УДК 519.67–519.24

А.А.Викентьев, А.В.Макеев

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

Исследование модельных расстояний между высказываниями экспертов с упорядоченными элементарными логическими высказываниями базы знаний

Результат работы — доказательство теоремы о метрике с учётом упорядочения элементарных высказываний каждым экспертом и степени разброса моделей. Приведены примеры, свидетельствующие о новизне метрики, и способ построения новой метрики по ранее полученным или имеющимся метрикам.

Ключевые слова: метрика, высказывания, модельные расстояния, конъюнкция, дизъюнкция, импликация, эксперт.

Посвящаем памяти профессора Г.С.Лбова

Введение

При анализе знаний, заданных в виде высказываний экспертов, для различия содержащейся в них информации и группирования их по похожести, возникает необходимость введения расстояния между высказываниями экспертов [1–8].

Ранее на множестве моделей базы знаний было введено расстояние ρ между высказываниями экспертов, как нормированная мера (мощность) симметрической разности между моделями этих высказываний, и изучены его свойства. Хотелось бы ввести подобное расстояние, чтобы оно учитывало упорядоченность элементарных высказываний, которые численно упорядочил эксперт. Такой подход кажется естественным, потому что симметрическая разность недостаточным образом отражает, насколько отличны множества моделей друг от друга.

База знаний состоит из высказываний экспертов, записанных в виде формул исчисления высказываний. Это могут быть, например:

1) элементарные высказывания A = (сегодня холодно) и их отрицания $\neg A$;

2) конъюнкции элементарных высказываний $B = B_1 \& B_2$ = (сейчас зима) & (идет снег);

3) дизъюнкции $C = C_1 \vee C_2$ = (мороз)$\vee$ (снег).

Но чаще всего эксперты высказывают свои предложения в виде импликаций (продукций) элементарных высказываний.

4) $B_1 \to C_1 =$ если (сейчас зима), то (мороз). Причем вместе эти высказывания экспертов могут содержать противоречия, повторы и т.д.

Наша цель — на множестве неэквивалентных высказываний экспертов с упорядоченными элементарными высказываниями каждым экспертом и весами элементарных высказываний $p(x)$ ввести расстояние ρ^*, которое учитывает упорядочение каждого эксперта и степень разброса моделей изучаемой пары высказываний. Проверить, обладает ли оно известными свойствами расстояния ρ. Изучить вопрос о выполнимости свойств меры опровержимости (информативности в книге Г.С.Лбова).

Необходимые определения. Пусть дана конечная база знаний Σ, состоящая из формул исчисления высказываний (суждений экспертов).

Определение 1. Множество $S(\varphi)$ элементарных высказываний, используемых при написании формулы ИВ ϕ, назовём *носителем формулы* φ.

Определение 2. Назовём *носителем совокупности знаний* $S(\Sigma)$ объединение носителей формул, входящих в Σ, т.е. $S(\Sigma) = \bigcup_{\phi\in\Sigma} S(\varphi)$.

Рассмотрим множество $P(S(\Sigma)) = 2^{S(\Sigma)}$ всевозможных подмножеств S(Σ). Элементы множества $P(S(\Sigma))$ назовём *моделями*. Известно, что $|P(S(\Sigma))| = 2^{|S(\Sigma)|}$.

Определение 3. Элементарная формула A *истинна на модели* M (M ⊨A) $\Leftrightarrow$ A$\in$M, полагаем, что:

1) M ⊨φ&ψ $\Leftrightarrow$ (M ⊨ φ) и (M ⊨ ψ);
2) M ⊨φ∨ψ $\Leftrightarrow$ (M ⊨ φ) или (M ⊨ ψ);
3) M ⊨ ¬φ $\Leftrightarrow$ M ⊨ φ;
4) M ⊨ φ→ψ $\Leftrightarrow$ (из M ⊨ φ следуетM ⊨ ψ)или (M ⊨ ¬φ).

