

УДК 519.67, 510.64, 519.24

Е.С. Кабанова, А.А. Викентьев

РАССТОЯНИЕ МЕЖДУ ФОРМУЛАМИ ПЯТИЗНАЧНОЙ ЛОГИКИ ЛУКАСЕВИЧА И МЕРА НЕДОСТОВЕРНОСТИ ВЫСКАЗЫВАНИЙ ЭКСПЕРТОВ В КЛАСТЕРИЗАЦИИ БАЗ ЗНАНИЙ

Определяются расстояние между формулами пятизначной логики Лукасевича и мера недоверности экспертных высказываний, записанных в виде логических формул, а также доказываются свойства данных величин. Также показывается возможное применение расстояния и меры на примере кластеризации группы формул пятизначной логики Лукасевича.

Ключевые слова: *многозначная логика, логика Лукасевича, расстояние между формулами, мера недоверности, кластеризация, иерархический алгоритм, базы знаний, экспертные высказывания, теория моделей.*

На сегодняшний день актуальной является проблема анализа многозначной экспертной информации, представленной в виде высказываний экспертов, которые можно записать в виде логических формул исчисления высказываний. В данной работе экспертные высказывания представлены в виде формул пятизначной логики Лукасевича [1]. Ясно, что различные высказывания и соответствующие им формулы несут разное количество информации. Тем самым возникает вопрос о сравнении экспертных высказываний по информативности и, как следствие, их ранжировании. Ясно, что информативность всего высказывания должна зависеть от информативности элементарных компонент и степени различия содержащейся в них информации. Следовательно, необходимо ввести «расстояние» между логическими формулами, соответствующими высказываниям, для которого, по определению расстояния, должны выполняться аксиома тождества, аксиома симметрии и неравенство треугольника, а также меру недоверности высказывания. В [2, 3] определены расстояние и мера недоверности для случая классической двузначной логики [5]. Помимо основных свойств введённое там расстояние обладает свойствами, учитывающими семантику совпадения и различия информации.

Основной задачей работы являлось введение расстояния и меры недоверности для формул пятизначной логики Лукасевича L_5 , причём таким образом, чтобы выполнялось как можно больше свойств, характерных для данных величин, введённых и доказанных для формул классической логики, и чтобы учитывались все возможные истинностные значения L_5 .

1. Расстояние между формулами L_5

Теоретико-модельные понятия для случаев классической логики и L_5 , используемые в данной работе, определены в [2–4].

Для краткости будем пользоваться следующими обозначениями:

$$M(\varphi_k) = \left| \text{Mod}_{S(\Sigma)}(\varphi_k) \right|_4,$$

то есть количество моделей, где формула φ принимает значение $\frac{k}{4}$, а $\psi - \frac{l}{4}$.

$$M\left(\frac{k}{4}, \frac{l}{4}\right) = \left| \text{Mod}_{S(\Sigma)}\left(\left(\varphi\right)_{\frac{k}{4}} \& \left(\psi\right)_{\frac{l}{4}}\right) \right|.$$

Расстояние между двумя формулами двужанной логики определяется как мощность симметрической разности моделей этих формул. Это расстояние задаёт метрику на классах эквивалентных высказываний [2,3].

Определение 1. [2, с. 89; 3] Расстоянием между формулами φ и ψ двужанной логики при $S(\varphi) \cup S(\psi) \subseteq S(\Sigma)$ на множестве $P(S(\Sigma))$ называется величина.

$$\rho_{S(\Sigma)}(\varphi, \psi) = \frac{|\text{Mod}_{S(\Sigma)}((\neg\varphi \& \psi) \vee (\varphi \& \neg\psi))|}{2^{|S(\Sigma)|}}. \quad (1)$$

Преобразуем формулу (1) к следующему виду:

$$\begin{aligned} \rho_{S(\Sigma)}(\varphi, \psi) &= \frac{1}{2^{|S(\Sigma)|}} |\text{Mod}_{S(\Sigma)}((\neg\varphi \& \psi) \vee (\varphi \& \neg\psi))| = M(0,1) + M(1,0) = \\ &= 1 \cdot M(0,1) + 1 \cdot M(1,0) + 0 \cdot M(0,0) + 0 \cdot M(1,1) = \frac{1}{2^{|S(\Sigma)|}} \sum_{k=0,1} \sum_{l=0,1} |k-l| M(k,l). \end{aligned}$$

Таким образом, предположим, что для определения расстояния можно учитывать разность между истинностными значениями формул φ и ψ на каждой модели.

