А.А.Викентьев

Модельдер теориясын пайдаланып, сарапшылық тұжырымдарды кластерлеу

Мақалада сарапшылардың пікірлері логикалық формулалар түрінде көрсетілген. Ертерек теоретикалық-модельдік ұғымдар қолданып, формулалар және тұжырымдарды терістеу (дұрыс еместік, ақпараттау) шамасы арасындағы арақашықтық енгізілді және тұжырымдардағы ақпараттың ұқсастығы мен айырмашылығы семантикасын ескеретін қасиеттері дәлелденген. Берілген шамалар білім қорларын кластерлеу үшін қолданылған. Тұжырымдардың ақырлы топтарын әр түрлі әдістермен кластерлеудің мысалдары қарастырылған.

A.A.Vikentyev

Clusterization of expert statements using models for theories

In this paper we represent experts' statements by means of logical formulas. Earlier, there were introduced distances between formulas and measures of refutation (of unreliability, informativeness) for propositions that were based on model-theoretic concepts; and their properties were established that take into account the semantics of similarity and difference between the information represented in the propositions. These notions are used for clusterization of data bases. We consider several examples of clusterization for a finite group of propositions.

УДК 519.67–519.24

А.А.Викентьев, Е.С.Кабанова

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

Расстояние между формулами пятизначной логики Лукасевича и мера недостоверности высказываний экспертов

В статье высказывания экспертов представлены в виде формул пятизначной логики Лукасевича. Аналогично случаю классической логики, используя теорию моделей, были введены понятия расстояния между формулами и меры недостоверности высказываний. Определены и доказаны свойства введённых понятий, учитывающие семантику сходства и различия информации, содержащейся в высказываниях. Данные величины могут быть использованы для кластеризации многозначных баз знаний. Рассмотрен пример кластеризации группы высказываний иерархическим методом. Мера недостоверности в данном случае выступает в качестве критерия остановки алгоритма.

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

Посвящается Учителю по теории моделей
— Туленды Гарифовичу Мустафину

Введение

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. Поэтому 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 недостоверности высказывания. В работах [2–7] определены расстояние и мера недостоверности для случая классической двузначной логики. Помимо основных свойств, введённое там расстояние обладает свойствами, учитывающими семантику совпадения и различия информации. В данной работе по-новому определяются следующие величины: расстояние между формулами пятизначной логики Лукасевича и мера недостоверности 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. Основной задачей работы являлось ввести расстояние и меру недостоверности для формул пятизначной логики Лукасевича Ł$_5$, причём так, чтобы выполнялось как можно больше свойств, характерных для данных величин, введённых и доказанных для формул классической логики, а также чтобы учитывались все возможные истинностные значения Ł$_5$.

1. Обзор логики Ł$_5$.

Определение 1. Пропозициональный язык L —

1) $x, y, z,...$ — пропозициональные переменные;

2) $\neg$, $\to$ — пропозициональные связки (отрицание и импликация соответственно);

3) $(,)$ — вспомогательные символы.

Определение 2. Формула —

1) $x, y, z,...$ — формулы;

2) если φ и ψ — формулы, то $\neg\varphi$ и $\varphi \to \psi$ — тоже формулы;

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

Определение 3. Пятизначная матрица Лукасевича — это логическая матрица вида $M_5^L = < V_5, \neg, \to, \{1\} >$, где $V_5 = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ — множество значений истинности; $\neg$, $\to$ — унарная операция отрицания и бинарная операция импликации соответственно, определённые на множестве V_5; $\{1\}$ — выделенное значение истины.

Определение 4. Функция ν — функция оценки (гомоморфизм) формул языка L в матрице M_5^L, если она удовлетворяет следующим условиям:

1) функция ν определена для каждой формулы φ;

2) если φ — пропозициональная переменная, то $\nu(\varphi) \in V_5$;

3) если φ и ψ — формулы, то

$$\nu(\neg\varphi) = \neg\nu(\varphi);$$
$$\nu(\varphi \to \psi) = \nu(\varphi) \to \nu(\psi),$$

причём в левые части равенств входят пропозициональные связки, а в правые — операции из логической матрицы.

Определение 5. Формула φ называется тавтологией в матрице M_5^L, если $\nu(\varphi) = 1$ для любой функции оценки в матрице M_5^L.

Определение 6. Пятизначная матричная логика Лукасевича Ł$_5$ — это множество тавтологий в матрице M_5^L.

