



УДК 004.85, 519.68

MSC 68T27, 68Q87

DOI <https://doi.org/10.26516/1997-7670.2017.21.33>

СЕМАНТИЧЕСКИЙ ВЕРОЯТНОСТНЫЙ ВЫВОД ПРЕДСКАЗАНИЙ*

Е. Е. Витяев

Институт математики им. С. Л. Соболева

Аннотация. Предсказание является одним из важнейших понятий в науке. Предсказания, получаемые на основе вероятностных знаний, описываются индуктивно-статистическим выводом I-S-выводом (Inductive-Statistical inference). Однако такой вывод сталкивается с проблемой синтеза логики и вероятности, состоящей в том, что оценки вероятности высказываний резко падают в процессе логического вывода. Рассматриваемые в рамках Probabilistic Logic Programming процедуры вычисления оценок не решают проблему. С нашей точки зрения, предсказание нельзя соединять с логическим выводом. Логический вывод следует заменить на вычисление. В работе предлагается семантический подход к вычислению предсказаний, когда вывод рассматривается не как проверка истинности запроса на модели, а как поиск фактов в модели, предсказывающих интересующее нас высказывание с максимальной вероятностью. Для этого в работе определяется семантический вероятностный вывод, осуществляющий вычисление предсказаний. В процессе семантического вероятностного вывода оценки предсказаний строго возрастают. В работе доказывается, что получаемые семантическим вероятностным выводом оценки предсказаний заведомо не хуже оценок, получаемых в рамках Probabilistic Logic Programming логическим выводом с параллельным вычислением этих оценок.

Ключевые слова: предсказание, вероятностный вывод, семантика, логические программы, probabilistic logic programming.

1. Введение

Для создания баз знаний используются методы машинного обучения (Machine Learning) и экспертные знания. Эти знания используются

* Работа выполнена за счет гранта Российского научного фонда (проект № 17-11-01176)

затем для предсказания и принятия решений с помощью логического вывода. Логический вывод абстрагируются от возможной недостоверности знаний и осуществляют вывод, как будто мы имеем достоверные знания.

Вывод предсказаний описывается покрывающими моделями (Covering Law Models), состоящими в том, чтобы вывести факт как частный случай закона. Выделяют две модели предсказания:

- 1) Дедуктивно-номологическую модель (Deductive-Nomological (D-N)), основанную на фактах и дедуктивных законах;
- 2) Индуктивно-статистическую модель (Inductive-Statistical (I-S)), основанную на фактах и вероятностных законах.

Эти модели могут быть представлены следующей схемой,

L_1, \dots, L_m
C_1, \dots, C_n
G

в которой L_1, \dots, L_m — множество дедуктивных законов для D-N вывода предсказаний и вероятностных для I-S вывода предсказаний, C_1, \dots, C_n — множество фактов, G — предсказываемое высказывание.

Оценка степени достоверности выводимых в I-S выводе предсказаний вычисляется различными методами параллельно процессу логического вывода. Есть работы, в которых степень достоверности рассматривается как значение истинности утверждений, а процесс логического вывода обобщается до так называемых «количественных дедукций» [5; 6; 16]. В работах [9; 10; 13] описываются довольно богатые формальные системы, содержащие как частные случаи основные «количественные дедукции».

Рассмотрим, как ведут себя оценки степени достоверности решений в процессе логического вывода. Возможны два случая: 1) знания получены методами машинного обучения на вероятностных данных и 2) знания описываются и оцениваются экспертом:

- 1) Анализ изменения вероятностных оценок утверждений в процессе логического вывода показывает, что они могут значительно уменьшаться. Как следует из работ по вероятностной логике [7; 14; 8; 15], полученные оценки не могут быть улучшены. Даже если ограничиться использованием правил с условной вероятностью не меньшей чем $1 - \varepsilon$, как это делается в [3], то это все равно не избавляет нас от существенного уменьшения вероятности в процессе вывода и, кроме того, это не соответствует условиям реально возникающих задач.

- 2) В работах по «количественной дедукции» истинностное значение заключения правила определяется как функция минимума или наибольшей нижней границы (для значений истинности в решетке) значений истинности атомов посылки. Соответствует ли это экспертным оценкам правила? Как правило, не соответствует. В этом случае ситуация по существу такая же, как и в предыдущем вероятностном случае, только проявляется она не в вероятностных терминах, а в терминах зависимости оценок эксперта от контекста, целостности восприятия ситуаций, адекватных и неадекватных (ситуациям) знаний и т. д. Если, например, атомы посылки правила описывают ситуацию, которая с точки зрения эксперта невозможна, то эксперт либо вообще откажется дать оценку заключению правила, либо присвоит ему значение близкое к нулю, хотя это правило по правилам вероятностной логики может иметь отличное от нуля значение.

Таким образом, несмотря на разнообразие формальных систем, вычисляющих оценки утверждений, адекватное вычисление оценок предсказаний отсутствует. В чем причина?

Причина в том, что, обобщая значения истинности, не обобщается сам процесс логического вывода. Следует осознать тот факт, что оценки утверждений делаются экспертом не в соответствии и не параллельно правилам логического вывода.

Предсказанием запроса некоторой логической программой PR в D-N выводе назовем такой путь вычисления запроса, на котором достигается максимум оценки условной вероятности запроса относительно подставленных в процессе вычисления фактов. Оценки условных вероятностей можно вычислить по вероятностным характеристикам правил и фактов, используя вероятностную логику (см. оценки в п. 4). Оценки не ухудшаются, если в процессе вывода используются правила, имеющие условную вероятность равную единице, и могут значительно ухудшаться, если используются правила с условной вероятностью строго меньшей 1.

Рассмотрим процесс вычисления с точки зрения «семантического» подхода к программированию [17]. Идея семантического программирования состоит в том, чтобы процесс вычисления рассматривать как проверку истинности утверждений (включая возможное использование логического вывода) на некоторой модели (моделью могут быть данные, представленные некоторой многосортной системой; некоторая специальная модель теории или абстрактного типа данных предметной области и т. д.). При таком взгляде на процесс вычисления, процедуру логического вывода можно обобщить, рассматривая более разнообразные взаимоотношения высказываний и модели — рассмотреть процесс вычисления как, например, определение вероятности, подтвержденно-

сти, достоверности, статистической значимости и т.д. высказываний на модели или классе моделей (выборка из некоторой генеральной совокупности или множества возможных миров G). Для этого вводится вероятность на множестве возможных миров или выборках из генеральной совокупности и таким образом определяется вероятностная модель \mathcal{M} данных. Такой обобщенный вывод будем вслед за работой [17] называть семантическим. В работе семантический подход к созданию баз знаний рассматривается для случая логических программ в языке первого порядка с вероятностной мерой μ [7; 14; 8; 15] и соответствующей вероятностной моделью данных \mathcal{M} .

Целью предсказания некоторого запроса A , с семантической точки зрения, является нахождение таких вероятностных фактов (множеств фактов) в вероятностной модели данных \mathcal{M} , из которых запрос A следовал бы с максимальной условной вероятностью. Предсказание, получаемое логической программой, не удовлетворяет этой цели. Во-первых, вероятностные оценки запроса могут существенно снижаться в процессе вычисления, а, во-вторых, вычисление не всегда может приводить к фактам, дающим максимальную оценку условной вероятности запроса.

