ЛОГИКА И СЕМАНТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ
(Вычислительные системы)
1992 год Выпуск 146

УДК 519.68

СЕМАНТИЧЕСКИЙ ПОДХОД К СОЗДАНИЮ БАЗ ЗНАНИЙ. СЕМАНТИЧЕСКИЙ ВЕРОЯТНОСТНЫЙ ВЫВОД НАИЛУЧШИХ ДЛЯ ПРЕДСКАЗАНИЯ ПРОЛОГ-ПРОГРАММ ПО ВЕРОЯТНОСТНОЙ МОДЕЛИ ДАННЫХ*)

Е.Е.Витяев

Введение

Методы машинного обучения (Mashine Learning) часто используются в экспертных системах и системах принятия решений для получения новых знаний из данных. Полученные знания используются далее для принятия решений с помощью методов логического вывода, которые абстрагируются от возможной недостоверности знаний и осуществляют вывод как будто бы мы имели достоверные знания. В результате решения имеют неопределенную степень достоверности и, строго говоря, непонятно в каком смысле являются решениями.

Для оценки степени достоверности решений разрабатываются различные методы их вычисления параллельно процессу логического вывода. Есть работы, в которых степень достоверности рассматривается как значение истинности утверждений, а процесс логического вывода обобщается до так называемой "количественной дедукции" [2-5]. В работах [5-7] описываются до-

*) Данная работа является расширенным изложением препринта [1]. Существенно расширено введение, подробно объясняющее цель и место работы. Основная часть практически не изменена.

вольно богатые формальные системы, содержащие как частные случаи основные известные "количественные дедукции".

В какой степени разработанные методы оценки достоверности обосновывают и придают смысл решениям?

Рассмотрим знания, полученные методами машинного обучения на вероятностных данных. Анализ изменения оценок вероятности утверждений в процессе логического вывода показывает, что они могут значительно уменьшаться. Как следует из работ по вероятностной логике [8-11], полученные оценки не могут быть улучшены. Даже если ограничиться использованием правил с условной вероятностью, не меньшей чем $1 - \varepsilon$, как это делается в [12], то это все равно не избавляет нас от существенного уменьшения вероятности в процессе вывода и, кроме того, это не соответствует условиям реально возникающих задач.

Рассмотрим знания, извлекаемые и оцениваемые экспертом. В работах по "количественной дедукции" [2-5] истинностное значение заключения правила вычисляется как функция минимума или наибольшей нижней границы (для значений истинности в решетке) значений истинности атомов посылки. Соответствует ли это экспертными оценкам правила? Как правило, не соответствует, так как и 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 решений 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 вообще откажется дать оценку заключению правила либо присвоит ему значение, близкое к нулю, хотя это правило может быть выведено по правилам вероятностной логики и иметь отличное от нуля значение.

Таким образом, несмотря на значительный прогресс в построении формальных систем, вычисляющих оценки утверждений, адекватное вычисление оценок решений отсутствует. В чем причина?

Причина в том, что при обобщении 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.

Можно более остро поставить проблему: идея создания баз знаний и экспертных систем основана на "аксиоматическом" подходе к знаниям - "извлечь" из эксперта и поместить в базу знаний основополагающие знания (аксиомы) так, чтобы остальные знания и решения получались логическим выводом с параллельным вычислением оценок достоверности. Невозможность адекватного вычисления оценок решений говорит о невозможности и самого аксиоматического подхода к построению баз знаний и необходимости его пересмотра. На какой основе это можно сделать?

Рассмотрим процесс вычисления с точки зрения "семантического" подхода к программированию [13]. Идея семантического программирования состоит в том, что процесс вычисления можно 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, подтвержденности, достоверности, статистической значимости и т.д. высказываний на модели. Такой обобщенный вывод будем называть семантическим.

В работе семантический подход к базам знаний разрабатывается для случая ПРОЛОГ-программ в языке первого порядка с вероятностной мерой μ [8-11], а также вероятностных данных (нам известна вероятностная модель данных $\mathfrak{M}$ [8,11] - вероятностная мера μ, заданная на множестве всех основных предложений, см. определение 1.2).

Наиболее важной вероятностной оценкой решений является оценка предсказательной силы высказываний. Высказывание вместе с такой оценкой назовем предсказанием.

Рассмотрим сначала стандартный процесс вычислений ПРОЛОГ-программ. Предсказанием запроса ПРОЛОГ-программой $\mathbf{PR}$ назовем такое вычисление 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 §3). 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.

Цель предсказания в общем случае состоит в нахождении таких фактов, из которых решение следовало бы с максимальной (условной) вероятностью. Предсказание, получаемое ПРОЛОГ-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 $\mathbf{A}$ в работе определяется <u>семантический процесс вычисления</u> - вероятностный вывод (см. определение 8.1), в котором вычисление осуществляется путем движения вдоль "уточняющего" графа [14,15]. В этом графе правила,начиная с $A \leftarrow$,"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 являются результирующая подстановка и достигнутая условная вероятность.

На уточняющие правила в вероятностном выводе можно наложить (без ограничения общности) дополнительное требование - чтобы каждый атом в посылке был "существенным" для предсказания атома A (удаление любого атома из посылки уменьшало бы условную вероятность атома A). Такие правила в работе называются вероятностными закономерностями (см. определение 6.2). Для получения любого вероятностного вывода, таким образом, достаточно иметь множество всех возможных вероятностных закономерностей данной вероятностной модели данных $\mathfrak{M}$. В работе это множество обозначается через $PR(\mathfrak{M})$.

Отметим, что для вероятностного вывода не нужны никакие правила вывода.Процесс вычисления вполне определяется требованием увеличения оценки предсказательной силы высказывания, определяемой по вероятностной модели данных $\mathfrak{M}$. Если в результате такого вероятностного вывода получена оценка условной вероятности, равная 1, что может означать получение тождественно истинного высказывания, и дальнейший вывод, опираясь только на оценку, уже невозможен, тогда вступают в силу правила логического вывода (например, резолюция), которые можно применять, используя только правила с условной вероятностью,равной 1. Таким образом, вероятностный вывод является естественным обобщением логического при его семантической интерпретации. Но такое обобщение, естественное с семантической точки зрения, невозможно и даже противоречит аксиоматическому подходу к знаниям, так как даже не нуждается в правилах вывода.

Множество $PR(\mathfrak{M})$ является в определенном смысле полным и минимальным множеством вероятностных знаний, обеспечивающим любой вероятностный вывод и максимальную оценку предсказаний и, таким образом, полностью удовлетворяющим поставленной цели - получению наилучших предсказаний для решений. В этом смысле множество $PR(\mathfrak{M})$ дает "вероятностную теорию" предметной области. Анализ множества $PR(\mathfrak{M})$ показывает, что его невозможно получить никакими правилами логического вывода. Таким образом, модификация процедуры вывода позволяет нам получить "вероятностную теорию" предметной области всю целиком, а не выводить ее из аксиом по правилам вывода.

Под "вероятностной теорией" понимается не формальная система, объединяющая теорию вероятностей и логику первого порядка, как это до некоторой степени делается в работах по вероятностной логике и "количественным дедукциям" [6,7,11], а ближе к логике, как множество утверждений с максимальными значениями истинности или оценок (в нашем случае с максимальными значениями оценок условных вероятностей).

Причина, по которой в работе рассматриваются только одноатомные запросы, обсуждается в заключении.

Пусть есть данные $D(\mathfrak{N})$ из некоторой модели $\mathfrak{N}$, случайно выбранной из множества возможных миров в соответствии с вероятностной моделью данных $\mathfrak{M}$ (см. §7).

Рассмотрим ПРОЛОГ-программу $PR(\mathfrak{M},\mathfrak{N}) = P(\mathfrak{M}) \cup D(\mathfrak{N})$, где $P(\mathfrak{M}) \subset PR(\mathfrak{M})$ - множество всех вероятностных закономерностей с непустой посылкой. В работе доказывается, что программа $PR(\mathfrak{M},\mathfrak{N})$ предсказывает лучше любой другой ПРОЛОГ-программы Pr, имеющей те же факты $D(\mathfrak{N})$ (теорема 9.1). Более того, предсказание любого атома A (данной сигнатуры) осуществляется "лучшим для предсказания атома A" правилом (см.определение 7.1) в один шаг (не считая подстановки фактов, см.доказательство теоремы 9.2). "Лучшее для предсказания атома $\mathbf{A}$ правило" является вероятностной закономерностью (см. теорему 7.1) и может быть получено вероятностным выводом.

