**Лекция 7. Продукционная модель представления знаний**

Форма представления знаний. Продукционная модель знаний. Механизм вывода в продукционной системе знаний. Стратегии управления выводом решения.

Эвристические принципы управления правилами. Достоинства и недостатки продукционной модели.

Если проследить за речью человека (не ее содержанием, а структурным построением фраз), то легко заметить, что значительная часть наших мыслей оформляется в виде правил типа «ЕСЛИ (далее следует ряд условий), ТО (далее следует что надо сделать)».

Примеры: «Если ты будешь хорошо учиться, то папа купит тебе велосипед», «Закаляйся, если хочешь быть здоров», «(Если) Не продается вдохновенье, но (то) можно рукопись продать» и т.п.

Несмотря на то, что такая форма представления знаний чрезвычайно популярна для человека, она является достаточно сложной для реализации в виде жестких алгоритмических схем. Проблемы появляются в связи с тем, что знания многообразны как по уровням, так и по спектру, допускают неоднозначное толкование, чаще всего имеют ассоциативный характер использования. Трудно четко разложить их на причины и следствия и описывать их, например, на языке исчисления высказываний и предикатов. Добавим еще, что знания постоянно меняются, пополняются, модифицируются и приспосабливаются к условиям среды.

**Форма представления знаний**

Поиск удобной модели для представления знаний типа «Если <условия>, то <действия>» привел к идее введения продукции в качестве эквивалента представления знаний и механизма работы СИИ.

**Продукционная модель знаний**

Термин «продукция» был введен американским логиком Е. Постом в 40-х годах в работах по обоснованию и формализации алгоритмических систем. Продукцией называется правило вида , где , - слова над некоторым алфавитом . Совокупность правил типа получила название системы продукций.

Рассмотрим этот механизм представления, интерпретируя левую и правую части правил продукций как логические функции. Наиболее популярная форма такая:

(1)

Здесь - логическое выражение; - заключение; - действие, то есть продукция - это пара <антецедент, консеквент>, а «» - знак секвенции. Для простоты изложения кванторы опущены. Выход за рамки формальной системы осуществляется путем добавления такой понятийной категории, как действия.

Если при замене и конкретными значениями принимает значение И, то выдается заключение и выполняется действие . Формы заключения могут быть различными:

(здесь - статическое высказывание, истинное всегда);

(здесь - одно или несколько действий, не требующих интерпретации);

(здесь - логическая функция, истинная при - действие, которое доопределяется путем замены на , соответственно).

В отличие от импликации случай здесь не рассматривается. Это значит, что при истинности левой части реализуется все, что рекомендуется в правой, то есть продукция истинна всегда.

В простейшем случае правая и левая части продукции могут быть простыми высказываниями. Главное отличие продукции от логических формул - они допускают более широкий спектр интерпретации.

В наиболее общем виде под продукцией понимается выражение:

- уникальное имя продукции, с помощью которого продукция выделяется из всего множества. В качестве имени может выступать любая цепочка символов или цифр, но чаще - лексема, отражающая суть (смысл) данной продукции.

- характеризует сферу применения данной продукции. Такие сферы легко выделяются в системах знаний человека по принципу «полки». На одной «полочке» хранятся знания о том, как добраться до работы, на другой - как варить кашу и т.п. Разделение знаний на отдельные сферы экономит время на поиск нужных в конкретных условиях знаний.

- называется ядром продукции. Это основной элемент. Допускается форма: если , то , иначе. При этом секвенция () истолковывается в обычном, логическом смысле как знак логического следования из истинного (в случае, если не является истинным, как мы уже говорили, то об ничего сказать нельзя).

- есть условие применимости ядра продукции. Обычно, если принимает значение И, то ядро продукции активизируется. В противном случае продукция не может быть применена.

Элемент описывает постусловия продукции. Они активизируются лишь в случае, если ядро продукции реализовалось. Постусловия описывают действия и процедуры, которые необходимо выполнить после реализации ядра.

Ядра продукции можно классифицировать по различным основаниям. Прежде всего все ядро делится на два больших типа: детерминированные (точно определенные) и недетерминированные (вероятостные). В последнем случае может не выполняться и при истинности . Ядро продукции интерпретируется тогда так: если , то, возможно, . Сама возможность определяется некоторыми оценками реализации ядра. Для формирования оценок используется вероятностный или какие-либо другие подходы. В простейших случаях продукции могут быть такими:

если , то с вероятностью р реализовать , или,

если , то с большой долей уверенности .

