Серия «Математика»
2009. Т. 2, № 2, С. 20—39

Онлайн-доступ к журналу:
http://isu.ru/izvestia

И З В Е С Т И Я
Иркутского государственного университета

УДК 510.67-519.24

О метриках на формулах и мере опровержимости логических формул УИП с вероятностями на измеримых классах моделей *

А. А. Викентьев, Р. А. Викентьев
Институт математики СО РАН

Аннотация. Рассматриваются логические высказывания экспертов, представленные формулами исчисления высказываний с вероятностями, формулами языка первого порядка без вероятностей и с вероятностями. Предлагаются способы задания метрик на таких высказываниях, определения мер информативности (как опровержимости) и вероятности этих формул. Изучаются свойства введенных метрик и связанных с ними мер, приводятся примеры. Исследование позволит решать вопросы связанные с согласованием экспертных высказываний, построением решающих функций, распознавания образов, а также при создании логических баз знаний, их кластеризации и разработки экспертных систем.

Ключевые слова: распознавание образов, расстояние между формулами, метрики, теория моделей, базы знаний

Введение

К настоящему времени достаточно хорошо развиты теория и методы построения решающих функций распознавания образов на основе анализа эмпирической информации, представленной в виде таблиц данных. Наряду с этим проявляется все больший интерес к анализу экспертной информации, заданной в виде вероятностных логических высказываний нескольких экспертов. С помощью подходящей процедуры высказывания экспертов можно записать в виде формул исчисления высказываний или формул языка первого порядка. Ясно, что различные высказывания экспертов (и соответствующие им формулы) несут в себе, вообще говоря, разное количество информации.

Предполагается, что эксперты не согласовывали свои суждения друг с другом, поэтому их информация может содержать противоречия, дуб-

* Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проекты 07-01-00331а, 08-07-00136а)

лирование. Зачастую повторное обращение к экспертам с целью устранения противоречий оказывается невозможным, поэтому возникает задача согласования подобных высказываний без привлечения дополнительной информации. Проблемы возникают не только из-за противоречивых, но и из-за просто пересекающихся высказываний, даже если данные экспертами вероятности совпадают.

В работе [1] поставлена задача о введении меры информативности I на множестве классов эквивалентных формул – высказываний экспертов (в частности, некоторого фиксированного языка), и сформулированы естественные требования, которым должна удовлетворять эта функция, выражаемая через некоторую метрику ρ на множестве классов эквивалентных формул, значения которой принадлежат интервалу $[0,1]$. Все желаемые свойства меры информативности [1] мы сформулируем на формулах языка первого порядка, точнее на представителях их классов, где они лежат:
1) Если $\varphi \equiv 1$, то $I(\varphi) = 0$ (информативность тождественно истинной формулы равна нулю).
2) Если $\varphi \equiv \psi$, то $I(\varphi \wedge \psi) = I(\varphi)$.
3) Если $\rho(\varphi, \psi) = 1$, то $I(\varphi \wedge \psi) = I(\varphi) + I(\psi)$.
4) В случае $\rho(\varphi, \psi) \in (0,1)$, $I(\varphi \wedge \psi) = f(I(\varphi), I(\psi), \rho(\varphi, \psi))$, где $f(x, y, z)$ некоторая функция от переменных x, y, z удовлетворяющая следующим свойствам:
а) $f(x, y, 1) = x + y$; б) $f(x, x, 0) = x$;
в) $f(x, y, z_1) \leq f(x, y, z_2)$, если $z_1 \leq z_2$.
5) $I(\varphi) = 1 - I(\neg\varphi)$.

Таким образом, для решения поставленной задачи (и применений) необходимо решить две подзадачи:
1) ввести достаточно адекватную метрику на классах формул;
2) найти такую функцию $f(x, y, z)$, которая обладает требуемыми свойствами.

В работах [2, 3] эта задача решена для формул языка исчисления высказываний. Мера информативности здесь (и ранее) определена как мера опровержимости формулы на моделях. В случае непротиворечивости формулы (если для нее существует хотя бы одна модель, на которой она истинна (выполнима)) эта мера отражает интуитивное представление о количестве информации, содержащейся в высказывании эксперта, и ее ценности (важности) с точки зрения эксперта. В [4] задача решена для формул языка первого порядка с использованием конечного класса моделей.

В настоящей работе предлагается некоторый общий подход с использованием измеримого класса моделей и его расширений для вычисления естественных расстояний, мер информативности и установления свойств, удовлетворяющих требованиям 1 - 5. В дальнейшем всегда вместо понятия мера информативности будем использовать слова мера опровержимости, поскольку это более соответствует сути вводимой меры. Рассматривается также задача введения расстояний на множестве вероятностных формул. Для предложенных расстояний также справедливы хорошие свойства метрики (как в предыдущих случаях) и меры опровержимости.

Предложенные метрики и меры опровержимости аппробированы на отдельных примерах и на задаче по интеграционному проекту СО РАН «Компьютерная система для анализа антропологической информации» (рук. д.т.н., профессор Г. С. Лбов), которые кратко приводятся в работе.

Отметим, что проблемой введения расстояний в классе специальных формул и предикатов занимается Н. Г. Загоруйко [5, 6]. Его подход более проработан с точки зрения практических применений. Наш подход отличается некоторой общностью (произвольные предикаты и формулы) и теоретической проработкой исследований как свойств предлагаемых метрик, так и решением задачи о введении информативности (как меры опровержимости).

В решении поставленных задач играет важную роль измеримость числа моделей и используется логическая теория моделей, как теория алгебраических систем по академику А. И. Мальцеву, что позволяет изучать вопрос не только с точки зрения высказываний экспертов, но и «знаний» экспертов (гипотез), выраженных в виде интерпретаций сигнатурных предикатов (возможно и многоместных)в моделях.

Исследование позволит решать вопросы связанные с согласованием экспертных высказываний, построением решающих функций на основе различных метрик, распознавания образов, а также при создании логических баз знаний, их кластеризации и разработке экспертных систем [2, 5]. Результаты статьи прошли аппробацию на Всероссийской конференции ММРО-14, Суздаль 21-25 сентября 2009г., а так же и других.

1. Расстояния между формулами языка первого порядка

Пусть $\Omega = \{P_1^{m_1}, ..., P_t^{m_t}\}$ — фиксированная сигнатура, состоящая из конечного числа предикатных символов, которые выбираются для записи и изучения имеющихся связей между переменными в конкретной прикладной области. Случай исчисления высказываний описан в [2, 3]. Пусть задано некоторое исходное множество переменных $X = \{x_1, ..., x_p\}$. Обозначим через D_{x_j} непустое множество с конечной мерой возможных значений переменной x_j. Пусть $A_n = \bigcup_{j=1}^{p} D_{x_j}$ —непустое множество мощности n, являющееся объединением всех значений рассматриваемых переменных, включенных в предикаты.

В сигнатуру Ω для каждой переменной x_j дополнительно включен одноместный предикат P_{x_j}, выделяющий область значений переменной

x_j в A_n, то есть предикат $P_{x_j}(a)$ истинен в модели A_n на элементах $a \in A_n$ тогда и только тогда, когда $a \in D_{x_j}$.

Определение 1. [7]. Под интерпретацией будем понимать отображение γ, ставящее в соответствие каждому m_i-местному предикатному символу $P_i^{m_i}$ из сигнатуры Ω m_i-местный предикат (отношение) $P_i^{A_n} \subseteq A_n^{m_i}$, заданный на множестве A_n.

Это позволяет говорить о модели $< A_n, \Omega >$ сигнатуры Ω. В модели $< A_n, \Omega >$ предикат $P_i(x_1, ..., x_{m_i})$ истинен на элементах $a_1, ..., a_{m_i}$ из A_n (записывается $< A_n, \Omega > \models P_i(a_1, ..., a_{m_i})$) тогда и только тогда, когда $< a_1, ..., a_{m_i} > \in P_i^{A_n}$ и $a_j \in D_{x_j}$. Будем рассматривать модели только конечной сигнатуры.

