УДК 510.67-519.24

А.А.Викентьев[1,2], В.В.Фефелова[2]

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

Введение полных расстояний и мер недостоверности для формул логик Лукасевича для автоматической кластеризации множеств логических высказываний из базы знаний

В статье рассмотрены логические высказывания экспертов, которые можно представить в виде логических формул n-значной логики Лукасевича. Используя теоретико-модельный подход, введены полные расстояния между формулами и меры недостоверности формул. Изучены свойства введенных величин. Также показаны вычисления и применения величин для кластеризации групп формул n-значной логики Лукасевича из базы знаний (экспертной системы).

Ключевые слова: многозначная логика, логика Лукасевича, расстояние между формулами, мера недостоверности, кластеризация, иерархический алгоритм, теория моделей.

1 Введение

На сегодняшний день задача анализа многозначной экспертной информации является актуальной [1–6]. В данной работе рассматриваются логические высказывания (экспертов), представленные в виде логических формул n-значной логики Лукасевича из базы знаний. Понятно, что различные высказывания несут в себе разное количество информации. Тем самым возникает вопрос о сравнении (экспертных) высказываний по информативности и, как следствие, их ранжировании и выделении близких высказываний. Для этого необходимо ввести расстояния между высказываниями, полно учитывающие всевозможные логические возможности, а также меру информативности, что и сделано. Полученные величины можно использовать в кластеризации множеств высказываний. Работа выполнена при поддержке грантов РФФИ, проекты 14–07–00851а, 14–07–00249 а.

2 Теоретические проблемы

2.1 Постановка задачи. Основной задачей данной работы являлось введение полных расстояний и мер недостоверности для формул n-значной логики Лукасевича, уточняющих и расширяющих [7–10], причем так, чтобы выполнялось как можно больше свойств, характерных для данных величин в известных случаях. Необходимо проиллюстрировать применение полученных величин в кластеризации.

2.2 Расстояния между формулами $\mathbf{Ł_n}$. Теоретико-модельные понятия, используемые в данной работе, определены в [3, 8, 10].

Обозначим через $M\left(\frac{k}{n-1}\right) = \left|Mod_{S(\Sigma)}(\varphi)_{\frac{k}{n-1}}\right|$ количество моделей, на которых формула φ принимает значение $\frac{k}{n-1}$, а через $M\left(\frac{k}{n-1}, \frac{l}{n-1}\right) = \left|Mod_{S(\Sigma)}((\varphi)_{\frac{k}{n-1}} \& (\psi)_{\frac{l}{n-1}})\right|$ обозначим количество моделей, на которых формула φ принимает значение $\frac{k}{n-1}$, а формула — $\psi - \frac{l}{n-1}$.

Для определения расстояния мы учитываем разницу между значениями двух формул на каждой модели. Объединим модели с одинаковыми модулями разности между значениями формул φ и ψ и возьмем их с некоторым весом, учитывающим близость логических значений.

$$\tilde{\rho}(\varphi,\psi) = \alpha_0\left(M(0,0) + M\left(\frac{1}{n-1},\frac{1}{n-1}\right) + \dots + M\left(\frac{n-2}{n-1},\frac{n-2}{n-1}\right) + M(1,1)\right) + \alpha_1(M\left(0,\frac{1}{n-1}\right) +$$
$$+M\left(\frac{1}{n-1},0\right) + M\left(\frac{1}{n-1},\frac{2}{n-1}\right) + M\left(\frac{2}{n-1},\frac{1}{n-1}\right) + \dots + M\left(\frac{n-2}{n-1},1\right) + M\left(1,\frac{n-2}{n-1}\right)) +$$
$$+ \dots + \alpha_{n-2}\left(M\left(0,\tfrac{n-2}{n-1}\right) + M\left(\tfrac{n-2}{n-1},0\right) + M\left(1,\tfrac{1}{n-1}\right) + M\left(\tfrac{1}{n-1},1\right)\right) + \alpha_{n-1}(M(0,1) + M(1,0)).$$

Очевидно, что модели, на которых значения формул совпадают, рассматривать не нужно, поэтому полагаем $\alpha_0 = 0$. Модели, на которых формула φ принимает значение 0, а формула ψ принимает значение 1 (и наоборот), берем с весом $\alpha_{n-1} = 1$. Полагаем, что чем меньше модуль разности между значениями формул φ и ψ на модели, тем они ближе на данной модели, и потому их нужно брать с меньшим весом, поэтому $\alpha_i \le \alpha_{i+1}$ для $i = 0, \dots, n-1$. Остается нормировать величину $\tilde{\rho}(\varphi, \psi)$.

