

ЛЕКЦИЯ 13. 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 $(\Rightarrow C_1), \dots, (\Rightarrow C_n)$.

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

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

Для вывода предсказаний из знаний, имеющих некоторую оценку вероятности, достоверности, нечеткости $[r]$ необходимо вычислять эти оценки, как было показано в *метаинтерпретаторе*.

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 согласованы.

Вычислению этих оценок посвящены *работы по вероятностной логике* ([Fitting M.C., 88], [Shapiro E., 83]; [Kifer M., V.S.Subrahmanian, 90], [Ng R.T., Subrahmanian V.S., 90a,b]; [Gaifman H., 64], [Nilsson N.J., 86], [Hailperin T., 84], [Halpern J.Y., 90], [Scott D.S., Krauss P., 66], [Adams E.W., 75], [Van Emden M.N., 86].

Есть работы, в которых *вероятность рассматривается как значение истинности утверждений*, а процесс логического вывода обобщается до так называемой “*количественных дедукций*” (дедуктивных систем, в которых значения истинности непрерывны и принимают значения в интервале $[0,1]$): [Shapiro E., 83], [Kifer M., V.S.Subrahmanian, 90], [Ng R.T., Subrahmanian V.S. 90a,b]; [Van Emden M.N., 86].

В работах ([Ng R.T., Subrahmanian V.S. 90a,b]; [Van Emden M.N., 86]) 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 случай-

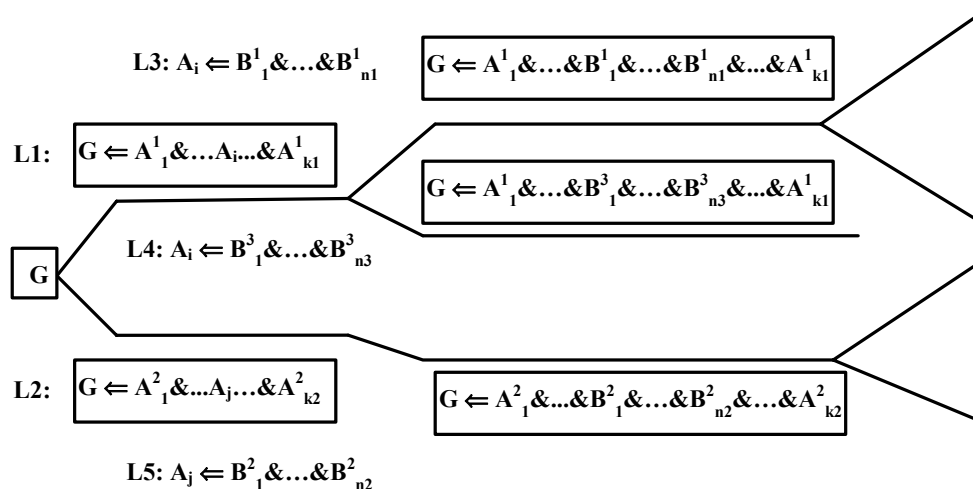


Рис. 1

но.

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 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.

Понятие предсказания для индуктивных знаний отсутствует.

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.

Для решения данной проблемы нужен более радикальный пересмотр самого подхода к выводу, чем это делается в работах workshop.

Каким образом можно обобщить вывод?

Рассмотрим процесс вывода с точки зрения “семантического” подхода к программированию [Goncharov S.S., Ershov Yu.L., Sviridenko D.I., 86].

Семантический подход к логическому программированию состоит в рассмотрении теоретико-модельной семантики логических программ, состоящей в том, что факты являются высказываниями некоторой модели, эмпирической системы $\mathfrak{S} = \langle A; \Omega_{\mathfrak{S}} \rangle$, представляющей предметную область.

В этом случае процесс *вывода* предсказания можно рассматривать как *вычисление* предсказания состоящее в обнаружении таких фактов C_1, \dots, C_n в эмпирической системе \mathfrak{S} , из истинности которых по законам L_1, \dots, L_m выводится предсказываемое высказывание. Тогда будет иметь место не просто вывод высказывания G , а еще и истинность $\mathfrak{S} \models \exists x_1, \dots, x_n G$, $\mathfrak{S} \models G[x_1/t_1, \dots, x_n/t_n]$.

Такой *процесс вывода* можно рассматривать как *вычисление истинности* предсказываемого факта G на модели представляющей наш мир.

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, подтвержденность, достоверность и т.д. имеют континуум значений. Поэтому, если использовать не значения истинности: истина и ложь, среди которых не имеет смысла искать “более истинное”, а континуум значений, то поиск наиболее вероятного, достоверного и т.д. утверждения уже имеют самостоятельный смысл, которого нет в обычном понимании вывода.

При таком выводе мы даже не нуждаемся в правилах вывода.

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*.