Замечание (доказательство в [2, с. 93; 4]): Для вычисления расстояния между формулами φ и ψ , такими, что $S(\varphi) \cup S(\psi) \subseteq S(\Sigma)$, достаточно рассматривать не всё $S(\Sigma)$, а только $S(\varphi) \cup S(\psi)$. Или, точнее, любое $S(\Sigma_0)$, такое, что $S(\varphi) \cup S(\psi) \subseteq S(\Sigma_0) \subseteq S(\Sigma)$.

Далее, нижний индекс $S(\Sigma)$ в записи $\rho_{S(\Sigma)}$ будем опускать.

Обобщим этот подход на пятизначный случай. Объединим модели с одинаковыми модулями разности между значениями φ и ψ . Естественно предположить, что чем меньше модуль разности между значениями φ и ψ , тем формулы более близки в данной модели. Следовательно, умножим количество моделей с одинаковыми модулями разности на коэффициент, учитывающий близость формул. В качестве таких коэффициентов возьмём пять истинностных значений для L_5 :

$$\begin{aligned} \tilde{\rho}(\varphi, \psi) &= 0 \cdot \left(M(0,0) + M\left(\frac{1}{4}, \frac{1}{4}\right) + M\left(\frac{1}{2}, \frac{1}{2}\right) + M\left(\frac{3}{4}, \frac{3}{4}\right) + M(1,1) \right) + \\ &+ \frac{1}{4} \left(M\left(0, \frac{1}{4}\right) + M\left(\frac{1}{4}, 0\right) + M\left(\frac{1}{4}, \frac{1}{2}\right) + M\left(\frac{1}{2}, \frac{1}{4}\right) + M\left(\frac{1}{2}, \frac{3}{4}\right) + M\left(\frac{3}{4}, \frac{1}{2}\right) + M\left(\frac{3}{4}, 1\right) + M\left(1, \frac{3}{4}\right) \right) + \\ &+ \frac{1}{2} \left(M\left(0, \frac{1}{2}\right) + M\left(\frac{1}{2}, 0\right) + M\left(\frac{1}{4}, \frac{3}{4}\right) + M\left(\frac{3}{4}, \frac{1}{4}\right) + M\left(\frac{1}{2}, 1\right) + M\left(1, \frac{1}{2}\right) \right) + \\ &+ \frac{3}{4} \left(M\left(0, \frac{3}{4}\right) + M\left(\frac{3}{4}, 0\right) + M\left(\frac{1}{4}, 1\right) + M\left(1, \frac{1}{4}\right) \right) + \\ &+ 1(M(0,1) + M(1,0)) = \sum_{k=0}^4 \sum_{l=0}^4 \frac{|k-l|}{4} M\left(\frac{k}{4}, \frac{l}{4}\right). \end{aligned} \quad (2)$$

Остаётся только нормировать величину $\tilde{\rho}$.

Определение 2. Расстоянием между формулами φ и ψ пятизначной логики L_5 при $S(\varphi) \cup S(\psi) \subseteq S(\Sigma)$ на множестве $P(S(\Sigma))$ назовём величину

$$\rho(\varphi, \psi) = \frac{1}{5^{|S(\Sigma)|}} \sum_{k=0}^4 \sum_{l=0}^4 \frac{|k-l|}{4} M\left(\frac{k}{4}, \frac{l}{4}\right). \quad (3)$$

Теорема 1. Расстояние между двумя формулами L_5 , определённое равенством (3), для любых $\varphi, \psi, \chi \in \Sigma$ удовлетворяет следующим свойствам:

- 1) $0 \leq \rho(\varphi, \psi) \leq 1$;
- 2) $\rho(\varphi, \psi) = 0 \Leftrightarrow \varphi \equiv \psi$;
- 3) $\rho(\varphi, \psi) = \rho(\psi, \varphi)$;
- 4) $\rho(\varphi, \psi) \leq \rho(\varphi, \chi) + \rho(\chi, \psi)$;
- 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) $\rho((\varphi \wedge \psi), (\varphi \vee \psi)) = \rho(\varphi, \psi)$;
- 8) $\rho(\varphi, \psi) = \rho(\varphi \rightarrow \psi, \psi \rightarrow \varphi)$.