Логические операции на множестве V_5 определяются следующим образом:

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

Через операции выражаются дизъюнкция и конъюнкция соответственно:

$$x \vee y = (x \to y) \to y = \max\{x, y\};$$
$$x \wedge y = \neg(\neg x \vee \neg y) = \min\{x, y\}.$$

2. Основные определения и обозначения.

Пусть база знаний $\sum$ состоит из формул Ł$_5$.

Определение 7. Множество $S^5(\varphi)$ элементарных высказываний, используемых при написании формулы φ логики Ł$_5$, назовём носителем формулы φ.

Определение 8. Назовём носителем базы знаний $S^5(\Sigma)$ объединение всех носителей формул, входящих в Σ, т.е. $S^5(\Sigma)=\bigcup_{\varphi\in\Sigma} S^5(\varphi)$.

Определение 9. Назовём множеством возможных значений носителя базы знаний величину

$$Q^5(\Sigma)=\{\varphi_{\frac{k}{4}}:\varphi\in\Sigma, k=0,...,4\}.$$

Определение 10. Назовём моделью M любое множество $Q^5(\Sigma)$ такое, что M не содержит одновременно $\varphi_{\frac{k}{4}}$ и $\varphi_{\frac{l}{4}}$ $\forall k\neq l$.

Обозначим множество всех моделей как $P(S^5(\Sigma))$. Ясно, что $\left|P\left(S^5(\Sigma)\right)\right|=5^{\left|S^5(\Sigma)\right|}$. Далее верхний индекс «5» будем опускать.

Определение 11. Элементарная формула A принимает на модели M значение $\frac{k}{4}, k=1,...,4,$ если $A_{\frac{k}{4}}\in M$, т.е. $M|=A_{\frac{k}{4}}\Leftrightarrow A_{\frac{k}{4}}\in M$.

Определение 12. Элементарная формула A принимает на модели M значение 0, если $A_{\frac{k}{4}}\notin M$.

Далее полагаем:

1) $M|=(A\wedge B)_{\frac{k}{4}}\Leftrightarrow(M|=A_{\frac{p}{4}}$ и $M|=A_{\frac{q}{4}}),\ \min\{p,q\}=k;$

2) $M|=(A\vee B)_{\frac{k}{4}}\Leftrightarrow(M|=A_{\frac{p}{4}}$ или $M|=A_{\frac{q}{4}}),\ \max\{p,q\}=k;$

3) $M|=(\neg A)_{\frac{k}{4}}\Leftrightarrow M|=A_{\frac{4-k}{4}}.$

Введём следующие обозначения:

$$Mod_{S(\Sigma)}(A)_{\frac{k}{4}}=\{M:M\in P(S(\Sigma)),M|=A_{\frac{k}{4}},\forall k=1,...4\};$$

$$Mod_{S(\Sigma)}(A)_0=\{M:M\notin P(S(\Sigma)),M|\neq A_{\frac{k}{4}},\forall k=1,...4\}.$$

То есть любой формуле φ такой, что $S(\varphi)\subseteq S(\Sigma)$, соответствует совокупность $Mod_{S(\Sigma)}(\varphi)_{\frac{k}{4}},\forall k=1,...,4,$ моделей из $P(S(\Sigma))$, на которых она принимает значения $\frac{k}{4}$, $k=1,...,4,$ соответственно.

Сформулируем некоторые свойства этого понятия:

$$Mod_{S(\Sigma)}\left((A\wedge B)_{\frac{k}{4}}\right)=\bigcup_{l=k}^{4}\left((Mod_{S(\Sigma)}(A)_{\frac{l}{4}}\cap Mod_{S(\Sigma)}(B)_{\frac{k}{4}}\right)\cup\left(Mod_{S(\Sigma)}(A)_{\frac{k}{4}}\cap Mod_{S(\Sigma)}(B)_{\frac{l}{4}})\right);$$

$$Mod_{S(\Sigma)}\left((A\vee B)_{\frac{k}{4}}\right)=\bigcup_{l=0}^{4}\left((Mod_{S(\Sigma)}(A)_{\frac{l}{4}}\cup Mod_{S(\Sigma)}(B)_{\frac{k}{4}}\right)\cup\left(Mod_{S(\Sigma)}(A)_{\frac{k}{4}}\cup Mod_{S(\Sigma)}(B)_{\frac{l}{4}})\right);$$

$$Mod_{S(\Sigma)}(\neg A)_{\frac{k}{4}}=Mod_{S(\Sigma)}(A)_{\frac{4-k}{4}}.$$

Определение 13. Назовём формулы φ и ψ эквивалентными $(\varphi\equiv\psi)$, если $Mod_{S(\Sigma)}(\varphi)_{\frac{k}{4}}=Mod_{S(\Sigma)}(\psi)_{\frac{k}{4}}$ для каждого $k=1,...,4,$ т.е. они имеют одно и то же множество моделей в каждом значении истинности. Это отношение — отношение эквивалентности.