Для получения наилучших предсказаний для любого одноатомного запроса A в работе определяется семантический процесс вычисления – *семантический вероятностный вывод* (см. определение 8.1), в котором вычисление осуществляется путем движения вдоль «уточняющего» графа [18; 12]. В этом графе правила, начиная с A , уточняются, либо добавлением произвольного атома (или конъюнкции атомов) в посылку, либо применением подстановки. Выбор уточнения, удлиняющего соответствующую ветвь графа, определяется требованием увеличения условной вероятности, определяемой по вероятностной модели данных. Результатом вычисления является результирующая подстановка и достигнутая условная вероятность. На уточняющие правила в семантическом вероятностном выводе налагается (без ограничения общности) дополнительное требование – чтобы каждый атом в посылке был «существенным» для предсказания запроса A (удаление любого атома из посылки уменьшает условную вероятность запроса A). Такие правила называются вероятностными закономерностями (см. определение 6.4). Для семантического вероятностного вывода, таким образом, достаточно иметь множество всех возможных вероятностных закономерностей данной вероятностной модели данных \mathcal{M} , обозначаемое через $PR(\mathcal{M})$. Это множество закономерностей составляет индуктивно выведенную из вероятностной модели данных \mathcal{M} базу знаний. Отметим, что для семантического вероятностного вывода не нужны правила вывода. Процесс вычисления вполне определяется требованием увеличением оценки условной вероятности (определяемой вероятностной моделью данных \mathcal{M}).

База знаний $PR(\mathfrak{M})$ является в определенном смысле полным и минимальным множеством вероятностных знаний, обеспечивающим лобой семантический вероятностный вывод и максимальную оценку предсказаний, и таким образом, удовлетворяет поставленной цели — получению наилучших предсказаний.

I-S предсказание новых фактов по базе знаний $PR(\mathfrak{M})$ на некоторых данных $D(N)$, принадлежащих случайно выбранной из множества возможных миров G модели N , в соответствии с вероятностной моделью данных \mathfrak{M} , может осуществляться логической программой $PR(\mathfrak{M}, N) = P(\mathfrak{M}) \cup D(N)$, где $P(\mathfrak{M}) \subset PR(\mathfrak{M})$ — множество всех вероятностных закономерностей с непустой посылкой. В работе доказывается, что программа $PR(\mathfrak{M}, N)$ предсказывает лучше любой другой логической программы, имеющей те же факты $D(N)$ (теорема 9.1). Более того, предсказание любого атома A осуществляется “лучшим для предсказания атома A правилом” (определение 8.2) в один шаг, не считая подстановки фактов. Таким образом, база знаний $PR(\mathfrak{M})$, рассматриваемая как логическая программа, осуществляет I-S предсказание на одних и тех же фактах лучше чем D-N предсказание любой другой логической программой.

В заключении отметим, что множество $PR(\mathfrak{M}, N)$ не является слишком большим. Понятие вероятностной закономерности было использовано для разработки метода обнаружения закономерностей [11; 1]. Этот метод был реализован и успешно применялся для решения ряда практических задач. Опыт решения задач показал, что множество $PR(\mathfrak{M})$ практически может быть найдено даже на средних персональных компьютерах.

В настоящее время, наиболее перспективным применением метода обнаружения закономерностей на множестве возможных миров G являются локально-простые информационные системы, описанные в [2], когда рассматривается сложная сеть из локально-простых [2] моделей, входящих в G , описывающих, например, рабочие места, склады, кассовые аппараты и т. д. Тогда можно осуществлять прогнозирование поведения этих локально-простых моделей.

2. Вероятность. Эрбрановы модели. Вероятностная модель данных

Зафиксируем язык первого порядка L с равенством не более чем счетной сигнатуры $\Omega = \langle P_1, P_2, \dots; f_1, f_2, \dots; C_1, C_2, \dots \rangle$, $C = \{C_{k \in K}\}$, $K \neq \emptyset$. Обозначим через U множество всех *основных термов* (не содержащих свободных переменных), X — множество *переменных*, T — множество *термов*, F — множество *формул*, F_0 — множество *формул без кванторов*, S — множество *предложений* (формул без свободных

переменных), $\mathfrak{R} = F_0 \cap S$ – множество всех *основных предложений* сигнатуры Ω .

Следуя [8] определим вероятность μ на подмножестве $F \subset \mathfrak{R}$, $F \neq \emptyset$ предложений, замкнутом относительно логических операций $\&$, (равенство не строгое, для строгого равенства необходимы дополнительные аксиомы см. [8]).

Определение 2.1. [8] Вероятностью μ на подмножестве $F \subset \mathfrak{R}$ называется отображение $\mu : F \rightarrow [0, 1]$, удовлетворяющее условиям:

- 1) Если $\vdash \varphi$, то $\mu(\varphi) = 1$;
- 2) Если $\vdash \neg(\phi \& \varphi)$, то $\mu(\phi \vee \varphi) = \mu(\phi) + \mu(\varphi)$.

Следствие 2.1. [8]. Если $\vdash (\varphi \equiv \phi)$, то $\mu(\varphi) = \mu(\phi)$. Если $\vdash \neg\varphi$, то $\mu(\varphi) = 0$.

Вероятность μ является конечно-аддитивной мерой на подалгебре $\{\phi / \equiv | \phi \in F\}$ булевой алгебры Линденбаума – Тарского.

Определение 2.2. Вероятностной Эрбрановой моделью *сигнатуры* Ω будем называть пару $M = \langle U, \mu \rangle$, где μ – вероятность на \mathfrak{R} . Функциональные символы интерпретируются на U обычным образом [4].

Определение 2.3. Эрбрановой моделью *сигнатуры* будем называть вероятностную Эрбранову модель $M = \langle U, \mu \rangle$, где $\mu : \mathfrak{R} \rightarrow \{0, 1\}$.

Рассмотрим множество S всех Эрбрановых моделей $M = \langle U, \mu \rangle$ сигнатуры Ω . Пусть дан некоторый класс Эрбрановых моделей $G \subset S$ (множество возможных миров) и вероятность μ на некотором подмножестве $F \subset \mathfrak{R}$ формул замкнутом относительно логических операций. Определим булеву подалгебру D подмножеств $G(\varphi) = \{M | M \in G, M \models \varphi\}$, $\varphi \in \mathfrak{R}$ множества G , где \models выполнимость утверждения φ на модели M при означивании μ (где 1 – истина, 0 – ложь).

Определение 2.4. Класс Эрбрановых моделей G будем называть согласованным с вероятностью на множестве формул F , если из $G(\varphi) = \emptyset$, $\varphi \in F$ следует $\mu(\varphi) = 0$.

Лемма 2.1. Величина $\eta(G(\varphi)) = \mu(\varphi)$, $\varphi \in F$ является конечно-аддитивной мерой на подалгебре D , если класс Эрбрановых моделей G согласован с μ на множестве формул F .