Пусть имеется конечное число s экспертов и области возможных значений всех переменных. Модели (в смысле теории моделей) задаются специалистами согласно утверждениям экспертов. С каждым экспертом связывается модель, согласно его знаниям интерпретируются сигнатурные предикаты. Каждый эксперт «задает» свою интерпретацию каждого предикатного символа $P_i^{m_i}$ сигнатуры Ω соответствующим отношением (предикатом) на множестве A_n. В результате имеем множество моделей $\{M_j\}_{j=1}^{s}$ (исходный класс моделей).

«Знания» экспертов можно записать в виде формул (возможно с кванторами) в некоторой сигнатуре первого порядка (возможно и предложениями, то есть формулами с кванторами без свободных переменных). Предложения либо истинны, либо ложны на модели, а формулы с переменными определяют формульные предикаты (подмножества) в каждой модели M_j по заданной интерпретации j-го эксперта [2].

Пусть F — система подмножеств множества $\bigcup\limits_k A_n^k$, образующая σ - алгебру, где $A_n^k = \underbrace{A_n \times ... \times A_n}_{k}$. Нас будут интересовать только такие подмножества S_j из F, для которых найдется формула ψ_j, отражающая "знания"экспертов, которая и определяет это подмножество S_j (формульное подмножество). То есть S_j — это множество кортежей из $\bigcup\limits_k A_n^k$, на которых выполняется формула ψ_j. Формула ψ_j либо отражает какое-то из высказанных «знаний» экспертов, либо является их булевой комбинацией и навешиванием кванторов на некоторые переменные. В дальнейшем каждому рассматриваемому нами множеству S_j соответствует некоторая формула ψ_j.

Пусть B — замыкание множества Ω относительно логических операций $\neg$, $\wedge$, $\vee$, $\rightarrow$ и кванторов $\forall$ и $\exists$ по переменным. Ясно, что рассматриваемое экспертами множество формул в B содержится.

Определение 2. [8]. Вероятностной мерой μ на множестве B называется отображение $\mu : B \longrightarrow [0, 1]$, удовлетворяющее для φ и $\psi \in B$

условиям:
1) если $\vdash \varphi \equiv \psi$, то $\mu(\varphi) = \mu(\psi)$, ($\vdash \varphi$ означает, что на всех моделях φ истинна);
2) если $\vdash \neg(\varphi \wedge \psi)$, то $\mu(\varphi \vee \psi) = \mu(\varphi) + \mu(\psi)$;
3) если $\vdash \varphi$, то $\mu(\varphi) = 1$;
4) если $\vdash \neg\varphi$, то $\mu(\varphi) = 0$;
5) $\mu(\neg\varphi) = 1 - \mu(\varphi)$.

Заметим, что в каждом конкретном случае мера μ выбирается исходя из задачи.

Далее предполагаем, что на множестве формул B задана вероятностная мера μ. Тем самым, вероятностная мера μ задана на элементах множества F.

По заданным «знаниям» экспертов мы построили класс моделей (исходный класс). Чтобы более полно использовать информацию экспертов расширим исходный класс моделей. Учитывая одновременно информацию нескольких экспертов, можно уточнить каждую модель (тем самым получить еще одну или более новых моделей) и эти новые модели добавить в исходный класс. Уточнять модели будем следующим образом. Вместо «знания», заданного экспертом в виде предиката P_i в модели M_j, далее будем рассматривать его уточнение — предикат $\widetilde{P}_i$.

Под уточнением $\widetilde{P}_i$ предиката P_i понимаем уточнение интерпретации этого предиката в модели M_j одним из способов:
1) оставить предикат P_i без изменений;
2) исключить из P_i те элементы, в истинности которых j-ый эксперт не совсем уверен;
3) в предикат P_i добавить новые элементы и исключить некоторые старые, например, с учетом «знаний» других экспертов;
4) выполнить пункты 2) и 3) одновременно.

Обозначим произвольный расширенный класс моделей как $\mathrm{Mod}_n(\Omega)$. Введем расстояние на множестве «знаний» экспертов с помощью более полного класса моделей $\mathrm{Mod}_n(\Omega)$. Модели различаются интерпретациями сигнатурных предикатов, входящих в «знания» экспертов.

Определим расстояние между формульными подмножествами (предикатами) от одних и тех же переменных в каждой модели $M_i \in \mathrm{Mod}_n(\Omega)$, как отнормированную меру их симметрической разности. Предикат $P_i^{M_j}$ в новой добавленной к исходному классу модели M_j это предикат $\widetilde{P}_i^{M_j}$.

Определение 3. Расстоянием между предикатами (формульными подмножествами) $P_k^{M_i}$ и $P_j^{M_i}$ от одних и тех же переменных, определенными в модели M_i, назовем величину

$$\rho_{M_i}(P_k^{M_i}, P_j^{M_i}) = \mu(P_k^{M_i} \triangle P_j^{M_i}).$$

Замечание 1. Это определение можно расширить на формулы с различными наборами свободных переменных. В этом случае, если рассматриваемые предикаты имеют разную местность или разный набор переменных, и эксперт считает отсутствующую в одной из формул переменную x_i несущественной, полагаем, что она принимает любое из возможных значений и добавляем конъюнктивно к нужной формуле предикат P_{x_i}, выделяющий область значений этой переменной. В противном случае (если она существенна) доопределяем эту переменную, добавив конъюнктивно к нужной формуле предикат P'_{x_i}, уточняющий значения этой переменной после повторного обращения к эксперту [2]. Если расстояние измеряется между формулами, у которых наборы переменных различаются, например, по переменной x_i, причем в одной из формул переменная x_i находится под действием квантора, а в другую формулу x_i не входит, тогда относительно переменой x_i в таких формулах никаких изменений делать не надо. Если же расстояние измеряется между формулами, у которых наборы переменных различаются по переменной x_i, и в одной из формул переменная x_i находится под действием квантора, а в другой формуле x_i является свободной переменной, тогда доопределяем эту переменную, добавив конъюнктивно к нужной формуле предикат P_{x_i} или P'_{x_i}, уточняющий значения этой переменной, в зависимости от способа доопределения переменной. Например, заменяем формулу $\psi(x_j) = \exists x_i \varphi(x_i, x_j)$ на формулу $\psi(x_i, x_j) = \exists x_i(\varphi(x_i, x_j) \wedge P'_{x_i}) \wedge P'_{x_i}$. Аналогично рассматривается случай с квантором $\forall$.

В дальнейшем, с учетом выше сказанного, будем изучать расстояния между формулами от одних и тех же переменных (одной арности).

Расстояние между формулами, определенными на множестве моделей $\mathrm{Mod}_n(\Omega)$, определим как среднее на множестве расстояний в моделях.

Определение 4. Расстоянием между формулами P_k и P_j от одних и тех же переменных, определенными на множестве $\mathrm{Mod}_n(\Omega)$, назовем величину

$$\rho_1(P_k, P_j) = \frac{\sum\limits_{M_i \in \mathrm{Mod}_n(\Omega)} \rho_{M_i}(P_k^{M_i}, P_j^{M_i})}{\mid \mathrm{Mod}_n(\Omega) \mid}.$$

Теперь рассмотрим способ определения расстояния между предложениями (замкнутыми формулами). Обозначим через $\mathrm{Mod}(\varphi)$ множество моделей из $\mathrm{Mod}_n(\Omega)$, на которых истинно предложение φ, т.е. $\mathrm{Mod}(\varphi) = \{M_i \in \mathrm{Mod}_n(\Omega) \mid \; M_i \models \varphi\}$.

Очевидно, существуют такие модели, на которых предложение (эксперта) истинно, и такие, на которых оно ложно. Естественно с семантической точки зрения измерять различие информации, содержащейся в предложениях, количеством моделей, на которых предложения принимают разные значения истинности.

