УДК 519.67

А.А.Викентьев

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

Изучение модельных расстояний на логических высказываниях с учетом экспертных интерпретаций для формул многозначных логик Лукасевича и автоматической кластеризации в базах знаний. II

В статье исследованы модельные расстояния на логических высказываниях с учетом экспертных интерпретаций для формул многозначных логик Лукасевича. Результатом работы являются теоремы 2–4. Приведены примеры с подсчетом расстояний между формулами, свидетельствующие о новизне метрики.

Ключевые слова: модельные расстояния, логические высказывания, логика Лукасевича, база знаний.

Расстояние между множествами моделей

Перейдем к вопросу введения расстояния между множествами моделей для определения степени разброса моделей двух формул с учетом упорядочения и фиксированным заданием $p(x)$ экспертом.

Определение 5. Расстояние между элементами x метрического пространства $< P(S(\Sigma)), \rho >$ и подмножеством (моделью) B этого метрического пространства с фиксированным заданием $p(x)$ (экспертом) зададим следующим образом:

$$d(x, B) = \min\{\rho(x, y) | y \in B\}.$$

Определение 6. Введем в рассмотрение функцию от множества A и B из метрического пространства $< P(S(\Sigma)), \rho >$ с фиксированным заданием $p(x)$ экспертом, определяемую как

$$d(A, B) = \max\{d(x, B) | x \in A\}.$$

Заметим, что $d(A, B)$ метрику, вообще говоря, не определяет (так как не выполняется, например, аксиома симметрии).

Определение 7. Расстояние $\hat{\rho}$ между двумя конечными множествами моделей A и B из метрического пространства $< P(S(\Sigma)), \rho >$ с фиксированным заданием $p_i(x)$ экспертом зададим формулой

$$\hat{\rho}(A, B) = \frac{d(A, B) + d(B, A)}{2}.$$

Покажем, что $\hat{\rho}$ действительно определяет метрику.

Теорема 2. Функция $\hat{\rho}$ является метрикой.

Доказательство. Требуется проверить выполнимость аксиом.

1. $\hat{\rho}(A, B) \geq 0$. Это следует из определения $\hat{\rho}$, так как величины d(A,B) и d(B,A) неотрицательны.

2. $\hat{\rho}(A, B) = 0 \Leftrightarrow A = B$. Если $A = B$, то очевидно, что $\hat{\rho}(A, B) = 0$. С другой стороны, если $\hat{\rho}(A, B) = 0$, то $d(A, B) = -d(B, A)$. Так как $d(A, B) \geq 0$, то $d(A, B) = d(B, A) = 0$. Вследствие свойств функции d получаем $A = B$.

3. $\hat{\rho}(A,B)=\hat{\rho}(B,A)$. Это утверждение следует из определения $\hat{\rho}$.

4. $\hat{\rho}(A,C)\le\hat{\rho}(A,B)+\hat{\rho}(B,C)$. Сначала покажем, что для любых $A,B,C\subseteq P(S(\Sigma))$ выполняется $d(A,C)\le d(A,B)+d(B,C)$. Докажем первое равенство. Пусть $a\in A$, тогда $d(a,C)=$ $=\min\{\rho(x,c)|c\in C\}$. Для каждого $b\in B$

$$d(a,C)\le\min\{d(a,b)+d(b,c)|c\in C\}\le d(a,b)+\min\{d(b,c)|c\in C\}\le$$

$$\le d(a,B)+\max\{d(b,C)|b\in B\}\le d(a,b)+d(B,C).$$

Так как это равенство верно при любом $a\in A$, то $d(A,C)\le d(A,B)+d(B,C)$. Далее суммируем два неравенства:

$$d(A,C)\le d(A,B)+d(B,C) \text{ и } d(C,A)\le d(B,A)+d(C,B),$$

а полученную сумму делим на 2. Получаем $\hat{\rho}(A,C)\le\hat{\rho}(A,B)+\hat{\rho}(B,C)$.