Доказательство. Так как D – булева подалгебра подмножеств G является кольцом множеств, то достаточно доказать, что $\eta(G(\varphi_1) \cup G(\varphi_2)) = \eta(G(\varphi_1)) + \eta(G(\varphi_2))$, если $G(\varphi_1) \& G(\varphi_2) = \emptyset$; $\varphi_1, \varphi_2 \in F$. Так как $\eta(G(\varphi_1) \cup G(\varphi_2)) = \eta(G(\varphi_1 \vee \varphi_2)) = \mu(\varphi_1 \vee \varphi_2)$; $\eta(G(\varphi_1)) = \mu(\varphi_1)$; $\eta(G(\varphi_2)) = \mu(\varphi_2)$; $G(\varphi_1) \cap G(\varphi_2) = G(\varphi_1 \& \varphi_2)$, то нам достаточно доказать, что $\mu(\varphi_1 \vee \varphi_2) = \mu(\varphi_1) + \mu(\varphi_2)$, если $G(\varphi_1 \& \varphi_2) = \emptyset$. Из определения меры μ

следует, что $\mu(\varphi_1 \vee \varphi_2) = \mu(\varphi_1) + \mu(\varphi_2) - \mu(\varphi_1 \& \varphi_2)$. Из условия леммы и определения 2.4 следует, что если $G(\varphi_1 \& \varphi_2) = \emptyset$, то $\mu(\varphi_1 \& \varphi_2) = 0$. \square

Если множество формул F совпадает с \mathfrak{R} , то будем говорить, что класс Эрбрановых моделей G согласован с вероятностной Эрбрановой моделью $M = \langle U, \mu \rangle$, а модель M является *вероятностной моделью множества возможных миров* G или выборкой из некоторой генеральной совокупности.

3. Логические программы

Обозначим через PR множество всех *правил* $A \leftarrow A_1, \dots, A_k$, $k \geq 0$ сигнатуры, где A, A_1, \dots, A_k – атомы сигнатуры Ω . Если атом A отсутствует, то правило $\leftarrow A_1, \dots, A_k$ называется *целью (запросом)*. Если $k = 0$, то правило $A \leftarrow$ называется *фактом*.

Логическая программа Pr есть конечная совокупность правил. *Подстановкой* называется отображение $\theta: X \rightarrow T$. Подстановка $\theta(x) = x$ называется тождественной. Обозначим через Θ множество всех подстановок. Подстановки естественным образом распространяются на произвольные выражения: так для терма $t = f(t_1, \dots, t_n)$ и атома $A = P(t_1, \dots, t_n)$ их подстановки соответственно равны

$$t\theta = f(t_1\theta, \dots, t_n\theta), A\theta = P(t_1\theta, \dots, t_n\theta).$$

Правило $A\theta \leftarrow A_1\theta, \dots, A_n\theta$ называется вариантом правила $A \leftarrow A_1, \dots, A_n$ если θ – перестановка множества X .

Зафиксируем правило вычисления R , определяющее в каждом запросе выделенный атом. Пусть

$$N = \leftarrow A_1, \dots, A_i, \dots, A_k, k \geq 1$$

запрос, в котором правилом R выделен атом A_i и

$$A \leftarrow B_1, \dots, B_l$$

вариант некоторого правила программы Pr , в котором все переменные отличны от переменных запроса. Пусть θ – наиболее общий унификатор атомов A_i и A . Тогда запросы

$$\begin{aligned} \leftarrow (A_1, \dots, B_1, \dots, B_l, \dots, A_k)\theta, l \geq 1, \\ \leftarrow (A_1, \dots, \underline{A}_i, \dots, A_k)\theta, l = 0 \end{aligned} \tag{3.1}$$

будем называть выводимыми из запроса N по правилу $A \leftarrow B_1, \dots, B_l$ с помощью подстановки θ и правила вычисления R . Как видно из определения, атом A_i не удаляется из запроса при его унификации с некоторым фактом программы. Такие атомы выделяются подчеркиванием. Будем предполагать, что правило R не выбирает для очередного шага вывода подчеркнутые атомы.

Пространством вычислений для программы Pr и правила вычисления R называется множество всех возможных запросов сигнатуры Ω

с заданным на нем отношением выводимости. SLDF-выводом (Linear resolution with Selection rule for Definite clauses and underlined Facts) цели N в некотором пространстве вычислений, назовем максимальной последовательностью запросов $N = N_0, N_1, N_2, \dots$ вместе с последовательностью правил C_0, C_1, \dots и унификаторов $\theta_0, \theta_1, \dots$, такую что запросы N_{i+1} выводимы из запросов N_i по правилам C_i с помощью подстановок θ_i и правила вычисления R . *SLDF-вывод* – максимальный путь в пространстве вычислений, начинающийся с N . SLDF-вывод, заканчивающийся запросом, в котором все атомы подчеркнуты, называется *успешным*. Конечный SLDF-вывод, не являющийся успешным – *тупиковым*. Множество всех SLDF-выводов, начинающихся с цели N , обычно представляют в виде дерева (префикс дерева SLDF-выводов) и называют SLDF-деревом вычислений запроса N . SLDF-дерево, содержащее успешный SLDF-вывод, называется успешным.

4. Оценки вероятностей и условных вероятностей запросов

Пусть $M = \langle U, \mu \rangle$ вероятностная Эрбранова модель. Рассмотрим успешный SLDF-вывод N, N_1, \dots, N_k запроса N с помощью последовательности правил C_0, C_1, \dots, C_{k-1} некоторой программы Pr , последовательности унификаций $\theta_0, \theta_1, \dots, \theta_{k-1}$; $\theta = \theta_0\theta_1\dots\theta_{k-1}$ и некоторого правила вычислений R .

Последовательность запросов $N\theta, N_1\theta, \dots, N_k, \theta = \theta_0\theta_1\dots\theta_{k-1}$ также будет SLDF-выводом запроса $N\theta$ с помощью последовательности правил $C_0\theta, C_1\theta, \dots, C_{k-1}\theta$ тождественных подстановок и правила вычислений R . Будем предполагать, что $N\theta, N_1\theta, \dots, N_k \in \mathfrak{R}$. В данном пункте факты $A \leftarrow$ будем представлять правилами $A \leftarrow \text{true}$, тогда $\mu(C) = \mu(A/\text{true}) = \mu(A)$, для факта $C = A \leftarrow$, $A \in \mathfrak{R}$.

Определим через $\&N_i$ конъюнкцию всех не подчеркнутых атомов запроса N_i . Если все атомы подчеркнуты (как в запросе N_k), то положим $\&N_k \equiv \text{true}$. Обозначим через $\&N_i F$ (F - facts) конъюнкцию всех подчеркнутых атомов запроса N_i . Тогда $\&N_i F$ – конъюнкция всех фактов, использованных в SLDF-выводе запроса $N\theta$.

Цель данного пункта – оценить вероятности $\mu(\&N\theta)$, $\mu(\&N\theta/\&N_k F)$ по SLDF-выводу запроса $N\theta$, предполагая, что нам известны только вероятности фактов и правил.

Рассмотрим вывод запросов (3.1) из запроса $N\theta = (\leftarrow A_1, \dots, A_i, \dots, A_k)\theta$, $k \geq 1$ по правилу $(A_i \leftarrow B_1, \dots, B_l)\theta$. Представим запросы (3.1) в виде $N_1\theta = (\leftarrow A_1, \dots, A_{i-1}, B, A_{i+1}, \dots, A_k)\theta$, $B = B_1, \dots, B_l$ или $N_1\theta = (\leftarrow A_1, \dots, \underline{A_i}, \dots, A_k)\theta$. Второй запрос является частным случаем первого, когда $B = \text{true}$. Оценим вероятности $\mu(\&N\theta)$, $\mu(\&N\theta/\&N_1\theta)$ в предположении, что нам известны только вероятности $\mu(\&N_1\theta)$, $\mu(A_i\theta)$, $\mu(\&B\theta)$, $p = \mu(A_i\theta/\&B\theta)$.