Доказательство. Для удобства доказательства перепишем формулу для нахождения (3), учитывая (2), в следующем виде:

$$\rho(\varphi, \psi) = \frac{1}{5^{|S(\Sigma)|}} \left(0 \cdot A + \frac{1}{4} \cdot B + \frac{1}{2} \cdot C + \frac{3}{4} \cdot D + 1 \cdot E \right).$$

1) В формуле для вычисления расстояния участвуют все модели с коэффициентами от 0 до 1. $\rho(\varphi, \psi) = 0$, если все модели содержатся в A , то есть, когда $\varphi \equiv \psi$; $\rho(\varphi, \psi) = 1$, если все модели содержатся в E , то есть $\varphi \equiv \neg\psi$ и φ и ψ принимают на моделях только значения 0 и 1. Значит, $0 \leq \rho(\varphi, \psi) \leq 1$.

2) *Необходимость.* Следует из доказательства свойства (1).

Достаточность. По определению эквивалентности [4, с. 181], если $\varphi \equiv \psi$, то их значения на всех моделях совпадают. Следовательно, все $Mod_{S(\Sigma)}((\varphi)_{\frac{k}{4}} \& (\psi)_{\frac{l}{4}})$ при $k = l$ содержатся в A , и, значит, $\rho(\varphi, \psi) = 0$.

3) Это свойство следует из того, что симметричные пары $M(k/4, l/4) \neq M(l/4, k/4)$ умножаются на один и тот же коэффициент.

4) Рассмотрим произвольную модель, на которой некоторые формулы φ, ψ, χ принимают соответственно значения $k/4, l/4$ и $i/4$. Считаем, что изначально расстояние равно нулю и это первая из рассматриваемых $5^{|S(\Sigma)|}$ моделей, которые по очереди добавляются в сумму. Распишем свойство для одной модели:

$$\frac{1}{5^{|S(\Sigma)|}} \frac{|k-l|}{4} \leq \frac{1}{5^{|S(\Sigma)|}} \frac{|k-i|}{4} + \frac{1}{5^{|S(\Sigma)|}} \frac{|i-l|}{4};$$

$$|k-l| \leq |k-i| + |i-l|.$$

Случай 1: $k \leq i \leq l$. Тогда $l-k = l-i + i-k$, свойство выполнено.

Случай 2: $k \leq l < i$. Тогда $l-k < i-k + i-l, l < i$, свойство выполнено.

Случай 3: $i < k \leq l$. Тогда $l-k < k-i + l-i, i < k$, свойство выполнено.

В силу симметричности, для $k \geq l$ – аналогично.

Так как свойство выполняется для каждой рассматриваемой модели, то для их общей суммы неравенство треугольника тоже будет выполнено.

5) Следует из определения эквивалентности двух формул.

6) Из соотношения $Mod_{S(\Sigma)}(\neg A)_{\frac{k}{4}} = Mod_{S(\Sigma)}(A)_{\frac{4-k}{4}}$ [4] следует, что

$$Mod_{S(\Sigma)}((\varphi)_{\frac{k}{4}} \& (\psi)_{\frac{l}{4}}) = Mod_{S(\Sigma)}((\neg \varphi)_{\frac{4-k}{4}} \& (\neg \psi)_{\frac{4-l}{4}}), \quad k, l = 0, \dots, 4.$$

Рассмотрим произвольную модель, как в доказательстве свойства (4). Тогда в левой части предполагаемого равенства будет стоять $|k-l|$, а в правой $|(4-k)-(4-l)| = |4-k-4+l| = |l-k| = |k-l|$. Значит, свойство выполняется.