Определение 5. Расстоянием между предложениями φ и ψ от одних и тех же переменных назовем величину

$$\rho_2(\varphi,\psi) = \frac{\left|\mathrm{Mod}((\neg\varphi \wedge \psi) \vee (\varphi \wedge \neg\psi))\right|}{\mid \mathrm{Mod}_n(\Omega) \mid}.$$

Рассмотрим еще один способ определения расстояния между формулами. Дополним сигнатуру Ω константами из множества A_n. Для этого множества рассмотрим произвольные кортежи $\bar{a}$ длины местности формул, равной $\ell(\bar{a})$. При подстановке кортежей в формулы, в предположении, что формулы имеют одинаковую местность (как этого добиться было показано выше), формулы становятся предложениями.

Под эквивалентностью двух формул $\varphi(\bar{x})$ и $\psi(\bar{x})$ ($\varphi(\bar{x}) \equiv \psi(\bar{x})$) будем понимать эквивалентность их в классе моделей $\mathrm{Mod}_n(\Omega)$, то есть в каждой модели $M_i \in \mathrm{Mod}_n(\Omega)$ соответствующие $\varphi(\bar{x})$ и $\psi(\bar{x})$ формульные подмножества (множество реализаций формул) совпадают.

Тогда расстояние между неэквивалентными формулами φ и ψ определим как минимум расстояний, определенных по всем кортежам, которые реализуют формулу φ, но не реализуют формулу ψ и наоборот, используя определенное выше расстояние для предложений. Расстояние между эквивалентными формулами равно нулю.

Определение 6. Расстоянием между формулами φ и ψ от одних и тех же переменных назовем величину

$$\rho_3(\varphi(\bar{x}),\psi(\bar{x})) = \begin{cases} \min\limits_{\bar{a}\in\chi(A_n^{\ell(\bar{a})})} \rho_2(\varphi(\bar{a}),\psi(\bar{a})), & \text{если } \varphi(\bar{x}) \text{ неэквивалентна } \psi(\bar{x}) \\ 0, & \text{если } \ \varphi(\bar{x}) \equiv \psi(\bar{x}) \end{cases}$$

где $\chi(\bar{x}) = (\neg\varphi(\bar{x}) \wedge \psi(\bar{x})) \vee (\varphi(\bar{x}) \wedge \neg\psi(\bar{x}))$ и $\chi(A_n^{\ell(\bar{a})}) = \{\bar{b} \in A_n^{\ell(\bar{a})} \mid \exists M_i \in \mathrm{Mod}_n(\Omega) \ : \ M_i \models \chi(\bar{b})\}$

Ранее доказана теорема [4], из которой следует, что предложенные расстояния действительно являются метриками. В теореме 1 обобщается этот результат и приведены некоторые свойства введенных для формул расстояний.

Далее вместо $\varphi(\bar{x})$, $\psi(\bar{x})$, $\chi(\bar{x})$ для краткости будем писать φ, ψ, χ.

Теорема 1. *Для любых формул φ, ψ, χ от одних и тех же переменных и любого конечного расширения исходного класса моделей для любого ρ_i $(i = 1, 2, 3)$ выполняются следующие свойства:*
1. $0 \le \rho_i(\varphi,\psi) \le 1$.
2. $\rho_i(\varphi,\psi) = \rho_i(\psi,\varphi)$ *(симметричность).*

3. Если $\rho_i(\varphi,\psi) = \rho_i(\varphi_1,\psi_1)$ *и* $\rho_i(\varphi_1,\psi_1) = \rho_i(\varphi_2,\psi_2)$*, то* $\rho_i(\varphi,\psi) = \rho_i(\varphi_2,\psi_2)$.
4. $\rho_i(\varphi,\psi) \le \rho_i(\varphi,\chi) + \rho_i(\chi,\psi)$.
5. $\varphi \equiv \psi \iff \rho_i(\varphi,\psi) = 0$.
6. $\varphi \equiv \neg\psi \iff \rho_i(\varphi,\psi) = 1$.
7. $\rho_i(\varphi,\psi) = 1 - \rho_i(\varphi,\neg\psi) = \rho_i(\neg\varphi,\neg\psi)$.
8. $\rho_i(\varphi,\psi) = \rho_i(\varphi\wedge\psi,\varphi\vee\psi)$.
9. $\rho_i(\varphi,\neg\varphi) = \rho_i(\varphi,\psi) + \rho_i(\psi,\neg\varphi)$.

Доказательство теоремы следует из определений, свойств вероятностной меры, теоретико-модельных вычислений и аналогично доказательству из [4]. Для наглядности приведем доказательство пункта 4.

Для $i = 1$ достаточно показать, что $\mu(\varphi^{M_i}\triangle\psi^{M_i}) \le \mu(\varphi^{M_i}\triangle\chi^{M_i}) + \mu(\chi^{M_i}\triangle\psi^{M_i})$ для каждой модели M_i. Рассмотрим $\mu(\varphi^{M_i}\triangle\psi^{M_i}) = \mu(\varphi^{M_i}\setminus\psi^{M_i}) + \mu(\psi^{M_i}\setminus\varphi^{M_i}) = \mu(\varphi^{M_i}\wedge\neg\psi^{M_i}) + \mu(\psi^{M_i}\wedge\neg\varphi^{M_i}) =$
$= \mu(\varphi^{M_i}\wedge\chi^{M_i}\wedge\neg\psi^{M_i}) + \mu(\varphi^{M_i}\wedge\neg\chi^{M_i}\wedge\neg\psi^{M_i}) + \mu(\psi^{M_i}\wedge\neg\varphi^{M_i}\wedge\chi^{M_i}) + \mu(\psi^{M_i}\wedge\neg\varphi^{M_i}\wedge\neg\chi^{M_i}) \le \mu(\chi^{M_i}\wedge\neg\psi^{M_i}) + \mu(\varphi^{M_i}\wedge\neg\chi^{M_i}) + \mu(\neg\varphi^{M_i}\wedge\chi^{M_i}) + \mu(\psi^{M_i}\wedge\neg\chi^{M_i}) = \mu(\varphi^{M_i}\triangle\chi^{M_i}) + \mu(\chi^{M_i}\triangle\psi^{M_i})$, что и требовалось получить.

Докажем свойство 4 для $i = 2$. Обозначим через $S_1 = Mod((\neg\varphi\wedge\psi)\vee(\varphi\wedge\neg\psi))$, $S_2 = Mod((\neg\varphi\wedge\chi)\vee(\varphi\wedge\neg\chi))$, $S_3 = Mod((\neg\chi\wedge\psi)\vee(\chi\wedge\neg\psi))$. Докажем, что $|S_1| \le |S_2| + |S_3|$. Достаточно доказать, что $S_1 \subset S_2\cup S_3$. Возьмем произвольно $M_j \in S_1$, тогда $(M_j \models (\neg\varphi\wedge\psi))\vee(M_j \models (\varphi\wedge\neg\psi))$. Без ограничения общности (используя симметричность) будем считать, что $M_j \models (\neg\varphi\wedge\psi)$. Значит, $M_j \models \neg\varphi$ и $M_j \models \psi$. Для произвольного χ ясно, что $(M_j \models \chi)\vee(M_j \models \neg\chi)$. Если $M_j \models \chi$, то $M_j \models (\neg\varphi\wedge\chi)$, т. е. $M_j \in S_2$. Если $M_j \models \neg\chi$, то $M_j \models (\neg\chi\wedge\psi)$, т. е. $M_j \in S_3$. Получили, что $(M_j \in S_2)\vee(M_j \in S_3)$, т.е. $M_j \in S_2\cup S_3$. В силу произвольности M_j имеем $S_1 \subset S_2\cup S_3$. Поэтому, $|S_1| \le |S_2\cup S_3| = |S_2|+|S_3|-|S_2\cap S_3| \le |S_2|+|S_3|$, что и требовалось. Для $i = 3$ свойство 4 доказывается аналогично.

2. Меры опровержимости и вероятности формул

С точки зрения важности информации, сообщенной экспертом, которому мы доверяем, естественно считать, что опровержимость непустой формулы тем выше, чем меньше число удовлетворяющих ей элементов для конечного случая, и чем меньше мера задаваемого предикатной формулой подмножества в общем случае. Поэтому введем меру опровержимости следующим образом.

