ПРЕДСКАЗАНИЕ КАК ВЫЧИСЛЕНИЕ[1]

Витяев Е.Е.

Введение

Предсказание является одним из важнейших понятий в науке. В настоящее время все чаще требуется осуществлять предсказания на основании индуктивно выведенных знаний. Обнаружением таких знаний занимаются интенсивно развиваемые направления Machine Learning (Машинное Обучение (МО)) и Knowledge discovery in Data Bases and Data Mining (Извлечение Знаний из Баз Данных (ИЗБД)). Знания, получаемые такими методами, имеют статус вероятностных законов. Предсказания, получаемые по вероятностным законам, обычно описываются индуктивно-статистическим выводом I-S-выводом (Inductive-Statistical inference).

Понятие предсказания также важно для моделирования когнитивных процессов, таких как антиципация, воображение, образ, образ мира, мышление, развиваемые в рамках теории деятельности. С точки зрения когнитивных процессов мозг – это предвосхищающее устройство, а не логическое.

Гемпелем [1, 2] отмечал, что предсказания по I-S выводу статистически двусмысленны. Во избежание двусмысленности он ввел дополнительное требование максимальной специфичности RMS (Requirement of Maximum Specificity). Однако, это требование не решает полностью проблемы статистической двусмысленности. Статистическая двусмысленность приводит к необходимости разрабатывать логики, работающие с противоречивыми знаниями, например, логики по умолчанию (default logic [3]) или паранепротиворечивые логики.

Другой проблемой существующей для I-S выводов является несогласованность вероятностных оценок с логическим выводом. Известно, что вероятностные оценки высказываний резко падают в процессе логического вывода и эти оценки нельзя улучшить. Вероятность и вывод по существу не согласованны. Вычислению этих оценок посвящены работы по вероятностной логике [4–15]. Есть работы, в которых вероятность рассматривается как значение истинности утверждений, а процесс логического вывода обобщается до так называемых “количественных дедукций” (дедуктивных систем, в которых значения истинности непрерывны и принимают значения в интервале [0,1]) [5–8, 14]. В работах [7–8, 15] описываются довольно богатые формальные системы, содержащие как частные случаи основные известные “количественные дедукции”. Но, несмотря на значительный прогресс в разработке формальных систем все они основаны на **логическом выводе знаний** - вероятностные оценки высказываний вычисляются **после** получения логического вывода. Анализ вероятностных оценок утверждений в процессе логического вывода показывает, что они всегда уменьшаются и, как правило, существенно. И это не случайно. Дело в том, что использование правил вывода неявно предполагает абсолютную достоверность (или гипотетичность) используемых в выводе знаний и отвечает требованиям сохранения истинности, а не вероятности. Только для достоверного знания можно применять правила вывода неограниченное число раз, и только в этом случае они действительно являются правилами вывода - сохраняют значения

[1] Работа поддержана грантом РФФИ 05-07-90185в, Интеграционными проектами СО РАН №1, №115, Программой президента Российской Федерации поддержки научных школ 4413.2006.1

истинности. Неограниченное применение правил вывода к вероятностным знаниям неприменимо, т.к. может приводить к знаниям со сколь угодно низкой оценкой вероятности, которые фактически не являются знаниями. Логический вывод не предназначен для сохранения значения вероятности.

Эта проблема в настоящее время обсуждается как необходимость **синтеза логики и вероятности**. В 2002 в королевском колледже Лондона было проведено рабочее совещание “Combining Probability and Logic”. В аннотации к нему говорится: “Artificial intelligence is one key discipline in which probability theory competes with other logics for application. It is becoming vitally important to evaluate and integrate systems that are based on very different approaches to reasoning, and there is strong demand for theoretical understanding of the relationships between these approaches”. Во введении к спецвыпуску «Special issue on Combining Probability and Logic» журнала Journal of Applied Logic (2003), посвященному этому совещанию, редакторы Jon Williamson и Dov Gabbay пишут, что существующие подходы к синтезу логики и вероятности можно разбить на две части: “One approach is to argue that ***probability is logic***, which requires showing that probability is a determinate relation between statements. Kyburg, Howson and Paris and Vencovská appeal to the concepts of frequency, consistency and entropy respectively to determine this relation. Alternatively one can explore other formalisms which ***interface between probability and logic***: argumentation in the case of Fox and Kohlas; default reasoning in the case of Bourne and Weydert.” Однако, эти исследования, с нашей точки зрения не решают проблемы синтеза логики и вероятности.

Первый шаг к синтезу логики и вероятности был сделан в “количественных дедукциях”, где значения истинности были обобщены до значений вероятности. Но в количественных дедукциях сохраняется очевидное несоответствие: при обобщении значений истинности, не обобщаются правила вывода. Правила вывода применяются для сохранения значений истинности, но если значения истинности обобщены, то и правила вывода должны быть обобщены так, чтобы сохранять эти обобщенные значения, а не значения истинности. Каким образом можно обобщить вывод? Для решения данной проблемы нужен более радикальный пересмотр самого подхода к выводу, чем это делается в работах workshop. С нашей точки зрения предсказание нельзя вывести и соединить с каким-то процессом вывода, его можно только **вычислить** без использования какого-либо вывода. Замена вывода предсказания на его вычисление является радикальной сменой парадигмы предсказания.

Рассмотрим процесс вывода с точки зрения “семантического” подхода к программированию [16]. Идея семантического программирования состоит в том, что **процесс вывода** можно рассматривать с семантической точки зрения как **вычисление** истинности утверждений на модели. При таком взгляде на вывод его можно обобщить, определяя новые взаимоотношения высказываний и модели. Можно рассмотреть вывод не только как проверку истинности на модели, но и как поиск фактов в модели, предсказывающих интересующее нас высказывание с максимальной вероятностью, или как поиск наиболее подтверждающих фактов, и т.д. Такие выводы будем называть семантическими. Такой вывод возможен потому, что истинность имеет только два значения, а вероятность, подтверждённость, достоверность и т.д. имеют континуум значений. Поэтому, если использовать не значения истинности: истина и ложь, среди которых не имеет смысла искать “более истин-

ное”, а континуум значений, то поиск наиболее вероятного, достоверного и т.д. утверждения уже имеют самостоятельный смысл, которого нет в обычном понимании вывода. Как мы покажем, при таком выводе мы даже не нуждаемся в правилах вывода.