Лемма 4.1. Если $\mu(\&N_1\theta) > 0$ и $\mu(\&B\theta) > 0$, то

- 1) $\mu(\&N\theta) \leq \mu(\&N_1\theta) + \mu(\neg \&B\theta)$;
- 2) $\mu(\&N\theta) \geq \mu(\&N_1\theta) - (1-p)\mu(\&B\theta)$;
- 3) $\mu(\&N\theta / \&N_1\theta) \leq p \cdot \mu(\&B\theta) / \mu(\&N_1\theta)$;
- 4) $\mu(\&N\theta / \&N_1\theta) \geq 1 - (1-p) / \mu(\&N\theta \setminus A_i\theta) / \&B\theta$.

Доказательство.

$$1. \mu(\&N\theta) = \mu(\&N\theta \wedge \&B\theta) + \mu(\&N\theta \wedge \neg \&B\theta) \leq \mu(\&N_1\theta \wedge A_i\theta) + \mu(\neg \&B\theta) \leq \mu(\&N_1\theta) + \mu(\neg \&B\theta).$$

$$2. \mu(\&N\theta / \&N_1\theta) = \mu(\&N\theta \wedge \&B\theta) / \mu(\&N_1\theta) \leq \mu(A_i\theta \wedge \&B\theta) / \mu(\&N_1\theta) = p \cdot \mu(\&B\theta) / \mu(\&N_1\theta).$$

3. $\mu(\&N\theta) \geq \mu(\&N\theta \wedge \&B\theta) \geq \mu(\&N\theta \wedge \&B\theta) - \mu(\neg \&N\theta \wedge \neg \&A\theta \wedge \&B\theta)$.
Выражение из правой части пункта 2 утверждения леммы равно этому же выражению:

$$\begin{aligned} \mu(\&N_1\theta) - (1-p)\mu(\&B\theta) &= \mu(\&N_1\theta) + \mu(A_i\theta \wedge \&B\theta) - \mu(\&B\theta) = \\ &= \mu(\&N_1\theta \wedge A_i\theta) + \mu(\&N_1\theta \wedge \neg A_i\theta) + \mu(A_i\theta \wedge \&B\theta \wedge \&N\theta) + \\ &+ \mu(A_i\theta \wedge \&B\theta \wedge \neg \&N\theta) - \mu(\&B\theta) = \\ &= \mu(\&B\theta \wedge \&N\theta \wedge A_i\theta) + (\mu(\&B\theta \wedge (\&N\theta \setminus A_i\theta) \wedge \neg A_i\theta) + \\ &+ \mu(\&B\theta \wedge A_i\theta \wedge \&N\theta) + \mu(\&B\theta \wedge A_i\theta \wedge \neg \&N\theta) - \mu(\&B\theta)) = \\ &= \mu(\&B\theta \wedge \&N\theta \wedge A_i\theta) - \mu(\&B\theta \wedge \neg \&N\theta \wedge \neg A_i\theta) = \\ &= \mu(\&N\theta \wedge \&B\theta) - \mu(\neg \&N\theta \wedge \neg A_i\theta \wedge \&B\theta). \end{aligned}$$

$$4. \mu(\&N\theta / \&N_1\theta) = \mu(\&N\theta \wedge \&B\theta) / \mu(\&N_1\theta) \geq (\mu(\&N_1\theta) - (1-p)\mu(\&B\theta)) / \mu(\&N_1\theta) \text{ (см. доказательство пункта 2).}$$

$$(\mu(\&N_1\theta) - (1-p)\mu(\&B\theta)) / \mu(\&N_1\theta) = 1 - (1-p) / \mu(\&N\theta \setminus A_i\theta) / \&B\theta,$$

где $\&N\theta \setminus A_i\theta = (\leftarrow A_1, \dots, A_{i-1}, A_{i+1}, \dots, A_k)\theta$ \square

Следствие 4.1. Если $\mu(\&N_1\theta) > 0$, $\mu(\&B\theta) > 0$ и $p = 1$, то:

- 1) $\mu(\&N_1\theta) \leq \mu(\&N\theta) \leq \mu(\&N_1\theta) + \mu(\neg \&B\theta)$.
- 2) $\mu(\&N\theta / \&N_1\theta) = 1$.

Следствие 4.2. Если $\mu(\&N_1\theta) > 0$ и правило $(A \leftarrow true)\theta$ является фактом, то:

- 1) $\mu(\&N_1\theta) + \mu(A\theta) - 1 \leq \mu(\&N\theta) \leq \mu(\&N_1\theta)$;
- 2) $1 - (1 - \mu(A\theta)) / \mu(\&N\theta \setminus A_i\theta) \leq \mu(\&N\theta / \&N_1\theta) \leq \mu(A\theta) / \mu(\&N_1\theta)$.

Доказательство. Следует из леммы 4.1 и равенства $p = \mu(A\theta)$ \square

Следствие 4.3. Если $\mu(\&B\theta) > 0$, то:

- 1) $\mu(\&N\theta \wedge \&B\theta) \leq \mu(\&N_1\theta) + \mu(\neg \&B\theta)$;

$$2) \mu(\&N\theta \wedge \&B\theta) \geq \mu(\&N_1\theta) - (1-p)\mu(\&B\theta).$$

Доказательство. Следует из доказательств пунктов 1, 2 леммы 4.1. \square

Рассмотрим SLDF-вывод $N\theta, N_1\theta, \dots, N_k$ запроса $N\theta$ посредством последовательности правил $C_i\theta = (A \leftarrow B^i, \dots, B^i)_{li}\theta$, $i = 0, 1, \dots, k-1$ и пустых унификаций. Положим $B^i\theta = (B^i_1 \wedge \dots \wedge B^i_{li})\theta$, $p_i = \mu(C_i\theta)$.

Теорема 4.1. *Если $\mu(B^i\theta) > 0$, $i = 0, 1, \dots, k-1$, то:*

$$\mu(\&N\theta \wedge A^0\theta \wedge A^1\theta \wedge \dots \wedge A^{k-1}\theta) \geq 1 - \sum_{i=0}^{k-1} (1-p_i)\mu(B^i\theta).$$