Определение 7. Мерой опровержимости формулы $\varphi(\bar{x})$ назовем величину

$$I_i(\varphi(\bar{x})) = \rho_i(\varphi(\bar{x}), 1),$$

где 1 — тождественно истинный предикат (например, $\bar{x} = \bar{x}$).

Для введенных расстояний ρ_i $(i = 1, 2, 3)$ получаем соответственно следующие меры опровержимости:

$$I_1(\varphi) = \frac{\sum\limits_{M_i \in \mathrm{Mod}_n(\Omega)} \mu(\neg\varphi^{M_i})}{|\mathrm{Mod}_n(\Omega)|},\ I_2(\varphi) = \frac{\left|\mathrm{Mod}(\neg\varphi)\right|}{|\mathrm{Mod}_n(\Omega)|},\ I_3(\varphi) = \frac{\min\limits_{\bar{a} \in \chi(A_n^{\ell(\bar{a})})} \left|\mathrm{Mod}(\neg\varphi(\bar{a}))\right|}{|\mathrm{Mod}_n(\Omega)|}.$$

Для мер опровержимости справедлива следующая

Теорема 2. *Для любых формул исчисления предикатов φ и ψ от одних и тех же переменных и любого конечного расширения исходного класса моделей справедливы для любого ρ_i $(i = 1, 2, 3)$ следующие свойства:*

1. $0 \leq I_i(\varphi) \leq 1$.
2. $I_i(1) = 0$.
3. $I_i(0) = 1$.
4. $I_i(\varphi) = 1 - I_i(\neg\varphi)$.
5. $I_i(\varphi) \leq I_i(\varphi \wedge \psi)$.
6. $I_i(\varphi) \geq I_i(\varphi \vee \psi)$.
7. $I_i(\varphi \wedge \psi) = \rho_i(\varphi, \psi) + I_i(\varphi \vee \psi)$.
8. Если $\varphi \equiv \psi$, *то* $I_i(\varphi) = I_i(\psi)$.
9. Если $\rho_i(\varphi, \psi) = 0$, *то* $I_i(\varphi \vee \psi) = I_i(\varphi \wedge \psi) = I_i(\varphi)$.
10. $I_i(\varphi \wedge \psi) = (I_i(\varphi) + I_i(\psi) + \rho_i(\varphi, \psi))/2$.
11. $I_i(\varphi \vee \psi) = (I_i(\varphi) + I_i(\psi) - \rho_i(\varphi, \psi))/2$.

Для доказательства теоремы используются введенные определения, свойства метрики из теоремы 1 и теоретико-модельные вычисления аналогичные доказательству из [4]. Приведем доказательство пункта 10.

Докажем для $i = 1$. $I_1(\varphi \wedge \psi) = 1 - I_1(\neg(\varphi \wedge \psi)) = 1 - I_1(\neg\varphi \vee \neg\psi) =$
$= 1 - \frac{\sum\limits_{M_i \in \mathrm{Mod}_n(\Omega)} \mu(\varphi^{M_i} \wedge \psi^{M_i})}{|\mathrm{Mod}_n(\Omega)|} = 1 + \frac{\sum\limits_{M_i \in \mathrm{Mod}_n(\Omega)} (\mu(\varphi^{M_i}) + \mu(\psi^{M_i}) - 2\mu(\varphi^{M_i} \wedge \psi^{M_i}))}{2|\mathrm{Mod}_n(\Omega)|} -$
$- \frac{\sum\limits_{M_i \in \mathrm{Mod}_n(\Omega)} (\mu(\varphi^{M_i}) + \mu(\psi^{M_i}))}{2|\mathrm{Mod}_n(\Omega)|} = 1 + \frac{\sum\limits_{M_i \in \mathrm{Mod}_n(\Omega)} \mu(\varphi^{M_i} \triangle \psi^{M_i})}{2|\mathrm{Mod}_n(\Omega)|} - \frac{I_1(\neg\varphi) + I_1(\neg\psi)}{2} = 1 +$
$+ \frac{1}{2}\rho_1(\varphi, \psi) - \frac{1}{2}(1 - I_1(\varphi) + 1 - I_1(\psi)) = \frac{1}{2}(I_1(\varphi) + I_1(\psi) + \rho_1(\varphi, \psi))$. Что и требовалось доказать. Для $i = 2, 3$ аналогично.

На практике же эксперт обычно задает высказывание с его «вероятностью». А вопрос состоит в изучении и согласовании таких высказываний, введении некоторой метрики на таких высказываниях. Первоочередной задачей, на наш взгляд, является определение вероятностей для формул с помощью модельного подхода.

Определение 8. Вероятностью формулы $\varphi(\bar{x})$ назовем величину

$$P_i(\varphi(\bar{x})) = \rho_i(\varphi(\bar{x}), 0).$$

Теорема 3. *Для любого конечного расширения исходного класса моделей для любых формул исчисления предикатов φ и ψ от одних и тех*

же переменных и для любого ρ_i $(i = 1, 2, 3)$ *справедливы следующие утверждения:*
1. $0 \le P_i(\varphi) \le 1$.
2. $P_i(1) = 1$.
3. $P_i(0) = 0$.
4. $P_i(\varphi) = 1 - P_i(\neg\varphi)$.
5. $P_i(\varphi) \ge P_i(\varphi \wedge \psi)$.
6. $P_i(\varphi) \le P_i(\varphi \vee \psi)$.
7. $P_i(\varphi \wedge \psi) = P_i(\varphi \vee \psi) - \rho_i(\varphi, \psi)$.
8. Если $\varphi \equiv \psi$, *то* $P_i(\varphi) = P_i(\psi)$.
9. Если $\rho_i(\varphi, \psi) = 0$, *то* $P_i(\varphi \vee \psi) = P_i(\varphi \wedge \psi) = P_i(\varphi)$.
10. $P_i(\varphi \wedge \psi) = (P_i(\varphi) + P_i(\psi) - \rho_i(\varphi, \psi))/2$.
11. $P_i(\varphi \vee \psi) = (P_i(\varphi) + P_i(\psi) + \rho_i(\varphi, \psi))/2$.

Для доказательства теоремы 3 последовательно используются результаты теорем 1 и 2 и то, что $P_i(\varphi(\bar{x})) = I_i(\neg\varphi(\bar{x}))$.

Так вычисленные с использованием модельного подхода вероятности позволяют уточнять «вероятности» экспертов и будут применяться в дальнейшем.

Разработан и реализован алгоритм для вычисления расстояний между формулами и мер опровержимости формул. Приведем один из тестовых примеров.

Пример 1. Пусть три эксперта высказываются о возрасте человека и его профессии, предсказывая наличие у него того или иного заболевания, например, радикулита и ухудшения зрения. То есть
$X = \{x_1(\text{профессия}), x_2(\text{возраст})\}$;
$D_{x_1} = \{\text{учитель, шофер, строитель, продавец}\}$;
$D_{x_2} = \{41, 42, 43, 44, 45\}$;
$\Omega = \{P_1 - \text{радикулит}, P_2 - \text{ухудшение \quad зрения}\}$.

Для краткости формулу $(x = a_1 \vee x = a_2 \vee x = a_3)$ будем записывать как $(x = a_1 \vee a_2 \vee a_3)$ и через $P(x, y) =$ и обозначать, что предикат истинен.

1-ый эксперт:
$P_1^1(x_1, x_2) = \text{и} \iff [(x_1 = \text{ш} \vee \text{стр} \vee \text{прод}) \wedge (x_2 = 42 \vee 43 \vee 44 \vee 45)]$
$P_2^1(x_1, x_2) = \text{и} \iff [(x_1 = \text{уч} \vee \text{ш} \vee \text{прод}) \wedge (x_2 = 41 \vee 42 \vee 43 \vee 44)]$

2-ой эксперт:
$P_1^2(x_1, x_2) = \text{и} \iff [(x_1 = \text{стр}) \wedge (x_2 = 44 \vee 45)]$
$P_2^2(x_1, x_2) = \text{и} \iff [(x_1 = \text{уч} \vee \text{стр} \vee \text{прод}) \wedge (x_2 = 43 \vee 44 \vee 45)]$

3-ий эксперт:
$P_1^3(x_1, x_2) = \text{и} \iff [(x_1 = \text{стр} \vee \text{прод})]$
$P_2^3(x_1, x_2) = \text{и} \iff [(x_1 = \text{уч} \vee \text{стр}) \wedge (x_2 = 41 \vee 43 \vee 45)]$

Вычислим расстояние между предикатами P_1 и P_2 и меры их опровержимости.