Таким образом, база знаний $PR(\mathfrak{M})$, рассматриваемая как ПРОЛОГ-программа, предсказывает на одних и тех же фактах лучше любой другой ПРОЛОГ-программы.

Почему множеству вероятностных закономерностей удается аппроксимировать по предсказанию значительно более разнообразный и богатый комбинационными возможностями логический вывод? Поясним это на примере шахматной игры. Целью игры является выигрыш, а правила игры можно представить как правила вывода. Опытный игрок никогда не использует чисто комбинационный анализ всех возможных ходов за себя и за противника, т.е. чисто логический вывод. Для достижения выигрыша и проведения глубокого анализа вариантов игрок использует некоторую оценку позиции, которую он стремится улучшить. Ведущей к цели - выигрышу - становится оценка, а перебор вариантов подчинен требованию улучшения оценки позиции. Логический вывод не должен быть самоцелью. Цель вывода должна определяться независимо от самого вывода и являться определяющей, а логический вывод должен быть подчинен поставленной цели.

Точный анализ цели доказательств в математических теориях осуществлен в [19,20]. Цель доказательств состоит в решении задач: "... мы понимаем задачу только тогда, когда ей сопоставили обоснованное чувство уверенности в том, что всякое состояние нашего сознания мы сумеем убедительным и безошибочным образом распознать как такое, когда решение найдено, или как такое, когда решение задачи не найдено" [19]. Формализация этого требования и его анализ показали,что оно накладывает существенные ограничения на формальные системы,в которых должны ставиться и решаться задачи.

В задачах искусственного интеллекта приведенное требование на осмысленность постановок задач также должно быть выполнено. Задача принятия решений осмысленна только тогда, когда мы не только можем вывести решение, но и всегда определить, является ли оно таковым. В работах [19,20] показано, что формальные системы для постановок и решения задач должны быть слабыми. Для этого подходит, в частности, логическое программирование. Как отмечается в [19], "... в рамках новой парадигмы выглядит весьма естественным так называемый "логический подход к программированию",... согласно которому следует создавать языки спецификаций не только программ, но и задач."

С точки зрения задач, в данной работе показывается, что если целью доказательства является не только решение некоторой задачи, но и достижение максимума некоторой оценки, то необходимо не только наложить существенные ограничения на используемые формальные системы и использовать, например, логическое программирование, но и пересмотреть само понятие вывода.

В заключение отметим, что множество $\mathrm{PR}(\mathfrak{M}, \mathfrak{N})$ не является слишком большим. Понятие вероятностной закономерности было ранее введено автором для разработки метода обнаружения закономерностей [16-18] - метода построения всех статистических аппроксимаций вероятностных закономерностей, т.е. метода построения статистической аппроксимации множества $\mathrm{PR}(\mathfrak{M})$. Этот метод был реализован и успешно применялся для решения ряда практических задач. Опыт решения задач показал, что множество $\mathrm{PR}(\mathfrak{M})$ практически приемлемо даже для малых ЭВМ.

§1. Вероятность. Эрбрановы модели. Вероятностная модель данных

Зафиксируем язык первого порядка $\mathbf{L}$ с равенством не более чем счетной *сигнатуры* $\sigma = \langle P_1, P_2, \ldots ; f_1, f_2, \ldots ; C_1, C_2, \ldots \rangle$, $C = \{C_k\}_{k \in K}$, $K \neq \emptyset$. 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), $\mathfrak{B} = F_0 \cap S$ - множество всех *основных предложений* сигнатуры $\mathfrak{S}$.

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

ОПРЕДЕЛЕНИЕ 1.1 [8]. *Вероятностью* μ на подмножестве $\Phi \subset \mathfrak{B}$ называется отображение $\mu : \Phi \to [0,1]$, удовлетворяющее условиям:

1) если $\vdash \varphi$, то $\mu(\varphi) = 1$,

2) если $\vdash \neg(\varphi \,\&\, \psi)$, то $\mu(\varphi \vee \psi) = \mu(\varphi) + (\psi)$.

СЛЕДСТВИЕ 1.1 [8].

Если $\vdash \varphi \equiv \psi$, *то* $\mu(\varphi) = \mu(\psi)$.

Если $\vdash \neg\varphi$, *то* $\mu(\varphi) = 0$.

Вероятность μ является конечно-аддитивной мерой на подалгебре $\{\psi/\equiv \,|\, \psi \in \Phi\}$ булевой алгебры Линденбаума-Тарского.

ОПРЕДЕЛЕНИЕ 1.2. *Вероятностной эрбрановой моделью* сигнатуры $\mathfrak{S}$ будем называть пару $\mathfrak{M} = \langle U, \mu \rangle$, где μ - вероятность на $\mathfrak{B}$. Функциональные символы интерпретируются на U обычным образом [21].

ОПРЕДЕЛЕНИЕ 1.3. *Эрбрановой моделью* сигнатуры $\mathfrak{S}$ будем называть вероятностную эрбранову модель $\mathfrak{M} = \langle U, I \rangle$, где $I: \mathfrak{B} \to \{0,1\}$.

Рассмотрим множество Σ всех эрбрановых моделей сигнатуры $\mathfrak{S}$. Пусть даны некоторый класс эрбрановых моделей $G \subset \Sigma$ (множество возможных миров) и вероятность μ на некотором подмножестве $\Phi \subset \mathfrak{B}$ формул, замкнутом относительно логических операций. Определим булеву подалгебру $\mathfrak{D}$ подмножеств $G(\varphi) = \{\mathfrak{M} \mid \mathfrak{M} \in G, \mathfrak{M} \models \varphi\}$, $\varphi \in \Phi$, множества G.

ОПРЕДЕЛЕНИЕ 1.4. Класс эрбрановых моделей G будем называть *согласованным с вероятностью* μ на множестве формул Φ, если из $G(\varphi) = \emptyset$, $\varphi \in \Phi$, следует $\mu(\varphi) = 0$.

ЛЕММА*) 1.1. *Величина* $\nu(G(\varphi)) = \mu(\varphi)$, $\varphi \in \Phi$, *является конечно-аддитивной мерой на подалгебре* $\mathfrak{D}$, *если класс эрбрановых моделей* G *согласован с* μ *на множестве формул* Φ.

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

§2. Логические программы

Обозначим через PR множество всех *правил* сигнатуры $\mathfrak{S}$ $A \leftarrow A_1, \ldots, A_k$, $k \geq 0$, где $A, A_1, \ldots, A_k$ - *атомы сигнатуры* $\mathfrak{S}$. Если атом A отсутствует, то правило $\leftarrow A_1, \ldots, A_k$ называется целью (запросом). Если $k = 0$, то правило $A \leftarrow$ называется *фактом*.

Логическая программа Pr есть конечная совокупность правил. *Подстановкой* называется отображение $\theta : X \to T$. Подстановка $\theta(x) = x$ называется *тождественной*. Подстановки естественным образом распространяются на произвольные выражения - так, для терма $t = f(t_1, \ldots, t_n)$ и атома $A =$

---

*) Доказательства лемм, теорем и следствий, отмеченных *, даны в приложении.

$= P(t_1,\ldots, t_n)$ их подстановки равны $t\theta = f(t_1\theta, \ldots$ $\ldots, t_n\theta)$, $A\theta = P(t_1\theta,\ldots,t_n\theta)$.Правило $A\theta \leftarrow A_1\theta, \ldots$ $\ldots, A_n\theta$ называется *вариантом правила* $A \leftarrow A_1, \ldots, A_n$, если θ - перестановка множества X .

