластеризациялау алгоритмдері құрылған; әр түрлі білім қорынан формулалардың ішкі жиындарына кластеризация жүргізіліп, кластеризацияның әр түрлі алгоритмдерінің нәтижелерін салыстыру тәсілдері ұсынылған.

A.A.Vikent'ev , M.S. Avilov

New comprehensive metrics and nontrivial steps to formulas multi-valued logic in the automatic clustering of logical formulas Knowledge Base. I

This article describes the following tasks: summarized previously entered distances and non-triviality of the measures on the n-digit case of any logic, removing restrictions on the parameters; with the new range, developed algorithms for clustering sets of formulas in the n-valued logic; conducted clustering subsets of formulas from different knowledge bases and provides methods of comparison adapted results of various clustering algorithms.

References

1 Ershov Yu.L., Palyutin E.A. Mathematical logic, Moscow: Nauka, 1987.

2 Vikent'ev A.A., Vikent'ev R.A. *Bull. of Novosibirsk State University*, Ser. Mathematics, mechanics, computer science, 2011, 11, p. 51–64.

3 Vikent'ev A.A. *Informatics Problems*, 2011, 3, p. 33–45.

4 Berikov V.B. *Bull. of Tomsk State University*, 2013, 2(23), p. 22–31.

5 Avilov M.S. *Student and scientific - technical progress*: proceedings of the 52nd International Scientific Student Conference, Novosibirsk, 2015.

6 Vikent'ev A.A., Kabanova E.S. *Bull. of Tomsk State University*, 2013, 2(23), p. 121–129.

7 Vikent'ev A.A., Lbov G.S. *Pattern Recognition And Image Analysis*, 1997, 7, 2, p. 175–183.

8 Cooke R., Goossens L. *Reliability Engineering and System Safety*, 2008, 93(5), p. 657–674.

9 Baskerville R., Dulipovici A. *Knowledge Management Research Practice*, 2006, 4, p. 83–105.

10 Zagoruiko N.G. *Applied methods of data analysis and knowledge*, Novosibirsk: Publ. of the Institute of Mathematics, 1999.

11 Lbov G.S., Berikov V.B. *Stability crucial functions in problems of pattern recognition and analysis of heterogeneous information*, Novosibirsk: Publ. of the Institute of Mathematics, 2005.

12 Lbov G.S., Startseva N.G. *The logical decision functions and statistical stability issues making*, Novosibirsk: Publ. of the Institute of Mathematics, 1999.

13 Vikent'ev A.A. *Knowledge-Dialogue-Solution*: proceedings of the XIII-th International Conference (June, 18-24, 2007, Varna (Bulgaria)), 2008, p. 179–188.

14 Karpenko A.S. *Logic Lukasiewicz and prime numbers*, Moscow: Nauka, 2000.