В модели M_1 первого эксперта имеем:
$\rho_{M_1}(P_1^1, P_2^1) = 0,6$; $I_{M_1}(P_1^1) = 0,4$; $I_{M_1}(P_2^1) = 0,4$.
В модели M_2 второго эксперта имеем:
$\rho_{M_2}(P_1^2, P_2^2) = 0,35$; $I_{M_2}(P_1^2) = 0,9$; $I_{M_2}(P_2^2) = 0,55$.
В модели M_3 третьего эксперта имеем:
$\rho_{M_3}(P_1^3, P_2^3) = 0,5$; $I_{M_3}(P_1^3) = 0,5$; $I_{M_3}(P_2^3) = 0,7$.

Тогда результирующее расстояние между предикатами P_1 и P_2 равно $\rho(P_1, P_2) = 0,4833$; а меры опровержимости предикатов соответственно равны $I(P_1) = 0,6$; $I(P_2) = 0,55$.

Можно вычислить расстояния между «знаниями» экспертов относительно одного и того же заболевания:
$\rho(P_1^1, P_1^2) = 0,5$; $\rho(P_1^1, P_1^3) = 0,3$; $\rho(P_1^2, P_1^3) = 0,4$;
$\rho(P_2^1, P_2^2) = 0,65$; $\rho(P_2^1, P_2^3) = 0,7$; $\rho(P_2^2, P_2^3) = 0,35$.

Приведем еще один пример.

Пример 2. ([6]). Два эксперта высказываются о влажности (X_1), которая измеряется четырьмя значениями, то есть $D_{X_1} = \{$нулевая, низкая, средняя, высокая$\}$, о количестве азотных удобрений (X_2), диапазон изменений которого от 0 до 320 кг/га, то есть $D_{X_2} = (0, 320)$ (X_2 - непрерывная переменная), о типе почвы (X_3), число которых 4, и видах культур-предшественников (X_4), число которых 8.

На вопрос, каковы условия для получения хорошего урожая пшеницы, эксперты дали такие ответы:
1: $(X_1 = \text{средняя}) \wedge (X_2 \in (160 - 200)) \wedge (X_3 = \text{чернозем})$,
2: $(X_1 = \text{средняя} \vee \text{высокая}) \wedge (X_2 \in (200 - 280)) \wedge (X_4 = \text{бобовые} \vee \text{травяные})$.

Тогда расстояние между высказываниями экспертов и мера опровержимости высказываний, вычисленные по введенным формулам с учетом замечания 1 следующие: $\rho_1(1, 2) = 0,039$, $I_1(1) = 0,992$, $I_1(2) = 0,969$.

Если не учитывать замечание 1, а вычислять расстояние только по общим переменным, то результат получается такой: $\rho_1^{'}(1, 2) = 0.156$.

Если вычислять расстояние между этими мнениями экспертов по формуле, предложенной Н. Г. Загоруйко в работах [5, 6], то получается $R = 0.18$.

3. Расстояние между вероятностными высказываниями экспертов

Как уже отмечалось, на практике эксперт обычно задает высказывание с его «вероятностью». В этом случае вопрос состоит в изучении и согласовании таких высказываний, введении некоторой метрики на таких высказываниях. Рассмотрим «знания» экспертов, представленные формулами исчисления высказываний с вероятностями (вероятностные высказывания), т. е. высказывания вида: «φ с вероятностью p_φ», где φ — формула исчисления высказываний. Используем следующую запись для таких высказываний: $B_i =< \varphi, p_\varphi >$, $B_j =< \psi, p_\psi >$.

Пусть Σ — база знаний, состоящая из формул исчисления высказываний (в Σ содержатся все те формулы, которые дали эксперты), $S(\varphi)$ — носитель формулы φ, т.е. множество элементарных высказываний, используемых при написании формулы φ, и $S(\Sigma) = \bigcup\limits_{\varphi\in\Sigma} S(\varphi)$ — носитель совокупности знаний.

Рассмотрим множество $P(S(\Sigma)) = 2^{S(\Sigma)}$ — множество всех подмножеств множества $S(\Sigma)$. Элементы множества $P(S(\Sigma))$ назовем моделями. Известно, что мощность множества $P(S(\Sigma))$ равна $2^{|S(\Sigma)|} = n$ (обозначим для простоты).

Предположим, что эксперты говорят о вероятностях формул на множестве всех моделей, и каждое высказывание присутствует только с одной вероятностью (случай, когда у высказывания не одна вероятность будет рассмотрен ниже).

Тогда интерпретируем вероятность, данную экспертом, следующим образом: $B =< \varphi, p_\varphi >$ означает, что высказывание φ истинно на $n_\varphi = \lfloor n \cdot p_\varphi \rfloor$ моделях, где $n = 2^{|S(\Sigma)|}$ — всего моделей.

Пусть даны два вероятностных логических высказывания $B_i =< \varphi, p_\varphi >$ и $B_j =< \psi, p_\psi >$, и требуется вычислить расстояние $\rho(B_i, B_j)$ между такими высказываниями.

Тогда, интерпретируя данные экспертами вероятности описанным выше способом, получаем, что высказывание φ истинно на $n_\varphi = \lfloor n \cdot p_\varphi \rfloor$ моделях, а высказывание ψ истинно на $n_\psi = \lfloor n \cdot p_\psi \rfloor$ моделях.

Отметим, однако, что неизвестно на каких именно моделях каждое высказывание истинно, а также число моделей, на которых эти высказывания истинны одновременно.

Рассмотрим следующую подзадачу. Пусть высказывание φ истинно на n_φ моделях, высказывание ψ истинно на n_ψ моделях и k — число моделей, на которых эти высказывания одновременно истинны. Требуется вычислить расстояние между высказываниями $B_i =< \varphi, p_\varphi >$ и $B_j =< \psi, p_\psi >$.

Возникающие в дальнейшем расстояния обозначим через $\rho_k(B_i, B_j)$, где $k = t, t+1, ..., \min(n_\varphi, n_\psi)$, здесь и далее $t = \max(0, n_\varphi + n_\psi - n)$. Заметим, что значение k для каждой пары высказываний свое.

Как и раньше [3–4], расстояние $\rho_k(B_i, B_j)$ определим как симметрическую разность, то есть

$$\rho_k(B_i, B_j) = \frac{1}{n}(n_\varphi - k + n_\psi - k) = \frac{1}{n}(n_\varphi + n_\psi - 2k), \qquad (1)$$

для каждого $k = t, t+1, ..., \min(n_\varphi, n_\psi)$.

Для расстояний $\rho_k(B_i, B_j)$ справедливы все утверждения теоремы 1. (Здесь $B_i \equiv B_j \iff \varphi \equiv \psi$ и $p_\varphi = p_\psi$, это означает, что формулы φ и ψ истинны на одних и тех же моделях.)

Предложим несколько способов вычисления расстояния $\rho(B_i, B_j)$ между вероятностными высказываниями $B_i =< \varphi, p_\varphi >$ и $B_j =< \psi, p_\psi >$.

Так как нам не известно число k (число моделей, на которых высказывания φ и ψ одновременно истинны), и если нет никаких предпочтений для значения k (предпочтение может быть высказано экспертами), то можем поступить следующим образом.

Предположим, что для нас все значения для числа k равновероятны. Тогда расстояние между вероятностными высказываниями $B_i =< \varphi, p_\varphi >$ и $B_j =< \psi, p_\psi >$ определим как усреднение расстояний $\rho_k(B_i, B_j)$ по всем значениям k, то есть

$$\rho_{\text{сред}}(B_i, B_j) = \frac{\sum\limits_{k=t}^{\min(n_\varphi, n_\psi)} \rho_k(B_i, B_j)}{\min(n_\varphi, n_\psi) + 1 - t}. \tag{2}$$

Для этого расстояния также справедлива теорема 1.