3. Расстояние между формулами в $Ł_5$.

Для краткости будем пользоваться следующими обозначениями:

$M(\varphi_{\frac{k}{4}})=\left|Mod_{S(\Sigma)}(\varphi)_{\frac{k}{4}}\right|$, т.е. количество моделей, на которых формула φ принимает значение $\frac{k}{4}$.

$M(\frac{k}{4},\frac{l}{4})=\left|Mod_{S(\Sigma)}((\varphi)_{\frac{k}{4}}\,\&\,(\psi)_{\frac{l}{4}})\right|$, т.е. количество моделей, на которых формула φ принимает значение $\frac{k}{4}$, а ψ – $\frac{k}{4}$.

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

Определение 14. Расстоянием между формулами φ и ψ двузначной логики при $S(\varphi)\cup S(\psi)\subseteq S(\Sigma)$ на множестве $P(S(\Sigma))$ называется величина

$$\rho_{S(\Sigma)}(\varphi,\psi)=\frac{\left|Mod_{S(\Sigma)}((\neg\varphi\,\&\,\psi)\vee(\varphi\,\&\,\neg\psi))\right|}{2^{|S(\Sigma)|}}. \qquad (1)$$

Преобразуем формулу (1) к следующему виду:

$$\rho_{S(\Sigma)}(\varphi,\psi)=\frac{1}{2^{|S(\Sigma)|}}\cdot\left|Mod_{S(\Sigma)}((\neg\varphi\,\&\,\psi)\vee(\varphi\,\&\,\neg\psi))\right|=M(0,1)+M(1,0)=$$

$$=1\cdot M(0,1)+1\cdot M(1,0)+0\cdot M(0,0)+0\cdot M(1,1)=\frac{1}{2^{|S(\Sigma)|}}\cdot\sum_{k=0,1}\sum_{l=0,1}|k-l|\cdot M(k,l).$$

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

Замечание (доказательство в [2; 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 $Ł_5$.

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

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

Определение 15. Расстоянием между формулами φ и ψ пятизначной логики $Ł_5$ при $S(\varphi)\cup S(\psi)\subseteq S(\Sigma)$ на множестве $P(S(\Sigma))$ назовём величину

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

Теорема. Расстояние между двумя формулами $Ł_5$, определённое равенством (3), для любых $\varphi,\psi,\chi\in\Sigma$ 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) $\rho(\varphi,\psi) \le \rho(\varphi,\chi) + \rho(\chi,\psi)$;
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) $\rho(\varphi,\psi) = \rho(\varphi \to \psi, \psi \to \varphi)$.

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

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

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, т.е. когда $\varphi \equiv \psi$. $\rho(\varphi,\psi) = 1$, если все модели содержатся в E, т.е. $\varphi \equiv \neg\psi$, и φ и ψ принимают на моделях только значения от 0 до 1. Значит,

$$0 \le \rho(\varphi,\psi) \le 1.$$

2. **Необходимость:** следует из доказательства свойства (1).

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

$$\rho(\varphi,\psi) = 0.$$

3. Следует из того, что симметричные пары $M\left(\frac{k}{4},\frac{l}{4}\right) \ne M\left(\frac{l}{4},\frac{k}{4}\right)$ умножаются на один и тот же коэффициент.

4. Рассмотрим произвольную модель, на которой некоторые формулы φ,ψ,χ принимают соответственно значения $\frac{k}{4},\frac{l}{4}$ и $\frac{i}{4}$. 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: $\frac{1}{5^{|S(\Sigma)|}} \cdot \frac{|k-l|}{4} \le \frac{1}{5^{|S(\Sigma)|}} \cdot \frac{|k-i|}{4} + \frac{1}{5^{|S(\Sigma)|}} \cdot \frac{|i-l|}{4}$;

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

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

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

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

В силу симметричности, для $k \ge l$ — 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

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

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 $\max\left\{\frac{k}{4}, \frac{l}{4}\right\}$, а ($\varphi \wedge \psi$) — значение $\min\left\{\frac{k}{4}, \frac{l}{4}\right\}$.