Обозначим через $\mathrm{Mod}_{S(\Sigma)}(A)$ множество моделей из S(Σ), на которых истинна A, т.е.

$$\mathrm{Mod}_{S(\Sigma)}(A)=\{M \in P(S(\Sigma)) \mid M \models A\}.$$

Свойства моделей:

а) $Mod_{S(\Sigma)}(A \,\&\, B) = Mod_{S(\Sigma)}(A) \cap Mod_{S(\Sigma)}(B)$;

б) $Mod_{S(\Sigma)}(A \vee B) = Mod_{S(\Sigma)}(A) \cup Mod_{S(\Sigma)}(B)$;

в) $Mod_{S(\Sigma)}(\neg A) = P(S(\Sigma)) \setminus Mod_{S(\Sigma)}(A)$.

Таким образом, любой формуле φ такой, что $S(\varphi) \subseteq S(\Sigma)$, соответствует совокупность $Mod_{S(\Sigma)}$ моделей из $P(S(\Sigma))$, на которых истинна φ.

Определение 4. Назовем формулу φ *эквивалентной* ψ (φ ≡ ψ), если $\mathrm{Mod}_{S(\Sigma)}(\varphi)=\mathrm{Mod}_{S(\Sigma)}(\psi)$. Это отношение является отношением эквивалентности.

Определение 5. Расстоянием $\rho_{S(\Sigma)}(\varphi,\psi)$ между φ и ψ при ($S(\varphi) \cap S(\psi) \subseteq S(\Sigma)$) на множестве $P(S(\Sigma))$ назовем величину

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

Очевидно, что

$$\rho_{S(\Sigma)} = \frac{|Mod_{S(\Sigma)}(\varphi)| + |Mod_{S(\Sigma)}(\psi)| - 2\,|Mod_{S(\Sigma)}(\varphi \,\&\, \psi)|}{2^{|S(\Sigma)|}}.$$

Определение 6. Мера информативности $\mu_{S(\Sigma)}(A)$ для формул из $\Phi(\Sigma) = \{\varphi \mid S(\varphi) \subset S(\Sigma)\}$ задаётся равенством $\mu_{S(\Sigma)}(A) = \rho_{S(\Sigma)}(A,1)$.

Формализация задачи и путь реализации

Пусть дано M — конечное число экспертов, и каждый из них независимо упорядочивает элементарные высказывания, входящие в базу знаний. Зададим упорядочения с помощью функций $p_i(x): P(S(\Sigma)) \to (0,1], i = 1,...,M$. Для построения расстояния, учитывающего мнения о «расстоянии» каждого эксперта, сначала введём расстояние для одного фиксированного эксперта. Для простоты обозначения пока опустим индекс. Расстояние будем строить поэтапно. Для начала введём расстояние между моделями с фиксированным упорядочением элементарных высказываний. Вторым этапом зададим расстояние между множествами моделей. Далее, пользуясь построенными расстояниями, введём расстояние между формулами ИВ с учётом расстояний для конечного множества экспертов.

Расстояния между моделями и его свойства

Для любых двух моделей A и B из множества всех конечных моделей $P(S(\Sigma))$ и фиксированного задания $p(x)$ экспертом определим расстояние между этими моделями A и B формулой

$$\rho(A,B) = \frac{\sqrt{\sum\limits_{x \in S(\Sigma)} p^2(x)(\chi_A(x) - \chi_B(x))^2}}{\sqrt{\sum\limits_{x \in S(\Sigma)} p^2(x)}}.$$

Утверждение 1. На классе $P(S(\Sigma))$ конечных моделей с фиксированным заданием $p(x)$ экспертом функция $\rho(A,B)$ удовлетворяет свойствам метрики, т.е. $< P(S(\Sigma)),\rho >$ является метрическим пространством.

Доказательство. 1. $\rho(A,B)=0 \Leftrightarrow A=B$. Пусть $\rho(A,B)=0$, докажем, что при этом $A=B$. $\rho(A,B)=0$, следовательно, либо $x\notin A, x\notin B$, либо $x\in A, x\in B$, значит, $A=B$. Теперь пусть $A=B$ докажем, что при этом $\rho(A,B)=0$. $A=B \Rightarrow \chi_A(x)=\chi_B(x) \Rightarrow \chi_A(x)-\chi_B(x)=0, \forall x\in S(\Sigma)$. Следовательно, $\rho(A,B)=0$.