Зафиксируем *правило вычисления* R ,определяющее в каждом запросе выделенный атом.Пусть $N = \leftarrow A_1 \& \ldots \& A_i \& \ldots$ $\ldots \& A_k$, $k \geq 1$, - запрос,в котором правилом R выделен атом A_i, и $A \leftarrow B_1 \& \ldots \& B_l$ - 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

$$\begin{aligned} &\leftarrow (A_1 \& \ldots \& B_1 \& \ldots \& B_l \& \ldots \& A_k)\ , \ l \geq 1, \\ &\leftarrow (A_1 \& \ldots \& A_i \& \ldots \& A_k)\theta,\ l = 0, \end{aligned} \tag{1}$$

будем называть *выводимыми* из запроса N по правилу $A \leftarrow$ $\leftarrow B_1 \& \ldots \& B_l$ 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 помечаются 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 σ с заданным на нем отношением выводимости . SLDF-*выводом* (Linear resolution with Selection rule for Definite clauses and underlined Facts) цели N в пространстве вычислений назовем максимальную последовательность запросов $N =$ $= N_0, N_1, N_2, \ldots$ вместе с последовательностью правил C_0, $C_1, \ldots$ и унификаций $\theta_0, \theta_1, \ldots$ такую, что запросы N_{i+1} выводимы из запросов N_i по правилам C_i с помощью подстановок θ_i и правила R. SLDF-вывод - максимальный путь в пространстве вычислений, начинающийся с $\mathbf{N}$. 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 $\mathbf{N}$, обычно представляют в виде дерева (префикс дерева SLDF-выводов) и называют SLDF-*деревом* вычислений запроса $\mathbf{N}$. SLDF-дерево, содержащее успешный SLDF-вывод, называется успешным.

§3. Оценки вероятностей и условных вероятностей запросов

Пусть $\mathfrak{M} = \langle U, \mu \rangle$ - вероятностная эрбранова модель. Рассмотрим успешный SLDF-вывод $N, N_1, \ldots, N_k$ запроса $\mathbf{N}$ с помощью последовательности правил $C_0, C_1, \ldots, C_{k-1}$ некоторой программы Pr , последовательности унификаций θ_0, $\theta_1, \ldots, \theta_{k-1}$; $\theta \overset{df}{=} \theta_0 \theta_1 \ldots \theta_{k-1}$ и некоторого правила вычисления $\mathbf{R}$.

Последовательность запросов $N\theta, N_1\theta, \ldots, N_k$ также будет SLDF-выводом запроса $N\theta$ с помощью последовательности правил $C_0\theta, C_1\theta, \ldots, C_{k-1}\theta$, тождественных подстановок и правила вычисления $\mathbf{R}$. Будем предполагать, что $N\theta, N_1\theta, \ldots$ $\ldots, N_k \in \mathfrak{B}$. В данном пункте факты $A \leftarrow$ представимы правилами $A \leftarrow \mathrm{true}$. Тогда $\mu(C) \overset{df}{=} \mu(A/\mathrm{true}) = \mu(A)$ для фактов $C = A \leftarrow; A \in \mathfrak{B}$.

Определим через $\widehat{N}_i$ конъюнкцию всех неподчеркнутых атомов запроса N_i . Если все атомы подчеркнуты (как в запросе N_k), то положим $\widehat{N}_k \equiv \mathrm{true}$. Обозначим через $\widehat{N_i F}$ конъюнкцию всех подчеркнутых атомов запроса N_i . Тогда $\widehat{N_k F}$ - конъюнкция всех фактов, использованных в SLDF-выводе запроса $N\theta$.

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

Рассмотрим вывод запросов (1) из запроса $N\theta = (\leftarrow A_1, \ldots, A_k)\theta$ по правилу $(A \leftarrow B_1, \ldots, B_l)\theta$.

Представим запросы (1) в виде $N_1\theta \overset{df}{=} (\leftarrow A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots, A_k)\theta$, $B \overset{df}{=} B_1, \ldots, B_l$ и $\underline{N}\theta \overset{df}{=} (\leftarrow A_1, \ldots, \underline{A}_i, \ldots, A_k)\theta$. Конъюнкция $\widehat{\underline{N}\theta}$ является частным случаем конъюнкции $\widehat{N_1\theta}$, когда $B = \mathrm{true}$. Оценим вероятности $\mu(\widehat{N\theta})$, $(\widehat{N\theta}/\widehat{N_1\theta})$ в предположении, что нам известны только вероятности $\mu(\widehat{N_1\theta})$, $\mu(A\theta)$, $\mu(\widehat{B\theta})$, $p = \mu(A\theta/\widehat{B\theta})$.

ЛЕММА* 3.1. *Если* $\mu(\widehat{N_1\theta}) > 0$ *и* $\mu(\widehat{B\theta}) > 0$, *то*

1) $\mu(\widehat{N\theta}) \le \mu(\neg\widehat{B\theta}) + \min\{\mu(\widehat{N_1\theta}), \mu(A\theta \mathbin{\&} \widehat{B\theta})\}$;

2) $\mu(\widehat{N\theta}) \ge \mu(\widehat{N_1\theta}) - (1-p)\mu(\widehat{B\theta})$;

3) $\mu(\widehat{N\theta}/\widehat{N_1\theta}) \le p/\mu(\widehat{\underline{N}\theta}/\widehat{B\theta})$;

4) $\mu(\widehat{N\theta}/\widehat{N_1\theta}) \ge 1 - (1-p)/\mu(\widehat{\underline{N}\theta}/\widehat{B\theta})$.

СЛЕДСТВИЕ* 3.1. *Если* $\mu(\widehat{N_1\theta}) > 0$, $\mu(\widehat{B\theta}) > 0$ *и* $p = 1$, *то*

1) $\mu(\widehat{N_1\theta}) \le \mu(\widehat{N\theta}) \le \min\{1, \mu(\neg\widehat{B\theta}) + \mu(\widehat{N_1\theta})\}$;

2) $\mu(\widehat{N\theta}/\widehat{N_1\theta}) = 1$.

СЛЕДСТВИЕ* 3.2. *Если* $\mu(\widehat{N_1\theta}) > 0$ *и правило является фактом* $(A \leftarrow \mathrm{true})\theta$, , *то*

1) $\mu(\widehat{N\theta}) \le \min\{\mu(\widehat{N_1\theta}), \mu(A\theta)\}$;

2) $\mu(\widehat{N\theta}) \ge \mu(\widehat{N_1\theta}) + \mu(A\theta) - 1$;

3) $\mu(\widehat{N\theta}/\widehat{N_1\theta}) \le \mu(A\theta)/\mu(\widehat{N_1\theta})$;

4) $\mu(\widehat{N\theta}/\widehat{N_1\theta}) \ge 1-(1-\mu(A\theta))/\mu(\widehat{N_1\theta})$.

ДОКАЗАТЕЛЬСТВО следует из леммы 3.1 и равенств $p = \mu(\widehat{A\theta})$, $\mu(\widehat{\underline{N}\theta}) = \mu(\widehat{N_1\theta})$.

СЛЕДСТВИЕ 3.3. *Если* $\mu(\widehat{B\theta}) > 0$, *то*

1) $\mu(\widehat{N\theta}\ \&\ \widehat{B\theta}) \le \min\{\mu(\widehat{N_1\theta}), \mu(A\theta\ \&\ \widehat{B\theta})\}$;

2) $\mu(\widehat{N\theta}\ \&\ \widehat{B\theta}) \ge \mu(\widehat{N_1\theta})-(1-p)\mu(\widehat{B\theta})$.

ДОКАЗАТЕЛЬСТВО следует из доказательств пп.1,2 леммы 3.1.

Рассмотрим SLDF-вывод $N\theta, N_1\theta, \ldots, N_k$ запроса $N\theta$ посредством последовательности правил $C_i\theta = (A \leftarrow B_1^i, \ldots$ $\ldots, B_{l_i}^i)\theta$, $i = 0,1,\ldots,k-1$, и пустых унификаций. Положим $B^i\theta = (B_1^i\ \&\ \ldots\ \&\ B_{l_i}^i)\theta$, $p_i = \mu(C_i\theta)$.

ТЕОРЕМА* 3.1. *Если* $\mu(B^i\theta) > 0$, $i = 0,1,\ldots,k-1$, *то*

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

СЛЕДСТВИЕ 3.4. *Если* $\mu(B^i\theta) > 0$, $i = 0,1,\ldots,k-1$, *то*

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

ДОКАЗАТЕЛЬСТВО следует из $\mu(\widehat{N\theta}) \ge \mu(\widehat{N\theta}\ \&\ A^0\theta\ \&\ldots$ $\ldots\ \&\ A^{k-1}\theta)$ и теоремы 3.1.