2.2.1 Определение. Расстоянием (полным) между формулами φ и ψ n-значной логики $Ł_n$ при $S(\varphi) \cup S(\psi) \subseteq S(\Sigma)$ на множестве $P(S(\Sigma))$ назовем величину

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

где α_i удовлетворяет условиям

$$\begin{cases} 0 = \alpha_0 \le \alpha_i \le \alpha_{n-1} = 1; \\ \alpha_i \le \alpha_{i+1}, i = 0, \dots, n-1; \\ \alpha_i + \alpha_{n-1-i} = 1, \forall i = 0, \dots, \dfrac{n-1}{2}. \end{cases}$$

2.2.2 Теорема. Расстояние между формулами $Ł_n$, определенное равенством (1), для любых $\varphi, \psi, \tau \in \Sigma$ удовлетворяет следующим свойствам:

1) $0 \le \rho(\varphi, \psi) \le 1$;
2) $\rho(\varphi, \psi) = 0 \Leftrightarrow \varphi \equiv \psi$;
3) $\rho(\varphi, \psi) = \rho(\psi, \varphi)$;
4) $\rho(\varphi, \psi) \le \rho(\varphi, \tau) + \rho(\tau, \psi)$;
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) $\rho\big((\varphi \wedge \psi), (\varphi \vee \psi)\big) = \rho(\varphi, \psi)$.

Доказательство. Для удобства доказательства перепишем формулу для нахождения расстояния в следующем виде:

$$\rho(\varphi, \psi) = \frac{1}{n^{|S(\Sigma)|}} (0 \cdot A_0 + \alpha_1 \cdot A_1 + \alpha_2 \cdot A_2 + \dots + \alpha_{n-2} \cdot A_{n-2} + 1 \cdot A_{n-1}).$$

1) В формуле для вычисления расстояния участвуют все модели с коэффициентами от 0 до 1. $\rho(\varphi, \psi) = 0$, если все модели лежат в A_0, то есть, когда $\varphi \equiv \psi$; $\rho(\varphi, \psi) = 1$, если все модели содержатся в A_{n-1}, то есть, когда $\varphi \equiv \neg\psi$ и φ и ψ, принимают на моделях только значения 0 и 1. Значит, $0 \le \rho(\varphi, \psi) \le 1$.
2) Необходимость. Следует из доказательства свойства (1).
Достаточность. Следует из определения эквивалентности [3, 8].
3) Следует из того, что пары $M\left(\frac{k}{n-1}, \frac{l}{n-1}\right) \ne M\left(\frac{l}{n-1}, \frac{k}{n-1}\right)$ умножаются на один и тот же коэффициент.
4) $\rho(\varphi, \psi) = \frac{1}{n^{|S(\Sigma)|}}(\alpha_0 \left(M(0,0) + M\left(\frac{1}{n-1}, \frac{1}{n-1}\right) + \dots + M\left(\frac{n-2}{n-1}, \frac{n-2}{n-1}\right) + M(1,1)\right) + \alpha_1(M\left(0, \frac{1}{n-1}\right) +$

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

$$+ \dots + \alpha_{n-2}\left(M\left(0, \frac{n-2}{n-1}\right) + M\left(\frac{n-2}{n-1}, 0\right) + M\left(1, \frac{1}{n-1}\right) + M\left(\frac{1}{n-1}, 1\right)\right) + \alpha_{n-1}(M(0,1) + M(1,0))) = \frac{1}{n^{|S(\Sigma)|}}(0 +$$

$$+\alpha_1\left(\left(\varphi_0 \Delta \psi_{\frac{1}{n-1}}\right) + \left(\varphi_{\frac{1}{n-1}} \Delta \psi_{\frac{2}{n-1}}\right) + \dots + \left(\varphi_{\frac{n-2}{n-1}} \Delta \psi_1\right)\right) + \dots + \alpha_{n-2}(\left(\varphi_0 \Delta \psi_{\frac{n-2}{n-1}}\right) + \left(\varphi_1 \Delta \psi_{\frac{1}{n-1}}\right) + ((\varphi_0 \Delta \psi_1))).$$

Каждое слагаемое является нормированной симметрической разностью, взятой с некоторым весом. В работе [8] доказано, что нормированная симметрическая разность, взятая с некоторым весом, является расстоянием. Так как сумма расстояний тоже является расстоянием, то неравенство треугольника выполнятся.