2. $\rho(A,B)=\rho(B,A)$.

Доказательство вытекает из определения.

3. $\rho(A,B)\le \rho(A,C)+\rho(C,B)$.

Докажем от противного. Пусть $\rho(A,B)>\rho(A,C)+\rho(C,B)$, тогда

$$\frac{\sqrt{\sum\limits_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_B(x))^2}}{\sqrt{\sum\limits_{x\in S(\Sigma)} p^2(x)}} > \frac{\sqrt{\sum\limits_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2}}{\sqrt{\sum\limits_{x\in S(\Sigma)} p^2(x)}} + \frac{\sqrt{\sum\limits_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2}}{\sqrt{\sum\limits_{x\in S(\Sigma)} p^2(x)}};$$

$$\sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_B(x))^2} > \sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2} + \sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2}.$$

Возведём неравенство в квадрат:

$$\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_B(x))^2 > \sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2 +$$
$$+\sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2}\sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2} + \sum_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2.$$

$$\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_B(x))^2 - \sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2 - \sum_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2 >$$
$$> \sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2}\sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2}.$$

$$\sum_{x\in S(\Sigma)} p^2(x)\left((\chi_A(x)-\chi_B(x))^2-(\chi_A(x)-\chi_C(x))^2-(\chi_C(x)-\chi_B(x))^2\right) >$$
$$> \sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2}\sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2}.$$

Далее, возводя в квадрат и приводя подобные,

$$\sum_{x\in S(\Sigma)} p^2(x)(\chi_A^2(x)-2\chi_A(x)\chi_B(x)+\chi_B^2(x)-\chi_A^2(x)+2\chi_A(x)\chi_C(x)-\chi_C^2(x)-$$
$$-\chi_B^2(x)+2\chi_B(x)\chi_C(x)-\chi_C^2(x)) > \sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2}\sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2}.$$

$$\sum_{x\in S(\Sigma)} p^2(x)\left(-2\chi_A(x)\chi_B(x)+2\chi_A(x)\chi_C(x)-2\chi_C^2(x)+2\chi_B(x)\chi_C(x)\right) >$$
$$> \sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2}\sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2}.$$

$$\sum_{x\in S(\Sigma)} p^2(x)\left(2\chi_B(x)(-\chi_A(x)+\chi_C(x))-2\chi_C(x)(-\chi_A(x)+\chi_C(x))\right) >$$
$$> \sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2}\sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2},$$

получаем

$$\sum_{x\in S(\Sigma)} 2p^2(x)\big((-\chi_A(x)+\chi_C(x))(\chi_B(x)-\chi_C(x))\big) >$$

$$> \sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_A(x)-\chi_C(x))^2}\sqrt{\sum_{x\in S(\Sigma)} p^2(x)(\chi_C(x)-\chi_B(x))^2}. \quad (*)$$

Рассмотрим одно из слагаемых $(-\chi_A(x)+\chi_C(x))(\chi_B(x)-\chi_C(x))$ и проверим, когда оно положительно. Это возможно в двух случаях. Первый:

$$\begin{cases}-\chi_A(x)+\chi_C(x)>0\\ \chi_B(x)-\chi_C(x)>0\end{cases} \Leftrightarrow \begin{cases}\chi_A(x)<\chi_C(x)\\ \chi_B(x)>\chi_C(x)\end{cases} \Leftrightarrow \begin{cases}x\notin A\wedge x\in C;\\ x\in B\wedge x\notin C.\end{cases}$$

Противоречие. Рассмотрим второй случай:

$$\begin{cases}-\chi_A(x)+\chi_C(x)<0\\ \chi_B(x)-\chi_C(x)<0\end{cases} \Leftrightarrow \begin{cases}\chi_A(x)>\chi_C(x)\\ \chi_B(x)<\chi_C(x)\end{cases} \Leftrightarrow \begin{cases}x\in A\wedge x\notin C\\ x\notin B\wedge x\in C\end{cases}.$$