Для каждого успешного SLDF-вывода $N\theta = N_0\theta, N_1\theta, \ldots$ $\ldots, N_k$ существует успешный SLDF'-вывод $N\theta = N_0\theta, N_1'\theta, \ldots$ $\ldots, N_i'\theta, \ldots, N_k$, в котором факты применяются последними и до запроса $N_i'\theta$ применяются правила C_j, с длиной $l_j \ge 1$; $j = 0,\ldots,i-1$. Тогда запрос $N_i'\theta$ будет иметь вид

$\leftarrow A_1, \ldots, A_m$, а запрос N_k – вид $\leftarrow \underline{A}_1, \ldots, \underline{A}_m$. Такой SLDF'-вывод будем называть *нормализованным.*

ТЕОРЕМА* 3.2. *Если* $\mu(B^j\theta) > 0$, $j = 0,1, \ldots$ $\ldots, i-1$, *и* $\mu(\widehat{N_k F}) > 0$, *то*

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

где p_j *– условные вероятности, а* $B^j\theta$ *– условия правил* C_j, $j = 1, \ldots, i-1$.

§4. Вероятностные оценки запросов

4.1. Определим вероятностные оценки $\eta(N)$ и $\nu(N)$ запросов пространства вычислений программы Pr по правилу R . Рассмотрим SLDF-дерево некоторого запроса N пространства вычислений. Если SLDF-дерево не успешно, то оценки $\eta(N)$ и $\nu(N)$ не определены. Для успешного SLDF-дерева рассмотрим множество $\{SLDF'_1, \ldots, SLDF'_m\}$ всех успешных нормализованных SLDF'-выводов целей $N\theta_1, \ldots, N\theta_m$, у которых конечные запросы $N^1_{k1}, \ldots, N^m_{km}$ не содержат переменных. Если это множество пусто, то оценки $\eta(N), \nu(N)$ не определены.

Вычислим оценки $\eta_1, \ldots, \eta_m$, равные правой части неравенств следствия 3.4, для вероятностей $\mu(\widehat{N\theta_1}) \geq \eta_1, \ldots$ $\ldots, \mu(\widehat{N\theta_m}) \geq \eta_m$ запросов $N\theta_1, \ldots, N\theta_m$. Вычислим также оценки $\nu_1, \ldots, \nu_m$, равные правой части неравенств теоремы 3.2 для условных вероятностей запросов $\widehat{N\theta_1}, \ldots, \widehat{N\theta_m}$. Положим (ср. с [2]) $\eta(N) = \sup\{\eta_1, \ldots, \eta_m\}$, $\nu(N) = \sup\{\nu_1, \ldots, \nu_m\}$. 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 $SLDF'_1, \ldots, SLDF'_m$ состоит только в применении фактов, то, как следует из теоремы 3.2, он будет иметь оценку $\nu(N) = 1$. 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 $\nu(N)$. *Оценкой предсказания* запроса N будем называть величину $\nu(N)$. Если предсказание не определено, то и оценка предсказания $\nu(N)$ не определена.

4.2. Пусть $\mathfrak{M} = \langle U, \mu \rangle$ - вероятностная эрбранова модель, согласованная с классом G.

ОПРЕДЕЛЕНИЕ 4.2.1. *Программа* Pr *истинна* на эрбрановой модели $\mathfrak{N}, \mathfrak{N} \models Pr$ тогда и только тогда, когда каждое правило истинно на $\mathfrak{N}$. *Правило истинно* на $\mathfrak{N}$ тогда и только тогда, когда оно истинно на $\mathfrak{N}$ при любых *состояниях* [21] (при любых отображениях $\rho: X \to U$).

ОПРЕДЕЛЕНИЕ 4.2.2. *Программа* Pr *истинна на классе моделей* G тогда и только тогда, когда $\mathfrak{N} \models Pr$, $\mathfrak{N} \in G$.

Распространим вероятность μ на множество формул со свободными переменными F_0. Для $\varphi \in F_0 \setminus S$ положим $\mu(\varphi) = \inf_{\theta \in \Theta G} \{\mu(\varphi\theta)\}$, где ΘG - множество всех подстановок основных термов вместо переменных. Основание для этого определения можно найти в §7. Для правил $C = A \leftarrow B_1, \ldots, B_k$, $k \geq 0$, определим условную вероятность равенством $\mu(C) \overset{df}{=} \mu(A)/\mu(B_1 \,\&\, \ldots \,\&\, B_k)$, если правило не содержит переменных, и равенствами $\mu(C) \overset{df}{=} \mu(A/B_1 \,\&\, \ldots \,\&\, B_k) \overset{df}{=} \inf_{\theta \in \Theta G}\{\mu(A\theta/(B_1 \,\&\, \ldots \,\&\, B_k)\theta)\}$, если правило содержит переменные. Если $\mu((B_1 \,\&\, \ldots \,\&\, B_k)\theta) = 0$ для некоторой подстановки $\theta \in \Theta G$ или $\mu(B_1 \,\&\, \ldots \,\&\, B_k) = 0$ для

$B_1 \,\&\, \ldots \,\&\, B_k \in \mathfrak{B}$, то значение $\mu(C)$ не определено. При $k = 0$ правило $A \leftarrow$ рассматривается как правило $A \leftarrow \mathrm{true}$ с вероятностью посылки $\mu(\mathrm{true}) = 1$. Далее запись $\mu(C)$ всегда означает, что вероятность $\mu(C)$ определена. Обозначим через PRO, $PRO \subset PR$, множество всех правил сигнатуры σ, для которых вероятность μ определена.

ЛЕММА 4.2.1. *Из определения распространения меры* μ *на формулы* F_0 *следует, что* $\mu(\varphi\theta) \geq \mu(\varphi)$, *где* $\varphi \in F_0$, θ *– некоторая подстановка.*

ЛЕММА* 4.2.2. *Если программа* Pr *истинна на классе моделей* G, *то* $\mu(C) = 1$, $C \in Pr$.

§5. Детерминированные закономерности

Определим на множестве PR отношение $\sqsupset$ - "быть более общим". Обозначим множество всех подстановок, не являющихся перестановками, через Θt (тождественная подстановка принадлежит Θt).

ОПРЕДЕЛЕНИЕ 5.1. *Отношение* $C \sqsupset C'$, $C = A \leftarrow B_1, \ldots, \ldots, B_n$; $C' = A' \leftarrow B'_1, \ldots, B'_{n'}$, $n, n' \geq 0$, имеет место тогда и только тогда, когда существует подстановка $\theta \in \Theta t$ такая, что $A\theta = A'$ и $\{B_1\theta, \ldots, B_n\theta\} \subset \{B'_1, \ldots, B'_{n'}\}$, и либо θ не тождественная подстановка, либо $n < n'$.

ЛЕММА 5.1. *Отношение* $\sqsupset$ *является отношением строгого частичного порядка на* PR.

Обозначим через $W(G) \subset PR$ множество всех правил, истинных на G.

СЛЕДСТВИЕ 5.1. *Если* $C \in W(G)$ *и* $C \sqsupset C'$, *то* $C' \in W(G)$.

Пусть $W'(G)$, $W'(G) \subset W(G)$, - множество всех максимальных по отношению к $\sqsupset$ правил из $W(G)$. Правила из $W'(G)$ нельзя обобщить, сохраняя их истинность на G. Среди правил $W'(G)$ могут быть такие, которые истинны на G только потому, что посылка правила всегда ложна.

ОПРЕДЕЛЕНИЕ 5.2. *Детерминированной закономерностью, или* D-*правилом*, будем называть правило $A \leftarrow B_1, \dots, B_n \in W'(G)$, для которого утверждение $\exists x(B_1 \,\&\, \dots \,\&\, B_n)$ истинно хотя бы на одной модели из G.

§6. Вероятностные закономерности

ОПРЕДЕЛЕНИЕ 6.1. *Отношением вероятностной выводимости* назовем отношение $C' > C \overset{df}{\Leftrightarrow} C' \supset C \,\&\, \mu(C') < \mu(C)$.

ОПРЕДЕЛЕНИЕ 6.2. *Вероятностной закономерностью* (P-правилом) будем называть правило $C \in PRO$ такое, что из $C' \supset C$, $C' \in PRO$, следует $C' > C$.