В работе мы определяем **семантический вероятностный вывод** (СВВ)**,** основанный на приведенной идее семантического программирования, который осуществляет вычисление предсказаний. Для СВВ вероятностные оценки высказываний не только не падают в процессе вывода, а наоборот строго возрастают. Кроме того, в нём синтезируется логика, вероятность и обучение, а также решается проблема статистической двусмысленности.

Проблема синтеза логики, вероятности и обучения также обсуждаются. Например, в широко цитируемой работе L.De Raedt and K.Kersting «Probabilistic logic learning» [17] говорится, что «One of the central open questions in data mining and artificial intelligence, concerns probabilistic logic learning, i.e. the integration of relational or logical representations, probabilistic reasoning mechanisms with machine learning and data mining principals». Мы покажем, как в семантическом вероятностном выводе синтезируется вероятность, логика и обучение.

§1. Проблема статистической двусмысленности.

В индуктивном выводе мы можем получить утверждения, из которых выводятся противоречивые утверждения. Приведем классический пример. Предположим, что в теории T есть следующие высказывания:

(Л1) - ‘Почти все случаи заболевания стрептококком быстро вылечиваются инъекцией пенициллина’;

(Л2) - ‘Почти всегда устойчивая к пенициллину стрептококковая инфекция не вылечивается после инъекции пенициллина’;

(С1) - ‘Джейн Джонс заболел стрептококковой инфекцией’;

(С2) - ‘Джейн Джонс получил инъекцию пенициллина’;

(С3) - ‘Джейн Джонс имеет устойчивую к пенициллину стрептококковую инфекцию’.

Из этой теории можно вывести два противоречивых утверждения: одно, объясняющее, почему Джейн Джонс выздоровеет быстро (E), и другое, объясняющее отрицание первого – почему Джейн Джонс не выздоровеет быстро (¬E).

<table>
<tr><td colspan="3">Объяснение 1</td><td colspan="2">Объяснение 2</td></tr>
<tr><td>L1</td><td rowspan="3">[r]</td><td rowspan="3"></td><td>L2</td><td rowspan="3">[r]</td></tr>
<tr><td>C1,C2</td><td>C2,C3</td></tr>
<tr><td>E</td><td>¬E</td></tr>
</table>

Условия обоих объяснений не противоречат друг другу, оба они могут быть истинны. Тем не менее, их выводы противоречат друг другу. Поэтому набор правил T приводит к противоречивым выводам.

Гемпель надеялся решить эту проблему, требуя, что бы статистические законы удовлетворяли требованию максимальной специфичности (они должны содержать всю относящуюся к рассматриваемому вопросу информацию). В нашем примере условие C3 второго объяснения опровергает условие первого объяснения в силу

того, что закон L1 не максимально специфичен по отношению ко всей информации теории T относительно Джонса. Потому теория T может объяснить только утверждение $\neg$E, но не E.

§2. Модели предсказания.

Вывод предсказания обычно 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.

Дедуктивно-номологическая модель может быть представлена следующей схемой.

$L_1,\ldots,L_m$		
$C_1,\ldots,C_n$		
G		

i) $L_1,\ldots,L_m$ - множество законов;
ii) $C_1,\ldots,C_n$ - множество фактов;
iii) G – предсказываемое высказывание;
iv) $L_1,\ldots,L_m, C_1,\ldots,C_n \vdash G$;
v) множество $L_1,\ldots,L_m, C_1,\ldots,C_n$ непротиворечиво;
vi) $L_1,\ldots,L_m \nvdash G$, $C_1,\ldots,C_n \nvdash G$;
vii) Законы $L_1,\ldots,L_m$ содержат только кванторы всеобщности. Множество фактов $C_1,\ldots,C_n$ – бескванторные формулы;

Индуктивно-статистическая модель аналогична предыдущей с тем отличием, что иначе формулируется свойство vii и добавляется свойство RMS:

$L_1,\ldots,L_m$		
$C_1,\ldots,C_n$	[r]	
G		

Удовлетворяет условиям i-vi дедуктивно-номологической модели.
vii) множество $L_1,\ldots,L_m$ содержит статистические законы. Множество фактов $C_1,\ldots,C_n$ – бескванторные формулы;
viii) RMS: Все законы $L_1,\ldots,L_m$ максимально специфичны.

По Гемпелю [1, 2] *требование максимальной специфичности* RMS определяется следующим образом: I-S вывод вида

p(G;F) = r		
F(***a***)	[r]	
G(***a***)		

является приемлемым при состоянии знания K, если для каждого класса H, для которого оба нижеследующих высказывания принадлежат K

$\forall x(H(x) \Rightarrow F(x))$,

H(***a***),

существует статистический закон p(G;H) = r’ в K такой, что r = r’.

Идея требования RMS состоит в том, что если F и H оба содержат объект ***a***, и H является подмножеством F, то H обладает более специфической информацией об объекте ***a***, чем F и, следовательно, закон p(G;H) должен предпочитаться закону p(G;F). Тем не менее, закон p(G;H) имеет ту же вероятность, что и закон p(G;F)

§3. Понятие закона для дедуктивно-номологического вывода

Определим понятие закона, которое, с одной стороны, будет использовано для вывода теории предметной области и может быть использовано в дедуктивно-номологическом выводе предсказаний из этой теории, а, с другой стороны, может быть получено, определяемым далее семантическим вероятностным выводом.

Представим предметную область *эмпирической системой* $M = \langle A,W\rangle$, *сигнатуры* Ω, где A – основное множество эмпирической системы, $W = \langle P_1,...,P_k\rangle$, $k > 0$ - множество *предикатов*, определенных на A, $\Omega = \langle P_1,...,P_k\rangle$ - множество *предикатных символов* местности $n_1,...,n_k$.