Доказательство. Используем оценку 2 следствия 4.3, примененную к последнему шагу вывода от $\&N_{k-1}\theta$ к N_k . Получим $\mu(\&N_{k-1}\theta \wedge B^{k-1}\theta) \geq \mu(\&N_k) - (1-p_{k-1})\mu(B^{k-1}\theta)$, где $\mu(\&N_k) = \mu(\text{true}) = 1$, так как все атомы выделены. Рассмотрим вывод запроса $\&N_{k-1}\theta \wedge B^{k-1}\theta$ из запроса $\&N_{k-2}\theta \wedge B^{k-1}\theta$ посредством правила $C_{k-2}\theta$. Снова применим оценку 2 Следствия 4.3. Получим: $\mu(\&N_{k-2}\theta \wedge B^{k-1}\theta \wedge B^{k-2}\theta) \geq \mu(\&N_{k-1}\theta \wedge B^{k-1}\theta) - (1-p_{k-2})\mu(B^{k-2}\theta)$. Рассмотрим вывод запроса $\&N_{k-2}\theta \wedge B^{k-1}\theta \wedge B^{k-2}\theta$ из запроса $\&N_{k-3}\theta \wedge B^{k-1}\theta \wedge B^{k-2}\theta$ посредством правила $C_{k-3}\theta$ и т.д. Получим: $\mu(\&N\theta \wedge B^0\theta \wedge B^1\theta \wedge \dots \wedge B^{k-1}\theta) \geq \mu(\&N_1\theta \wedge B^1\theta \wedge \dots \wedge B^{k-1}\theta) - (1-p_0)\mu(B^0\theta)$. Подставляя левые части неравенств в их правые части, получим оценку

$$\mu(\&N\theta \wedge B^0\theta \wedge B^1\theta \wedge \dots \wedge B^{k-1}\theta) \geq 1 - \sum_{i=0}^{k-1} (1-p_i)\mu(B^i\theta)$$

Покажем, что если из конъюнкции

$$\mu(\&N\theta \wedge B^0\theta \wedge B^1\theta \wedge \dots \wedge B^{k-1}\theta)$$

удалить все константы true, то получим конъюнкцию

$$\mu(\&N\theta \wedge A^0\theta \wedge A^1\theta \wedge \dots \wedge A^{k-1}\theta).$$

Заметим, что каждый атом конъюнкции $B^i\theta$ (true — не атом) в процессе вывода обязательно унифицируется с левой частью одного из правил. Следовательно, каждый атом конъюнкции $B^0\theta \wedge B^1\theta \wedge \dots \wedge B^{k-1}\theta$ содержится в конъюнкции $A^0\theta \wedge A^1\theta \wedge \dots \wedge A^{k-1}\theta$. С другой стороны, каждый атом $A^i\theta$, $i = 0, 1, \dots, k-1$ содержится либо в $\&N\theta$, либо в правой части одного из правил $C_i\theta$, $i = 0, 1, \dots, k-1$. \square

Следствие 4.4. *Если $\mu(B^i\theta) > 0$, $i = 0, 1, \dots, k-1$, то:*

$$\mu(\&N\theta) \geq 1 - \sum_{i=0}^{k-1} (1-p_i)\mu(B^i\theta).$$

Доказательство. Следует из $\mu(\&N\theta) > \mu(\&N\theta \wedge A^0\theta \wedge \dots \wedge A^{k-1}\theta)$ и теоремы 4.1. \square

Для каждого успешного SLDF-вывода $N\theta = N_0\theta, N_1\theta, \dots, N_k$ существует успешный SLDF'-вывод $N\theta = N_0\theta, N'_1\theta, \dots, N'_i\theta, \dots, N_k$, в котором факты применяются последними и до запроса $N'_i\theta$ применяются правила C_j с длиной $l_j \geq 1$; $j = 0, \dots, i-1$. Тогда запрос $N'_i\theta$ будет иметь вид $\leftarrow A_1, \dots, A_m$, а запрос N_k — вид $\leftarrow \mathbf{A}_1, \dots, \mathbf{A}_m$. Такой SLDF'-вывод будем называть *нормализованным*.

Теорема 4.2. Если $\mu(B^j\theta) > 0$, $j = 0, 1, \dots, i-1$, и $\mu(\&N_k F) > 0$, то $\mu(\&N\theta / \&N_k F) \geq 1 - \sum_{j=1}^{i-1} (1 - p_j)\mu(B^j\theta) / \mu(\&N_k F)$, где p_j — условные вероятности, а $B^j\theta$ — условия правил C_j , $j = 1, \dots, i-1$.

Доказательство. Проводится аналогично доказательству теоремы 4.1, но для нормализованного вывода и начинается с запроса i . Первое неравенство имеет вид: $\mu(\&N_{i-1}\theta \wedge B^{i-1}\theta) \geq \mu(\&N_i\theta) - (1 - p_{i-1})\mu(B^{i-1}\theta)$, где $\mu(\&N_i\theta) = \mu(\&N_k F)$. Далее, рассуждая как в теореме 4.1, получим неравенство

$$\mu(\&N\theta \wedge B^0\theta \wedge \dots \wedge B^{i-1}\theta) \geq \mu(\&N_k F) - \sum_{j=1}^{i-1} (1 - p_j)\mu(B^j\theta).$$

Так как $\mu(\&N\theta \wedge B^0\theta \wedge \dots \wedge B^{i-1}\theta) \leq \mu(\&N\theta \wedge \&N_k F)$, то

$$\mu(\&N\theta / \&N_k F) = \mu(\&N\theta \wedge \&N_k F) / \mu(\&N_k F) \geq (\mu(\&N_k F) - \sum_{j=1}^{i-1} (1 - p_j)\mu(B^j\theta)) / \mu(\&N_k F).$$

□

5. Вероятностные оценки запросов логических программ

5.1. Определим вероятностные оценки $\nu(N)$, $\eta(N)$ запросов для пространства вычислений программы P_T по правилу R . Рассмотрим SLDF-дерево некоторого запроса N пространства вычислений. Если SLDF-дерево не успешно, то оценки $\nu(N)$, $\eta(N)$ не определены. Для успешного SLDF-дерева рассмотрим множество $\{SLDF'_1, \dots, SLDF'_m\}$ всех успешных нормализованных SLDF'-выводов целей $N\theta_1, \dots, N\theta_m$ у которых конечные запросы $N^1_{k_1}, \dots, N^m_{k_m}$ не содержат переменных. Если это множество пусто, то оценки $\nu(N)$, $\eta(N)$ не определены.

Вычислим оценки ν_1, \dots, ν_m , равные правой части неравенств следствия 4.4 вероятностей $\mu(\&N\theta_1)$, $\nu_1, \dots, \mu(\&N\theta_m)$, ν_m запросов $N\theta_1, \dots, N\theta_m$. Вычислим также оценки η_1, \dots, η_m , равные правой части неравенств теоремы 4.2 для условных вероятностей

$$\mu(\&N\theta_1 / \&N^1_{k_1} F) > \eta_1, \dots, \mu(\&N\theta_m / \&N^m_{k_m} F) > \eta_m$$

запросов $N\theta_1, \dots, N\theta_m$. Положим

$$\nu(N) = \sup\{\nu_1, \dots, \nu_m\}, \eta(N) = \sup\{\eta_1, \dots, \eta_m\}.$$

Выбор операции \sup не регламентируется чисто логическими соображениями. В данном случае автор исходит из желания объединить такие понятия как логический вывод (с вероятностными оценками) и предсказание.

Если один из выводов $SLDF'_1, \dots, SLDF'_m$ состоит только в применении фактов, то, как следует из теоремы 4.2, он будет иметь оценку $\eta(N) = 1$. Назовем такой SLDF-вывод проверкой истинности запроса N (по аналогии с семантическим программированием [13]). Предсказанием запроса N будем называть такой SLDF-вывод запроса $N\theta$, на котором достигается оценка $\eta(N)$. Оценкой предсказания запроса N будем называть величину $\eta(N)$. Если предсказание не определено, то оценка предсказания $\eta(N)$ не определена.

5.2. Пусть $M = \langle U, \mu \rangle$ – вероятностная Эрбранова модель, согласованная с классом G .