Случай 1: $k > l$. Тогда ($\varphi \vee \psi$) принимает значение $\frac{k}{4}$, а ($\varphi \wedge \psi$) — значение $\frac{l}{4}$.

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 $x \to y = \min\{1, 1 - x + y\}$. 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 $\rho(\varphi, \psi) = 1 \Leftrightarrow \varphi \equiv \neg\psi$, т.е. расстояние между двумя формулами максимально тогда и только тогда, когда эти формулы противоположны. В случае определённого выше расстояния между формулами *Ł*$_5$ 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 $\frac{1}{4}$, а формула ψ — только $\frac{3}{4}$. Ясно, что $\varphi \equiv \neg\psi$. Но в семантическом смысле (см. интерпретации многозначной логики Лукасевича [1; 76–90]) они различаются меньше, чем, например, тождественно истинная и тождественно ложная формулы. Поэтому выполнение свойства $\varphi \equiv \neg\psi \Rightarrow \rho(\varphi, \psi) = 1$ необязательно.

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 $\frac{3}{4}$).

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

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

При этом свойства расстояния (3) выполняются и для (4). Доказательства полностью аналогичны, так как коэффициент $\frac{1}{m_1 \cdot ... \cdot m_p}$, 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 $m_1 = ... = m_p = 5$. Если все $m_1 \cdot ... \cdot m_p$ моделей занумеровать, то формулу (4) можно переписать в виде

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

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

4. 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

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

Но в случае $Ł_5$ существует не одно, а четыре истинностных значений, отличных от 1.

Вместо термина «мера информативности» будем использовать термины «мера недостоверности» в силу соображений о том, 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:

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

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

Определение 16. Мера недостоверности $I(\varphi)$ для формул φ пятизначной логики $Ł_5$, таких что $S(\varphi)\subseteq S(\Sigma)$, на множестве $P(S(\Sigma))$ задаётся равенством

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

Теорема. Мера недостоверности, определённая равенством (5), для любых формул $\varphi,\psi,\chi\in\Sigma$ 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) $I(\varphi\vee\psi)+I(\varphi\wedge\psi)\ge I(\varphi)+I(\psi)$;
6) $I(\varphi\wedge\psi)=\rho(\varphi,\psi)+I(\varphi\vee\psi)$;
7) $\rho(\varphi,\psi)\le I(\varphi)+I(\psi)$;
8) $I(\varphi)\ge\rho(\varphi\to\psi,\psi)$;
9) $I(\varphi\to\psi)\le\rho(\varphi,\psi)$;
10) $I(\varphi\vee\psi)\le\rho(\varphi\to\psi,\varphi\wedge\psi)$.

Доказательство:

1. Очевидно, так как $\left|\bigcup_{i=0}^{4}Mod_{S(\Sigma)}(\varphi)_{\frac{i}{4}}\right|=5^{|S(\Sigma)|}$ и $\frac{4-i}{4}\le 1$ для любых $i=0,\ldots,4$.

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

$$I(\varphi)+I(\neg\varphi)=\frac{1}{5^{|S(\Sigma)|}}\cdot\left(\frac{1}{4}\cdot M\left(\varphi_{\frac{3}{4}}\right)+\frac{1}{2}\cdot M\left(\varphi_{\frac{1}{2}}\right)+\frac{3}{4}\cdot M\left(\varphi_{\frac{1}{4}}\right)+1\cdot M(\varphi_0)+\right.$$

$$\left.+\frac{3}{4}\cdot M\left(\varphi_{\frac{3}{4}}\right)+\frac{1}{2}\cdot M\left(\varphi_{\frac{1}{2}}\right)+\frac{1}{4}\cdot M\left(\varphi_{\frac{1}{4}}\right)+1\cdot M(\varphi_1)\right)=$$

$$=\frac{1}{5^{|S(\Sigma)|}}\cdot\left(M(\varphi_0)+M\left(\varphi_{\frac{1}{4}}\right)+M\left(\varphi_{\frac{1}{2}}\right)+M\left(\varphi_{\frac{3}{4}}\right)+M(\varphi_1)\right)=\frac{1}{5^{|S(\Sigma)|}}\cdot 5^{|S(\Sigma)|}=1.$$

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 $\frac{k}{4}$, а формула ψ — значение $\frac{l}{4}$. Учитывая, что $x\vee y=\max\{x,y\}$, $x\wedge y=\min\{x,y\}$, свойство для одной модели примет вид $4-l\le|k-l|+4-k$ или $4-k\le|k-l|+4-l$, в зависимости от k и l. Рассмотрим случаи:

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

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

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

Случай 4: $k\ne 4$, $l=4$. 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 недостоверности, 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 $4-k\ge|\min\{4,4-k+l\}-l|$.