Рассмотрим задачу обнаружения *теории* Th(M) *эмпирической системы* M сигнатуры Ω (совокупность всех истинных на M высказываний). Будем предполагать, что теория Th(M) представляет собой совокупность универсальных формул.

Для дальнейших рассмотрений необходимо выделить фрагмент *языка первого порядка* $L(\Omega)$ сигнатуры Ω включающий: множество X *свободных переменных*; множество $U(\Omega)$ всех *атомарных формул*, (атомов) вида $P(x_1,...x_n)$, $x_1,...x_n \in X$; множество $\Re(\Omega)$ *утверждений* языка $L(\Omega)$, полученное замыканием множества $U(\Omega)$ относительно логических операций &,v,¬.

Известно, что совокупность универсальных формул логически эквивалентна совокупности формул следующего вида, которые будем называть правилами

$$\forall x_1,\ldots,x_2(A_1\&...\&A_k \Rightarrow A_0),\ k \geq 0, \quad (1)$$

где $A_0,A_1,\ldots, A_k$ – *литеры* (атомарные формулы или их отрицания).

Поэтому без ограничения общности можно считать, что теория Th(M) представляет собой совокупность правил вида (1).

Проанализируем задачу обнаружения теории Th(M). Что можно сказать об истинности высказываний вида (1) теории Th(M) на эмпирической системе M, опираясь только на их логический анализ. Во-первых, правило $C = (A_1\&...\&A_k \Rightarrow A_0)$ может быть истинным на эмпирической системе только потому, что посылка правила всегда ложна. Во-вторых, правило C может быть истинно на эмпирической системе только потому, что некоторое логически более сильное "подправило", содержащее только часть посылки и то же заключение, истинно на эмпирической системе. Поэтому теория Th(M) может быть истинна на эмпирической системе M потому, что на ней истинна некоторая система подправил из истинности которых следуют все остальные высказывания теории. Найдем эти правила.

Теорема 1 [18]. Правило $C = (A_1\&...\&A_k \Rightarrow A_0)$ логически следует из любого правила вида:

$(A_{i1}\&...\&A_{ih} \Rightarrow A_0)$,

где $\{A_{i1},...,A_{ih}\} \subset \{A_1,...,A_k\}$, $0 \leq h < k$;

$(A_{i1}\&...\&A_{ih} \Rightarrow A_0) \vdash (A_1\&...\&A_k \Rightarrow A_0)$.

Определение 1. *Подправилом* некоторого правила C вида (1) будем называть любое логически более сильное правило, определенное в теореме 1.

Легко видеть, что любое подправило также имеет вид (1).

Следствие 1. Если некоторое подправило правила C истинно на эмпирической системе M, то и само правило C истинно на M.

Из логики и методологии науки хорошо известно, что те высказывания следует считать **законами**, которые, при одинаковой их подтвержденности на экспериментальных данных, наиболее фальсифицируемы, просты и содержат наименьшее число параметров. В нашем случае все эти свойства, которые обычно трудно определить, следуют из определения логической силы высказывания. Подправило является одновременно и логически более сильным высказыванием, чем само правило; и более фальсифицируемым, так как содержит более слабую посылку и, следовательно, применимо к большему объему данных и тем самым в большей степени подвержено фальсификации; и более простым, так как содержит меньшее число атомарных высказываний, чем правило; и содержит меньшее число "параметров", так как лишние атомарные высказывания тоже можно считать параметрами "подстройки" высказывания под данные. Поэтому резонно ввести понятие закона следующим образом.

Определение 2. *Законом* эмпирической системы $M = \langle A,W\rangle$ будем называть истинное на M правило C вида (1), для которого каждое из его подправил уже не истинно на M.

Обозначим через L множество всех законов эмпирической системы M.

Теорема 2 [18]. $L \vdash Th(M)$.

Таким образом, теория эмпирической системы выводится из множества всех законов. Тем самым из множества L выводятся и все предсказания, которые можно сделать используя теорию Th(M) в дедуктивно-номологическом выводе.

§4. Определение вероятности на высказываниях. Вероятностные законы.

Обобщим понятие закона на вероятностный случай. Как известно [12] вероятность в языке первого порядка может быть введена различным образом: как вероятность на модели, либо как вероятность на множестве возможных миров. Мы определим её простейшим образом как вероятность на модели путем определения вероятности объектов модели. *Вероятность* на модели $M = \langle A,W\rangle$ определим как отображение $\mu: A \to [0,1]$, удовлетворяющее условиям:

$$\sum_{a\in A} \mu(a)=1 \ \text{ и } \ \mu(a) \neq 0,\ a \in B,$$
$$\mu(D)=\sum_{b\in D} \mu(b),\ D \subseteq A, \qquad (2)$$
$$\mu^n(a_1,...,a_n)=\mu(a_1)\times...\times\mu(a_n)$$

Под формулой $\hat{\varphi}$, мы будем понимать интерпретацию формулы $\varphi \in \Re(\Omega)$ на модели $M = \langle A,W\rangle$, когда все предикатные символы сигнатуры Ω заменены предикатами из W. Под $\nu\hat{\varphi}$ мы будем понимать формулу $\hat{\varphi}$ в которой вместо переменных подставлен набор объектов из множества A некоторым состоянием $\nu: X \to A$. В частности, $\nu\hat{P}(x_1,...,x_n)^{\varepsilon} = P(a_1,...,a_n)^{\varepsilon}$, $\nu(x_1) = a_1, \ldots, \nu(x_n) = a_n$. Если $x_1,\ldots,x_n$ – все переменные выказывания $\varphi \in \Re(\Omega)$, то его вероятность η определяется следующим образом [12]:

$\eta(\varphi) = \mu^n(\{(a_1,\dots,a_n) \mid \nu\,\hat{\varphi}$ истинно на M, $\nu(x_1) = a_1,\dots, \nu(x_n) = a_n\})$.