Расстояние между двумя формулами исчисления высказываний

Определение 8. Расстояние между двумя формулами высказываний эксперта с фиксированным заданием $p(x)$ экспертом зададим формулой

$$\tilde{\rho}(\varphi,\psi)=\frac{\rho(\varphi,\psi)+\hat{\rho}(Mod_{S(\Sigma)}(\varphi),Mod_{S(\Sigma)}(\psi))}{2}.$$

Утверждение. На множестве классов эквивалентности высказываний экспертов с фиксированным заданием $p(x)$ $\tilde{\rho}$ определяет метрику.

Доказательство получается из свойства, что линейная комбинация метрик является метрикой.

Пример. Пусть база знаний экспертов состоит из трех элементарных высказываний $S(\Sigma)=\{A,B,C\}$, упорядочение для фиксированного эксперта — $A<B<C$, $P(A)=1$, $P(B)=1/2, P(C)=1/3$.

φ	ψ	$\rho(\varphi,\psi)$	$\tilde{\rho}(\varphi,\psi)$
$A\wedge B$	$A\to B$	1/3	0,406245
$A\wedge B$	$A\vee B$	2/9	0,325397
$A\vee B$	$A\to B$	2/9	0,325397
$A\to(B\wedge C)$	$A\to(B\vee C)$	2/27	0,181216
$A\wedge(B\to C)$	$\sim A\vee C$	10/27	0,411062
$(A\vee B)\wedge C$	$A\to C$	5/27	0,221362
$\sim A$	$B\to C$	7/27	0,408301
$A\wedge(B\vee C)$	$(A\wedge B)\vee C$	3/27	0,269841

Формула для расстояния при наличии нескольких различных экспертов

Теорема 3. На множестве классов эквивалентности экспертных высказываний можно задать расстояние ρ^*(а также ρ^{**} с весами), являющееся метрикой с учётом упорядочения элементарных высказываний каждым экспертом и степени разброса каждого эксперта:

$$\rho^*(\varphi,\psi)=\frac{\sum\limits_{i\in\{1,\ldots,M\}}\tilde{p}_i(\varphi,\psi)}{M}=\frac{\rho(\varphi,\psi)}{2}+\frac{\sum\limits_{i\in\{1,\ldots,M\}}\tilde{p}_i(Mod_{S(\Sigma)}(\varphi),Mod_{S(\Sigma)}(\psi))}{2M},$$

а также с весами

$$\rho^{**}(\varphi,\psi)=\alpha\rho(\varphi,\psi)+\frac{\sum\limits_{i\in\{1,\ldots,M\}}\gamma_i\tilde{p}_i(Mod_{S(\Sigma)}(\varphi),Mod_{S(\Sigma)}(\psi))}{M},$$

где $\alpha+\sum\limits_{i\in\{1,\ldots,M\}}\gamma_i=1$.

Заметим, что коэффициенты α и γ_i можем выбрать методом наименьших квадратов, если будем иметь дополнительные сведения о желаемом расстоянии ρ^{**}. ρ^* и ρ^{**} будут расстояниями, так как являются линейной комбинацией расстояний.

Пример. Пусть база знаний экспертов состоит из трех элементарных высказываний $S(\Sigma)=\{A,B,C\}$ и двух экспертов. Для первого возьмём упорядочение $P(A)=1$; $P(B)=1/2, P(C)=1/3$. Для второго — $P(A)=1/2; P(B)=2/3; P(C)=1/4$.