Определим семантический вероятностный вывод (СВВ), основанный на приведенной идее семантического программирования.

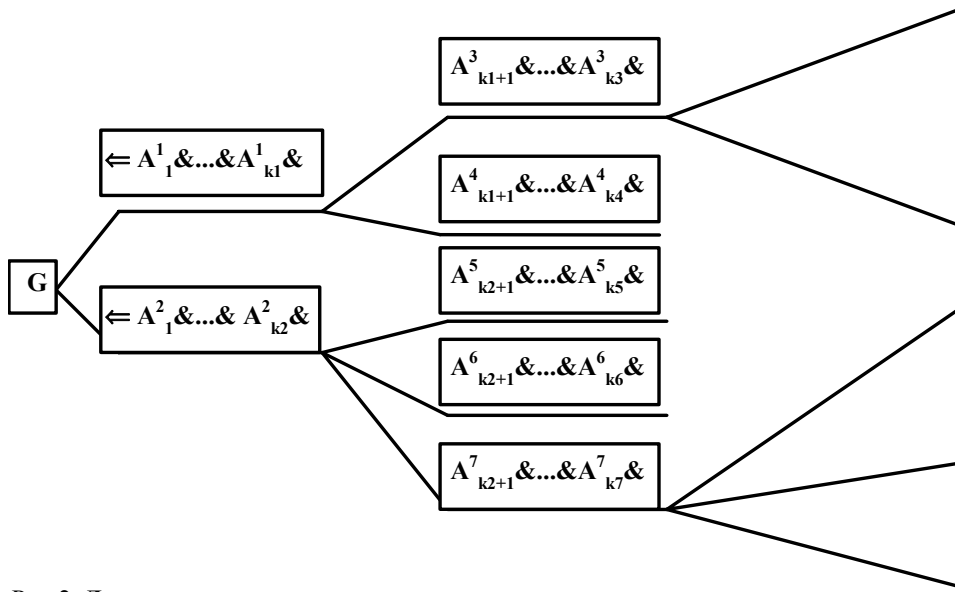


Рис 2. Дерево семантического вероятностного вывода

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.

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

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

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

Следствие 1. 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 .

Определение 7. Максимально Специфическим Законом $MC3(G)$ для вывода факта 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 $MC3(G)$ для всех атомов $G \in U(\Omega)$ обозначим через $MC3$.

Предложение 3. $L \subset MC3 \subset CB3 \subset LP$.

Требование максимальной специфичности.

Решение проблемы статистической двусмысленности.

Определим Требование Максимальной Специфичности (ТМС) и докажем, что все максимально специфические законы из $MC3$ удовлетворяют требованию максимальной специфичности – содержат всю информацию требуемую для предсказания некоторого факта G и эти законы нельзя уточнить выбирая подмножества объектов.

Будем предполагать, что класс H объектов в определении RMS является предложением $H \in \mathfrak{X}(\mathfrak{S})$. Тогда ТМС говорит о том, что должно быть выполнено равенство $\eta(G/H) = \eta(G/F) = r$ для любого $H \in \mathfrak{X}(\mathfrak{S})$.

Определение 8 Требование максимальной специфичности (ТМС):

При добавлении любого предложения $H \in \mathfrak{X}(\mathfrak{S})$ к посылке правила $C = (F \Rightarrow G)$, если выполнено условие $F(a) \& H(a)$, то должно выполняться равенство $\eta(G/F \& H) = \eta(G/F) = r$.

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

Лемма 1. Если высказывание $H \in \mathfrak{X}(\mathfrak{S})$ 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 $C = (B_1 \& \dots \& B_t \Rightarrow A_0)$, $\eta(B_1 \& \dots \& B_t) > 0$ существует вероятностный закон $C' = (A_1 \& \dots \& A_k \Rightarrow A_0)$, являющийся подправилом правила C такой что $\eta(C') \geq \eta(C)$.

Теорема 4. Любое максимально специфическое правило $MC3(G)$ удовлетворяет ТМС.

Следствие. Любой закон из L удовлетворяет требованию ТМС.

Теорема 5. I-S вывод непротиворечив для любой теории $T \subset MC3$.

Проиллюстрируем эту теорему на примере статистической двусмысленности, приведенном в §1. Максимально специфичными правилами для высказываний E и $\neg E$ будут следующие правила $MC3(E)$ и $MC3(\neg E)$:

(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 $MC3(E)$ 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. **синтеза логики, вероятности и обучения:** 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.

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. Покажем, что **главной задачей мозга является предсказание**.

Займемся естественным интеллектом.