Если экспертами высказано, какое значение для k для пары B_i и B_j предпочтительнее, то в качестве $\rho(B_i, B_j)$ берется это $\rho_k(B_i, B_j)$, то есть

$$\rho(B_i, B_j) = \rho_k(B_i, B_j). \tag{3}$$

Подойдем к этому вопросу с вероятностной точки зрения: для каждого k построим вероятностную модель и вычислим вероятность того, что ровно на k моделях высказывания φ и ψ одновременно истинны.

Рассмотрим следующую задачу. Найти вероятность p_k того, что в выбранных n_φ и n_ψ моделях (выбираются из n моделей) будет ровно k моделей, на которых высказывания φ и ψ одновременно истинны, где $k = t, t+1, ..., \min(n_\varphi, n_\psi)$.

Сначала определим вероятностное пространство $< \Omega, A, p >$, где Ω — пространство элементарных исходов, $\Omega = \{t, t+1, ..., \min(n_\varphi, n_\psi)\}$ — множество всевозможных совпадений моделей в наборах из n_φ и n_ψ моделей, A — система подмножеств множества Ω, образующая σ-алгебру событий, и p — вероятность на $< \Omega, A >$. Определим на Ω такую случайную величину ξ что $\xi(k) = \rho_k(B_i, B_j)$.

Вероятность того, что ровно k моделей совпало в наборах из n_φ и n_ψ моделей вычисляется следующим образом:

$$p_k = \binom{n}{k}\binom{n-k}{n_\varphi - k}\binom{n-n_\varphi}{n_\psi - k}\Big/\left\{\binom{n}{n_\varphi}\binom{n}{n_\psi}\right\}.$$

В результате имеем $\rho_k(B_i, B_j)$ с вероятностями p_k, где $k = t, t+1, ...,$ $\min(n_\varphi, n_\psi)$.

Зная вероятности p_k для каждого расстояния $\rho_k(B_i, B_j)$, в качестве расстояния между вероятностными высказываниями $B_i =< \varphi, p_\varphi >$ и $B_j =< \psi, p_\psi >$ можно взять расстояние $\rho_k(B_i, B_j)$ с наибольшей вероятностью, то есть

$$\rho_{\max\ p}(B_i, B_j) = \rho_m(B_i, B_j), \text{ где } \quad p_m = \max_k p_k. \tag{4}$$

Очевидно, что для расстояния $\rho_{\max\ p}(B_i, B_j)$ справедлива теорема 1.

Можно в качестве расстояния между вероятностными высказываниями $B_i =< \varphi, p_\varphi >$ и $B_j =< \psi, p_\psi >$ взять величину, равную математическому ожиданию или среднему значению случайной величины ξ, то есть

$$\rho_{\text{ср}}(B_i, B_j) = M\xi = \sum_{k=t}^{\min(n_\varphi, n_\psi)} \rho_k(B_i, B_j) \cdot p_k. \tag{5}$$

Для этого расстояния $\rho_{\text{ср}}(B_i, B_j)$ справедливы все утверждения теоремы 1.

В предложенных формулах (1)—(5) расстояний не учитывается вид формул, между которыми вычисляется расстояние. Поэтому естественно предложить расстояние, которое учитывает и сами формулы.

Применяя модельный подход [3–4] к элементам множества $P(S(\Sigma))$ (моделям), найдем вероятности $P(\varphi)$ и $P(\psi)$ и расстояние $\rho(\varphi, \psi)$ (модельные вероятности и расстояние). Затем по свойству, аналогичному свойству 10 теоремы 3, справедливому и для формул исчисления высказываний, вычислим $P(\varphi \wedge \psi) = \frac{1}{2}(P(\varphi) + P(\psi) - \rho(\varphi, \psi))$. Тогда можно найти число $k_0 = \lfloor n \cdot P(\varphi \wedge \psi) \rfloor$ — число моделей, на которых высказывания φ и ψ одновременно истинны.

Имея число k_0 (вычисленное по моделям), можно уменьшить число возможных значений для k. Здесь возможны 3 случая:

1) если $t < k_0 < \min(n_\varphi, n_\psi)$, то $k = k_0 - 1, k_0, k_0 + 1$;

2) если $k_0 = t$ или $k_0 = \min(n_\varphi, n_\psi)$, то, например, $k = k_0, k_0 + 1$ или $k = k_0 - 1, k_0$ соответственно;

3) если $k_0 > \min(n_\varphi, n_\psi)$ или $k_0 < t$, то $k = \min(n_\varphi, n_\psi)$ или $k = t$ соответственно.

И уже к этим значениям для числа k применять предложенные выше формулы (1)—(5). По необходимости возможно некоторое расширение числа значений для k.

Теперь рассмотрим случай, когда одно и то же высказывание присутствует с разными вероятностями. В этом случае предлагаем два варианта введения расстояния:

1) Пусть имеется s экспертов, которые указали вероятности p_1, p_2,..., p_s высказывания φ. Тогда в качестве вероятности p_φ можно взять $p_\varphi = \frac{1}{s} \sum_{i=1}^{s} p_i$. Проделав такую процедуру для каждого высказывания, полу-

чим, что каждое высказывание присутствует только с одной вероятностью. Далее следует уменьшить число возможных значений для k и применять формулы (1)—(5).

2) Рассмотрим r-того эксперта. Имеем высказывание φ с вероятностью p_r и высказывание ψ с вероятностью q_r, данные r-тым экспертом. Тогда вычислим расстояние $\rho(B_i, B_j)$ для r-ого эксперта как в случае высказываний с одной вероятностью. Это расстояние обозначим через $\rho^r(B_i, B_j)$. Такую процедуру проделаем с каждым из имеющихся s экспертов, а за $\rho(B_i, B_j)$ примем $\rho(B_i, B_j) = \frac{1}{s} \sum\limits_{r=1}^{s} \rho^r(B_i, B_j)$.

Замечание 2. Предложенными способами можно также вычислять расстояние между следующими высказываниями: $B_i =< \varphi, p_{\varphi_i} >$ — информация, полученная от i-го эксперта, и $B_j =< \varphi, p_{\varphi_j} >$ — информация, полученная от j-го эксперта.

Результаты предыдущих рассуждений переносятся на формулы исчисления предикатов с вероятностями и произвольного конечного класса конечных моделей фиксированной сигнатуры. Остановимся на существенных моментах, отличающих этот случай от предыдущего.

Рассмотрим «знания» экспертов, представленные формулами исчисления предикатов с вероятностями, т. е. высказывания вида: «$\varphi(\bar{x})$ с вероятностью p_φ», где $\varphi(\bar{x})$ — формула исчисления предикатов. Используем следующую запись для таких высказываний: $B_i =< \varphi(\bar{x}), p_\varphi >$, $B_j =< \psi(\bar{x}), p_\psi >$.

Напомним, что $\Omega = \{P_1^{m_1}, ..., P_t^{m_t}\}$ — сигнатура, $X = \{x_1, ..., x_p\}$ — множество переменных, $D_{x_1}, ..., D_{x_p}$ — конечные множества возможных значений переменных, s — число экспертов. Каждый эксперт «задает» свою интерпретацию каждого предикатного символа сигнатуры Ω соответствующим отношением с приписанной ему вероятностью, т. е. «знания» i-го эксперта могут быть записаны в виде: $< P_1^i(\bar{x}), p_1^i >, ..., < P_t^i(\bar{x}), p_t^i >$. $\{M_i\}_{i=1}^s$ — исходное множество моделей. Высказывания экспертов могут быть представлены и формулами исчисления предикатов данной сигнатуры.

Зададим для простоты расстояние между предикатами P_l и P_j. Для этого сначала вычислим расстояние между вероятностными интерпретациями $B_l^i =< P_l^i(\bar{x}), p_l^i >$ и $B_j^i =< P_j^i(\bar{x}), p_j^i >$ предикатов в каждой модели M_i. Расстояния вычисляются между предикатами одинаковой местности и от одних и тех же переменных (см. замечание 1).