φ	ψ	$\tilde{\rho}_1(\varphi,\psi)$	$\tilde{\rho}_2(\varphi,\psi)$	$\rho^*(\varphi,\psi)$
$A\wedge B$	$A\to B$	0,406245	0,406123	0,406184
$A\wedge B$	$A\vee B$	0,325397	0,302676	0,314037
$A\vee B$	$A\to B$	0,325397	0,398459	0,361928
$A\to(B\wedge C)$	$A\to(B\vee C)$	0,181216	0,180711	0,180964
$A\wedge(B\to C)$	$\sim A\vee C$	0,411062	0,345825	0,378444
$(A\vee B)\wedge C$	$A\to C$	0,221362	0,297184	0,259270
$\sim A$	$B\to C$	0,408301	0,34514	0,37672
$A\wedge(B\vee C)$	$(A\wedge B)\vee C$	0,269841	0,19923	0,2345355

Поиск коэффициентов

Известны расстояния:

$$\rho(A\wedge B,A\to B)=0,4;$$

$$\rho(A\vee B,A\to B)=0,3;$$

$$\rho(A\to(B\wedge C),A\to(B\vee C))=0,2;$$

$$\rho(A\wedge(B\to C),\sim A\vee C)=0,4;$$

$$\rho(A\wedge(B\vee C),(A\wedge B)\vee C)=0,25.$$

Находим коэффициенты методом наименьших квадратов для первого эксперта:

$$\rho_i x+\hat{\rho}_i(1-x)-\tilde{\rho}_i=0;$$

$$x=0,53;\quad y=0,47\ \rho(A\wedge B,A\vee B)=0,319206;$$

$$\rho((A\vee B)\wedge C,A\to C)=0,219192;\quad \rho(\sim A,B\to C)=0,399358.$$

Для второго эксперта:

$$x=0,5583;\quad y=0,4417\ \rho(A\wedge B,A\vee B)=0,293295;$$

$$\rho((A\vee B)\wedge C,A\to C)=0,284125;\quad \rho(\sim A,B\to C)=0,335127.$$

С оптимальными коэффициентами для обоих экспертов (с равной степенью доверия) расстояния получатся:

$$\rho(A \wedge B, A \vee B) = 0,3062505; \rho((A \vee B) \wedge C, A \to C) = 0,251583; \rho(\sim A, B \to C) = 0,36724.$$

Стоит заметить, что исходя из известных расстояний мы можем выбрать коэффициенты доверия эксперту. В нашем примере они равняются $\gamma_1 = 3/4; \gamma_2 = 1/4$.

$$\rho(A \wedge B, A \vee B) = 0,31273;\ \rho((A \vee B) \wedge C, A \to C) = 0,235425;\ \rho(\sim A, B \to C) = 0,3833.$$

Меры информативности и недостоверности (нетривиальности)

Определим информативность и недостоверность в терминах теории моделей для трехзначной логики. Под информативностью высказывания будем понимать относительное число моделей, на которых это высказывание ложно, т.е. нормированное расстояние от высказывания до тождественно истинной формулы 1. Понятно, что высказывание тем информативней, чем меньше моделей из данного класса, на которых оно истинно. При этом предполагаем, что оно имеет модели

$$\mu_{S(\Sigma)}(A) = \rho_{S(\Sigma)}(A, 1) = \frac{|Mod_{S(\Sigma)}(\sim A)|}{3^{|S(\Sigma)|}}.$$

Для трехзначной логики можно ввести меру недостоверности — относительное число моделей, на которых высказывание не истинно:

$$\eta_{S(\Sigma)}(A) = 1 - \frac{|Mod_{S(\Sigma)}(A)|}{3^{|S(\Sigma)|}}.$$

Аналогично можно ввести новую меру информативности $\tilde{\mu}_{S(\Sigma)}(\varphi)$ как расстояние от искомой формулы до тождественной формулы, используя при этом введенное ранее расстояние, т.е.

$$\tilde{\mu}_{S(\Sigma)}(A) = \tilde{\rho}_{S(\Sigma)}(A, 1).$$

Теорема 4(свойства мер информативности и недостоверности):

1) $0 \leq \mu_{S(\Sigma)}(\varphi) \leq 1$;

2) $\mu_{S(\Sigma)}(\sim \varphi) \neq 1 - \mu_{S(\Sigma)}(\varphi)$;