Обобщим понятие закона на вероятностный случай. Сделаем это таким образом, что бы понятие закона являлось частным случаем понятия вероятностного закона. Вспомним определение закона на M, законом является такое истинное на M правило, все подправила которого ложны на M. Можно иначе переформулировать понятие закона на M, законами являются такие правила истинные на M, которые нельзя более упростить и логически усилить, сохраняя их истинность. Это свойство неупрощаемости позволяет сформулировать закон не только в терминах истинности, но и вероятности.

Для правила $C = (A_1\&...\&A_k \Rightarrow A_0)$ вида (1) определим его вероятность как $\eta(C) = \eta(A_0/A_1\&...\&A_k) = \eta(A_0\&A_1\&...\&A_k)/\eta(A_1\&...\&A_k)$, $\eta(A_1\&...\&A_k) > 0$. В частном случае правила $C = \Rightarrow A_0$ его вероятностью будет $\eta(A_0)$.

Теорема 3 [18]. Для правила $C = (A_1\&...\&A_k \Rightarrow A_0)$ вида (1) следующие два условия эквивалентны:

1. правило C является законом на M;
2. a) вероятность правила C определена ($\eta(A_1\&...\&A_k) > 0$) и $\eta(C) = 1$;
   b) вероятность $\eta(C)$ правила строго больше вероятностей каждого из его подправил.

Данная теорема дает нам эквивалентное определение закона на M в терминах вероятностей.

Определение 3. *Вероятностным законом в детерминированном случае* будем называть правило $C = (A_1\&...\&A_k \Rightarrow A_0)$ вида (1), удовлетворяющее условиям:

a) вероятность правила C определена ($\eta(A_1\&...\&A_k) > 0$) и $\eta(C) = 1$;
b) вероятность $\eta(C)$ правила C строго больше условных вероятностей каждого из его подправил.

Следствие 2. Правило $C = (A_1\&...\&A_k \Rightarrow A_0)$ вида (1) является вероятностным законом на M в детерминированном случае тогда и только тогда, когда оно является законом на эмпирической системе M.

Обобщим определение 3.

Определение 4 [18]. *Вероятностным законом* на M будем называть правило $C = (A_1\&...\&A_k \Rightarrow A_0)$ вида (1), удовлетворяющее условию б): вероятность $\eta(C)$ правила C определена и строго больше вероятностей каждого из его подправил.

Обозначим через LP множество всех вероятностных законов.

Определение 5. *Сильнейшим Вероятностным Законом* (СВЗ) будем называть вероятностный закон C, который не является подправилом никакого другого вероятностного закона.

Обозначим через СВЗ множество всех сильнейших вероятностных законов.

Предложение 1 [18]. $L \subset$ СВЗ $\subset LP$.

§5. Вывод предсказаний в логическом программировании.

Представим процесс предсказания I-S выводом в рамках логического программирования. В логическом программировании вывод предсказания можно рассматривать как вычисление. Предсказание в логическом программировании формулируется как запрос G к множеству законов $L_1,\dots, L_m$ вида (1) и фактов $C_1,\dots,C_n$ представленных правилами $(\Rightarrow C_1),\dots,(\Rightarrow C_n)$.

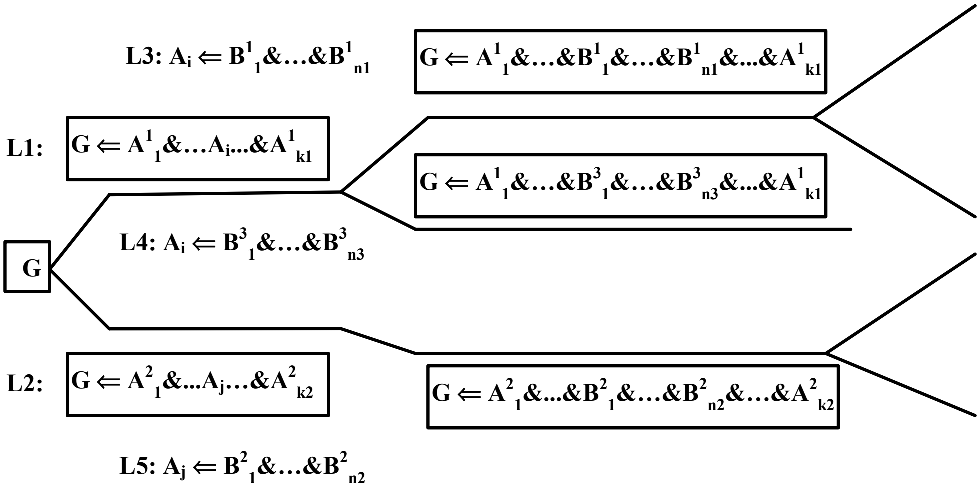

Рис. 1

В процессе вычисления ответа на запрос $G(x_1,\ldots,x_n)$ вычисляется:

1. вывод $\{L_1,\ldots, L_m, C_1,\ldots,C_n\} \vdash \exists x_1,\ldots,x_nG$;

2. набор термов $t_1,\ldots,t_n$ для которых $\{L_1,\ldots, L_m\ C_1,\ldots,C_n\} \vdash G[x_1/t_1,\ldots,x_n/t_n]$.

Процесс вычисления (вывода) предсказания для хорновых формул можно представить в виде дерева рис. 1. В нём запрос G унифицируется с законами L1 и L2 так, чтобы заключение правила совпало с запросом G. В этом случае запрос выводится из истинности атомов посылки $A^1{}_1\&\ldots A_i\ldots\&A^1{}_{k1}$ или посылки $A^2{}_1\&\ldots A_j\ldots\&A^2{}_{k2}$. Если среди законов $L_1,\ldots, L_m$ есть правила L3, L4, L5, которые унифицируются с некоторыми атомами A_i или A_j, то посылки этих правил $B^1{}_1\&\ldots\&B^1{}_{n1}$ или $B^2{}_1\&\ldots\&B^2{}_{n2}$ или $B^3{}_1\&\ldots\&B^3{}_{n3}$ подставляются вместо соответствующих атомов A_i или A_j . Если какие-то правила L3, L4 или L5 являются фактами вида $\Rightarrow A_i$, то соответствующий атом после унификации удаляется из посылки правила L1. В результате получаются правила, стоящие в правой части дерева вывода. Вывод (вычисление) заканчивается, когда найдена такая ветвь дерева вывода. которая содержит правило ($G \Leftarrow$).