Противоречие. Неравенство $(-\chi_A(x)+\chi_C(x))(\chi_B(x)-\chi_C(x))>0$ невозможно, значит, $\forall x\in S(\Sigma)$ $(-\chi_A(x)+\chi_C(x))(\chi_B(x)-\chi_C(x))\le 0$, отсюда $\sum_{x\in S(\Sigma)} 2p^2(x)\big((-\chi_A(x)+\chi_C(x))(\chi_B(x)-\chi_C(x))\big)\le 0$, противоречие с $(*)$, так как справа стоит неотрицательная величина.

Расстояние между классами моделей

Далее перейдём к введению расстояния между множествами моделей для определения степени разброса моделей двух формул с учётом упорядочения и фиксированным заданием $p(x)$ экспертом.

Определение 7. Расстояние между элементом x метрического пространства $< P(S(\Sigma)),\rho >$ и множеством (моделью) B из этого метрического пространства с фиксированным заданием $p(x)$ (экспертом) зададим следующим образом: $d(x,B)=\min\{\rho(x,y)\mid y\in B\}$.

Определение 8. Введем в рассмотрение функцию от множеств A и B из метрического пространства $< P(S(\Sigma)),\rho >$ с фиксированным заданием $p(x)$ (экспертом), определяемую как

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

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

Определение 9. 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

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

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

Утверждение 2. Функция $\hat\rho_i$ является метрикой.

Доказательство. Для простоты обозначения опустим индекс *i*.

Для доказательства необходимо проверить выполнимость аксиом.

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

$\hat\rho(A,B)=0\Leftrightarrow A=B$. 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 $d(A,B)\ge 0$, то $d(A,B)=d(B,A)=0$. Вследствие свойств функции d получаем $A=B$.

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

$\hat\rho(A,C)\le\hat\rho(A,B)+\hat\rho(B,C)$. 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 $d(a,C)=\min\{\rho(x,c)\mid c\in C\}$. Для каждого $b\in B$:

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

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: $d(A,C) \le d(A,B) + d(B,C)$ и $d(C,A) \le d(B,A) + d(C,B)$, делим на 2. Получаем

$$\hat{\rho}(A,C) \le \hat{\rho}(A,B) + \hat{\rho}(B,C).$$

Расстояние между двумя формулами исчисления высказываний.

Определение 10. Расстояние между двумя формулами высказывания эксперта с фиксированным заданием $p_i(x)$ экспертом зададим формулой

$$\tilde{\rho}_i(\varphi,\psi) = \frac{\rho(\varphi,\psi) + \hat{\rho}_i(\mathrm{Mod}_{S(\Sigma)}(\varphi), \mathrm{Mod}_{S(\Sigma)}(\psi))}{2}.$$

Утверждение 3. На множестве классов эквивалентности высказываний (экспертов) с фиксированным заданием $p_i(x)$ (экспертом) $\tilde{\rho}_i$ определяет метрику.

Доказательство вытекает из проверяемого свойства, что линейная комбинация метрик является метрикой.

Далее рассмотрим несколько примеров и сформулируем основной результат работы.

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

Таблица 1

Формула ϕ	Формула ψ	Расстояние между ϕ,ψ без учёта упорядочения	Расстояние между ϕ,ψ с учётом упорядочения
$A \& B$	$A \to B$	1/2	$\frac{1}{4} + \frac{1}{4}\sqrt{\frac{45}{49}} \approx 0,489578712$
$A \& B$	$A \vee B$	1/2	$\frac{1}{4} + \frac{1}{4}\frac{6}{7} \approx 0,464285714$
$A \vee B$	$A \to B$	1/2	$\frac{1}{4} + \frac{1}{2}\frac{3}{7} \approx 0,464285714$
$A \to (B \& C)$	$A \to (B \vee C)$	1/4	$\frac{1}{8} + \frac{1}{4}\frac{6}{7} \approx 0,339285714$

Пример показывает, что введённое расстояние более тонко различает неэквивалентные формулы с помощью расстояния.