3) $\mu_{S(\Sigma)}(\varphi) \geq \mu_{S(\Sigma)}(\varphi \wedge \psi)$;

4) $\mu_{S(\Sigma)}(\varphi \vee \psi) \leq \mu_{S(\Sigma)}(\varphi)$;

5) если $\rho_{S(\Sigma)}(\varphi, \psi) = 1$, то $\mu_{S(\Sigma)}(\varphi \wedge \psi) = 1$ и $\mu_{S(\Sigma)}(\varphi \vee \phi) = 0$;

6) если $\rho_{S(\Sigma)}(\varphi, \psi) = 0$, то $\mu_{S(\Sigma)}(\varphi \wedge \psi) = \mu_{S(\Sigma)}(\varphi)$ и $\mu_{S(\Sigma)}(\varphi \vee \psi) = \mu_{S(\Sigma)}(\varphi)$;

7) $\mu_{S(\Sigma)}(\varphi \wedge \psi) = \rho_{S(\Sigma)}(\varphi, \psi) + \mu_{S(\Sigma)}(\varphi \vee \psi)$;

8) $\min\{\mu_{S(\Sigma)}(\varphi), \mu_{S(\Sigma)}(\psi)\} \geq \mu_{S(\Sigma)}(\varphi \vee \psi) \geq \max\{\mu_{S(\Sigma)}(\varphi), \mu_{S(\Sigma)}(\psi)\} - \rho_{S(\Sigma)}(\varphi, \psi)$
и $\min\{\mu_{S(\Sigma)}(\varphi), \mu_{S(\Sigma)}(\psi)\} + \rho_{S(\Sigma)}(\varphi, \psi) \geq \mu_{S(\Sigma)}(\varphi \wedge \psi) \geq \max\{\mu_{S(\Sigma)}(\varphi), \mu_{S(\Sigma)}(\psi)\}$;

9) если $\mu_{S(\Sigma)}(\varphi) = \mu_{S(\Sigma)}(\varphi_1), \mu_{S(\Sigma)}(\psi) = \mu_{S(\Sigma)}(\psi_1), \rho_{S(\Sigma)}(\varphi, \psi) \leq \rho_{S(\Sigma)}(\varphi_1, \psi_1)$,
то $\mu_{S(\Sigma)}(\varphi \wedge \psi) \leq \mu_{S(\Sigma)}(\varphi_1 \wedge \psi_1)$;

10) если $\mu_{S(\Sigma)}(\varphi) \leq \mu_{S(\Sigma)}(\varphi_1), \mu_{S(\Sigma)}(\psi) = \mu_{S(\Sigma)}(\psi_1), \rho_{S(\Sigma)}(\varphi, \psi) = \rho_{S(\Sigma)}(\varphi_1, \psi_1)$,
то $\mu_{S(\Sigma)}(\varphi \wedge \psi) \leq \mu_{S(\Sigma)}(\varphi_1 \wedge \psi_1)$;

11) $\mu_{S(\Sigma)}(\varphi \wedge \psi) = f(\mu_{S(\Sigma)}(\varphi), \mu_{S(\Sigma)}(\psi), \rho_{S(\Sigma)}(\varphi, \psi)$, где $f(x, y, z) = \dfrac{(x + y + z)}{2}$;

12) $\mu_{S(\Sigma)}(\varphi \vee \psi) = \dfrac{\mu_{S(\Sigma)}(\varphi) + \mu_{S(\Sigma)}(\psi) - \rho_{S(\Sigma)}(\varphi, \psi)}{2}$.

Доказательство.

1. Очевидно, что свойство 1 выполняется для $\mu_{S(\Sigma)}(A)$ и $\eta_{S(\Sigma)}(A)$.