Семантический подход к логическому программированию состоит в рассмотрении теоретико-модельной семантики логических программ, состоящей в том, что факты являются высказываниями некоторой модели, эмпирической системы M, представляющей предметную область. В этом случае вывод (вычисление) предсказания состоит в нахождении таких фактов в модели (нашем мире), из истинности которых выводится по законам $L_1,\ldots, L_m$ предсказываемое высказывание. Такой процесс вывода можно рассматривать как вычисление истинности предсказываемого факта G на модели представляющей наш мир.

§6. Семантический вероятностный вывод предсказаний

Как упоминалось во введении, процесс вывода рис. 1 можно обобщить и рассмотреть не только проверку истинности на модели, а, например, поиск фактов в модели предсказывающих высказывание G с максимальной вероятностью.

Определим семантический вероятностный вывод, реализующий эту идею и следующий семантическому подходу к программированию [16].

Определение 6 [19]. *Семантическим Вероятностным Выводом* некоторого сильнейшего вероятностного закона C_n будем называть такую последовательность вероятностных законов $C_1 \sqsubset C_2 \sqsubset \ldots \sqsubset C_n$, что:

$$\begin{gathered}C_1, C_2, \ldots, C_n \in LP,\ C_i = (A^i_1 \& \ldots \& A^i_{ki} \Rightarrow G),\ i = 1,2,\ldots n,\ n \geq 1, \\ \text{правило } C_i \text{ является подправилом правила } C_{i+1}, \\ \eta(C_i) < \eta(C_{i+1}),\ i = 1,2,\ldots n-1, \\ C_n - \text{СВЗ-правило.}\end{gathered} \tag{1}$$

Предложение 2. Любой вероятностный закон принадлежит некоторому семантическому вероятностному выводу. Любой сильнейший вероятностный закон находится в конечной ветке некоторого семантического вероятностного вывода.

Следствие 3. Для любого закона из L существует его семантический вероятностный вывод.

Рассмотрим множество всех семантических вероятностных выводов некоторого факта G. Это множество можно представить семантическим вероятностным деревом вывода факта G (см. Рис 2). Сравнение рис. 1 и рис. 2 показывает, что по структуре семантический вероятностный вывод полностью аналогичен выводу предсказания в логическом программировании за исключением того, что для проведения семантического вероятностного вывода не нужны правила L1, L2, L3, L4, L5 и, значит, не нужен логический вывод, представленный на рис. 1. Единственно, что нужно – это уточнение посылки правил путем добавления дополнительных условий в посылку так, чтобы оценка вероятности предсказания атома G строго увеличивалась. Поскольку вероятность является числом, то для увеличения оценки вероятности предсказания не нужен логический вывод, можно просто искать в модели факты увеличивающие вероятность предсказания атома G. Как мы докажем далее, такое уточнение правил позволяет найти: (1) максимально специфические правила (определение приведено ниже) (2) решение проблемы статистической двусмысленности (3) определить предсказание без противоречий (4) доказать, что максимально специфические правила удовлетворяют требованию максимальной спе-

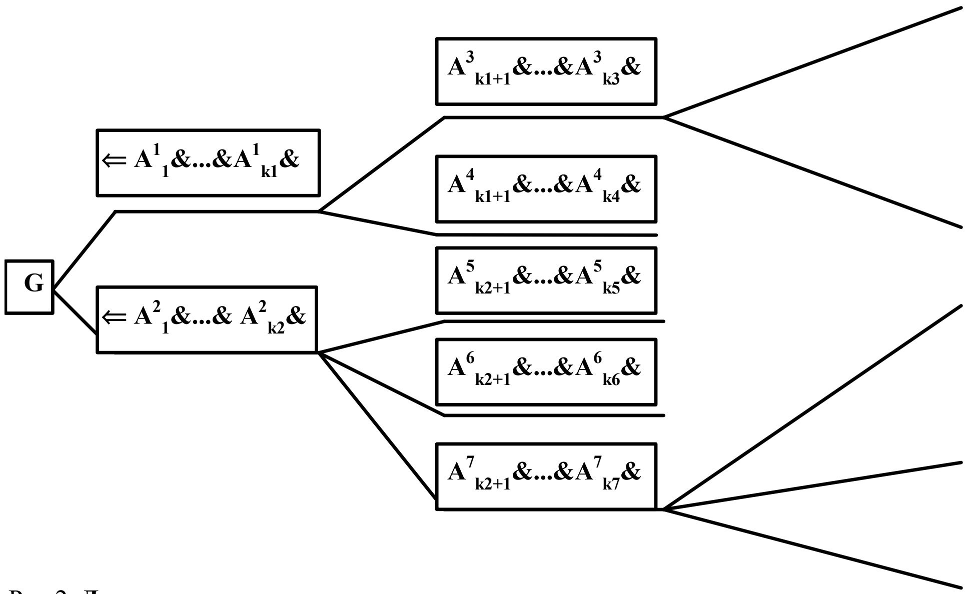

Рис 2. **Дерево семантического вероятностного вывода**

цифичности (определение приведено ниже).

Определение 7 [20]. *Максимально Специфическим Законом* для вывода факта G (МСЗ(G)) будем называть сильнейший вероятностный закон, принадлежащий семантическому вероятностному дереву вывода факта G, имеющий максимальное значение вероятности среди всех правил дерева.

Множество всех максимально специфических законов МСЗ(G) для всех атомов $G \in U(\Omega)$ обозначим через МСЗ.

Предложение 3. $L \subset$ МСЗ $\subset$ СВЗ $\subset LP$.