Теорема 1. 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:

$$\rho^*(\varphi,\psi) = \frac{\sum_{i=1}^{M} \tilde{\rho}_i(\varphi,\psi)}{M} = \frac{\rho(\varphi,\psi)}{2} + \frac{\sum_{i=1}^{M} \hat{\rho}_i(\mathrm{Mod}_{S(\Sigma)}(\varphi), \mathrm{Mod}_{S(\Sigma)}(\psi))}{2M},$$

а также с весами

$$\rho^{**}(\varphi,\psi) = \alpha\rho(\varphi,\psi) + \frac{\sum_{i=1}^{M} \gamma_i \hat{\rho}_i(\mathrm{Mod}_{S(\Sigma)}(\varphi), \mathrm{Mod}_{S(\Sigma)}(\psi))}{M}, \text{ где } \alpha + \sum_{i=1}^{M} \gamma_i = 1.$$

Доказательство аналогично доказательству утверждения 3.

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

Таблица 2

Формула ϕ	Формула ψ	Расстояние между ϕ, ψ без учёта упорядочения	Расстояние между ϕ, ψ с учётом упорядочения
$A\&B$	$A \to B$	1/2	$\frac{1}{4}+\frac{1}{4}\sqrt{\frac{45}{49}} \approx 0,489578712$
$A\&B$	$A \vee B$	1/2	$\frac{1}{4}+\frac{1}{4}\frac{6}{7} \approx 0,464285714$
$A \vee B$	$A \to B$	1/2	$\frac{1}{4}+\frac{1}{2}\frac{3}{7} \approx 0,464285714$
$A \to (B\&C)$	$A \to (B \vee C)$	1/4	$\frac{1}{8}+\frac{1}{4}\frac{6}{7} \approx 0,339285714$

Пример 2. Даны: база знаний экспертов, состоящая из трёх элементарных высказываний $P(S(\Sigma)) = \{A, B, C\}$, и два эксперта. Для первого возьмём упорядочение и задание из примера 1, для второго — упорядочение и задание $p_i(x)$ — $B < A < C$, $p(A) = 1/2, p(B) = 2/3, p(C) = 1/4$ (табл. 3).

Таблица 3

φ	ψ	$\rho(\varphi,\psi)$	$\tilde{\rho}_1(\varphi,\psi)$	$\tilde{\rho}_2(\varphi,\psi)$	$\rho^*(\varphi,\psi)$
$A\&B$	$A \to B$	1/2	$\frac{1}{4}+\frac{1}{4}\sqrt{\frac{45}{49}} \approx$ $\approx 0,489578712$	$\frac{1}{4}+\frac{1}{4}\sqrt{\frac{100}{109}} \approx$ $\approx 0,489456571$	$\frac{1}{4}+\frac{1}{8}\left(\sqrt{\frac{45}{49}}+\sqrt{\frac{100}{109}}\right) \approx$ $\approx 0,489517642$
$A\&B$	$A \vee B$	1/2	$\frac{1}{4}+\frac{1}{4}\frac{6}{7} \approx$ $\approx 0,464285714$	$\frac{1}{4}+\frac{1}{4}\sqrt{\frac{64}{109}} \approx$ $\approx 0,441565257$	$\frac{1}{4}+\frac{1}{8}\left(\frac{6}{7}+\sqrt{\frac{64}{109}}\right) \approx$ $\approx 0,452925486$
$A \vee B$	$A \to B$	1/2	$\frac{1}{4}+\frac{1}{2}\frac{3}{7} \approx$ $\approx 0,464285714$	$\frac{1}{4}+\frac{1}{2}\sqrt{\frac{36}{109}} \approx$ $\approx 0,537347886$	$\frac{1}{4}+\frac{1}{4}\left(\frac{3}{7}+\sqrt{\frac{36}{109}}\right) \approx$ $\approx 0,5008168$
$A \to (B\&C)$	$A \to (B \vee C)$	1/4	$\frac{1}{8}+\frac{1}{4}\frac{6}{7} \approx$ $\approx 0,339285714$	$\frac{1}{8}+\frac{1}{4}\sqrt{\frac{36}{109}} \approx$ $\approx 0,268673943$	$\frac{1}{8}+\frac{1}{8}\left(\frac{3}{7}+\sqrt{\frac{36}{109}}\right) \approx$ $\approx 0,303979829$