7) Рассмотрим произвольную модель, как в доказательстве свойства (4). В многозначной логике Лукасевича $x \vee y = \max\{x, y\}$, $x \wedge y = \min\{x, y\}$. Тогда на этой модели формула $(\varphi \vee \psi)$ принимает значение $\max\{k/4, l/4\}$, а $(\varphi \wedge \psi)$ – значение $\min\{k/4, l/4\}$.

Случай 1: $k > l$. Тогда $(\varphi \vee \psi)$ принимает значение $k/4$, а $(\varphi \wedge \psi)$ – значение $l/4$.

Тогда и в левой, и в правой части предполагаемого равенства будет стоять $|k-l|$.

Случай 2: $k < l$. Аналогично первому случаю.

Случай 3: $k = l$. В обеих частях предполагаемого равенства будет стоять ноль.

Других случаев нет. Равенство доказано.

8) Рассмотрим произвольную модель, как в доказательстве свойства (4). В многозначной логике Лукасевича $x \rightarrow y = \min\{1, 1-x+y\}$. Значит, нам надо доказать, что верно равенство $|k-l| = |\min\{4, 4-k+l\} - \min\{4, 4-l+k\}|$.

Случай 1: $k = l$. Тогда $0 = 0$.

Случай 2: $k > l$. Тогда $|k-l| = |(4-k+l)-4|$, $k-l = |l-k|$, $k-l = k-l$.

Случай 3: $k < l$. Тогда $l-k = |4-(4-l+k)|$, $l-k = l-k$.

Других случаев нет. Равенство доказано. **Теорема доказана.**

Замечание 1. Свойства 2) – 4) – это свойства метрики. Таким образом, мы получили метрическое пространство на классах эквивалентных высказываний.

Замечание 2. Для расстояния между формулами классической логики выполняется интуитивно понятное свойство $\rho(\varphi, \psi) = 1 \Leftrightarrow \varphi \equiv \neg\psi$, то есть расстояние между двумя формулами максимально тогда и только тогда, когда эти формулы противоположны. В случае определённого выше расстояния между формулами $\mathbb{L}_5$, достаточность выполняется только тогда, когда φ и ψ на всех моделях принимают значения только 0 и 1. Посмотрим, почему в многозначной логике это допустимо. Пусть φ на всех моделях принимает только значение $1/4$, а формула ψ – только $3/4$. Ясно, что $\varphi \equiv \neg\psi$. Но в семантическом смысле (см. интерпретации многозначной логики Лукасевича [1, с. 76–90]) они различаются меньше, чем, например, тождественно истинная и тождественно ложная формулы. Поэтому выполнения свойства $\varphi \equiv \neg\psi \Rightarrow \rho(\varphi, \psi) = 1$ не обязательно.

Замечание 3. Расстояние, заданное формулой (3) – это расстояние для случая, когда все значения переменных заранее не известны. Пусть теперь стали известны

истинностные значения некоторых переменных (например, $x_1 = 0$ или x_1 точно не равно 1 и $3/4$).

Пусть переменные $x_1, \dots, x_p$, $x_i \in S(\varphi) \cup S(\psi)$, $i = 1, \dots, p$, $p = |S(\varphi) \cup S(\psi)|$ соответственно принимают $m_1, \dots, m_p$, $m_i \leq 5$ истинностных значений. Тогда формула для нахождения расстояния между формулами φ и ψ выглядит следующим образом:

$$\rho(\varphi, \psi) = \frac{1}{m_1 \cdot \dots \cdot m_p} \sum_{k=0}^4 \sum_{l=0}^4 \frac{|k-l|}{4} M\left(\frac{k}{4}, \frac{l}{4}\right). \quad (4)$$

При этом свойства расстояния (3) выполняются и для (4). Доказательства полнотой аналогичны, так как коэффициент $\frac{1}{m_1 \cdot \dots \cdot m_p}$, так же, как и $\frac{1}{5^{|S(\Sigma)|}}$ из выражения (3), необходим только для нормирования расстояния.