§7. Требование максимальной специфичности

Определим Требование Максимальной Специфичности (ТМС). Будем предполагать, что класс H объектов в определении RMS является предложением $H \in \Re(\Im)$. Тогда ТМС говорит о том, что должно быть выполнено равенство $p(G;H) = p(G;F) = r$. В терминах вероятности это означает, что должны быть выполнено равенство $\eta(G/H) = \eta(G/F) = r$ для любого $H \in \Re(\Im)$.

Определение 8 [20]. Требование максимальной специфичности (ТМС):

a) при добавлении любого предложения $H \in \Re(\Im)$ к посылке правила $C = (F \Rightarrow G)$ (можно заметить, что тогда высказывание $\forall x(F(x)\&H(x) \Rightarrow F(x))$ истинно);

b) и выполнении условия $F(a)\&H(a)$ (тогда $\eta(F\&H) > 0$),

должно выполняться равенство $\eta(G/F\&H) = \eta(G/F) = r$.

Другими словами ТМС означает, что не существует высказывания $H \in \Re(\Im)$, которое увеличивает (или уменьшает, смотри нижеследующую лемму) условную вероятность $\eta(G/F) = r$ путем добавления его в посылку правила.

Лемма 1. Если высказывание $H \in \Re(\Im)$ уменьшает условную вероятность $\eta(G/F\&H) < \eta(G/F)$, то высказывание $\neg H$ увеличивает её и $\eta(G/F\&\neg H) > \eta(G/F)$.

Лемма 2. Для любого правила $C = (B_1\&...\&B_t \Rightarrow A_0)$, $\eta(B_1\&...\&B_t) > 0$ вида (1) существует вероятностный закон $C' = (A_1\&...\&A_k \Rightarrow A_0)$, являющийся подправилом правила C такой что $\eta(C') \geq \eta(C)$.

Теорема 4 [20]. Любое МСЗ(G) правило удовлетворяет требованию ТМС.

Предложение 4. Любой закон из L удовлетворяет требованию ТМС.

§8. Решение проблемы статистической двусмысленности

Теорема 5 [20]. I-S вывод непротиворечив для любой теории $T \subset$ МСЗ.

Проиллюстрируем эту теорему на примере статистической двусмысленности, приведенном в §1. Максимально специфичными правилами для высказываний E и ¬E будут следующие правила МСЗ(E) и МСЗ(¬E):

(Л1)' : 'Почти все случаи заболевания стрептококком, **который не является устойчивым к пенициллину**, быстро вылечиваются инъекцией пенициллина';

(Л2): 'Почти всегда устойчивая к пенициллину стрептококковая инфекция не вылечивается после инъекции пенициллина'.

Правило (Л1)' имеет большую условную вероятность, чем исходное правило (Л1) и, следовательно, оно должно быть максимально специфичным МСЗ(E) правилом для высказывания E. Правила (Л1)' и (Л2) уже не могут быть выполнены на одних и тех же данных и поэтому не противоречат друг другу.

Выводы

Итак мы определили семантический вероятностный вывод, который обладает следующими свойствами:

1. в нем синтезируется логика, вероятность и обучение для вычисления (вывода) предсказаний следующим образом:
    a. вывод заменяется на вычисление;
    b. истинность обобщается до вероятности;
    c. процесс вывода-вычисления состоит в нахождении высказываний, имеющих максимальную оценку условной вероятности с целью наилучшего предсказания. Нами доказано, что получающиеся правила максимально специфичны, т.е. они содержат максимум информации, требующийся для максимально точного и непротиворечивого предсказания. Этот процесс в точности совпадает с целью обучения и индуктивного вывода знаний. Специфика обучения в семантическом вероятностном выводе только в том, что ищутся общие (вероятностные) и в тоже время максимально специфические (конкретные) высказывания, которые позволяют предсказывать определенный факт G.
2. предсказания, получающиеся семантическим вероятностным выводом по максимально специфическим правилам обладают следующими преимуществами по сравнению с традиционными:
    a. предсказания непротиворечивы;
    b. для них не возникает проблема статистической двусмысленности;
    c. они получаются на основании максимально специфических правил.
3. с точки зрения методов Machine Learning и Knowledge Discovery from Data Bases and Data Mining (KDD&DM), семантический вероятностный вывод позволяет индуктивно выводить из данных (или, как говорят в области Data Mining, извлекать из данных) в определенном смысле «полное» множество знаний. «Полнота» извлекаемых знаний состоит в том, что (1) извлекаемые знания содержат теорию Th(M) (см. теорему 2 и предложения 3 и 4) и (2) извлекаемые знания соодержат все максимально специфические знания. На основании семантического вероятностного вывода разработана программная система Discovery, реализующая семантический вероятностный вывод и успешно примененная для решения большого числа практических задач.
4. совместно с Борисом Ковалерчуком разработан реляционный подход к методам KDD&DM, опубликованный в монографии [21] и в главах в монографиях [22, 23]. Более подробная информация представлена на сайте [24]). В этом подходе на основании семантического вероятностного вывода и реализующей его системы Discovery разработан оригинальный подход к извлечению знаний. В реляционном подходе к извлечению знаний следующим образом снимаются практически все ограничения с существующих ML, KDD&DM-методов:
    a. ограничения с используемых типов данных за счет использования Теории Измерений и многосортных эмпирических систем;
    b. использование Теории Измерений позволяет извлекать всю информацию из данных, что не делают другие методы;

c. ограничения в использовании априорного знания путем представления априорного знания в логике первого порядка;
d. ограничения с классов проверяемых гипотез за счет введения типа обнаруживаемых знаний Rule Type в языке первого порядка;
e. разработана система Discovery, обнаруживающая:
   i. множество законов L на эмпирической системе M;
   ii. множество MCЗ максимально специфических правил;
   iii. множество правил с максимальными оценками условной вероятности;
f. множество обнаруживаемых знаний полно в двух смыслах:
   i. в смысле полноты извлечения информации из данных за счет использования Теории Измерений;
   ii. полноты обнаруживаемых множеств правил (i), (ii), (iii);
g. система Discovery применялась для решения большого числа практических задач и практически всегда система Discovery давала лучшие результаты, чем другие методы (подробности сравнения приведены на сайте [Scientific Discovery])