5) Следует из определения эквивалентности двух формул.
6) Следует из соотношения:
$Mod_{S(\Sigma)}\left((\neg\varphi)_{\frac{k}{n-1}} \& (\neg\psi)_{\frac{l}{n-1}}\right) = Mod_{S(\Sigma)}\left((\varphi)_{\frac{(n-1)-k}{n-1}} \& (\psi)_{\frac{(n-1)-l}{n-1}}\right)$, $k, l = 0, \dots, 1$ [9].
7) Следует из соотношений работы [9]

$$Mod_{S(\Sigma)}(\varphi \& \psi)_{\frac{k}{n-1}} = \cup_{l=k}^{n-1}((Mod_{S(\Sigma)}(\varphi)_{\frac{l}{n-1}} \cap Mod_{S(\Sigma)}(\psi)_{\frac{k}{n-1}}) \cup (Mod_{S(\Sigma)}(\varphi)_{\frac{k}{n-1}} \cap Mod_{S(\Sigma)}(\psi)_{\frac{l}{n-1}}));$$

$$Mod_{S(\Sigma)}(\varphi \vee \psi)_{\frac{k}{n-1}} = \cup_{l=k}^{n-1}((Mod_{S(\Sigma)}(\varphi)_{\frac{l}{n-1}} \cup Mod_{S(\Sigma)}(\psi)_{\frac{k}{n-1}}) \cup (Mod_{S(\Sigma)}(\varphi)_{\frac{k}{n-1}} \cup Mod_{S(\Sigma)}(\psi)_{\frac{l}{n-1}})).$$

Теорема доказана.

2.2.3. Замечание. Свойства 2)–4) — это аналоги свойства метрики. Таким образом, мы получили метрическое пространство на классах эквивалентности высказываний.

2.2.4 Замечание. Расстояние, заданное формулой (1), — это расстояние для случая, когда все значения переменных заранее не известны. Допустим, что теперь нам известны истинностные значения некоторых переменных. Пусть переменные $x_1, \dots, x_r,\ x_i \in S(\varphi) \cup S(\psi),\ i = 1, \dots, r,\ r = |S(\varphi) \cup S(\psi)|$ соответственно принимают $d_1, \dots, d_r, d_i \le n$ истинностные значения. Тогда формула для нахождения расстояния между формулами φ и ψ имеет вид

$$\hat{\rho}(\varphi, \psi) = \frac{1}{d_1 \cdot \ldots \cdot d_r} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \alpha_{|k-l|} M\left(\frac{k}{n-1}, \frac{l}{n-1}\right), \quad (2)$$

где α_i удовлетворяет условиям

$$\begin{cases} 0 = \alpha_0 \le \alpha_i \le \alpha_{n-1} = 1; \\ \alpha_i \le \alpha_{i+1}, i = 0, \dots, n-1; \\ \alpha_i + \alpha_{n-1-i} = 1, \forall i = 0, \dots, \dfrac{n-1}{2}. \end{cases}$$

Теорема 2.2.2 верна для $\hat{\rho}(\varphi, \psi)$.

2.2.5 Пример. Рассмотрим формулы $\varphi = (x \longrightarrow y) \vee z$, $\psi = (x \wedge y) \longrightarrow z$. Для начала рассмотрим случай $n = 3$. Возьмем $\alpha_0 = 0, \alpha_1 = \frac{1}{3}, \alpha_2 = 1$. Тогда $\rho(\varphi, \psi) = 0{,}16$.

Пусть теперь переменные, входящие в эти формулы, принимают следующие значения:

$$x \in \left\{\frac{1}{2}, 1\right\};\ y \in \{1\};\ z \in \left\{0, \frac{1}{2}, 1\right\}.\ \text{Тогда } \hat{\rho}(\varphi, \psi) = 0{,}28.$$

Теперь рассмотрим случай, когда $n = 7$. Возьмем $\alpha_0 = 0, \alpha_1 = \frac{1}{7}, \alpha_2 = \frac{1}{5}, \alpha_3 = \frac{1}{2}, \alpha_4 = \frac{4}{5};$ $\alpha_5 = \frac{6}{7}, \alpha_6 = 1$. Тогда $\rho(\varphi, \psi) = 0{,}163$.