Кластеризация по расстоянию с учетом упорядочения примера 1. Объединение кластеров:

Шаг 1: $\min\limits_{i \neq j} \rho(\varphi_i, \varphi_j) = \rho(\varphi_2, \varphi_4)$ = 0,1339285. Кластеры: $\varphi_1, \varphi_{24}, \varphi_3, \varphi_5$. F = 1,267.

Шаг 2: $\min\limits_{i \neq j} \rho(\varphi_i, \varphi_j) = \rho(\varphi_5, \varphi_{24})$ = 0,1696428. Кластеры: $\varphi_1, \varphi_{245}, \varphi_3$. F = 1,073.

Шаг 3: $\min\limits_{i \neq j} \rho(\varphi_i, \varphi_j) = \rho(\varphi_3, \varphi_{245})$ = 0,3660714. Кластеры: $\varphi_1, \varphi_{2345}$. F = 0,732.

Шаг 4: Кластер: φ_{12345} .

Пусть выборка совпадает с выборкой из примера 1, поэтому обозначения оставим те же. Недостающие расстояния для второго эксперта $\tilde{\rho}_2(\varphi, \psi)$ таковы

$$\tilde{\rho}_2(\varphi_1, \varphi_4) = 0{,}6590448;$$

$$\tilde{\rho}_2(\varphi_1, \varphi_5) = 0{,}5519565;$$

$$\tilde{\rho}_2(\varphi_2, \varphi_4) = \tilde{\rho}_2(\varphi_2, \varphi_5) = 0{,}089272;$$

$$\tilde{\rho}_2(\varphi_3, \varphi_4) = 0{,}5998478;$$

$$\tilde{\rho}_2(\varphi_3, \varphi_5) = 0{,}3579459.$$

Расстояние с учетом упорядочения и задания расстояния между двумя экспертами $\rho^*(\varphi,\psi)$:

$$\rho^*(\varphi_1,\varphi_4) = 0{,}641276;$$
$$\rho^*(\varphi_1,\varphi_5) = 0{,}5520175;$$
$$\rho^*(\varphi_2,\varphi_4) = 0{,}1116002;$$
$$\rho^*(\varphi_2,\varphi_5) = 0{,}1294574;$$
$$\rho^*(\varphi_3,\varphi_4) = 0{,}6168881;$$
$$\rho^*(\varphi_3,\varphi_5) = 0{,}4472316.$$

Кластеризация по расстоянию $\tilde{\rho}_2(\varphi,\psi)$.

Объединение кластеров:

Шаг 1: $\min\limits_{i\neq j}\rho(\varphi_i,\varphi_j) = \rho(\varphi_2,\varphi_4) = \rho(\varphi_2,\varphi_5) = 0{,}089272$. Кластеры: $\varphi_1, \varphi_{245}, \varphi_3$. F = 1,782.

Шаг 2: $\min\limits_{i\neq j}\rho(\varphi_i,\varphi_j) = \rho(\varphi_3,\varphi_{245}) = 0{,}3579459$. Кластеры: $\varphi_1, \varphi_{2345}$. F = 0,917.

Шаг 3: Кластер: φ_{12345}.

Кластеризация по расстоянию $\rho^*(\varphi,\psi)$

Объединение кластеров:

Шаг 1: $\min\limits_{i\neq j}\rho(\varphi_i,\varphi_j) = \rho(\varphi_2,\varphi_4) = 0{,}1116002$. Кластеры: $\varphi_1, \varphi_{24}, \varphi_3, \varphi_5$. F = 1,674.

Шаг 2: $\min\limits_{i\neq j}\rho(\varphi_i,\varphi_j) = \rho(\varphi_5,\varphi_{24}) = 0{,}1294574$. Кластеры: $\varphi_1, \varphi_{245}, \varphi_3$. F = 1,49.