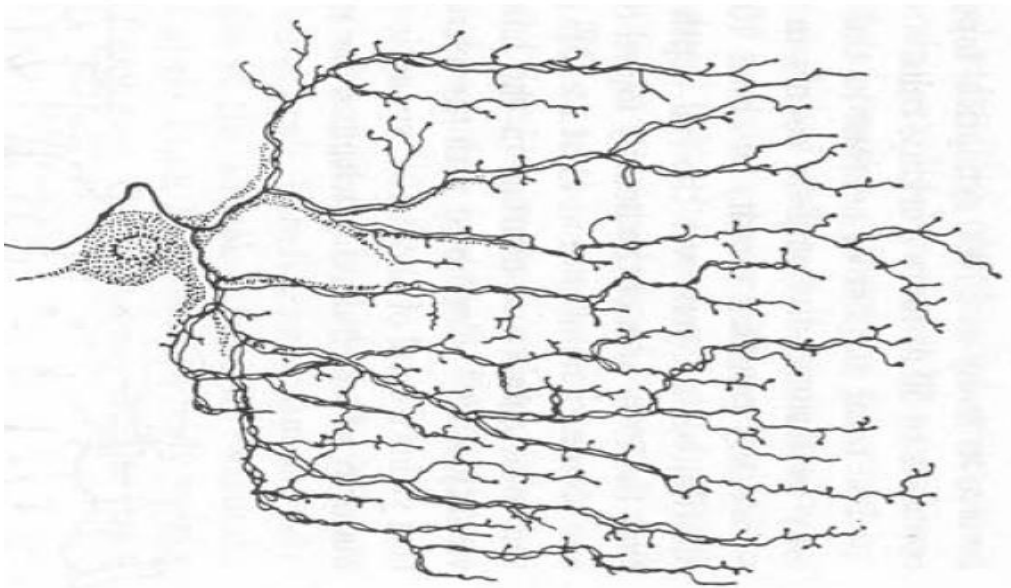
Нейрон занимается предсказанием.

Формальная модель нейрона – есть семантический вероятностный вывод.
физиологические свойства нейрона:

1. Когда человек обучается, то по **принципу условного рефлекса** 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 увеличивают вероятность предсказания результата. На уровне нейронов **условный рефлекс** проявляется в виде эффектов замыкания условных связей на уровне нейронов.

2. Тот факт, что нейрон занимается предсказанием подтверждает свойство нейрона использовать в своей работе только **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 срабатывают по максимально специфическим правилам и именно эти нейроны в первую очередь возбуждают или тормозят связанные с ними нейроны.

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.



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 $L \subset MC3$ множество максимально специфических законов является вероятностным непротиворечивым расширением теории $Th(M)$.

Назовем теорию МЗП *теорией предсказания*. С когнитивной точки зрения теория предсказания является тем **множеством предвосхищений**, которые проявляются в *анперцепции* и *мышлении* (как оно описано в части III работы [Брушлинский А.В., 1996] озаглавленной «мышление и прогнозирование»).

Реляционный подход к извлечению знаний – реализация логического пути познания

Реализация логического пути познания осуществлена нами в виде реляционного подхода к извлечению знаний и теорий. Нами разработана программная система Discovery, реализующая семантический вероятностный вывод и позволяющая обнаружить на данных все упомянутые в предыдущем параграфе множества:

- а) все правила, истинные на эмпирической системе;
- б) все правила, имеющие максимальные значения условной вероятности;
- с) все максимально специфические правила.

В настоящее время обнаружением теорий и знаний занимаются в направлениях: машинного обучения Machine Learning (ML) и извлечения знаний из данных Knowledge Discovery in Data Bases and Data Mining.

Любой ML, KDD&DM-метод явно или неявно предполагает заданным:

- i) типы данных с которыми работает метод;
- ii) язык обработки и интерпретации данных (онтологию KDD&DM-метода);
- iii) класс гипотез, сформулированных в онтологии метода, которые он проверяет на данных (тип знаний KDD&DM-метода);

В рамках реляционного подхода снимаются все ограничения с ML-, KDD&DM-методов за счет использования теории измерений для представления онтологии метода и использования логики первого порядка для представления типа знаний метода.

В реляционном подходе к извлечению знаний снимаются следующие ограничения с существующих ML-, KDD&DM-методов:

- (1) ограничения с используемых типов данных за счет использования теории измерений и многосортных эмпирических систем;
- (2) использование теории измерений позволяет извлекать всю информацию из данных, что не делают другие методы;
- (3) ограничения в использовании априорного знания путем представления априорного знания в логике первого порядка;
- (4) ограничения с классов проверяемых гипотез за счет введения типа обнаруживаемых знаний Rule Type в языке первого порядка;
- (5) разработана система Discovery, обнаруживающая виды множеств (а), (б), (с) для заданного типа гипотез RuleType, которые не обнаруживаются другими методами;
- (6) база знаний, обнаруживаемая системой Discovery полна в двух смыслах:
 - а. в смысле полноты извлечения информации из данных за счет использования теории измерений;
 - б. полноты обнаруживаемых множеств правил (а), (б), (с).