Определение 5.2.1. Правило $C = A \leftarrow B_1, \dots, B_k$, $k > 0$ программы Pr применимо к классу Эрбрановых моделей G тогда и только тогда, когда $M \models (B_1, \dots, B_k)\theta$ для некоторой подстановки $\theta \in \Theta G$ и модели $M \in G$, где ΘG – множество всех подстановок основных (Ground) термов вместо переменных. Если правило не содержит переменных, то оно применимо к классу Эрбрановых моделей G тогда и только тогда, когда существует модель $M \in G$, $M \models (B_1, \dots, B_k)$.

Далее будем предполагать, что все правила $C = A \leftarrow B_1, \dots, B_k$, $k > 0$ программы Pr применимы к классу моделей G . Обозначим через $PR(G) \subseteq PR$ множество всех правил сигнатуры применимых к классу Эрбрановых моделей G .

Определение 5.2.2. Правило $C = A \leftarrow B_1, \dots, B_k$, $k \geq 0$ программы Pr истинно на Эрбрановой модели $M \in G$ тогда и только тогда, когда $M \models (A\theta \leftarrow (B_1, \dots, B_k)\theta)$, $\theta \in \Theta G$, $M \in G$.

Определение 5.2.3. Программа Pr истинна на Эрбрановой модели $M \in G$, $M \models Pr$ тогда и только тогда, когда каждое правило программы истинно на M .

Определение 5.2.4. Программа Pr истинна на классе моделей G тогда и только тогда, когда $M \models Pr$, $M \in G$.

Распространим вероятность μ на множество формул со свободными переменными F_0 . Для $\varphi \in F_0 \setminus S$ положим

$$\mu(\varphi) = \inf_{\theta \in \Theta G} \{ \mu(\varphi\theta) \}.$$

Распространим вероятность μ на множество правил программы Pr . Для правил $C = A \leftarrow B_1, \dots, B_k$, $k \geq 0$, не содержащих переменных, определим условную вероятность равенством:

$$\mu(C) = \mu(A) / \mu(B_1 \& \dots \& B_k).$$

Эта вероятность определена, т.к. правила программы применимы к классу Эрбрановых моделей G и $\mu(B_1 \& \dots \& B_k) \neq 0$.

Для правил $C = A \leftarrow B_1, \dots, B_k$, $k \geq 0$ с переменными определим условную вероятность равенством:

$$\mu(C) = \mu(A / B_1 \& \dots \& B_k) = \frac{\inf_{\theta \in \Theta G} \{ \mu(A\theta / (B_1 \& \dots \& B_k)\theta) \}}{\inf_{\theta \in \Theta G} \{ \mu((B_1 \& \dots \& B_k)\theta) \}}.$$

Эта вероятность также определена, т.к. правила программы применимы к классу Эрбрановых моделей G и $\mu((B_1 \& \dots \& B_k)\theta) \neq 0$, $\theta \in \Theta G$. При $k = 0$ правило $C = A \leftarrow$ рассматривается как правило $A \leftarrow \text{true}$ с вероятностью посылки $\mu(\text{true}) = 1$.

Лемма 5.2.1. $\mu(\varphi\theta) > \mu(\varphi)$, $\varphi \in F_0$, θ – некоторая подстановка.

Лемма 5.2.2. Если программа Pr истинна на классе моделей G , то $\mu(C) = 1$, $C \in Pr$.

Доказательство. Пусть $C = A \leftarrow B_1, \dots, B_k$; $C \in Pr$, $k > 0$;

$$\mu(C) = \frac{\inf}{\theta \in \Theta G} \{ \mu(A\theta / (B_1 \& \dots \& B_k)\theta) \}.$$

Так как программа Pr истинна на G , то $G(A\theta \leftarrow (B_1, \dots, B_k)\theta) = G$ для любой подстановки $\theta \in \Theta G$. Так как мера согласована с классом моделей G , то $\mu(A\theta \leftarrow (B_1, \dots, B_k)\theta) = 1$, $\theta \in \Theta G$ и, следовательно, $\mu(A\theta / (B_1, \dots, B_k)\theta) = 1$. \square

6. Детерминированные и вероятностные закономерности

Определим на множестве PR отношение \blacktriangleright – «быть более общим». Обозначим множество всех подстановок, не являющихся перестановками, кроме тождественной, через Θt .

Определение 6.1. Отношение $C \blacktriangleright C'$, $C = A \leftarrow B_1, \dots, B_n$; $C' = A' \leftarrow B'_1, \dots, B'_{n'}$, $n, n' \geq 0$ имеет место тогда и только тогда, когда существует подстановка $\theta \in \Theta t$ такая, что $A\theta = A'$, $\{B_1\theta, \dots, B_n\theta\} \subset \{B'_1, \dots, B'_{n'}\}$ и либо θ не тождественная подстановка, либо $n < n'$.

Обозначим через $W(G) \subset PR$ множество правил, применимых к G и истинных на G (определение 5.2.2). Тогда $\mu(C) = 1$, $C \in W(G)$.

Лемма 6.1. Отношение \blacktriangleright – строгий частичный порядок на PR .

Лемма 6.2. Если $C \in W(G)$ и $C \blacktriangleright C'$, то $C' \in W(G)$.

Пусть $WD(G) \subset W(G)$ – множество всех максимальных по отношению \blacktriangleright правил из $W(G)$. Правила из $WD(G)$ нельзя обобщить, сохраняя их истинность на G .

Определение 6.2. Правила из $WD(G)$ будем называть детерминированными закономерностями или D-правилами.

Определение 6.3. Назовем отношение $C \sqsubset C' \Leftrightarrow (C \blacktriangleright C') \& (\mu(C) < \mu(C'))$, $C, C' \in PR(G)$ отношением вероятностной выводимости.

Определение 6.4. Вероятностной закономерностью (*P-правилом*) будем называть правило $C \in PR(G)$, такое, что из $C' \blacktriangleright C$, $C' \in PR(G)$ следует $C' \sqsubset C$.

Если детерминированные закономерности нельзя обобщить, сохраняя их истинность на классе моделей G , то вероятностные закономерности нельзя обобщить, не уменьшая их условную вероятность. Обозначим через $PR(G)$ множество всех *P-правил*.

Лемма 6.3. Если существует C' , $C' \blacktriangleright C$, $\mu(C') \geq \mu(C)$, то $C \notin PR(M)$.

Лемма 6.4. *D-правило* C , $C \in PR(G)$ является *P-правилом*, если из $C' \blacktriangleright C$, $C' \in PR(G)$ следует что $\mu(C') < 1$.

7. Предсказание и индуктивный синтез логических программ

Полный набор фактов для класса моделей G составляет совокупность множеств $F(N) = \{A \mid A - \text{atom}, N \models A \text{ для любого состояния атома } A\}$, $N \in G$. Любую конечную совокупность D конечных подмножеств $D(N) \subset F(N)$, $N \in G$ будем называть данными. Вероятностную Эрбранову модель M , согласованную с классом G , будем называть *вероятностной моделью данных* D .