«Знание» $B_l^i =< P_l^i(x_1, ..., x_p), p_l^i >$ означает, что предикат $P_l^i(x_1, ..., x_p)$ истинен на $n_{P_l^i} = \lfloor n \cdot p_l^i \rfloor$ кортежах длины p в модели M_i, где $n = \prod\limits_{j=1}^{p} | D_{x_j} |$.

Аналогично случаю исчисления высказываний пусть предикат $P_l^i(\bar{x})$ истинен на $n_{P_l^i}$ кортежах в модели M_i, предикат $P_j^i(\bar{x})$ истинен на $n_{P_j^i}$ кортежах в модели M_i и k^i — число кортежей, на которых эти предикаты одновременно истинны, где $k^i = t, t+1, ..., \min(n_{P_l^i}, n_{P_j^i})$, $t = \max(0, n_{P_l^i} + n_{P_j^i} - n)$. Значение k^i для каждой пары предикатов свое. Тогда для каждого k^i расстояние $\rho_{k^i}(B_l^i, B_j^i)$ зададим формулой

$$\rho_{k^i}(B_l^i, B_j^i) = \frac{1}{n}(n_{P_l^i} + n_{P_j^i} - 2k^i).$$

Для расстояний $\rho_{k^i}(B_l^i, B_j^i)$ справедливы все утверждения теоремы 1.

Применяя модельный подход [3–4] и теорему 3, вычислим вероятности $P_{M_i}(P_l^i)$, $P_{M_i}(P_j^i)$ и расстояние $\rho_{M_i}(P_l^i, P_j^i)$ в модели M_i, затем вычислим вероятность $P_{M_i}(P_l^i \wedge P_j^i)$ и найдем число $k_0^i = \lfloor P_{M_i}(P_l^i \wedge P_j^i) \cdot n \rfloor$ — число кортежей, на которых предикаты одновременно истинны, вычисленное по моделям.

Далее поступаем как в случае исчисления высказываний: в каждой модели M_i вычислим расстояния $\rho^i(B_l^i, B_j^i)$ с учетом вероятностей и уменьшенного числа значений для k^i, и в качестве $\rho_{\text{вер}}(P_l, P_j)$ возьмем величину

$$\rho_{\text{вер}}(P_l, P_j) = \frac{1}{s}\sum_{i=1}^{s} \rho^i(B_l^i, B_j^i). \tag{6}$$

Для расстояния $\rho_{\text{вер}}(P_l, P_j)$ также справедлива теорема 1.

Таким же образом вводится расстояние между двумя формулами языка первого порядка с вероятностями.

Мера опровержимости формул с вероятностями (в случае исчисления высказываний и в случае языка первого порядка) вводится так же, как в пункте 2, то есть

$$I_{\text{вер}}(\varphi) = \rho(\varphi, 1),$$

где 1 — тождественно истинная формула, а ρ — одно из введенных по формулам (2)—(6) расстояний на вероятностных высказываниях.

Для так введенных мер опровержимости справедливы все свойства теоремы 2, доказанные нами для формул исчисления высказываний и формул исчисления предикатов с использованием исходного конечного класса моделей.

Приведем один из тестовых примеров для формул исчисления предикатов.

Пример 3. Рассмотрим высказывания экспертов, приведенные в примере 1, но с приписанной экспертами вероятностью (вероятностные высказывания).

1-ый эксперт:
$P_1^1(x_1, x_2) =$ и $\iff$ $[(x_1 =$ ш $\vee$ стр $\vee$ прод$) \wedge (x_2 = 42 \vee 43 \vee 44 \vee 45)]$, вероятность $p_1^1 = 0,5$.
$P_2^1(x_1, x_2) =$ и $\iff$ $[(x_1 =$ уч $\vee$ ш $\vee$ прод$) \wedge (x_2 = 41 \vee 42 \vee 43 \vee 44)]$, вероятность $p_2^1 = 0,7$.

2-ой эксперт:
$P_1^2(x_1, x_2) =$ и $\iff$ $[(x_1 =$ стр$) \wedge (x_2 = 44 \vee 45)]$, вероятность $p_1^2 = 0,8$.
$P_2^2(x_1, x_2) =$ и $\iff$ $[(x_1 =$ уч $\vee$ стр $\vee$ прод$) \wedge (x_2 = 43 \vee 44 \vee 45)]$, вероятность $p_2^2 = 0,6$.

3-ий эксперт:
$P_1^3(x_1, x_2) =$ и $\iff$ $[(x_1 =$ стр $\vee$ прод$)]$, вероятность $p_1^3 = 0,6$.
$P_2^3(x_1, x_2) =$ и $\iff$ $[(x_1 =$ уч $\vee$ стр$) \wedge (x_2 = 41 \vee 43 \vee 45)]$, вероятность $p_2^3 = 0,7$.

Вычислим расстояние между вероятностными высказываниями экспертов, выраженными предикатами P_1 и P_2, и меры их опровержимости.

В модели M_1 первого эксперта имеем:
$n_{P_1^1} = 10$; $n_{P_2^1} = 14$; $k^1 = 4, ..., 10$; $P^1(P_1^1 \wedge P_2^1) = 0,3$; тогда $k_0^1 = 6$ и, следовательно, $k^1 = 5, 6, 7$; $\rho_5^1(P_1^1, P_2^1) = 0,7$; $\rho_6^1(P_1^1, P_2^1) = 0,6$; $\rho_7^1(P_1^1, P_2^1) = 0,5$; тогда расстояние $\rho^1(P_1^1, P_2^1) = 0,6$ и меры опровержимости $I^1(P_1^1) = 0,5$ и $I^1(P_2^1) = 0,3$.

В модели M_2 второго эксперта имеем::
$n_{P_1^2} = 16$; $n_{P_2^2} = 12$; $k^2 = 8, ..., 12$; $P^2(P_1^2 \wedge P_2^2) = 0,1$; тогда $k_0^2 = 2$ и, следовательно, $k^2 = 8$; $\rho_8^2(P_1^2, P_2^2) = 0,6$; тогда расстояние $\rho^2(P_1^2, P_2^2) = 0,6$ и меры опровержимости $I^2(P_1^2) = 0,2$ и $I^2(P_2^2) = 0,4$.

В модели M_3 третьего эксперта имеем:
$n_{P_1^3} = 12$; $n_{P_2^3} = 14$; $k^3 = 6, ..., 12$; $P^3(P_1^3 \wedge P_2^3) = 0,15$; тогда $k_0^3 = 3$ и, следовательно, $k^3 = 6$; $\rho_6^3(P_1^3, P_2^3) = 0,7$; тогда расстояние $\rho^3(P_1^3, P_2^3) = 0,7$ и меры опровержимости $I^3(P_1^3) = 0,4$ и $I^3(P_2^3) = 0,3$.

Усредняя, получаем расстояние $\rho_{\text{вер}}(P_1, P_2) = 0,6333$ и меры опровержимости $I_{\text{вер}}(P_1) = 0,3667$ и $I_{\text{вер}}(P_2) = 0,3333$.

Пример 4. Под руководством Гл. Н. С. Института математики, профессора Г.С. Лбова решалась задача из области археологии в рамках интеграционного проекта "Компьютерная система для анализа антропологической информации". В.Б. Бериковым и Т.И. Лучшевой независимо былиполучены закономерности, характеризующие антропологические промеры черепов той или иной определенной эпохи. В анализируемый антропометрический комплекс входили более 20 непрерывных переменных, представляющие собой важнейшие характеристики, дифференцирующие основные расы и их локальные варианты. Множество рассматриваемых объектов (черепов) было разбито на 24 образа, отвечающих той или иной культуре.

Для хорошей кластеризации (таксономии) множества объектов и хорошего распознавания образов желательно, чтобы расстояние между своими представителями каждого кластера были малыми, а расстояния до представителей других кластеров по возможности большими.

Для примера рассмотрим полученные закономерности для двух кластеров.