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

ЛЕММА 6.1. *Если существует* C', $C' \supset C$, $\mu(C') \geq \mu(C)$, *то* $C \notin PR(\mathfrak{M})$.

ЛЕММА 6.2. *Если для правила* $C \in W(G) \setminus W'(G)$ *существует правило* $C' \in W(G)$, $C' \supset C$, $C' \in PRO$, *то* $C \notin PR(\mathfrak{M})$.

ЛЕММА* 6.3. D-*правило* C, $C \in PRO$, *является* P-*правилом, если из* $C' \supset C$, $C' \in PRO$, *следует* $\mu(\{\mathfrak{M} \mid \mathfrak{M} \in G, \mathfrak{M} \models \neg C'\}) > 0$.

§7. Предсказание и индуктивный синтез логических программ

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

Как следует использовать правила $C = A \leftarrow B_1, \ldots, B_k$, $k \geq 1$, для предсказания? Если при некоторой подстановке $\theta \in \Theta G$ посылка правила $(B_1 \,\&\, \ldots \,\&\, B_k)\theta$ истинна на некоторой случайно выбранной из G в соответствии с мерой μ модели $\mathfrak{N}$, т.е. $\{B_1\theta, \ldots, B_k\theta\} \subset D(\mathfrak{N})$, то заключение $A\theta$ истинно на $\mathfrak{N}$ с вероятностью $\mu(A\theta/(B_1 \,\&\ldots\&\, B_k)\theta) \geq$ $\geq \mu\,(A/B_1 \,\&\, \ldots \,\&\, B_k) = \mu(C)$. Вероятность $\mu(C)$, определенная в §4 для правил со свободными переменными, дает нам нижнюю границу вероятностей предсказания атома $A\theta$. Заметим, что предсказание нужно делать по данным $D(\mathfrak{N})$ какой-то одной, случайно выбранной из G, модели $\mathfrak{N}$. Обозначим множество всех Р-правил с посылкой, содержащей хотя бы один атом, через $P(\mathfrak{M}) \subset PR(\mathfrak{M})$.

ОПРЕДЕЛЕНИЕ 7.1. Для атома A сигнатуры σ и некоторых данных $D(\mathfrak{N})$ правило $C = A' \leftarrow B_1, \ldots, B_l$; $l \geq 1$, $C \in PRO$, не содержащее одинаковых переменных с атомом A, будем называть *наилучшим для предсказания атома* A по данным $D(\mathfrak{N})$ в вероятностной модели данных $\mathfrak{M}$, если:

1) существует подстановка $\theta \in \Theta G$ такая, что $\{B_1\theta, \ldots$ $\ldots, B_l\theta\} \subset D(\mathfrak{N})$, $A\theta = A'\theta$ и $\mu(C) > \mu(A\theta)$;

2) на правиле достигается максимум условной вероятности среди правил, удовлетворяющих условию 1;

3) правило C максимально по отношению к $\sqsupset$ среди правил, удовлетворяющих условиям 1,2.

ТЕОРЕМА 7.1. *Все наилучшие для предсказания какого-либо атома* A *сигнатуры* σ *по некоторым данным* $D(\mathfrak{N})$, $\mathfrak{N} \in G$, *в вероятностной модели данных* $\mathfrak{M}$ *правила являются вероятностными закономерностями с непустой посылкой, т.е. принадлежат множеству* $P(\mathfrak{M})$.

ДОКАЗАТЕЛЬСТВО. Пусть правило $C = A' \leftarrow B_1, \ldots, B_k$, $k \geq 1$, $C \in PRO$, является наилучшим для предсказания атома A по данным $D(\mathfrak{N})$ и для некоторой подстановки

$\theta' \in \Theta G$ выполняются соотношения $\{B_1\theta', \ldots, B_k\theta'\} \subset D(\mathfrak{N})$, $A\theta' = A'\theta'$, $\mu(C) > \mu(A'\theta')$. Предположим противное, что $C \notin P(\mathfrak{M})$ и, значит, $C \notin PR(\mathfrak{M})$. Отсюда следует, что существуют правило $C' \supset C$, $C' \in PRO$, $C' = A'' \leftarrow B_1', \ldots, B_l'$, и подстановка θ'', $A''\theta'' = A'$, $\{B_1'\theta'', \ldots, B_l'\theta''\} \subset \{B_1, \ldots, B_k\}$ такие, что $\mu(C') \geq \mu(C) > \mu(A'\theta') \geq \mu(A')$. Так как $A''\theta'' = A'$, то $\mu(A') \geq \mu(A'')$ и, следовательно, $\mu(C') > \mu(A'')$. Отсюда следует, что посылка правила C' не пуста и $l \geq 1$. Покажем, что правило C' лучше правила C для предсказания атома A, что противоречит условию. Включение $\{B_1'\theta''\theta', \ldots, B_l'\theta''\theta'\} \subset \{B_1\theta', \ldots, B_k\theta'\} \subset D(\mathfrak{N})$, равенство $A'\theta' = A''\theta''\theta'$ и неравенство $\mu(C') \geq \mu(C) > \mu(A'\theta')$ говорят о выполнении условия 1 определения 7.1. Соотношения $\mu(C') \geq \mu(C)$, $C' \supset C$, противоречат выполнению условий 2,3 для правила C.

ОПРЕДЕЛЕНИЕ 7.2. ПРОЛОГ-*программой, индуктивно синтезированной по данным* D *и вероятностной модели данных* $\mathfrak{M}$, будем называть множество правил $PR(\mathfrak{M}, \mathfrak{N}) = P(\mathfrak{M}) \cup D(\mathfrak{N})$, где $D(\mathfrak{N}) \in D$, $\mathfrak{N}$ - некоторая модель, случайно выбранная из G в соответствии с вероятностной моделью данных $\mathfrak{M}$.

§8. Вероятностный вывод

ОПРЕДЕЛЕНИЕ 8.1. Вероятностным выводом (Р-выводом) произвольного атома A сигнатуры $\mathfrak{S}$ будем называть максимальную последовательность правил $C_1 > C_2 > \ldots$; $C_1, C_2, \ldots \in P(\mathfrak{M})$; $C_i = A_i \leftarrow B_1^i, \ldots, B_{l_i}^i$, $i = 1, 2, \ldots$, такую, что атом A унифицируем с атомами $A_1, A_2, \ldots$ Если такой последовательности не существует, то Р-вывод пуст.

Каждому P-выводу соответствует последовательность подстановок $\theta_1, \theta_2, \ldots$ определения 5.1 отношения $\sqsupset$. Подстановку $\theta \overset{df}{=} \theta_1\theta_2\ldots$ будем называть *результатом вероятностного вывода*. Последнее правило в конечном P-выводе будем называть *результирующим*.

ЛЕММА 8.1. *D-правило в P-выводе может быть только результирующим.*

P-деревом вероятностного вывода атома A будем называть совокупность всех P-выводов (возможно, пустую) цели A.

ОПРЕДЕЛЕНИЕ 8.2. *P-предсказанием некоторого атома* A *сигнатуры* $\mathfrak{S}$ *программой* $PR(\mathfrak{M}, \mathfrak{N}) = P(\mathfrak{M}) \cup \cup D(\mathfrak{N})$ будем называть такой P-вывод $C_1 > C_2 > \ldots \ldots > C_i > \ldots$; $C_1, C_2, \ldots, C_i, \ldots \in P(\mathfrak{M})$ цели A, в котором:

1) существуют правило $C_i = A_i \leftarrow B_1^i, \ldots, B_{l_i}^i$ и подстановка θ такие, что $\{B_1^i\theta, \ldots, B_{l_i}^i\theta\} \subset D(\mathfrak{N})$; $A\theta = A_i\theta$; $\mu(A_i\theta) < \mu(C_i)$;

2) на правиле C_i достигается максимум условной вероятности $\mu(C_i)$ среди всех правил, удовлетворяющих условию 1, всех P-выводов цели A;

3) если P-дерево вывода цели A пусто или требуемой подстановки не существует, то P-предсказание не определено;

4) *результатом P-предсказания* будем называть подстановку $\theta_p \overset{df}{=} \theta_1\theta_2\ldots\theta_{i-1}\theta$, где $\theta_1, \theta_2, \ldots, \theta_{i-1}$ - подстановки P-вывода $C_1 > C_2 > \ldots > C_i$;

5) *оценкой P-предсказания* будем называть величину $\nu_p(A) \overset{df}{=} \mu(C_i)$. Если P-предсказание не определено, то оценка $\nu_p(A)$ не определена.