5. семантический вероятностный вывод был положен в основу формальной модели нейрона и некоторых моделей когнитивных процессов., Работа мозга представляется в [25-28] как непрерывное во времени предсказание мозгом событий окружающей среды с одновременным тотальным контролем акцептором результатов действия правильности сделанных предсказаний. В этих же работах приведено детальное объяснение Теории Функциональных Систем работы мозга П.К.Анохина [29] и Информационной Теории Эмоций П.В.Симонова [30]. В последующих работах [31—33] эти теории, а также Теория Движений Н.А.Берштейна [34] и Принятие Решений представленs в виде схем. В последних работах проведены машинные эксперименты по моделированию анимата (искусственного автономного интеллектуального агента) с целью сравнения его работы с существующими подходами. В работе [35] показано, что анимат, основанный на семантическом вероятностном выводе, обучается намного эффективнее, чем аниматы, основанные на нейронных сетях и Reinforcement Learning.
6. В приведенных выше работах показано, что семантический вероятностный вывод достаточно хорошо формализует такие когнитивные процессы как:
   a. иерархическую организацию движений;
   b. принятие решений;
   c. персептивный цикл и восприятие;
   d. целенаправленное поведение.

Каков ***смысл полученного синтеза логики, вероятности и обучения*** в семантическом вероятностном выводе? Лучше всего это объяснить в терминах когнитивных процессов, поскольку в логике подходящих понятий нет.

По принципу условного рефлекса в процессе обучении человека происходит тот же процесс накопления навыков, что и в семантическом вероятностном выводе: сигналы, которые ассоциируются с результатом, начинают входить в систему обучения и связываться с результатом. В терминах семантического вероятностного вывода это те признаки, которые проявляют свою связь с результатом и увеличи-

вают вероятность предсказания результата. На уровне нейронов условный рефлекс проявляется в виде эффектов замыкания условных связей на уровне нейронов (см. более детальное описание формальной модели нейрона, основанной на семантическом вероятностном выводе в [28]).

В когнитивных процессах главную роль играют максимально специфические правила ввиду следующего свойства нейронных процессов. Известно, что нейроны максимально быстро срабатывают на сигналы, имеющие максимальные значения условной вероятности. Поэтому нейроны в первую очередь срабатывает по максимально специфическим правилам.

Обучение заканчивается тогда, когда результат прогнозируется с вероятностью 1 и действия по его достижению становятся автоматизированными. В семантическом вероятностном выводе пределом обучением является теория Th(M), включающая высказывания с условной вероятностью 1. Поэтому пределу обучения соответствует множество L законов и выводимая из них теория Th(M). В силу предложения 3 $L \subset$ МСЗ множество максимально специфических законов является вероятностным непротиворечивым расширением теории Th(M). Назовем её ***теорией предсказания***. С когнитивной точки зрения теория предсказания является тем множеством предвосхищений, которые проявляются в ***апперцепции*** и ***мышлении*** (как оно описано в части III работы [36] озаглавленной «мышление и прогнозирование»).

С эмпирической точки зрения теория предсказания является той промежуточной теорией между формальными теориями и реальностью, которая, с одной стороны, связана с реальностью и экспериментом, а, с другой стороны, содержит (аппроксимирует) теории. Например, в Теории Измерений [37] до сих пор сохраняется явный разрыв между теорией и экспериментом. С одной стороны, эмпирические системы определяются как удовлетворяющие некоторой системе аксиом, которая apriory полагается известной. С другой стороны, эти системы аксиом надо находить в результате экспериментов над реальностью представленной эмпирической системой.

ЛИТЕРАТУРА

1. *Hempel, C. G.* Aspects of Scientific Explanation, In: C. G. Hempel, Aspects of Scientific Explanation and other Essays in the Philosophy of Science. – The Free Press. – New York, 1965.
2. *Hempel, C. G.* ‘Maximal Specificity and Lawlikeness in Probabilistic Explanation’, Philosophy of Science **35**, 1968. – P. 16–33.
3. *Yao-Hua Tan.* Is default logic a reinvention of inductive-statistical reasoning? *Synthese*, 110: 357–379, 1997. Kluwer Academic Publishers.
4. *M.C.Fitting*. Logic Programming on a Topological Bilattices // *Fundamenta Informatica*. – V.11, 1988. – P. 209-218.
5. *E.Shapiro*. Logic Programs witn Uncertainties: A Tool for Implementing Expert Systems // Proc. IJCAI '83, Williams Kauffman, 1983. – P. 529-532.
6. *M.Kifer., V.S.Subrahmanian*. Theory of Generalized Annotated Logic Programming and its Applications // Research Report, University of Maryland, USA, 1990.
7. *R.T.Ng, V.S.Subrahmanian*. Probabilistic reasoning in Logic Programming // Proc. 5th Symposium on Methodologies for Intelligent Systems, Knoxville, North-Holland, 1990. – P. 9-16.
8. *R.T.Ng, V.S.Subrahmanian*. Annotation Variables and Formulas in Probabilistic Logic Programming // Technical report CS TR-2563, University of Maryland. 1990.
9. *H.Gaifman*. Concerning measure in first order calculi // Israel journal of Math. v.2, N1, 1964 – P. 1–18.
10. Nils J. Nillson. Probability logic // *Artif. Intell*., v.28, N1, 1986. – P. 71-87.