2. $\mu_{S(\Sigma)}(\sim \varphi) \neq 1 - \mu_{S(\Sigma)}(\varphi)$;

   а) $\dfrac{|Mod_{S(\Sigma)}(\varphi)|}{3^{|S(\Sigma)|}} \neq 1 - \dfrac{|Mod_{S(\Sigma)}(\sim \varphi)|}{3^{|S(\Sigma)|}}$,
   $\dfrac{|Mod_{S(\Sigma)}(\varphi)| + |Mod_{S(\Sigma)}(\sim \varphi)|}{3^{|S(\Sigma)|}} \neq 1$.
   Пусть $\varphi = A \wedge B$, тогда $1 \neq \dfrac{1+5}{9} = \dfrac{2}{3}$. В формуле можно поставить только $\leq$;

   б) $1 - \dfrac{|Mod_{S(\Sigma)}(\sim \varphi)|}{3^{|S(\Sigma)|}} \neq 1 - 1 + \dfrac{|Mod_{S(\Sigma)}(\varphi)|}{3^{|S(\Sigma)|}}$
   $1 \neq \dfrac{|Mod_{S(\Sigma)}(\varphi)| + |Mod_{S(\Sigma)}(\sim \varphi)|}{3^{|S(\Sigma)|}}$.
   Аналогично а), в формуле можно поставить только $\geq$.

3. $\mu_{S(\Sigma)}(\varphi) \geq \mu_{S(\Sigma)}(\varphi \wedge \psi)$;

   а) $\dfrac{|Mod_{S(\Sigma)}(\sim (\varphi \wedge \psi))|}{3^{|S(\Sigma)|}} = \dfrac{|Mod_{S(\Sigma)}(\sim (\varphi \wedge \psi))|}{3^{|S(\Sigma)|}} = \dfrac{|Mod_{S(\Sigma)}(\sim \varphi \vee \sim \psi)|}{3^{|S(\Sigma)|}} =$
   $= \dfrac{|Mod_{S(\Sigma)}(\sim \varphi) \cup Mod_{S(\Sigma)}(\sim \psi)|}{3^{|S(\Sigma)|}} \geq \dfrac{|Mod_{S(\Sigma)}(\sim \varphi)|}{3^{|S(\Sigma)|}}$;

   б) $1 - \dfrac{|Mod_{S(\Sigma)}(\varphi \wedge \psi)|}{3^{|S(\Sigma)|}} = 1 - \dfrac{|Mod_{S(\Sigma)}(\varphi) \cap Mod_{S(\Sigma)}(\psi)|}{3^{|S(\Sigma)|}} \geq 1 - \dfrac{|Mod_{S(\Sigma)}(\varphi)|}{3^{|S(\Sigma)|}}$.

4. $\mu_{S(\Sigma)}(\varphi \vee \psi) \leq \mu_{S(\Sigma)}(\varphi)$;

   а) $\dfrac{|Mod_{S(\Sigma)}(\sim (\varphi \vee \psi))|}{3^{|S(\Sigma)|}} = \dfrac{|Mod_{S(\Sigma)}(\sim \varphi \wedge \sim \psi)|}{3^{|S(\Sigma)|}} =$
   $= \dfrac{|Mod_{S(\Sigma)}(\sim \varphi) \cap Mod_{S(\Sigma)}(\sim \psi)|}{3^{|S(\Sigma)|}} \leq \dfrac{|Mod_{S(\Sigma)}(\sim \varphi)|}{3^{|S(\Sigma)|}}$;

   б) $1 - \dfrac{|Mod_{S(\Sigma)}(\varphi \vee \psi)|}{3^{|S(\Sigma)|}} = 1 - \dfrac{|Mod_{S(\Sigma)}(\varphi) \cup Mod_{S(\Sigma)}(\psi)|}{3^{|S(\Sigma)|}} \leq 1 - \dfrac{|Mod_{S(\Sigma)}(\varphi)|}{3^{|S(\Sigma)|}}$.