ТЕОРЕМА 8.1. *Р-предсказание атома* $\mathbf{A}$ *сигнатуры* $\mathfrak{S}$ *програмой* $\mathrm{PR}(\mathfrak{M}, \mathfrak{N}) = \mathrm{P}(\mathfrak{M}) \cup \mathrm{D}(\mathfrak{N})$ *определено тогда и только тогда, когда существует наилучшее для предсказания атома* $\mathbf{A}$ *правило* C *по данным* $\mathrm{D}(\mathfrak{N})$ *в вероятностной модели данных* $\mathfrak{M}$. *Если Р-предсказание атома* $\mathbf{A}$ *программой* $\mathrm{PR}(\mathfrak{M}, \mathfrak{N})$ *определено, то оно осуществляется Р-выводом, содержащим наилучшее для предсказания атома* $\mathbf{A}$ *правило* C. *Оценкой Р-предсказания является величина* $\nu_p = \mu(C)$.

ДОКАЗАТЕЛЬСТВО. Пусть C - наилучшее для предсказания атома $\mathbf{A}$ правило $C = \mathbf{A}' \leftarrow B_1, \dots, B_l$. Тогда, по теореме 7.1, $C \in \mathrm{P}(\mathfrak{M}) \subset \mathrm{PR}(\mathfrak{M}, \mathfrak{N})$. В силу свойства 1 определения 7.1 атом $\mathbf{A}$ унифицируем с атомом $\mathbf{A}'$. Отсюда следует, что существует Р-вывод, содержащий правило C. Из свойства 1 определения 7.1 следует свойство 1 определения 8.1. Следовательно, Р-предсказание атома $\mathbf{A}$ определено.

Если Р-предсказание определено, то существуют по крайней мере одно правило $C = \mathbf{A}' \leftarrow B_1, \dots, B_l$, $l \geq 1$, $C \in \mathrm{PRO}$ (так как $C \in \mathrm{PR}(\mathfrak{M}, \mathfrak{N})$) и подстановка θ такие, что $\mathbf{A}\theta = \mathbf{A}'\theta$, $\{B_1\theta, \dots, B_l\theta\} \subset \mathrm{D}(\mathfrak{N})$, $\mu(C) > \mu(\mathbf{A}'\theta)$. Таким образом, необходимые условия наилучшего для предсказания правила выполнены и, следовательно, наилучшее для предсказания правило существует.

Докажем вторую часть теоремы. Из первой части доказательства следует, что существует Р-вывод, содержащий наилучшее для предсказания атома $\mathbf{A}$ правило C. В силу свойства 2 определения 7.1 на этом правиле достигается максимум условной вероятности среди правил, удовлетворяющих условию 1 определения 7.1. Но, как показано в первой части доказательства, этому условию удовлетворяют все правила Р-дерева вывода цели $\mathbf{A}$, которые могут использоваться для предсказания (удовлетворяют условию 1 определения 8.2). Отсюда следует свойство 2 определения 8.2 Р-предсказания.

§9. Взаимосвязь вероятностного и логического выводов

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

ТЕОРЕМА 9.1. *Если атом* A *предсказывается программой* Pr *с оценкой* $\nu(A) > \mu(A\theta)$ *для любой подстановки* $\theta \in \Theta G$ *, то он P-предсказывается программой* $PR(\mathfrak{M},\mathfrak{N})$ *с оценкой P-предсказания* $\nu_p(A) \geq \nu(A)$.

ДОКАЗАТЕЛЬСТВО. По условию существует успешный SLDF-вывод $A\theta, N_1\theta, \ldots, N_k, \widehat{N_k F} \in \mathfrak{F}$, цели $A\theta$ в пространстве вычислений программы Pr такой, что $\mu(A\theta/\widehat{N_k F}) \geq \nu(A) > \mu(A\theta)$, $\theta \in \Theta G$, $N_k = \leftarrow \underline{B}_1, \ldots, \underline{B}_l$; $\{B_1\leftarrow, \ldots, B_l\leftarrow\} \subset Pr$, $l \geq 1$ (см. определение оценки ν и теорему 3.2).

Рассмотрим правило $C = A\theta \leftarrow B_1, \ldots, B_l$. Из условия $\nu(A) > \mu(A\theta)$, $\theta \in \Theta G$, следует, что $l \geq 1$. Так как $\mu(\widehat{N_k F}) > 0$, то $C \in PRO$. Кроме того, $\mu(C) \geq \nu(A) > \mu(A\theta)$, и, следовательно, выполнено условие 1 определения 7.1 наилучшего для предсказания атома A правила. Отсюда следует, что существует наилучшее для предсказания атома A правило CB и по теореме 8.1 P-предсказание атома A определено и $\nu_p(A) = \mu(CB)$. Так как правило C удовлетворяет условию 1 определения 7.1, то $\nu_p(A) = \mu(CB) \geq \mu(C)$ по условию 2 этого же определения.

Рассмотрим P-предсказание $C_1 > \ldots > C_i > \ldots$ цели A программой $PR(\mathfrak{M},\mathfrak{N}) = P(\mathfrak{M}) \cup D(\mathfrak{N})$ по наилучшему для предсказания атома A правилу $C_i = A_i \leftarrow B_1^i, \ldots, B_{l_i}^i$ и подстановке θ, $\{B_1^i\theta, \ldots, B_{l_i}^i\theta\} \subset D(\mathfrak{N})$, $A\theta = A_i\theta$, $\mu(A_i\theta) < \mu(C_i)$. Этому P-предсказанию поставим в соответствие нормализованный SLDF'-вывод, который будем обозначать как

SLDP(A)-вывод, $\leftarrow A\theta$; $\leftarrow B_1^i\theta, \ldots, B_{1_i}^i\theta; \ldots ; \leftarrow \underline{B}_1^i\theta, \ldots$

$\ldots, \underline{B}_{1_i}^i\theta$ цели $A\theta$ по правилам $C_i\theta$, $B_1^i\theta\leftarrow, \ldots, B_{1_i}^i\theta\leftarrow$.

По теореме 3.2 найдем оценку ν полученного SLDP(A)-вывода:

$$\mu(A\theta/\widehat{N_k F}) \geq \nu \overset{df}{=} 1-(1-p)\mu(B_1^i\theta \& \ldots \& B_{1_i}^i\theta)/\mu(\widehat{N_k F}) =$$

$$= 1-(1-p) = p, \quad \text{где } p = \mu(C_i).$$

Таким образом, $\nu(A) \geq \nu = \mu(C_i) = \nu_p(A)$ (по теореме 8.1). SLDP(A)-вывод цели A состоит в использовании наилучшего для предсказания атома A правила C_i и фактов $D(\mathfrak{N})$ программы.

ТЕОРЕМА 9.2. *Если атом* A *предсказывается программой* $PR(\mathfrak{M}, \mathfrak{N})$ *с оценкой* $\nu(A) > \mu(A\theta)$, $\theta \in \Theta G$, *и Р-предсказывается этой же программой с оценкой* $\nu_p(A)$, *то* $\nu(A) = \nu_p(A)$.

ДОКАЗАТЕЛЬСТВО. Выше, при введении понятия SLDP(A)-вывода, было доказано: если Р-предсказание атома A определено, то существует SLDP(A)-вывод атома A такой, что $\nu(A) \geq \nu_p(A)$. Обратное неравенство $\nu_p(A) \geq \nu(A)$ следует из теоремы 9.1, если в качестве программы Pr взять программу $PR(\mathfrak{M}, \mathfrak{N})$.

ТЕОРЕМА 9.3. *Если атом* A *предсказывается некоторой программой* Pr *с оценкой* $\nu(A) > \mu(A\theta)$, $\theta \in \Theta G$, *то он предсказывается программой* $PR(\mathfrak{M}, \mathfrak{N})$ *с оценкой* $\nu'(A) \geq \nu(A)$.