Как следует использовать правила $C = A \leftarrow B_1, \dots, B_k$, $k \geq 1$ из $W(G)$ или $PR(G)$ для предсказания? Если посылка правила $(B_1 \& \dots \& B_k)\theta$ истинна на некоторой случайно выбранной из G в соответствии с мерой μ модели N (при некоторой подстановке $\theta \in \Theta G: \{B_1\theta, \dots, B_k\theta\} \subset F(N)$), то заключение $A\theta$ должно быть истинным на N с вероятностью $\mu(A\theta / (B_1 \& \dots \& B_k)\theta) \geq \mu(A / B_1 \& \dots \& B_k) = \mu(C)$. Вероятность $\mu(C)$, определенная в параграфе 5 для правил со свободными переменными, дает нам нижнюю границу вероятностей предсказания атома $A\theta$. Заметим, что предсказание нужно делать по данным $D(N)$ какой-то одной, случайно выбранной из G , модели N . Обозначим множество всех *P-правил* с посылкой, содержащей хотя бы один атом, через $P(G) \subset PR(G)$.

Определение 7.1. Логической программой индуктивно синтезированной по данным D в вероятностной модели данных M будем называть множество правил $PR(M, N) = P(G) \cup D(N)$, где $D(N) \in D$, N – некоторая модель, случайно выбранная из G в соответствии с вероятностной моделью данных M .

8. Семантический вероятностный вывод и предсказание

Определение 8.1. Семантическим вероятностным выводом (СВВ) произвольного атома A сигнатуры Ω будем называть максимальную последовательность правил $C_1 \sqsubset C_2 \sqsubset \dots$; $C_1, C_2, \dots \in P(G)$; $C_i = A_i \leftarrow B^i_1, \dots, B^i_{l_i}$, $i = 1, 2, \dots$ такую, что атом A унифицируем с атомами A_1, A_2, \dots . Каждому выводу соответствует последовательность подстановок $\theta_1, \theta_2, \dots$ из определения отношения \blacktriangleright . Подстановку $\theta = \theta_1\theta_2 \dots$ будем называть результатом семантического вероятностного вывода. Последнее правило в конечном выводе будем называть результирующим.

Лемма 8.1. D -правило в СВВ выводе может быть только результирующим.

Деревом вероятностного семантического вывода атома A будем называть совокупность всех СВВ выводов запроса A .

Определение 8.2. P -предсказанием некоторого атома A сигнатуры Ω программой $PR(M, N) = P(G) \cup D(N)$ будем называть такой СВВ вывод $C_1 \sqsubset C_2 \sqsubset \dots \sqsubset C_i \sqsubset \dots$; $C_1, C_2, \dots, C_i, \dots \in P(G)$ запроса A , в котором:

- 1) Существует правило $C_i = A_i \leftarrow B^i_1, \dots, B^i_{l_i}$ и подстановка θ , такие что $\{B^i_1\theta, \dots, B^i_{l_i}\theta\} \subset D(N)$; $A\theta = A_i\theta$; $\mu(A_i\theta) < \mu(C_i)$;
- 2) Если P -дерево вывода цели A пусто или требуемой подстановки не существует, то P -предсказание не определено.
- 3) На правиле C_i достигается максимум условной вероятности $\mu(C_i)$ среди всех правил, удовлетворяющих условию 1, всех P -выводов цели A .
- 4) Результатом P -предсказания будем называть подстановку $\theta_p = \theta_1\theta_2 \dots \theta_{i-1}\theta$, где $\theta_1, \theta_2, \dots, \theta_{i-1}$ - подстановки СВВ вывода $C_1 \sqsubset C_2 \sqsubset \dots \sqsubset C_i$, удовлетворяющего 1 и 2.
- 5) Оценкой P -предсказания будем называть величину $\eta_p(A) = \mu(C_i)$. Если P -предсказание не определено, то оценка $\eta_p(A)$ не определена.
- 6) Правило C_i будем называть наилучшим для предсказания атома A правилом по данным $D(N)$ в вероятностной модели M .

9. Взаимосвязь семантического вероятностного и логического выводов

Пусть Pr – некоторая логическая программа, факты которой содержатся среди фактов $D(N)$ программы $PR(M, N) = P(G) \cup D(N)$.

Теорема 9.1. *Если атом A предсказывается программой Pr с оценкой $\eta(A) > \mu(A\theta)$, для любой подстановки $\theta \in \Theta G$, то он P -предсказывается программой $PR(M, N)$ с оценкой P -предсказания $\eta_p(A) \geq \eta(A)$.*

Доказательство. По условию существует успешный SLDF-вывод $A\theta$, $N_1\theta, \dots, N_k, \&N_kF \in \mathfrak{R}$ цели $A\theta$ в пространстве вычислений программы Pr такой, что $\mu(A\theta/\&N_kF) \geq \eta(A) > \mu(A\theta)$, $\mu(\&N_kF) > 0$, $N_k = \leftarrow V_1, \dots, V_l$; $\{V_1 \leftarrow, \dots, V_l \leftarrow\} \subset Pr$, $l > 1$ (см. определение предсказания, оценки η и теорему 4.2).

Рассмотрим правило $C = A\theta \leftarrow V_1, \dots, V_l$. Из условия $\eta(A) > \mu(A\theta) \geq 0$, следует что $l \geq 1$. Так как $\mu(\&N_kF) > 0$, то $C \in P(G)$. Кроме того из $\mu(A\theta/\&N_kF) \geq \eta(A)$ следует, что $\mu(C) \geq \eta(A) > \mu(A\theta)$ и значит выполнено условие 1 определения 8.2 и P -предсказание атома A определено. Тогда существует наилучшее для предсказания атома A правило CB и $\eta_p(A) = \mu(CB)$. Из условия 2 этого определения 8.2 следует, что $\eta_p(A) = \mu(CB) \geq \mu(C) \geq \eta(A)$. \square

Список литературы

1. Витяев Е. Е. Извлечение знаний из данных. Компьютерное познание. Моделирование когнитивных процессов / Е. Е. Витяев. – Новосибирск : НГУ, 2006. – 293 с.
2. Малых А. А. Документное моделирование / А. А. Малых, А. В. Манцивода // Изв. Иркут. гос. ун-та. Сер. Математика. – 2017. – Т. 21.
3. Adams Er. W. The logic of conditionals: An application of probability to deductive logic / Er. W. Adam // Synthese Library. – 1975. – Vol. 86.
4. Apt K. R. Introduction to logic programming / K. R. Apt // Computer Science / Department of Software Technology, Report CS-R874.
5. Van Emden M. N. Quantitative deduction and its fixpoint theory / M. N. Van Emden // J. Logic Programming. – 1986. – Vol. 3, N 1. – P. 37–53. [https://doi.org/10.1016/0743-1066\(86\)90003-8](https://doi.org/10.1016/0743-1066(86)90003-8)
6. Fitting M. C. Logic Programming on a Topological Bilattices / M. C. Fitting // Fundamenta Informatica. – 1988. – Vol. 11. – P. 209–218.
7. Gaifman H. Concerning measure in first order calculi / H. Gaifman // Israel journal of Math. – 1964. – Vol. 2, N 1. – P. 1–18. <https://doi.org/10.1007/BF02759729>
8. Hailperin T. Probability Logic / T. Hailperin // Notre Dame J. of Formal Logic. – 1998. – Vol. 25, N 3. – P. 198–212. <https://doi.org/10.1305/ndjfl/1093870625>
9. Kifer M. Theory of Generalized Annotated Logic Programming and its Applications / M. Kifer, V. S. Subrahmanian // Research Report. – University of Maryland, USA, 1990.