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

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

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

9. Рассмотрим произвольную модель. Надо доказать неравенство $4-\min\{4,4-k+l\}\le|k-l|$.

Случай 1: $k=l$. Тогда $0=0$.

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

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

10. Рассмотрим произвольную модель. Надо доказать неравенство

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

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

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

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

Теорема доказана.

Замечание. Попробуем подойти к определению меры недостоверности для $Ł_5$ с другой стороны. Так как в $Ł_5$ истинностных значений, отличных от 1, не одно, а четыре, то имеет смысл учитывать все модели, на которых формула φ принимает значения $\frac{k}{4}$ при $k=0,...,3$. Естественно предположить, что при каждом $M\left(\varphi_{\frac{k}{4}}\right)$ должен стоять коэффициент, учитывающий, насколько близко истинностное значение формулы φ к 1. Ясно, что модели, на которых формула принимает значение $\frac{1}{4}$, должны учитываться с большим коэффициентом, чем те, на которых формула принимает значение $\frac{3}{4}$, так как $\frac{3}{4}$ ближе к единице. А поскольку истинностные значения в логике Лукасевича распределены равномерно между 0 и 1, то можно определить значения коэффициентов как $\frac{4-i}{4}$, где $i=0,...,3$.

4) Приложения введенных мер.

По формулам выше рассчитаем расстояния и меры недостоверности формул:

φ	ψ	$\rho(\varphi,\psi)$
x	y	0,4
$x \wedge y$	$x \vee y$	0,4
$x \wedge y$	$\neg(x \wedge y)$	0,6
$x \to y$	$\neg(x \to y)$	0,76
$x \to (y \to (z \to w))$	$w \to (z \to (y \to x))$	0,0448
$x \wedge y \wedge z \wedge w$	$x \to w$	0,2576

φ	$I(\varphi)$
x	0,5
$x \to y$	0,2
$\neg(x \to y)$	0,8
$(x \to y) \wedge (y \to z) \vee (z \to x)$	0,044
$x \vee y \vee z$	0,2
$x \vee y \vee z \vee w$	0,1416

Применим к формулам 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 для уже $n-1$ 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:

$$\varphi_1 = x \to y;\ \varphi_2 = \neg(x \to y);\ \varphi_3 = (x \vee z) \to y;\ \varphi_4 = \neg((x \wedge y) \vee z) \to w;\ \varphi_5 = y \to (x \wedge z);$$
$$\varphi_6 = (\neg y \vee (x \to z)) \to w;\ \varphi_7 = ((x \to y) \to z) \to w;\ \varphi_8 = (w \to z) \wedge (y \to x).$$

Их меры недостоверности 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: $\min\limits_{i \neq j} \rho(\varphi_i,\varphi_j) = 0,0508 = \rho(\varphi_4,\varphi_6)$. Кластеры: $\varphi_1, \varphi_2, \varphi_3, \varphi_{46}, \varphi_5, \varphi_7, \varphi_8$.

Шаг 2: $\min\limits_{i \neq j} \rho(\varphi_i,\varphi_j) = 0,1000 = \rho(\varphi_1,\varphi_3)$. Кластеры: $\varphi_{13}, \varphi_2, \varphi_{46}, \varphi_5, \varphi_7, \varphi_8$.

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

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

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

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

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

Мера недостоверности 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 недостоверности 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 недостоверности элементов одного кластера. Соответственно, алгоритм останавливается, когда разница между мерами недостоверности достигает допустимого значения.

Заключение

Введённое расстояние ρ между формулами $Ł_5$ удовлетворяет не только свойствам расстояния (2)–(4), но и свойствам, учитывающим семантику совпадения и различия информации в высказываниях.

В дальнейшем, наряду с мерой недостоверности, можно ввести некую меру неопределённости, определив её как расстояние от произвольной формулы φ до формулы, тождественно равной $\frac{1}{2}$.

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

Работа поддержана грантом РФФИ, номера проектов 10–01–00113a, 11–07–00345a.

References

1 *Karpenko A.S.* Lukasiewicz logic and prime numbers. — Moscow: Nauka, 2000. — 319 p.

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

3 *Lbov G.S., Startseva Н.Г.* Logical decision functions and aspects of statistical stability of the solutions. — Novosibirsk: Izd-vo In-ta matematiki, 1999. — 212 p.