5. Если $\rho_{S(\Sigma)}(\varphi, \psi) = 1$, то $\mu_{S(\Sigma)}(\varphi \wedge \psi) = 1$ и $\mu_{S(\Sigma)}(\varphi \vee \phi) = 0$; свойство верно, если φ и ψ не принимают значения $1/2$, т.е. $\varphi = \sim \psi$, поскольку они двузначные.

6. Если $\rho_{S(\Sigma)}(\varphi, \psi) = 0$, то $\mu_{S(\Sigma)}(\varphi \wedge \psi) = \mu_{S(\Sigma)}(\varphi)$ и $\mu_{S(\Sigma)}(\varphi \vee \psi) = \mu_{S(\Sigma)}(\varphi)$; очевидно, так как из $\rho_{S(\Sigma)}(\varphi, \psi) = 0$ следует, что $\varphi \equiv \psi$.

7. $\mu_{S(\Sigma)}(\varphi \wedge \psi) = \rho_{S(\Sigma)}(\varphi, \psi) + \mu_{S(\Sigma)}(\varphi \vee \psi)$;
   $\dfrac{|Mod_{S(\Sigma)}(\sim (\varphi \wedge \psi))|}{3^{|S(\Sigma)|}} = \dfrac{|Mod_{S(\Sigma)}((\sim \varphi \wedge \psi) \vee (\varphi \wedge \sim \psi))|}{3^{|S(\Sigma)|}} + \dfrac{|Mod_{S(\Sigma)}(\sim (\varphi \vee \psi))|}{3^{|S(\Sigma)|}}$;

$$\frac{|Mod_{S(\Sigma)}(\sim(\varphi\wedge\psi))|}{3^{|S(\Sigma)|}}-\frac{|Mod_{S(\Sigma)}(\sim(\varphi\vee\psi))|}{3^{|S(\Sigma)|}}=\frac{|Mod_{S(\Sigma)}((\sim\varphi\wedge\psi)\vee(\varphi\wedge\sim\psi))|}{3^{|S(\Sigma)|}};$$

слева из числа моделей, на которых истинна φ, или ψ или обе ложны, вычитается число моделей, на которых и φ и ψ ложны. Получается число моделей, на которых истинна или φ, или ψ, но не $\varphi\wedge\psi$, а это и есть симметрическая разность, записанная справа.

8. $\min\{\mu_{S(\Sigma)}(\varphi),\mu_{S(\Sigma)}(\psi)\}\geq\mu_{S(\Sigma)}(\varphi\vee\psi)\geq\max\{\mu_{S(\Sigma)}(\varphi),\mu_{S(\Sigma)}(\psi)\}-\rho_{S(\Sigma)}(\varphi,\psi)$ и $\min\left\{\mu_{S(\Sigma)}(\varphi),\mu_{S(\Sigma)}(\psi)\right\}+\rho_{S(\Sigma)}(\varphi,\psi)\geq\mu_{S(\Sigma)}(\varphi\wedge\psi)\geq\max\left\{\mu_{S(\Sigma)}(\varphi),\mu_{S(\Sigma)}(\psi)\right\}$;

   а) $\mu_{S(\Sigma)}(\varphi\vee\psi)=\dfrac{|Mod_{S(\Sigma)}(\sim\varphi)\cap Mod_{S(\Sigma)}(\sim\psi)|}{3^{|S(\Sigma)|}}\leq\min\left\{\dfrac{|Mod_{S(\Sigma)}(\sim\varphi)|}{3^{|S(\Sigma)|}},\dfrac{|Mod_{S(\Sigma)}(\sim\psi)|}{3^{|S(\Sigma)|}}\right\}$ очевидно.

   $\mu_{S(\Sigma)}(\varphi\vee\psi)+\rho(\varphi,\psi)\geq\max\{\mu_{S(\Sigma)}(\varphi),\mu_{S(\Sigma)}(\psi)\}$, слева — все модели, кроме тех, где $\varphi\wedge\psi$ истинна, справа — все модели, кроме минимального из множеств $Mod_{S(\Sigma)}(\sim\varphi)$ и $Mod_{S(\Sigma)}(\psi)$.