Как видно, формула (3) является частным случаем данной, если $m_1 = \dots = m_p = 5$ и $p = |S(\varphi) \cup S(\psi)|$.

Если все $m_1 \cdot \dots \cdot m_p$ моделей занумеровать, то формулу (4) можно переписать в виде

$$\rho(\varphi, \psi) = \frac{1}{m_1 \cdot \dots \cdot m_p} \sum_{i=1}^{m_1 \cdot \dots \cdot m_p} |M_i(\varphi) - M_i(\psi)|,$$

где $M_i(\varphi)$ – значение формулы φ на модели M_i , $i = 1, \dots, m_1 \cdot \dots \cdot m_p$.

2. Мера недостоверности

В классической логике под информативностью высказывания подразумевается относительное число моделей, на которых данное высказывание эксперта ложно, или, что то же самое, нормированное расстояние от высказывания до тождественно истинной формулы. Чем меньше моделей, на которых высказывание истинно, тем оно информативней [2, 3] или менее достоверно.

Мера информативности для формул из $\Phi(\Sigma) = \{\varphi : S(\varphi) \subseteq S(\Sigma)\}$ в случае двузначной логики задаётся равенством

$$\mu(\varphi) = \rho_{S(\Sigma)}(\varphi, 1) = \frac{|Mod_{S(\Sigma)}(\neg\varphi)|}{2^{|S(\Sigma)|}}.$$

Но в случае L_5 существует не одно, а четыре истинностных значений, отличных от 1.

Вместо термина «мера информативности» будем использовать термин «мера недостоверности» из тех соображений, что чем больше моделей, на которых данная формула не истинна, тем она менее достоверна.

Обобщим случай $n = 2$ на $n = 5$.

Распишем подробно, что есть расстояние от формулы φ до тождественно истинной формулы:

$$\rho_{S(\Sigma)}(\varphi, 1) = \frac{1}{5^{|S(\Sigma)|}} \left(\frac{1}{4} M\left(\varphi_3, \frac{1}{4}\right) + \frac{1}{2} M\left(\varphi_1, \frac{1}{2}\right) + \frac{3}{4} M\left(\varphi_1, \frac{1}{4}\right) + 1 M(\varphi_0, 1) \right) =$$

$$= \frac{1}{5^{|S(\Sigma)|}} \left(\frac{1}{4} M(\varphi_{\frac{3}{4}}) + \frac{1}{2} M(\varphi_{\frac{1}{2}}) + \frac{3}{4} M(\varphi_{\frac{1}{4}}) + 1 M(\varphi_0) \right) = \frac{1}{5^{|S(\Sigma)|}} \sum_{i=0}^3 \frac{4-i}{4} M(\varphi_{\frac{i}{4}}).$$

Определение 3. Мера недоверности $I(\varphi)$ для формул φ пятизначной логики L_5 , таких, что $S(\varphi) \subseteq S(\Sigma)$, на множестве $P(S(\Sigma))$ задаётся равенством

$$I(\varphi) = \sum_{i=0}^3 \frac{4-i}{4} \cdot \frac{M(\varphi_{\frac{i}{4}})}{5^{|S(\Sigma)|}}. \quad (5)$$

Теорема 2. Мера недоверности, определённая равенством (5), для любых формул $\varphi, \psi, \chi \in \Sigma$ удовлетворяет следующим свойствам:

- 1) $0 \leq I(\varphi) \leq 1$;
- 2) $I(\varphi) + I(\neg\varphi) = 1$;
- 3) $I(\varphi \wedge \psi) \geq \max\{I(\varphi), I(\psi)\}$;
- 4) $I(\varphi \vee \psi) \leq \min\{I(\varphi), I(\psi)\}$;
- 5) $I(\varphi \vee \psi) + I(\varphi \wedge \psi) \geq I(\varphi) + I(\psi)$;
- 6) $I(\varphi \wedge \psi) = \rho(\varphi, \psi) + I(\varphi \vee \psi)$;
- 7) $\rho(\varphi, \psi) \leq I(\varphi) + I(\psi)$;
- 8) $I(\varphi) \geq \rho(\varphi \rightarrow \psi, \psi)$;
- 9) $I(\varphi \rightarrow \psi) \leq \rho(\varphi, \psi)$;
- 10) $I(\varphi \vee \psi) \leq \rho(\varphi \rightarrow \psi, \varphi \wedge \psi)$.