ДОКАЗАТЕЛЬСТВО. В силу теоремы 9.1 атом A Р-предсказывается программой $PR(\mathfrak{M}, \mathfrak{N})$ с оценкой $\nu_p(A) \geq \nu(A)$. Из предыдущих рассуждений следует, что в этом случае существует SLDP(A)-вывод атома A программой $PR(\mathfrak{M}, \mathfrak{N})$ с оценкой $\nu = \nu_p(A) \geq \nu(A) > \mu(A\theta)$, $\theta \in \Theta G$. Отсюда следует, что предсказание атома A программой $PR(\mathfrak{M}, \mathfrak{N})$ определено и для оценки предсказания $\nu'(A)$ имеет место соотношение $\nu'(A) = \nu_p(A)$ (теорема 9.2).

Процесс организации вычислений запросов $\leftarrow A_1, \ldots, A_k$, $k \geq 2$, в работе не рассматривается, хотя его можно было бы охватить, обобщив понятие вероятностной закономерности на утверждения $A_1 \& \ldots \& A_k \leftarrow B_1, \ldots, B_l$. Однако при этом множество вероятностных закономерностей увеличилось бы настолько, что предпочтительней было бы предсказывать запросы $\leftarrow A_1, \ldots, A_k$, $k \geq 2$, программой $PR(\mathfrak{M}, \mathfrak{N})$.

Причиной того, что в работе не рассматриваются запросы $\leftarrow A_1 \& \ldots \& A_k$, $k \geq 2$, является убеждение автора в том, что предсказание таких запросов является принципиально более сложной задачей, которая должна решаться по-другому.

В настоящее время на основе программ, реализующих метод обнаружения закономерностей [16,18], и на основе метода построения уточняющего графа [14,15] разрабатывается программная система PROL + PRED + SINT, для ЭВМ IBM PC, реализующая приведенные в работе идеи.

Литература

1. ВИТЯЕВ Е.Е. Предсказание и индуктивный синтез ПРОЛОГ-программ по вероятностной модели данных (Prediction and inductive synthesis of PROLOG-programms by a probabilistic model of data) //Препринт № 24 (Preprint №24),Новосибирск (Novosibirsk), 1990. - 34 с.

2. EMDEN M.N.,van.Quantitative deduction and its fixpoint theory //J.Logic Programming. - 1986. - Vol. 3, N 1.-P. 37-53.

3. FITTING M.C. Logic Programming on a Topological Bilattices //Fundamenta Informatica. - 1988. - Vol. 11. -P. 209-218.

4. SHAPIRO E. Logic Programs with Uncertainties: A Tool for Implementing Expert Systems //Proc. IJCAI'83, Williams Kauffman, 1983. -P. 529-532.

5. KIFER M.,, SUBRAHMANIAN V.S. Theory of Generalized Annotated Logic Programming and its Applications //Research Report, University of Maryland, USA, 1990.

6. NG R.T., SUBRAHMANIAN V.S. Probabilistic reasoning in Logic Programming //Proc. 5 th Symposium on Methodologies for Intelligent Systems, Knoxville, 1990. - P. 9-16.

7. NG R.T., SUBRAHMANIAN V.S. Annotation Variables and Formulas in Probabilistic Logic Programming //Technical report CS TR-2563. University of Maryland, 1990.

8. GAIFMAN H. Concerning measure in first order calculi //Israel journal of Math. - 1964. - Vol.2, N 1. - P. 1-18.

9. NILLSON N.J. Probability logic //Artif. Intell. - 1966. - Vol. 28, N 1. - P. 71-87.

10. HAILPERIN T. Probability Logic //Norte Dame J. of Formal Logic. - 1984. -Vol. 25, N 3. - P. 198-212.

11. SCOTT D.S., KRAUSS P. Assigning Probabilities to Logical Formulas //Aspects of Inductive Logic /Ed. J. Hintikka, P.Suppes. - North-Holland, 1966. - P. 219-264.

12. ADAMS Er.W. The logic of conditionals //An application of probability to deductive logic //Synthese Library. - 1975. - Vol. 86.

13. GONCHAROV S.S. ERSHOV Yu.L., SVIRIDENKO D.I. Semantic programming //10th World Congress Information Processing · 86, Dublin, Oct., 1986. - Amsterdam, 1986. - P. 1093-1100.

14. SHAPIRO E. Algorithmic Program Debugging //MIT Press, 1983. - P.204.

15. MATTHEW M. Huntbach An improved vershion of Shapiro's Model Inference system //Third International conference on Logic Programming (Lecture Nótes in Computer Science. Vol. 225), p. 180-187.

16. ВИТЯЕВ Е.Е. Метод обнаружения закономерностей и метод предсказания //Эмпирическое предсказание и распознавание образов. - Новосибирск, 1976. - Вып. 67: Вычислительные системы. - С. 54-68.

17. ВИТЯЕВ Е.Е., ПОДКОЛОДНЫЙ Н.Л. От экспертных систем к системам, создающим теории предметных областей //Компьютерный анализ структуры,функции и эволюции генетических макромолекул. - Новосибирск, 1989. - С. 264-282.

18. ВИТЯЕВ Е.Е. Обнаружение закономерностей (методология, метод, программная система SINTEZ). 1. Методология //Методологические проблемы науки. - Новосибирск, 1991. - Вып. 138: Вычислительные системы. - С. 26-60.

19. ЕРШОВ Ю.Л., САМОХВАЛОВ К.Ф. О новом подходе к философии математики //Структурный анализ символьных последовательностей. - Новосибирск, 1984. - Вып. 101: Вычислительные системы. - С. 141-148.

20. ЕРШОВ Ю.Л., САМОХВАЛОВ К.Ф. О новом подходе к методологии математики //Закономерности развития современной математики. - М.: 1987. - С. 85-105.

21. APT K.R. Introduction to Logic programming //Computer Science /Departament of Software Technology, Report CS-R874.

Поступила в ред.-изд.отд.
21 августа 1992 года

Доказательства лемм, теорем и следствий

<u>Доказательство леммы 2.1</u>. Так как $\mathfrak{D}$-булева подалгебра подмножества $\mathfrak{G}$ является кольцом множеств, то достаточно доказать, что $\nu(G(\varphi_1) \cup G(\varphi_2)) = \nu(G(\varphi_1)) + \nu(G(\varphi_2))$, если $G(\varphi_1) \cap G(\varphi_2) = \emptyset$; $\varphi_1, \varphi_2 \in \Phi$. Так как $\nu(G(\varphi_1) \cup G(\varphi_2)) = \nu(G(\varphi_1 \vee \varphi_2)) = \mu(\varphi_1 \vee \varphi_2)$, $\nu(G(\varphi_1)) = \mu(\varphi_1)$, $\nu(G(\varphi_2)) = \mu(\varphi_2)$, $G(\varphi_1) \cap G(\varphi_2) = G(\varphi_1 \& \varphi_2)$, 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 $\mu(\varphi_1 \& \varphi_2) = 0$.

<u>Доказательство леммы 4.1</u>.

1. $\mu(\widehat{N\theta}) = \mu(\widehat{N\theta} \& \widehat{B\theta}) + \mu(\widehat{N\theta} \& \neg\widehat{B\theta}) \le \mu(\neg\widehat{B\theta}) + \mu(\widehat{N\theta} \& \widehat{B\theta}) \le \mu(\neg\widehat{B\theta}) + \min\{\mu(\widehat{N_1\theta}), \mu(A\theta \& \widehat{B\theta})) .$

2. $\mu(\widehat{N\theta}/\widehat{N_1\theta}) = \mu(\widehat{N\theta} \& \widehat{B\theta})/\mu(\widehat{N_1\theta}) \le \mu(A\theta \& \widehat{B\theta})/\mu(\widehat{N_1\theta}) = p^{*}\mu(\widehat{B\theta})/\mu(\widehat{N_1\theta}) = p/\mu(\widehat{\underline{N}\theta}/\widehat{B\theta})$.