Пусть теперь переменные, входящие в эти формулы, принимают следующие значения:

$$x \in \left\{\frac{3}{6}\right\}, y \in \left\{\frac{1}{6}, \frac{2}{6}\right\}, z \in \left\{0, \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, 1\right\}.\ \text{Тогда } \hat{\rho}(\varphi, \psi) = 0{,}129.$$

2.3 Мера недостоверности. 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 достоверно. Поэтому вместо термина «мера информативности» будем использовать термин «мера недостоверности».

В работе [11] введена формула для меры недостоверности в случае $n = 6$. Обобщим эту формулу для случая конечнозначной логики $Ł_n$. Так как в $Ł_n$ истинностных значений, отличных от 1, (n–2), то нужно учитывать модели, на которых формула принимает значение $\frac{k}{n-1}$, $k = 0, \dots, n-2$. Также нужно учесть, насколько близко значение формулы к 1. Понятно, 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.

2.3.1 Определение. Мера недостоверности $I(\varphi)$ для формул n-значной логики Лукасевича при $S(\varphi) \subseteq S(\Sigma)$ на множестве $P(S(\Sigma))$ задается следующим образом:

$$I(\varphi) = \frac{1}{n^{|S(\Sigma)|}} \sum_{k=0}^{n-1} \alpha_{|(n-1)-k|} M\left(\frac{k}{n-1}\right), \quad (3)$$

где α_i удовлетворяет условиям:

$$\begin{cases} 0 = \alpha_0 \le \alpha_i \le \alpha_{n-1} = 1; \\ \alpha_i \le \alpha_{i+1}, i = 0, \dots, n-1; \\ \alpha_i + \alpha_{n-1-i} = 1, \forall i = 0, \dots, \dfrac{n-1}{2}. \end{cases}$$

2.3.2 Теорема. Мера недостоверности, определенная равенством (3), для любых формул $\varphi, \psi, \tau \in \Sigma$ удовлетворяет следующим свойствам:

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

Доказательство:

1) Очевидно, так как $I(\varphi) = \rho(\varphi, 1)$.
2) $I(\varphi) + I(\neg\varphi) = \frac{1}{n^{|S(\Sigma)|}} \Big(\alpha_{n-1} M(0) + \alpha_{n-2} M\left(\frac{1}{n-1}\right) + \alpha_{n-3} M\left(\frac{2}{n-1}\right) + \cdots + \alpha_2 M\left(\frac{n-3}{n-1}\right) + \alpha_1 M\left(\frac{n-2}{n-1}\right) +$ $+ \alpha_0 M(1) + \alpha_{n-1} M(1) + \alpha_{n-2} M\left(\frac{n-2}{n-1}\right) + \alpha_{n-3} M\left(\frac{n-3}{n-1}\right) + \cdots + \alpha_2 M\left(\frac{2}{n-1}\right) + \alpha_1 M\left(\frac{1}{n-1}\right) +$

$$+\alpha_0 M(0)) = \frac{1}{n^{|S(\Sigma)|}}\left(M(0) + M\left(\frac{1}{n-1}\right) + M\left(\frac{2}{n-1}\right) + \cdots + M\left(\frac{n-2}{n-1}\right) + M(1)\right) = \frac{1}{n^{|S(\Sigma)|}} n^{|S(\Sigma)|} = 1.$$

Доказательства свойств 3)–5) аналогичны доказательствам свойств 3)–5) меры недостоверности работ [9, 13].

2.3.3 Пример. Рассмотрим формулы $\varphi = (x \longrightarrow y) \vee z$, $\psi = (x \wedge y) \longrightarrow z$. Рассмотрим случай $n = 3$. Возьмем $\alpha_0 = 0, \alpha_1 = \frac{1}{3}, \alpha_2 = 1$. Тогда $I(\varphi) = 0{,}099, I(\psi) = 0{,}086$.

Рассмотрим случай, когда n=7. Возьмем $\alpha_0 = 0, \alpha_1 = \frac{1}{7}, \alpha_2 = \frac{1}{5}, \alpha_3 = \frac{1}{2}, \alpha_4 = \frac{4}{5}, \alpha_5 = \frac{6}{7}, \alpha_6 = 1$. Тогда $I(\varphi) = 0{,}117, I(\psi) = 0{,}087$.