   $\mu_{S(\Sigma)}(\varphi\wedge\psi)-\rho_{S(\Sigma)}(\varphi,\psi)$ — число моделей, на которых φ и ψ ложны. Это число не больше, чем $\min\{\mu_{S(\Sigma)}(\varphi),\mu_{S(\Sigma)}(\psi)\}$.

   $\mu_{S(\Sigma)}(\varphi\wedge\psi)\geq\max\{\mu_{S(\Sigma)}(\varphi),\mu_{S(\Sigma)}(\psi)\}$ очевидно;

   б) аналогично, с использованием свойств 3, 4, 7.

Пример кластеризации формул

Пусть база знаний экспертов состоит из трех элементарных высказываний $S(\Sigma)=\{A,B,C\}$ и есть упорядочение эксперта $P(A)=1; P(B)=2/3; P(C)=1/3$.

$\varphi_1=A\wedge B$;
$\varphi_2=A\vee B$;
$\varphi_3=A\to B$;
$\varphi_4=\sim A$;
$\varphi_5=B\to C$.

	φ_1	φ_2	φ_3	φ_4	φ_5
φ_1		0,31156	0,407573	0,67267	0,496536
φ_2	0,31156		0,365195	0,55638	0,27815
φ_3	0,407573	0,365195		0,25600	0,26497
φ_4	0,67267	0,55638	0,25600		0,36348
φ_5	0,496536	0,27815	0,26497	0,36348	

	φ_1	φ_2	φ_3	φ_4	φ_5
μ	5/9	1/9	1/9	3/9	1/9

Используем объединение кластеров по методу «ближайшего соседа».

Описание алгоритма. Применим иерархический алгоритм кластеризации к некоторой группе n высказываний. Сначала считаем, что у нас есть n кластеров. Построим матрицу расстояний для группы из n высказываний, 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 и будем повторять действия до тех пор, пока все высказывания не объединятся в один кластер. Кластеры будем соединять по методу «ближайшего соседа», т.е. $\rho(\varphi_k,\varphi_{ij})=\min\{\rho(\varphi_k,\varphi_i),\rho(\varphi_k,\varphi_j)\}$.

Шаг 1. $\min\rho(\varphi_i,\varphi_j)=0,25600=\rho(\varphi_3,\varphi_4)$. Кластеры: $\varphi_1,\varphi_2,\varphi_{34},\varphi_5$.

Шаг 2. $\min\rho(\varphi_i,\varphi_j)=0,26497=\rho(\varphi_{34},\varphi_5)$. Кластеры: $\varphi_1,\varphi_2,\varphi_{345}$.

Шаг 3. $\min \rho(\varphi_i, \varphi_j) = 0,27815 = \rho(\varphi_2, \varphi_{345})$. Кластеры: $\varphi_1, \varphi_{2345}$.

Шаг 4. $\min \rho(\varphi_i, \varphi_j) = 0,31156 = \rho(\varphi_1, \varphi_{2345})$. Кластеры: φ_{12345}.

В случае, если оптимальное число кластеров заранее не известно, в качестве критерия остановки алгоритма объединения можно взять меру информативности высказываний. Например, если перед началом кластеризации задать максимально допустимую разницу между мерами информативности элементов одного кластера, то алгоритм будет продолжаться до достижения этого значения.

На шаге 1 максимальная разница между мерами информативности одного кластера: $\max \|\mu(\varphi_i), \mu(\varphi_j)\| = 2/9$; на шаге 2: 2/9; на шаге 3: 2/9; на шаге 4: 4/9.