10. Ng R. T. Probabilistic reasoning in Logic Programming / R. T. Ng, V. S. Subrahmanian // Proc. 5th Symposium on Methodologies for Intelligent Systems. – Knoxville, North-Holland, 1990. – P. 9–16.
11. Kovalerchuk B. Data Mining in Finance: Advances in Relational and Hybrid Methods/ Boris Kovalerchuk, Evgenii Vityaev. – Kluwer Acad. Publ., 2000. – 308 p.
12. Matthew M. Huntbach An improved version of Shapiro's Model Inference system / M. Matthew // Third International conference on Logic Programming (Lecture Notes in Computer Science ; vol. 225). – P. 180–187.
13. Ng R.T. Annotation Variables and Formulas in Probabilistic Logic Programming / R. T. Ng, V. S. Subrahmanian // Technical report CS TR-2563. – University of Maryland, 1990.
14. Nilsson Nils J. Probability logic / Nils J. Nilsson // Artif. Intell. – 1986. – Vol. 28, N 1. – P. 71–87. [https://doi.org/10.1016/0004-3702\(86\)90031-7](https://doi.org/10.1016/0004-3702(86)90031-7)
15. Scott D. S. Assigning Probabilities to Logical Formulas / D. S. Scott, P. Krauss // Aspects of Inductive Logic / eds.: J. Hintikka, P. Suppes. – N. Holland, 1966. – P. 219–264. [https://doi.org/10.1016/S0049-237X\(08\)71672-0](https://doi.org/10.1016/S0049-237X(08)71672-0)
16. Shapiro E. Logic Programs with Uncertainties: A Tool for Implementing Expert Systems / E. Shapiro // Proc. IJCAI '83, Williams Kauffman. – 1983. - P. 529–532.
17. Goncharov S. S. Semantic programming / S. S. Goncharov, Yu. L. Ershov, D. I. Sviridenko // 10th World Congress Information Processing 86, Dublin, Oct., 1986. – Amsterdam, 1986. – P. 1093–1100.
18. Shapiro E. Algorithmic Program Debugging / E. Shapiro. – MIT Press., 1983. – 204 p.

Витяев Евгений Евгеньевич, доктор физико-математических наук, профессор, Институт математики им. С. Л. Соболева, 630090, Новосибирск, пр-т Акад. Коптюга, 4б, тел.: (383)3297691
(e-mail: vityaev@math.nsc.ru)

E. E. Vityaev

Semantic Probabilistic Inference of Predictions

Abstract. Prediction is one of the most important concepts in science. Predictions obtained from probabilistic knowledge, are described by an inductive-statistical inference (I-S inference). However, such an inference encounters a problem of synthesis the logic and probability that consists in the rapid decreasing of the probability estimates of predictions in the process of logical inference. The procedures for calculating estimates in the Probabilistic Logic Programming do not solve the problem. From our point of view, prediction can not be well combined with a logical inference. Logical inference should be replaced by calculations. The paper proposes a semantic approach to the calculation of prediction, when the inference is considered not as verification of the truth of some statement on the model, but as a search for facts in the model, predicting the statement with a maximum probability. To do this, the work defines a semantic probabilistic inference for calculating the predictions. In the process of semantic probabilistic inference, estimates of predictions strictly increase. We prove in the paper that prediction estimates obtained by the semantic probabilistic inference are certainly not worse than the estimates obtained by the logical inference with the parallel calculation of these estimates.

Keywords: prediction, probabilistic inference, semantics, logical programs, probabilistic logic programming.

References

1. Vityaev E.E. *Izvolechenie znaniy iz dannykh. Komp'yuternoe poznanie. Modelirovanie kognitivnykh protsessov*[Extracting knowledge from data. Computer cognition. Modeling of cognitive processes]. Novosibirsk, NSU, 2006. 293 p.
2. Malykh A., Mantsivoda A. Document modeling. *Izv. Irkutsk. Gos. Univ. Ser. Mat.*, 2017, vol. 21. (in Russian)
3. Adams Er.W. The logic of conditionals. An application of probability to deductive logic. Synthese Library, vol. 86, 1975.
4. Apt K.R. Introduction to logic programming. Computer Science. Department of Software Technology, Report CS-R874.
5. Van Emden M.N. Quantitative deduction and its fixpoint theory. *J. Logic Programming*, 1986, vol. 3, no 1, pp. 37-53. [https://doi.org/10.1016/0743-1066\(86\)90003-8](https://doi.org/10.1016/0743-1066(86)90003-8)
6. Fitting M.C. Logic Programming on a Topological Bilattices. *Fundamenta Informatica*, 1988, vol. 11, pp. 209-218.
7. Gaifman H. Concerning measure in first order calculi. *Israel journal of Math*, 1964, vol. 2, no 1, pp. 1-18. <https://doi.org/10.1007/BF02759729>
8. T.Hailperin, Probability Logic. *Notre Dame J. of Formal Logic*, vol. 25, no 3, 1984, pp.198-212. <https://doi.org/10.1305/ndjfl/1093870625>
9. Kifer M., Subrahmanian V.S. Theory of Generalized Annotated Logic Programming and its Applications. Research Report, University of Maryland, USA, 1990.
10. Ng R.T., Subrahmanian V.S. Probabilistic reasoning in Logic Programming. *Proc. 5th Symposium on Methodologies for Intelligent Systems, Knoxville, North-Holland*, 1990, pp. 9-16.
11. Kovalerchuk Boris, Vityaev Evgenii. Data Mining in Finance: Advances in Relational and Hybrid Methods. Kluwer Acad. Publ., 2000. 308 p.
12. Matthew M. Huntbach An improved version of Shapiro's Model Inference system. *Third International conference on Logic Programming (Lecture Notes in Computer Science, vol. 225)*, pp.180-187.
13. Ng R.T., Subrahmanian V.S. Annotation Variables and Formulas in Probabilistic Logic Programming. Technical report CS TR-2563, University of Maryland, 1990.
14. Nils J. Nilsson. Probability logic. *Artif. Intell.*, 1986, vol. 28, no 1, pp. 71-87. [https://doi.org/10.1016/0004-3702\(86\)90031-7](https://doi.org/10.1016/0004-3702(86)90031-7)
15. Scott D.S., Krauss P. Assigning Probabilities to Logical Formulas. Aspects of Inductive Logic, (ed. J. Hintikka, P. Suppes), N. Holland, 1966, pp. 219-264. [https://doi.org/10.1016/S0049-237X\(08\)71672-0](https://doi.org/10.1016/S0049-237X(08)71672-0)
16. Shapiro E. Logic Programs with Uncertainties: A Tool for Implementing Expert Systems. *Proc. IJCAI '83, Williams Kauffman*, 1983, pp. 529-532.
17. Goncharov S.S., Ershov Yu.L., Sviridenko D.I. Semantic programming. *10th World Congress Information Processing 86, Dublin, Oct., 1986*, Amsterdam, 1986, pp. 1093-1100.
18. Shapiro E. Algorithmic Program Debugging. MIT Press, 1983, 204 p.

Vityaev Evgenii Evgenievich, Doctor of Sciences (Physics and Mathematics), Professor, Novosibirsk State University, 4, Koptug, Novosibirsk, 630090 tel.: (383)3297691 (e-mail: vityaev@math.nsc.ru)