Закономерности для 1-го образа (кластера):
1. ($X1$ в $[165; 188.50]$) и ($X3$ в $[123.50; 134.05]$) и ($X4$ в $[93.50; 103.25]$) и ($X8$ в $[68.94; 86]$) и($X11$ в $[24.95; 32]$) и ($X14$ в $[19.95; 25.20]$) и ($X15$ в $[2.04; 3.84]$) и ($X20$ в $[83.80; 89.15]$).
2. ($X3$ в $[123.50; 134.05]$) и ($X8$ в $[68.94; 86]$) и ($X10$ в $[28.75; 36.25]$) и ($X11$ в $[24.95; 32]$) и($X15$ в $[2.04; 3.84]$) и ($X16$ в $[5.14; 9.14]$) и ($X18$ в $[140.89; 150.44]$) и ($X20$ в $[83.80; 89.15]$).
3. ($X3$ в $[123.50; 134.05]$) и ($X8$ в $[68.94; 86]$) и ($X11$ в $[24.95; 32]$) и ($X13$ в $[8.94; 12.14]$) и($X15$ в $[2.04; 3.84]$) и ($X16$ в $[5.14; 9.14]$) и ($X19$ в $[124.84; 143.14]$) и ($X20$ в $[83.80; 89.15]$).

Закономерности для 2-го кластера:
4. ($X1$ в $[174.50; 188.50]$) и ($X3$ в $[125.50; 139.50]$) и ($X12$ в $[53.90; 58.09]$) и ($X14$ в $[19.45; 23.34]$)и ($X17$ в $[78.75; 92.50]$) и ($X20$ в $[84.90; 90.15]$) и ($X21$ в $[20.89; 34]$).
5. ($X9$ в $[40.84; 46.04]$) и ($X12$ в $[53.90; 58.09]$) и ($X14$ в $[19.45; 23.34]$) и ($X17$ в $[78.75; 92.50]$) и($X18$ в $[137.10; 153.85]$) и ($X19$ в $[127.90; 146.39]$) и ($X20$ в $[84.90; 90.15]$) и ($X21$ в $[20.89; 34]$).
6. ($X7$ в $[125.50; 148.50]$) и ($X8$ в $[69.90; 82.34]$) и ($X9$ в $[40.84; 46.04]$) и ($X12$ в $[53.90; 58.09]$) и($X14$ в $[19.45; 23.34]$) и ($X17$ в $[78.75; 92.50]$) и ($X20$ в $[84.90; 90.15]$) и ($X21$ в $[20.89; 34]$).

Расстояния между закономерностями 1-го кластера, вычисленные по предложенным нами формулам следующие:
$\rho(1, 2) = 0,0006643$; $\rho(1, 3) = 0,0007205$; $\rho(2, 3) = 0,0006942$.

Расстояния между закономерностями 2-го кластера соответственно следующие:
$\rho(4, 5) = 0,0006458$; $\rho(4, 6) = 0,0006411$; $\rho(5, 6) = 0,0007238$.

И расстояния между закономерностями из разных кластеров:
$\rho(1, 4) = 0,0095527$; $\rho(1, 5) = 0,0381027$; $\rho(1, 6) = 0,0270584$;
$\rho(2, 4) = 0,0644439$; $\rho(2, 5) = 0,0810485$; $\rho(2, 6) = 0,0825482$;
$\rho(3, 4) = 0,0645001$; $\rho(3, 5) = 0,0971838$; $\rho(3, 6) = 0,0826044$.

Если вычислить меру информативности (опровержимости) каждой закономерности, то получаем следующие результаты:
$I(1) = 0,999655$; $I(2) = 0,999681$; $I(3) = 0,999625$; $I(4) = 0,999718$; $I(5) = 0,999636$; $I(6) = 0,99964$.

Оценим диаметр каждого кластера через среднее расстояние между всеми парами представителей одного кластера. Получаем, что $diam_1 = 0,007207$ и $diam_2 = 0,007239$. И оценим расстояние между кластерами (разнесенность кластеров) через среднее расстояние между всеми парами представителей из разных кластеров. Получаем, что $\rho_{1,2} = 0,0606527$.

Аналогичные результаты характерны и для всех остальных кластеров и закономерностей.

Таким образом, введенное в работе расстояние позволяет посмотреть на закономерности с точки зрения близости по метрике. Результаты говорят о том, что эти закономерности обладают высокой информативностью, и полученные кластеры образуют «компактные» сгустки [5], что говорит о хорошей кластеризации закономерностей.

Заключение

Рассмотрены логические высказывания экспертов, представленные формулами исчисления высказываний и формулами языка первого порядка. Предложен некоторый общий подход с использованием конечного класса моделей и его расширений для вычисления естественных расстояний, мер информативности, вероятностей формул и установления свойств, которым должны удовлетворять расстояния и мера информативности. Рассмотрена также задача введения расстояний на множестве вероятностных формул. Для предложенных расстояний справедливы хорошие свойства метрики и меры опровержимости.

Разработан и реализован алгоритм для вычисления расстояний между формулами, мер опровержимости и вероятностей формул с учетом меры разброса в метрических моделях. Теоремы распространяются на измеримые классы метрических моделей.

Исследование позволит решать вопросы связанные с согласованием экспертных высказываний, построением адаптивных решающих функций, распознаванием образов, а также при создании логических баз знаний, их кластеризации и разработки экспертных систем.

Авторы благодарят профессора Г. С. Лбова за постоянный интерес к вопросам этого направления.

Список литературы

1. Блощицын В.Я., Лбов Г.С. О мерах информативности логических высказываний. // Доклады Республиканской Школы-Семинара "Технология разработки экспертных систем". Кишинев, 1978. С.12-14.
2. Лбов Г.С., Старцева Н.Г. Логические решающие функции и вопросы статистической устойчивости решений. Новосибирск: Изд-во Института математики, 1999.
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, N. 2. P. 175–189.
4. Vikent'ev A.A., Koreneva L.N. Setting the metric and measures of informativity in predicate formulas corresponding to the statements of experts about hierarchical objects// Pattern Recognition and Image Analysis. 2000. V. 10, N. 3. P.303–308.

5. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Института математики, 1999.
6. Загоруйко Н.Г., Бушуев М.В. Меры расстояния в пространстве знаний.// Анализ данных в экспертных системах. Новосибирск, 1986. Вып. 117: Вычислительные системы. С. 24-35.
7. Ершов Ю.Л., Палютин Е.А. Математическая логика. М.: Наука, 1991.
8. Gaifman H. Concerning measures in the first order calculi. // Israel Math. 1964. V. 2, N. 1. P. 1 -18.

A. A. Vikentiev, R. A. Vikentiev

A distance and the measure refutations on logical formulas (statements of experts with probability) in the measure class metric models

Abstract. In the paper we discuss statements of experts about objects represented as the formulas in language logics and we offer techniques for introducing metrics on such statements and measure of refutability. The research will find a use to problems of the best matching of the statements, of construction the decision functions of pattern recognition and development of expert's systems. We have studied the properties of entered metrics and the measures connected to. We also give examples. The work was also supported by the program «Mathematical logika» of Novosibirsk State University.
Keywords: pattern recognition, distance between statements, theory of models, metrics, know base

Александр Александрович Викентьев, ИМ СО РАН (пр. Академика Коптюга, д.4.), Ст.н.с. лаборатории Анализа Данных, канд. физ-мат. наук, Новосибирский госуд. университет, доцент, (vikent@math.nsc.ru)

Руслан Александрович Викентьев, Институт математики СО РАН, инженер, Лаборатория Анализа данных. Новосибирский госуд. университет; Преподаватель, (ruslan.vikentiev@gmail.com)

A.A. Vikentiev, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences, Laboratory of Data Analysis, Senior Researcher, Ph.D. in Physics and Mathematics; Novosibirsk State University, professor, e-mail: vikent@math.nsc.ru

R.A. Vikentiev, Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences, Laboratory of Data Analysis, Engineer; Novosibirsk State University, Assistant,
e-mail: ruslan.vikentiev@gmail.com