Доказательство. Очевидно, так как $\left| \bigcup_{i=0}^4 \text{Mod}_{S(\Sigma)}(\varphi)_{\frac{i}{4}} \right| = 5^{|S(\Sigma)|}$ и $\frac{4-i}{4} \leq 1$ для

любых $i = 0, \dots, 4$.

- 1) $\text{Mod}_{S(\Sigma)}(\neg\varphi)_{\frac{k}{4}} = \text{Mod}_{S(\Sigma)}(\varphi)_{\frac{4-k}{4}}$ [4]. Значит,

$$\begin{aligned} I(\varphi) + I(\neg\varphi) &= \frac{1}{5^{|S(\Sigma)|}} \left(\frac{1}{4} M(\varphi_{\frac{3}{4}}) + \frac{1}{2} M(\varphi_{\frac{1}{2}}) + \frac{3}{4} M(\varphi_{\frac{1}{4}}) + 1 M(\varphi_0) + \right. \\ &\quad \left. + \frac{3}{4} M(\varphi_{\frac{3}{4}}) + \frac{1}{2} M(\varphi_{\frac{1}{2}}) + \frac{1}{4} M(\varphi_{\frac{1}{4}}) + 1 M(\varphi_1) \right) = \\ &= \frac{1}{5^{|S(\Sigma)|}} \left(M(\varphi_0) + M(\varphi_{\frac{1}{4}}) + M(\varphi_{\frac{1}{2}}) + M(\varphi_{\frac{3}{4}}) + M(\varphi_1) \right) = \frac{1}{5^{|S(\Sigma)|}} 5^{|S(\Sigma)|} = 1. \end{aligned}$$

Свойства 3) – 5) доказаны в работе [4].

6) Рассмотрим произвольную модель, как в доказательстве свойства (4) расстояния между формулами. Пусть на ней формула φ принимает значение $k/4$, а формула ψ – значение $l/4$. Учитывая, что $x \vee y = \max\{x, y\}$, $x \wedge y = \min\{x, y\}$, свойство для одной модели примет вид $4-l \leq |k-l| + 4-k$ или $4-k \leq |k-l| + 4-l$, в зависимости от k и l . Рассмотрим случаи:

Случай 1: $k, l \neq 4, k > l$. Тогда $4 - l = k - l + 4 - k$, $4 - l = 4 - l$, свойство выполняется.

Случай 2: $k, l \neq 4, k < l$. Тогда $4 - k = l - k + 4 - l$, $4 - k = 4 - k$.

Случай 3: $k = 4, l \neq 4$. Тогда $4 - l = 4 - l + 4 - 4$, $4 - l = 4 - l$.

Случай 4: $k \neq 4, l = 4$. Тогда, симметрично третьему случаю, $4 - k = 4 - k$.

Случай 5: $k = l$. Тогда $4 - k = 4 - k$, свойство выполнено.

Других случаев нет.

7) Следует из того, что мера недоверности по определению равна расстоянию от данной формулы до тождественно истинной, а для расстояния между формулами выполняется правило треугольника.

8) Рассмотрим произвольную модель. Надо доказать неравенство

$$4 - k \geq |\min\{4, 4 - k + l\} - l|.$$

Случай 1: $k = l$. Тогда $4 - k \geq |4 - l|$, $4 - k = 4 - l$.

Случай 2: $k > l$. Тогда $4 - k \geq |4 - k + l - l|$, $4 - k = 4 - k$.

Случай 3: $k < l$. Тогда $4 - k \geq |4 - l|$, $4 - k > 4 - l$.

9) Рассмотрим произвольную модель. Надо доказать неравенство

$$4 - \min\{4, 4 - k + l\} \leq |k - l|.$$

Случай 1: $k = l$. Тогда $0 = 0$.