11. *T.Hailperin*. Probability Logic // *Notre Dame J. of Formal Logic*, v.25, N3, 1984. – P. 198-212.
12. *Halpern, J.Y.* An analysis of first-order logic of probability. *Artificial Intelligence*. v. 46, 1990. – P. 311-350.
13. *D.S.Scott, P.Krauss*. Assigning Probabilities to Logical Formulas // Aspects of Inductive Logic / ads. J.Hintikka, P.Suppes, N.Holland, 1966. – P. 219-264.
14. *Er.W.Adams*. The logic of conditionals // An application of probability to deductive logic // *Synthese Library*, v.86. 1975.
15. *M.N.Van Emden*. Quantitative deduction and its fixpoint theory // *J. Logic Programming*, v.3, N.1. 1986, – P. 37-53.
16. *S.S.Goncharov, Yu.L.Ershov, D.I.Sviridenko*. Semantic programming // 10th World Congress Information Processing 86, Dublin, Oct.,1986. - Amsterdam, 1986. – P. 1093-1100.
17. *De Raedt L., Kersting K*. Probabilistic logic learning // *ACM-SIGKDD Explorations* / special issue on Multi-Relational Data Mining. Vol. 5(1). 2004. – P. 31–48, July.
18. *Evgenii Vityaev, Boris Kovalerchuk*. Empirical Theories Discovery based on the Measurement Theory. *Mind and Machine*, v.14, #4, 2004. – P. 551-573
19. *Витяев Е.Е*. Семантический подход к созданию баз знаний. Семантический вероятностный вывод наилучших для предсказания ПРОЛОГ-программ по вероятностной модели данных. // Логика и семантическое программирование (Выч. сист., вып. 146). Новосибирск, 1992. – С. 19-49.
20. *Evgenii Vityaev*. The logic of prediction. In: Mathematical Logic in Asia. Proceedings of the 9th Asian Logic Conference (August 16-19, 2005, Novosibirsk, Russia), World Scientific, Singapore, 2006. – P. 263-276
21. *Kovalerchuk, B., Vityaev*, E. Data Mining in finance: Advances in Relational and Hybrid Methods, Kluwer Academic Publishers, 2000. – P. 308.
22. *Kovalerchuk, B., Vityaev, E., Ruiz, J.F.* Consistent and Complete Data and 'Expert' Mining in Medicine // *Medical Data Mining and Knowledge Discovery*, Springer, 2001. – P. 238-280.
23. *Evgenii Vityaev, Boris Kovalerchuk*. Data Mining For Financial Applications. In: O. Maimon and L. Rokach (eds.), Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, Springer 2005, – P. 1203-1224.
24. Scientific Discovery. http://www.math.nsc.ru/AP/ScientificDiscovery
25. *Витяев Е.Е.* Принципы работы мозга. 2003. http://wsni2003.narod.ru/Papers/Vityaev.htm
26. *Витяев Е.Е.* Целеполагание как принцип работы мозга // Модели когнитивных процессов (Выч. системы, 158), Труды ИМ СО РАН, Новосибирск, 1997. – С. 9-52.
27. *Витяев Е.Е.* Вероятностное прогнозирование и предсказание как принцип работы мозга // Измерение и модели когнитивных процессов (Выч. системы, 162), Новосибирск, 1998. – С. 14-40.
28. *Витяев Е.Е.* Формальная модель работы мозга, основанная на принципе предсказания // Модели Когнитивных Процессов. (Выч. сист., 164), Новосибирск, 1998. – С. 3-61
29. *Анохин П.К.* Принципиальные вопросы теории функциональных систем // Философские аспекты теории функциональных систем. М.: Наука, 1978. – С. 49 - 106.
30. *Симонов П.В.* Эмоциональный мозг. М.: Наука, 1981. – С. 140.
31. *Михиенко Е.В., Витяев Е.Е.* Моделирование работы функциональной системы, VI Всероссийская научно-техническая конференция «Нейроинформатика-2004». Сборник научных трудов. В 2-х частях. Ч.2., М.: МИФИ, 2004. – С. 124-129.
32. *Витяев Е.Е.* Объяснение Теории Движений Н.А.Бернштейна. VII Всероссийская научно-техническая конференция «Нейроинформатика-2005», Сборник научных трудов, часть 1, Москва, 2005. – С. 234-240
33. *Витяев Е.Е.* Принятие решений. Переключающая и подкрепляющая функции эмоций // VIII Всероссийская научно-техническая конференция «Нейроинформатика-2006», Сборник научных трудов, Москва, 2006. – С. 24-30

34. *Бернштейн Н.А.* Биомеханика и физиология движений. // Избранные психологические труды, Москва-Воронеж, 1997. – С. 605
35. *Демин А.В., Витяев Е.Е.* Реализация модели анимата на основе семантического вероятностного вывода // VIII Всероссийская научно-техническая конференция «Нейроинформатика-2006», Сборник научных трудов т.2, Москва, 2006. – С. 16-24
36. *Брушлинский А.В.* Субъект: мышление, учение, воображение. (избранные психологические труды). М., 1996. – С. 387
37. *Krantz DH, Luce RD, Suppes P, and Tversky A*. Foundations of Measurement V.1-3, Acad. Press, NY, London. 1971, 1989, 1990.

Институт математики им. С. Л. Соболева СО РАН,
630090, Новосибирск, пр. Коптюга 2
тел.: 333-28-94
факс: 333-25-98
vityaev@math.nsc.ru
с.н.с., к.т.н.

Предсказание как вычисление.
Витяев Евгений Евгеньевич

В работе рассматривается проблема логического вывода знаний. Показывается, что логический вывод знаний и, в частности, предсказание сталкивается с проблемами статистической двусмысленности и падением вероятностных оценок знаний в процессе вывода. В работе предлагается рассмотреть логический вывод как вычисление в духе семантического подхода к программированию. Для этого определяется понятие семантического вероятностного вывода. Показывается, что семантический вероятностный вывод определят индуктивный вывод знаний и предсказание лишенное упомянутых недостатков и, в частности, решающее проблему статистической двусмысленности.

Prediction as computation
Evgenii Vityaev

The problem of logical inference of knowledge is considered in the paper. Denote that the logical inference of knowledge and, in particular, predictions encounter with the problem of statistical ambiguous and decreasing of the probabilistic estimations of the knowledge in the inference process. In the paper we propose to consider the logical inference as computation as it done in semantic approach to programming. For that purpose the notion of semantic probabilistic inference is defined. It proposed that semantic probabilistic inference produce the inductive inference of knowledge and predictions avoiding the mentioned problems and, in particular, the problem of statistical ambiguous.