3. $\mu(\widehat{N\theta}) \ge \mu(\widehat{N\theta} \& \widehat{B\theta}) \ge \mu(\widehat{N\theta} \& \widehat{B\theta}) - \mu(\neg\widehat{\underline{N}\theta} \& \neg A\theta \& \widehat{B\theta})$. Выражение из п.2 утверждения леммы равно этому же выражению: $\mu(\widehat{N_1\theta}) - (1-p)\mu(\widehat{B\theta}) = \mu(\widehat{N_1\theta}) + \mu(A\theta \& \widehat{B\theta}) - \mu(\widehat{B\theta}) = \mu(\widehat{N_1\theta} \& A\theta) + \mu(\widehat{N_1\theta} \& \neg A\theta) + \mu(A\theta \& \widehat{B\theta} \& \underline{N}\theta) + \mu(A\theta \& \widehat{B\theta} \& \neg\widehat{\underline{N}\theta}) - \mu(\widehat{B\theta}) = \mu(\widehat{B\theta} \& \widehat{\underline{N}\theta} \& A\theta) + \mu(\widehat{B\theta} \& \widehat{\underline{N}\theta} \& \neg A\theta) + \mu(\widehat{B\theta} \& A\theta \& \widehat{\underline{N}\theta}) + \mu(\widehat{B\theta} \& A\theta \& \neg\widehat{\underline{N}\theta}) - \mu(\widehat{B\theta}) = \mu(\widehat{B\theta} \& \widehat{\underline{N}\theta} \& A\theta) - \mu(\widehat{B\theta} \& \neg\widehat{\underline{N}\theta} \& \neg A\theta) = \mu(\widehat{N\theta} \& \widehat{B\theta}) - \mu(\neg\widehat{\underline{N}\theta} \& \neg A\theta \& \widehat{B\theta})$.

4. $\mu(\widehat{N\theta}/\widehat{N_1\theta}) = \mu(\widehat{N\theta} \,\&\, \widehat{B\theta})/\mu(\widehat{N_1\theta}) \geq (\mu(\widehat{N_1\theta}) - (1-p)\mu(\widehat{B\theta}))/\mu(\widehat{N_1\theta}) = 1 - (1-p)/\mu(\underline{N}\theta/\widehat{B\theta})$ (см. доказательство п.2).

Доказательство следствия 4.1. Подставим в лемму 4.1 значение $p = 1$. Второе из неравенств 1 следует из того, что величина $\min\{\mu(\widehat{N_1\theta}),\ \mu(A\theta \,\&\, \widehat{B\theta})\}$ равна либо $\mu(\widehat{N_1\theta})$, либо $\mu(\widehat{B\theta})$. Во втором случае $\mu(\neg\widehat{B\theta}) + \mu(\widehat{B\theta}) = 1$.

Доказательство теоремы 4.1. Используем оценку 2 следствия 4.3, примененную к последнему шагу вывода от $\widehat{N_{k-1}\theta}$ к N_k . Получим $\mu(\widehat{N_{k-1}\theta} \,\&\, B^{k-1}\theta) \geq \mu(\widehat{N_k}) - (1-p_{k-1})\mu(B^{k-1}\theta)$, где $\mu(\widehat{N_k}) = \mu(\text{true}) = 1$,так как все атомы подчеркнуты. Рассмотрим вывод запроса $\widehat{N_{k-1}}\theta \,\&\, B^{k-1}\theta$ из запроса $\widehat{N_{k-2}}\theta \,\&\, B^{k-1}\theta$ посредством правила $C_{k-2}\theta$. Снова применим оценку 2 следствия 4.3. Получим: $\mu(\widehat{N_{k-2}}\theta \,\&\, B^{k-1}\theta \,\&\, B^{k-2}\theta) \geq \mu(\widehat{N_{k-1}}\theta \,\&\, B^{k-1}\theta) - (1-p_{k-2}) * \mu(B^{k-2}\theta)$. . Рассмотрим вывод запроса $\widehat{N_{k-2}}\theta \,\&\, B^{k-1}\theta \,\&\, B^{k-2}\theta$ из запроса $\widehat{N_{k-3}}\theta \,\&\, B^{k-1}\theta \,\&\, B^{k-2}\theta$ посредством правила $C_{k-3}\theta$ и т.д. Получим: $\mu(\widehat{N\theta} \,\&\, B^0\theta \,\&\, B^1\theta \,\&\ldots\, \&\, B^{k-1}\theta) \geq \mu(\widehat{N_1\theta} \,\&\, B^1\theta \,\&\ldots \,\&\, B^{k-1}\theta) - (1-p_0)\mu(B^0\theta)$. Подставляя левые части неравенств в правые, получим оценку

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

Покажем, что если из конъюнкции $\widehat{N\theta} \,\&\, B^0\theta \,\&\, \ldots \,\&\, B^{k-1}\theta$ удалить все константы true, то получим конъюнкцию $\widehat{N\theta} \,\&\, A^0\theta \,\&\, A^1\theta \,\&\ldots\, \&\, A^{k-1}\theta$. 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 $B^0\theta \,\&\ldots\, \&\, B^{k-1}\theta$ содержится в конъюнкции $A^0\theta \,\&\ldots\,\&\, A^{k-1}\theta$. С другой стороны, каждый атом

$A^i\theta$, $i=0,1,\dots,k-1$, содержится либо в $\widehat{N\theta}$, либо в правой части одного из правил $C_i\theta$, $i=0,1,\dots,k-1$.

Доказательство теоремы 4.2. Проводится аналогично доказательству теоремы 4.1, но для нормализованного вывода и начинается с запроса i. Первое неравенство имеет вид: $\mu(\widehat{N_{i-1}\theta}\ B^{i-1}\theta) \geq \mu(\widehat{N_i\theta}) - (1-p_{i-1})\mu(B^{i-1}\theta)$, где $\mu(\widehat{N_i\theta}) = \mu(\widehat{N_k F})$. Далее, рассуждая как в теореме 4.1, получим неравенство $\mu(\widehat{N\theta}\,\&\,B^0\theta\ \&\ \dots\ \&\ B^{i-1}\theta) \geq \mu(\widehat{N_k F}) - \sum_{j=1}^{i-1}(1-p_j)\mu(B^j\theta)$. Так как $\mu(\widehat{N\theta}\ \&\ B^0\theta\ \&\ \dots\ \&\ B^{i-1}\theta) \leq \mu(\widehat{N\theta}\ \&\ \widehat{N_k F})$, то $\mu(\widehat{N\theta}/\widehat{N_k F}) = \mu(\widehat{N\theta}\ \&\ \widehat{N_k F})/(\widehat{N_k F}) \geq (\mu(\widehat{N_k F}) - \sum_{j=1}^{i-1}(1-p_j)\mu(B^j\theta))/\mu(\widehat{N_k F})$.

Доказательство леммы 5.2.2. Пусть $C = A \leftarrow B_1,\dots,B_k$; $C \in Pr$, $k \geq 0$; $\mu(C) = \inf\limits_{\theta\in\Theta G}\{\mu(A\theta/(B_1\ \&\ \dots\ \&\ B_k)\theta)\}$. Рассмотрим правило $A\theta \leftarrow (B_1,\dots,B_k)\theta$, $\theta\in\Theta G$, и условную вероятность $\mu(A\theta/(B_1\ \&\ \dots\ \&\ B_k)\theta)$. Каждая подстановка $\theta\in\Theta G$ однозначно определяет некоторое состояние $\rho: X \to U$. Так как программа Pr истинна на G для любого состояния, то для любой подстановки $\theta\in\Theta G$ будем иметь $G(A\theta \leftarrow (B_1\ \&\dots\&\ B_k)\theta) = G$. Так как мера μ согласована с классом моделей G, то из $G(\varphi_1) = G(\varphi_2)$ следует $\mu(\varphi_1) = \mu(\varphi_2)$ и, следовательно, $\mu(A\theta \leftarrow (B_1\ \&\ \dots\ \&\ B_k)\theta) = 1$, откуда $\mu(A\theta/(B_1\ \&\ \dots\ \&\ B_k)\theta) = 1$. Поэтому $\mu(C) = 1$, если $\mu(C)$ определена.

Доказательство леммы 7.3. В силу леммы 5.2.2, $\mu(C) = 1$. Докажем, что из $C' \supset C$, $C' \in PRO$, следует $\mu(C') < \mu(C) = 1$. По условию $\mu(\{\mathfrak{M} \mid \mathfrak{M}\in G,\ \mathfrak{M}\models \neg C'\}) > 0$. Отсюда следует, что $\mu(B'\,\&\,A') < \mu(A')$ и $\mu(C') < 1$, $C' = B' \leftarrow A'$.