Случай 2: $k > l$. Тогда $4 - (4 - k + l) \leq k - l$, $k - l = k - l$.

Случай 3: $k < l$. Тогда $4 - 4 \leq l - k$, $0 < l - k$.

10) Рассмотрим произвольную модель. Надо доказать неравенство

$$4 - \max\{k, l\} \leq |\min\{4, 4 - k + l\} - \min\{k, l\}|.$$

Случай 1: $k = l$. Тогда $4 - k \leq |4 - k|$, $4 - k = 4 - k$.

Случай 2: $k > l$. Тогда $4 - k \leq |4 - k + l - l|$, $4 - k = 4 - k$.

Случай 3: $k < l$. Тогда $4 - l \leq |4 - k|$, $4 - l < 4 - k$. **Теорема доказана.**

Замечание: Попробуем подойти к определению меры недоверности для $\mathbb{L}_5$ с другой стороны. Так как в $\mathbb{L}_5$ истинностных значений, отличных от 1, не одно, а четыре, то имеет смысл учитывать все модели, на которых формула φ принимает значения $k/4$ при $k = 0, \dots, 3$. Естественно предположить, что при каждом $M(\varphi_{k/4})$ должен стоять коэффициент, учитывающий, насколько близко истинностное значение формулы φ к 1. Ясно, что модели, на которых формула принимает значение $1/4$, должны учитываться с большим коэффициентом, чем те, на которых формула принимает значение $3/4$, так как $3/4$ ближе к единице. А так как истинностные значения в логике Лукасевича распределены равномерно между 0 и 1, то можно определить значения коэффициентов как $(4 - i)/4$, $i = 0, \dots, 3$.

3. Примеры

Применим иерархический алгоритм кластеризации с объединением кластеров по методу ближайшего соседа к некоторой группе высказываний. Пусть есть n высказываний. Сначала считаем, что у нас есть n кластеров. Построим матрицу расстояний для группы из n высказываний, потом выделим наименьшее расстояние между формулами φ_i и φ_j и объединим формулы φ_i и φ_j в один кластер. Затем

пересчитаем матрицу расстояний уже для $n-1$ высказывания по правилу $\rho(\varphi_k, \varphi_{ij}) = \min \{\rho(\varphi_k, \varphi_i), \rho(\varphi_k, \varphi_j)\}$ и будем повторять действия до тех пор, пока все высказывания не объединятся в один кластер.

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

Их меры недоверности соответственно равны: $I(\varphi_1) = 0,2000$; $I(\varphi_2) = 0,8000$; $I(\varphi_3) = 0,3000$; $I(\varphi_4) = 0,3584$; $I(\varphi_5) = 0,3000$; $I(\varphi_6) = 0,4092$; $I(\varphi_7) = 0,2716$; $I(\varphi_8) = 0,3416$.

Шаг 1: $\min_{i \neq j} \rho(\varphi_i, \varphi_j) = 0,0508 = \rho(\varphi_4, \varphi_6)$. Кластеры: $\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_7, \varphi_8$.

Шаг 2: $\min_{i \neq j} \rho(\varphi_i, \varphi_j) = 0,1000 = \rho(\varphi_1, \varphi_3)$. Кластеры: $\varphi_{13}, \varphi_2, \varphi_4, \varphi_5, \varphi_7, \varphi_8$.

Шаг 3: $\min_{i \neq j} \rho(\varphi_i, \varphi_j) = 0,1300 = \rho(\varphi_7, \varphi_{46})$. Кластеры: $\varphi_{13}, \varphi_2, \varphi_{467}, \varphi_5, \varphi_8$.

Шаг 4: $\min_{i \neq j} \rho(\varphi_i, \varphi_j) = 0,1416 = \rho(\varphi_5, \varphi_8)$. Кластеры: $\varphi_{13}, \varphi_2, \varphi_{467}, \varphi_{58}$.

Шаг 5: $\min_{i \neq j} \rho(\varphi_i, \varphi_j) = 0,2460 = \rho(\varphi_{13}, \varphi_{467})$. Кластеры: $\varphi_2, \varphi_{58}, \varphi_{13467}$.