Таким образом, если мы зададим максимальное значение разности мер информативности одного кластера, равное 2/9, то алгоритм остановится после первого шага и результатом будут выступать кластеры $\varphi_1, \varphi_2, \varphi_{34}, \varphi_5$. Если зададим 3/9, то результат — $\varphi_1, \varphi_{2345}$. Если 4/9, то φ_{12345}.

Пусть нам надо разбить множество наших формул на определенное число кластеров, допустим на 2. Возьмем формулы, расстояние между которыми максимальное: в нашем случае это φ_1 и φ_4, $\rho(\varphi_1, \varphi_4) = 0,67267$. Логично будет поместить эти формулы в разные кластеры. Искать разбиение на 2 кластера будем так: для каждого элемента найдем ближайшего соседа из тех высказываний, которые уже приписаны к какому-нибудь кластеру, и припишем его к тому же кластеру:

Шаг 1. $\min \rho(\varphi_2, \varphi_i) = 0,31156 = \rho(\varphi_2, \varphi_1)$. Кластеры: φ_{12}, φ_4.

Шаг 2. $\min \rho(\varphi_3, \varphi_i) = 0,25600 = \rho(\varphi_3, \varphi_4)$. Кластеры: $\varphi_{12}, \varphi_{34}$.

Шаг 3. $\min \rho(\varphi_5, \varphi_i) = 0,26497 = \rho(\varphi_5, \varphi_3)$. Кластеры: $\varphi_{12}, \varphi_{345}$.

Получаем 2 класса $K_1 = \{\varphi_1, \varphi_2\}; K_2 = \{\varphi_3, \varphi_4, \varphi_5\}$.

Заключение

Результатом работы являются теорема 1, о метрике с учетом упорядочения элементарных высказываний каждым экспертом, теорема 2, о метрике, построенной с помощью степени разброса (даны примеры с подсчетом расстояний между формулами, свидетельствующие о новизне метрики), теорема 3 о построении новой (коллективной) метрики по уже имеющимся (введены меры информативности и недостоверности), теорема 4 о свойствах этих мер (приведены алгоритмы кластеризации многозначных формул при помощи введенных расстояний и примеры).

Работа выполнена при поддержке грантов РФФИ, проект 14–07–00851а, 14–7–00249а, а также кафедр ДМИ ММФ НГУ и АМЛ НГТУ.

Список литературы

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

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

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

4 *Викентьев А.А., Лбов Г.С.* Setting the metric and informativeness on statements of experts // Pattern Recognition and Image Analysis. — 1997. — Vol. 7. — No. 2. — P. 175–183.

5 *Daisuke Kachi.* Bourne on future contingents and three-valued logic // Logic and Logical Philosophy. — 2009. — Vol. 18. — P. 33–43.

А.А.Викентьев

Білім қорында автоматты кластерлеу мен Лукасевичтің көпмәнді формулалары үшін сараптамалық түсіндіру негізінде логикалық тұжырымдарда модельдік қашықтықтарды зерттеу. II

Мақалада Лукасевичтің көпмәнді формулалары үшін сараптамалық түсіндіру негізінде логикалық тұжырымдарда модельдік қашықтықтар зерттелген. Осы жұмыстың нәтижесі 2–4-теоремалар болып табылады. Енгізілген метриканың жаңалығын дәлелдейтін формулалар арасындағы қашықтықты есептеу мысалдары келтірілген.

A.A.Vikent'ev

Study of model distances on logical statements based on expert interpretation of the formulas of many-valued logics of Lukasiewicz and automatic clustering of knowledge bases. II

In article the model range on logical statements based on expert interpretation of the formulas of many-valued logics of Lukasiewicz. The result of Theorem 2–4. Examples with the calculation of the distances between the formulas showing novelty metrics.

References

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

2 Lbov G.S., Startseva N.G. *Logical decision functions and problems of statistical stability of solutions,* Novosibirsk: Publ. House of the Institute of Mathematics, 1999.

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

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

5 Daisuke Kachi. *Logic and Logical Philosophy*, 2009, 18, p. 33–43.