2.4 Кластеризация множеств высказываний. Кластеризация — это разбиение исходного множества объектов на подмножества (кластеры), при котором каждый объект может быть отнесен к одному или нескольким заранее неизвестным классам. Внутри каждого кластера должны оказаться схожие объекты, а объекты разных кластеров должны как можно больше отличаться [12].

Для множеств высказываний известны только расстояния между формулами и меры недостоверности. В данной работе используется иерархический алгоритм кластеризации, для реализации которого достаточно знать попарные расстояния между объектами.

2.4.1. Иерархический алгоритм кластеризации для формул $\mathbf{Ł_n}$. Для начала задаем конечное множество формул $Ł_n$.

1) Строим матрицу расстояний для набора формул.
2) Ищем наименьшее расстояние и объединяем эти 2 формулы в один кластер.
3) Объединяем кластеры по методу ближайшего соседа. Матрица расстояний пересчитывается по правилу: $\rho(\varphi_k, \varphi_{ij}) = \min\{\rho(\varphi_k, \varphi_i), \rho(\varphi_k, \varphi_j)\}$.
4) Повторяем пункты 2 и 3, пока не выполнится критерий остановки.

Критерий остановки: работа алгоритма останавливается, когда максимальная разница между мерами недостоверности элементов одного кластера (обозначается d) достигает значения, заданного перед началом работы алгоритма.

2.4.2. Пример. Рассмотрим множество формул:

1) $\varphi_1 = x \longrightarrow y$;
2) $\varphi_2(x \longrightarrow y) \vee (z \wedge w)$;
3) $\varphi_3 = \neg(x \longrightarrow y)$;
4) $\varphi_4 = x \vee (y \vee z)$;
5) $\varphi_5 = ((x \wedge y) \wedge z) \longrightarrow w$;
6) $\varphi_6 = (\neg x \wedge y) \vee z$;
7) $\varphi_7 = (x \vee y) \wedge z$;
8) $\varphi_8 = \neg(x \wedge (y \wedge z))$.

В первом случае возьмем $n = 4$, $\alpha_0 = 0, \alpha_1 = \frac{1}{3}, \alpha_2 = \frac{2}{3}, \alpha_3 = 1$ (табл. 1, 2).

Таблица 1

Матрица расстояний для первого случая

	1	2	3	4	5	6	7	8
1	0	0,035	0,792	0,302	0,240	0,344	0,521	0,302
2		0	0,770	0,267	0,204	0,309	0,493	0,275
3			0	0,604	0,732	0,542	0,344	0,604
4				0	0,224	0,208	0,417	0,292
5					0	0,398	0,594	0,128
6						0	0,208	0,396
7							0	0,563
8								0

Таблица 2

Матрица мер недостоверности для первого случая

i	1	2	3	4	5	6	7	8
$I(\varphi_i)$	0,208	0,173	0,792	0,188	0,060	0,396	0,604	0,188

Во втором случае возьмем $n = 4$, $\alpha_0 = 0, \alpha_1 = \frac{1}{45}, \alpha_2 = \frac{44}{45}, \alpha_3 = 1$ (табл. 3, 4).

Таблица 3

Матрица расстояний для второго случая

	1	2	3	4	5	6	7	8
1	0	0,024	0,694	0,239	0,206	0,285	0,501	0,239
2		0	0,689	0,200	0,159	0,250	0,478	0,212
3			0	0,624	0,752	0,547	0,300	0,624
4				0	0,157	0,174	0,378	0,224
5					0	0,369	0,597	0,067
6						0	0,160	0,333
7							0	0,504
8								0

Таблица 4

Матрица мер недостоверности для второго случая

i	1	2	3	4	5	6	7	8
$I(\varphi_i)$	0,189	0,143	0,811	0,129	0,037	0,376	0,624	0,129

В третьем случае возьмем $n = 7$, $\alpha_0 = 0, \alpha_1 = \frac{1}{6}, \alpha_2 = \frac{2}{6}, \alpha_3 = \frac{3}{6}, \alpha_4 = \frac{4}{6}, \alpha_5 = \frac{5}{6}, \alpha_6 = 1$ (табл. 5, 6).