Шаг 6: $\min_{i \neq j} \rho(\varphi_i, \varphi_j) = 0,4032 = \rho(\varphi_{58}, \varphi_{13467})$. Кластеры: $\varphi_2, \varphi_{1345678}$.

Шаг 7: $\rho(\varphi_2, \varphi_{1345678}) = 0,5000$. Кластер $\varphi_{12345678}$.

Мера недоверности здесь может выступать в качестве критерия остановки процесса объединения формул в кластеры, если количество кластеров заранее не известно. Например, на шаге 4 максимальная разница между мерами недоверности элементов одного кластера равна 0,1376, а на шаге 5 – 0,2092. В реальной задаче значение 0,2092 может оказаться неприемлемым (слишком большим). Поэтому перед началом кластеризации можно задать допустимую разницу между мерами недоверности элементов одного кластера. Соответственно, алгоритм останавливается, когда разница между мерами недоверности достигает допустимого значения.

Заключение

Введённое расстояние ρ между формулами L_5 удовлетворяет не только свойствам расстояния (2) – (4), но и свойствам, учитывающим семантику совпадения и различия информации в высказываниях.

В дальнейшем, наряду с мерой недоверности, можно ввести некую меру неопределённости, определив её как расстояние от произвольной формулы φ до формулы, тождественно равной 0,5.

Помимо иерархического алгоритма кластеризации были рассмотрены и другие, но наиболее подходящим был выбран указанный в данной работе.

Расстояние между высказываниями и меру недоверности можно использовать при анализе баз знаний, их кластеризации, согласовании высказываний, создании экспертных систем, а также при построении логических решающих функций в распознавании [6; 7].

ЛИТЕРАТУРА

1. Карпенко А.С. Логика Лукасевича и простые числа. М.: Наука, 2000. 319 с.
2. Лбов Г.С., Старцева Н.Г. Логические решающие функции и вопросы статистической устойчивости решений. Новосибирск: Изд-во Ин-та математики, 1999. 212 с.
3. Vikent'ev A.A., Lbov G.S. Setting the metric and informativeness on statements of experts // Pattern Recognition and Image Analysis. 1997. V. 7. No. 2. P. 175–183.
4. Викентьев А.А. Мера опровержимости высказываний экспертов, расстояния в многозначной логике и процессы адаптации // XIV International Conference “Knowledge-Dialogue-Solution” KDS 2008. Varna, Bulgaria, 2008. С. 179–188.
5. Еришов Ю.Л., Палютин Е.А. Математическая логика. 2-е изд. М.: Наука, 1987. 336 с.
6. Лбов Г.С., Бериков В.Б. Устойчивость решающих функций в задачах распознавания образов и анализа разнотипной информации. Новосибирск: Изд-во ИМ СО РАН, 2005.
7. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. Новосибирск: ИМ СО РАН, 1999. 270 с.

Кабанова Елизавета Сергеевна

Новосибирский государственный университет

Викентьев Александр Александрович

Институт математики им. С.Л. Соболева СО РАН (г. Новосибирск),

Новосибирский государственный университет

E-mail: iameye@mail.ru, vikent@math.nsc.ru

Поступила в редакцию 15 мая 2012 г.

Kabanova Elizaveta S., Vikentiev Alexander A. (Novosibirsk State University, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences). **Distance between formulas of the five-valued Lukasiewicz logic and unreliability measure of expert statements on the clustering of knowledge databases.**

Keywords: many-valued logic, Lukasiewicz logic, distance between formulas, unreliability measure, clustering, hierarchical algorithm, knowledge databases, expert statements, model theory.

In this paper statements of experts are represented as formulas of the five-valued Lukasiewicz logic. Likewise the case of the classical logic, using model theory, distance between formulas and unreliability measure were defined. The properties of introduced notions are defined and proved. These properties take into account semantics of similarity and differences of information contained in statements. These notions can be applied for clustering of many-valued knowledge databases. The example of grouping a set of statements using the hierarchical clustering algorithm is considered. In this case the unreliability measure is the stopping criterion of clustering procedure.