4 *Vikentyev A.A., Lbov G.S.* Setting the metric and informativeness on statements of experts // Pattern Recognition and Image Analysis. — 1997. — Vol. 7. — № 2. — P. 175–183

5 *Vikentyev A.A.* Measure falsifiability statements of experts, distances in multi-valued logic and the process of adaptation // XIV International Conference «Knowledge-Dialogue-Solution» KDS 2008. — Bulgaria: Varna, 2008. — P. 179–188.

6 *Berikov V.B., Lbov G.S.* Construction of decision functions and questions of stability. — Novosibirsk: Izd-vo In-ta matematiki, 2006. — 217 p.

7 *Zagoruyko N.G.* Applied methods of data analysis and knowledge. — Novosibirsk: IM SO RAN, 1999. — P. 270.

А.А.Викентьев, Е.С.Кабанова

Лукасевичтің бес таңбалы логикасының формулалар арасындағы қашықтығы және сарапшылық тұжырымдардың ақиқат емес өлшемі

Мақалада сарапшылардың пікірлері Лукасевичтің бес таңбалы логикасының формуласы түрінде көрсетілген. Классикалық логика жағдайына ұқсас модельдер теориясын пайдалана отырып, «формулалар арасындағы қашықтық» және «тұжырымдардың ақиқат еместік шамасы» түсініктері енгізілген. Тұжырымдарда қамтылған мәліметтердің ұқсастықтары мен айырмашылықтарының семантикасы есепке алынып, енгізілген ұғымдардың қасиеттері анықталған және дәлелденген. Бұл шамалар көп мағына беріп, білім қорларын кластерлеу үшін қолданылды. Тұжырым топтарын иерархиялық әдіспен кластерлеудің мысалы қарастырылған. Бұл жағдайда ақиқат еместік өлшемі алгоритмді тоқтату шегі ретінде келтірілді.

A.A.Vikentyev, E.S.Kabanova

The distance between the formulas of five digit Lukasiewicz logic and unreliable measure of expert's statements

In this work 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. In this work 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 a stopping criterion of clustering procedure.

УДК 517.947

С.Ж.Игисинов

Евразийский национальный университет им. Л.Н.Гумилева, Астана
(E-mail: igisinovsabit@mail.ru)

О замыкаемости и об ограниченной обратимости дифференциального оператора смешанного типа в неограниченной области

В статье рассмотрен один класс сингулярного дифференциального оператора смешанного типа в неограниченной области. Первоначально доказана замыкаемость данного оператора, так как в данном случае не существует априорной оценки, как в случае с оператором, коэффициенты которого при слагаемых $u_x(x,y)$ и $u(x,y)$ зависят только от переменной y. Доказано существование резольвенты рассматриваемого оператора в неограниченной области. В работе использованы методы локализации и теории линейных операторов.

Ключевые слова: оператор смешанного типа, замыкание, резольвента, ограниченная обратимость, обратный оператор, неограниченная область, полоса, покрытие, ядро, сопряженный оператор, преобразование Фурье.

Часть 1. Введение. Формулировка основных результатов

Пусть $\Omega=\{(x,y):-\infty<x<\infty,-1<y<1\}$. Рассмотрим дифференциальный оператор

$$L_0u=k(y)\mathrm{u}_{xx}-u_{yy}+a(x)u_x+c(x)u, \qquad (0.1)$$

первоначально определенный на $C_0^\infty(\overline{\Omega})$ — множестве, состоящем из бесконечно дифференцируемых функций, удовлетворяющих условию $u(x;-1)=u(x;1)=0$ и финитных по переменной x, где коэффициенты $a(x)$ и $c(x)$ — непрерывные функции в $R=(-\infty;\infty)$, а $k(y)$ — непрерывная и ограниченная функция, удовлетворяющая условию: $yk(y)>0$ при $y\neq 0$ и $k(0)=0$.

Здесь отметим, что коэффициенты оператора (0.1) $a(x)$ и $c(x)$ могут быть неограниченными функциями.

В случае $\kappa(y)=-1$ оператор (0.1) принадлежит эллиптическому типу. Как известно, вопрос о существовании резольвенты сингулярных эллиптических операторов, заданных в неограниченных областях, достаточно хорошо изучен, например, в работах [1, 2].

Дифференциальный оператор гиперболического типа во всем пространстве E^n (евклидово пространство размерности n) исследован в работе [3], где коэффициенты оператора — непрерывные и ограниченные функции в E^n.