Шаг 3: $\min\limits_{i\neq j}\rho(\varphi_i,\varphi_j) = \rho(\varphi_3,\varphi_{245}) = 0{,}4472316$. Кластеры: $\varphi_1, \varphi_{2345}$. F = 0,887.

Шаг 4: Кластер: φ_{12345}.

Приведенные методы будут работать и в конечнозначных высказываниях.

Замечание 1. Расстояние ρ^{**} позволяет определить такое расстояние, которое удовлетворяет некоторым дополнительным свойствам, при разработке алгоритма (например, учёт степени важности каждого эксперта), которое, в свою очередь, обеспечивает адаптивное построение решающих функций и лучшую кластеризацию формул. Применение в распознавании экстренных событий для построения необходимых расстояний.

Замечание 2. Аналогично тому, как была введена мера информативности ранее, можно ввести новую меру информативности $\tilde{\mu}_{S(\Sigma)}(\varphi)$, как расстояния от искомой формулы до тождественно истинной формулы. При этом можно использовать введенные выше расстояния. Однако введённая таким образом мера информативности не будет обладать теми хорошими свойствами, которые имеет ранняя мера информативности (опровержимости) в [2,3,7].

Заключение

Главные результаты работы — теорема 1 о метрике с учётом упорядочения элементарных высказываний каждого эксперта и степени разброса моделей. Приведены примеры, свидетельствующие о новизне метрики, и способ построения новой метрики по ранее полученным или имеющимся нескольким метрикам. Планируется использование таких метрик для кластеризации формул из базы знаний для поиска наилучшей кластеризации.

Список литературы

1 *Блощицын В.Я., Лбов Г.С.* О мерах информативности логических высказываний // Докл. респ. школы-семинара «Технология разработки экспертных систем». — Кишинёв, 1987. — С. 12–14.

2 *Viken`tev 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.

3 *Викентьев А.А., Лбов Г.С.* О метризациях булевой алгебры предложений и информативности высказываний экспертов // Докл. РАН. — 1998. — Т. 361. — № 2. — С. 174–176.

4 *Загоруйко Н.Г.* Эмпирическое предсказание. — Новосибирск: Наука, 1979.

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

6 *Загоруйко Н. Г., Бушуев М.В.* Меры расстояний в пространстве знаний // Анализ данных в экспертных системах. — 1986. — Вып. 117. Вычислительные системы. — С. 24–35.

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

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

А.А.Викентьев, А.В.Макеев

Реттелген элементарлық логикалық тұжырымдарды оқу базасының модельдік қашықтықта сарапшылар тұжырымдары арасында зерттеу

Жұмыстың нәтижесі — реттелген элементарлық логикалық тұжырымды әрбір сарапшымен және модельдердің дәрежесі шашырауын ескере отырып, метрика туралы теореманы дәлелдеу. Метриканың жаңашылдығын көрсететін мысалдар және берілген метрика немесе бұрын алынған жаңа метриканы құрудың тәсілі келтірілген.

A.A.Vikent’ev, A.V.Makeev

Study model of distances between the statements of experts ordered elementary logical statements Knowledge Base

The result of — proof of the metric considering ordering elementary statements by each expert and the degree of dispersion models. Examples showing the novelty of the metric, and is given a new way to build metrics on the previously obtained or available metrics.

References

1 Bloshchitsyn V.Ya., Lbov G.S. Dokl. Rep. school-seminar «Technology development of expert systems.», Kishinev, 1987, p. 12–14.

2 Viken`tev A.A., Lbov G.S. *Pattern recognition and Image Analysis,* 1997, 7, 2, p. 175–183.

3 Vikent'ev A.A., Lbov G.S. *Dokl. Russian Academy of Sciences,* 1998, 361, 2, p. 174–176.

4 Zagoruiko N.G. *Empirical prediction,* Novosibirsk: Nauka, 1979.

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

6 Zagoruiko N.G., Bushuyev M.V. *Analysis of the data in expert systems,* Novosibirsk, 1986, 117. Computer systems, p. 24–35.

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

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