Таблица 5

Матрица расстояний для третьего случая

	1	2	3	4	5	6	7	8
1	0	0,026	0,755	0,284	0,216	0,337	0,498	0,284
2		0	0,743	0,258	0,190	0,310	0,478	0,265
3			0	0,595	0,753	0,516	0,337	0,595
4				0	0,234	0,190	0,381	0,280
5					0	0,394	0,573	0,158
6						0	0,190	0,356
7							0	0,501
8								0

Таблица 6

Матрица мер недостоверности для третьего случая

i	1	2	3	4	5	6	7	8
$I(\varphi_i)$	0,190	0,164	0,810	0,214	0,056	0,405	0,595	0,214

В четвертом случае возьмем $n = 7$, $\alpha_0 = 0, \alpha_1 = \frac{1}{32}, \alpha_2 = \frac{1}{5}, \alpha_3 = \frac{3}{6}, \alpha_4 = \frac{4}{5}, \alpha_5 = \frac{31}{32}, \alpha_6 = 1$ (табл. 7, 8).

Таблица 7

Матрица расстояний для четвертого случая

	1	2	3	4	5	6	7	8
1	0	0,019	0,755	0,238	0,190	0,293	0,492	0,238
2		0	0,740	0,208	0,156	0,268	0,473	0,214
3			0	0,612	0,771	0,517	0,301	0,612
4				0	0,181	0,168	0,348	0,223
5					0	0,373	0,583	0,104
6						0	0,153	0,324
7							0	0,490
8								0

Таблица 8

Матрица мер недостоверности для четвертого случая

i	1	2	3	4	5	6	7	8
$I(\varphi_i)$	0,174	0,141	0,826	0,165	0,039	0,388	0,612	0,165

Результаты кластеризации иерархическим алгоритмом запишем в таблицы 9 и 10.

Таблица 9

Результаты кластеризации иерархическим алгоритмом

Номер итерации	Кластеры (случай 1)	d	Кластеры (случай 2)	d
1	$(\varphi_1,\varphi_2),\varphi_3,\varphi_4,\varphi_5,\varphi_6,\varphi_7,\varphi_8$	0.035	$(\varphi_1,\varphi_2),\varphi_3,\varphi_4,\varphi_5,\varphi_6,\varphi_7,\varphi_8$	0.046
2	$(\varphi_1,\varphi_2),(\varphi_5,\varphi_8),\varphi_3,\varphi_4,\varphi_6,\varphi_7$	0.128	$(\varphi_1,\varphi_2),(\varphi_5,\varphi_8),\varphi_3,\varphi_4,\varphi_6,\varphi_7$	0.092
3	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8),\varphi_3,\varphi_4,\varphi_6,\varphi_7$	0.148	$(\varphi_1,\varphi_2),(\varphi_5,\varphi_8,\varphi_4),\varphi_3,\varphi_6,\varphi_7$	0.092
4	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8),(\varphi_6,\varphi_7),\varphi_3,\varphi_4$	0.208	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8,\varphi_4),\varphi_3,\varphi_6,\varphi_7$	0.152
5	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8),(\varphi_6,\varphi_7,\varphi_4),\varphi_3$	0.416	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8,\varphi_4),(\varphi_6,\varphi_7),\varphi_3$	0.248
6	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8,\varphi_6,\varphi_7,\varphi_4),\varphi_3$	0.544	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8,\varphi_4,\varphi_6,\varphi_7),\varphi_3$	0.587
7	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8,\varphi_6,\varphi_7,\varphi_4,\varphi_3)$	0.732	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8,\varphi_4,\varphi_6,\varphi_7,\varphi_3)$	0.774

Таблица 10

Результаты кластеризации иерархическим алгоритмом

Номер итерации	Кластеры (случай 3)	d	Кластеры (случай 4)	d
1	$(\varphi_1,\varphi_2),\varphi_3,\varphi_4,\varphi_5,\varphi_6,\varphi_7,\varphi_8$	0.026	$(\varphi_1,\varphi_2),\varphi_3,\varphi_4,\varphi_5,\varphi_6,\varphi_7,\varphi_8$	0.033
2	$(\varphi_1,\varphi_2),(\varphi_5,\varphi_8),\varphi_3,\varphi_4,\varphi_6,\varphi_7$	0.158	$(\varphi_1,\varphi_2),(\varphi_5,\varphi_8),\varphi_3,\varphi_4,\varphi_6,\varphi_7$	0.126
3	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8),\varphi_3,\varphi_4,\varphi_6,\varphi_7$	0.158	$(\varphi_1,\varphi_2),(\varphi_5,\varphi_8),(\varphi_6,\varphi_7),\varphi_3,\varphi_4$	0.224
4	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8),(\varphi_4,\varphi_6),\varphi_3,\varphi_7$	0.191	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8),(\varphi_6,\varphi_7),\varphi_3,\varphi_4$	0.224
5	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8),(\varphi_4,\varphi_6,\varphi_7),\varphi_3$	0.381	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8),(\varphi_6,\varphi_7,\varphi_4),\varphi_3$	0.447
6	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8,\varphi_4,\varphi_6,\varphi_7),\varphi_3$	0.539	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8,\varphi_6,\varphi_7,\varphi_4),\varphi_3$	0.573
7	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8,\varphi_4,\varphi_6,\varphi_7,\varphi_3)$	0.754	$(\varphi_1,\varphi_2,\varphi_5,\varphi_8,\varphi_6,\varphi_7,\varphi_4,\varphi_3)$	0.787

Возьмем $d = 0.15$. Тогда получим следующие кластеры:

Случай 1: $(\varphi_1,\varphi_2,\varphi_5,\varphi_8),\varphi_3,\varphi_4,\varphi_6,\varphi_7$.

Случай 2:$(\varphi_1,\varphi_2),(\varphi_5,\varphi_8),\varphi_3,\varphi_4,\varphi_6,\varphi_7$.

Случай 3: $(\varphi_1,\varphi_2),\varphi_3,\varphi_4,\varphi_5,\varphi_6,\varphi_7,\varphi_8$.

Случай 4: $(\varphi_1,\varphi_2),(\varphi_5,\varphi_8),\varphi_3,\varphi_4,\varphi_6,\varphi_7$.

Как видно из примера, для разных n и разных весов мы получаем различные кластеры. Следовательно, мы можем выбрать наилучшую кластеризацию.

Под наилучшей кластеризацией будем понимать следующее: элементы одного кластера должны быть как можно ближе друг к другу, а расстояние между кластерами должно быть наибольшим [13]. Пусть s_1 — это сумма диаметров кластеров, а s_2 — сумма расстояний между кластерами. Обозначим $s = \frac{s_1}{s_2}$. Чем меньше s, тем кластеризация является наиболее наилучшей.

Посчитаем s для нашего примера. В первом случае $s = 0{,}076$, во втором $s = 0{,}016$, в третьем $s = 0{,}003$, в четвертом $s = 0{,}022$. Следовательно, при $d = 0{,}15$ кластеризация $(\varphi_1, \varphi_2), \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8$, полученная в третьем случае, является наилучшей.

3 Заключение

Введены новые полные расстояния и меры недостоверности для формул n-значной логики Лукасевича, также доказаны свойства полученных величин. Сложность алгоритма вычисления расстояний между формулами — экспоненциальная. Адаптирован иерархический алгоритм кластеризации, также показан выбор наилучшей кластеризации.

Полученные величины можно использовать при анализе баз знаний, их кластеризации, создании экспертных систем, а также при построении логических решающих функций в распознавании.

Список литературы

1 *Ершов Ю.Л., Палютин Е.А.* Математическая логика — 2-е изд. — М.: Наука, 1987. — 336 с.

2 *Карпенко А.С.* Логики Лукасевича и простые числа. — М.: Наука, 2000. — 319 с.

3 *Лбов Г.С., Старцева Н.Г.* Логические решающие функции и вопросы статистической устойчивости решений. — Новосибирск: Изд-во Ин-та математики, 1999. — 212 с.

4 *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.

5 *Загоруйко Н.Г.* Прикладные методы анализа данных и знаний. — Новосибирск: Изд-во Ин-та математики, 1999. — 270 с.

6 *Викентьев А.А.* Мера опровержимости высказываний экспертов, расстояния в многозначной логике и процессы адаптации // XVI Internat. Conf. «Knowledge — Dialogue — Solution» KDS 2008. Varna, Bulgaria, 2008. — С. 179–188.

7 *Кабанова Е.С.* Расстояния между формулами пятизначной логики Лукасевича и мера недостоверности высказываний экспертов // Материалы 50-й Юбилейной МНСК «Студент и научно-технический прогресс». — Новосибирск: Изд-во НГУ, 2012.

8 *Викентьев А.А., Викентьев Р.А.* Расстояния и меры недостоверности на высказываниях n-значной логики // Вестн. НГУ. Сер. математика, механика, информатика. — 2011. — Т. 11. — Вып. 2. — С. 51–64.

9 *Викентьев А.А.* О возможных расстояниях и степенях недостоверности в многозначных высказываниях экспертов и приложение этих понятий в проблемах кластеризации и распознавания // Проблемы информатики. — 2011. — № 3 (11). — С. 33–45.

10 *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.

11 *Фефелова В.В.* Расстояния между формулами шестизначной логики Лукасевича и мера недостоверности высказываний в кластеризации множеств высказываний // Выпускная квалификационная работа бакалавра. — Новосибирск: Изд-во НГУ, 2014. — 40 с.

12 *Миркин Б.Г.* Методы кластер-анализа для поддержки принятия решений: Обзор. — М.: Изд. Дом ВШЭ, 2011. — 88 с.

13 *Лбов Г.С., Бериков В.Б.* Устойчивость решающих функций в задачах распознавания образов и анализа разнотипной информации. — Новосибирск: Изд-во Ин-та мат. СО РАН, 2005. — 200 с.

А.А.Викентьев, В.В.Фефелова

Білім қорынан автоматты кластеризацияланған логикалық мәлімдеме жиындары үшін Лукасевичтің логикалық формулалары үшін толық арақашықтықты және сенімсіз өлшемді енгізу

Мақалада Лукасевичтің n-мәнді логикасының логикалық формуласы арқылы келтіруге болатын сарапшылардың логикалық мәлімдемесі қарастырылды. Теориялық-модельді әдісті қолдана отырып, сенімсіз формулалардың өлшемі мен формулалары арасында толық арақашықтықтар енгізілді. Ол өлшемдердің қасиеттері зерттелді. Сонымен қатар білім қорынан Лукасевичтің n-мәнді логика формулаларының группасын кластеризациялау үшін өлшемдерді есептеу мен қолдану көрсетілді.

A.A.Vikent’ev, V.V.Fefelova

The introduction of the full distance and measures of uncertainty for formulas Lukasiewicz logics for automatic clustering of sets of logical statements from the knowledge base

This article discusses logical statements of experts who can be represented by logical formulas of *n*-valued logic of Lukasiewicz. Using a theoretical model approach, enter the full distance between the formulas and measures of uncertainty formulas. We study the properties of these variables. Also showing calculation and application variables for clustering groups of formulas of *n*-valued logic of Lukasiewicz Knowledge Base (expert systems).

References

1 Ershov Yu.L., Palyutin E.A. *Mathematical logic*, 2nd ed., Moscow: Nauka, 1987, 336 p.

2 Karpenko A.S. *Logic Lukasiewicz and prime numbers*, Moscow: Nauka, 2000, 319 p.

3 Lbov G.S., Startseva N.G. *The logical decision functions and the issues of the statistical stability of the solutions*, Novosibirsk: Publ. house of the Institute of Mathematics, 1999, 212 p.

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

5 Zagoruiko N.G. *Applied methods of data analysis and knowledge*, Novosibirsk: Publ. house of the Institute of Mathematics, 1999, 270 p.

6 Vikent'ev A.A. *XVI International Conference «Knowledge — Dialogue — Solution» KDS*, 2008, Varna, Bulgaria, 2008, p. 179–188.

7 Kabanova E.S. *Proceedings of the 50th Anniversary ISSC «Student and technological progress»*, Novosibirsk: NSU publ., 2012.

8 Vikent'ev A.A., Vikent'ev R.A. *Bull. NSU, Ser. mathematics, mechanics, computer science*, Novosibirsk: Publ. house of the Novosibirsk State University, 2011, 11, 2, p. 51–64.

9 Vikent'ev A.A. *Problems of Informatics*, Novosibirsk: SB RAS, 2011, 3 (11), p. 33–45.

10 Vikent'ev A.A. *Pattern Recognition and Image Analysis*, 2014, 24, 4, p. 489–501.

11 Fefelova V.V. *Final qualifying work of the bachelor*, Novosibirsk: NSU publ., 2014, 40.

12 Mirkin B.G. *Methods for cluster analysis to support decision making: a review*, Moscow: Publ. house. Home School of Economics, 2011, 88 p.

13 Lbov G.S., Berikov V.B. *Stability of decision functions in problems of pattern recognition and analysis reznotipnoy information*, Novosibirsk: Publ. house of IM SB RAS, 2005, 200 p.