Мы столкнулись здесь с нечеткостью определения оценок, что и стало предметом изучения нечеткой логики.

Детерминированные продукции могут быть однозначными и альтернативными (связка «или»). Здесь тоже нужен свой механизм выбора альтернативы из нескольких возможных. Это может быть, например, механизм присваивания весов, вероятностных или экспертных оценок. Например: если Р, то чаще надо делать А1, реже А2. Для решения задач прогнозирования могут вводиться оценки типа «с вероятностью р можно ожидать А1».

Способности выбора и конструирования форм продукций в каждом конкретном случае определяют культуру и возможности специалиста по проектированию интеллектуальных систем (инженера по знаниям).

Механизм вывода в продукционной системе знаний

Для того чтобы «запустить» машину вывода при наличии системы продукционных правил, очевидно, необходимы следующие компоненты:

• задание значений переменных, используемых в продукциях;

• сами правила-продукции;

• механизм интерпретации правил.

Значения переменных обычно задаются в так называемой рабочей базе данных - РБД (рабочая память) в виде набора конкретных фактов, формат которых согласуется с форматом правил. Совокупность правил-продукций является сердцем системы и называется базой знаний или правил (БЗ). Механизм интерпретации правил как раз и является механизмом вывода (MB).

Рассмотрим механизм вывода в продукционной системе. Как следует из сказанного, под продукционной системой (ПС) понимается совокупность:

ПС = <РБД, БЗ, МВ>. (2)

В каждом цикле интерпретации выполняются следующие процессы:

• определяется множество означиваний, то есть выделяется набор продукций БЗ и фактов из РБД, согласующихся между собой;

• выполняются сами означивания (сопоставления), то есть реализация указаний из правой части продукций.

В продукционной модели легко просматриваются аналогии: РБД - наборы фактов, описывающих текущие состояния предметной области;

БЗ - модель предметной области;

MB - модель механизма принятия решений (вывода) - прямой аналог человеческим рассуждениям.

Таким образом, обобщенный алгоритм работы ПС состоит из:

• задания модели текущего состояния предметной области (прямой диалог, моделирование, генерация, интерполяция и т.д.);

• интерпретации текущего состояния предметной области (то есть РБД) на БЗ и выработки вариантов решения;

• выбора варианта решения и выдачи его на выход системы для пользователя;

• изменения состояния модели предметной области (РБД) путем выполнения действий и процедур, рекомендованных в правых частях правил.

Самое замечательное в этом механизме то, что управление процессом вывода осуществляется путем изменения состояния РБД и не затрагивает БЗ.

**Стратегии управления выводом решения**

Это управляющий компонент ПС. Он определяет порядок применения правил и устанавливает, имеются ли еще факты в РБД, которые могут быть изменены для продолжения работы. Основных стратегий вывода - две: прямая - «ЛЧП ПЧП» (левая часть правила правая часть правила) и обратная (ПЧП ЛЧП). В каждом цикле может срабатывать только одно правило. Процесс управления состоит из выполнения четырех основных операций:

• сопоставления (образец правила сопоставляется с имеющимися фактами в РБД);

• выбора (если в конкретной ситуации могут быть применены сразу несколько правил, то есть правила конфликтуют за возможность первого применения, то из них выбирается наиболее подходящее);

• срабатывания (если образец правила при сопоставлении совпал с какими-либо фактами из РБД, то соответствующее правило срабатывает);

• изменения состояния РБД на новое (РБД изменяется путем добавления в нее заключения сработавшего правила. Если в правой части содержится еще и указание на какие-либо действия, то они выполняются настолько, насколько это возможно). Прямая стратегия «ЛЧП ПЧП» соответствует стратегии вывода «снизу-вверх» и осуществляется от ситуации к цели. Алгоритм вывода при этом следующий:

1) определяется ситуация, в которой выполняется условие в левой части (если), и делается логический вывод согласно констатирующей части (то);

2) полученный вывод понимается как условие, выполнение которого следует определить по условной части другого правила и т.д.

Пример 1. Рыбак собирается на рыбалку и рассуждает по следующей схеме.

1. Если вода в реке поднимается, то рыба выходит на мель.

2. Если вода в реке опускается, то рыба уходит в глубину.

3. Если лето влажное, то вода в реке поднимается.

4. Если лето сухое, то вода в реке опускается.

5. Если рыба выходит на мель, то следует применять легкую блесну.

6. Если рыба уходит в глубину, то следует применять тяжелую блесну.

Лето выдалось влажное - какую блесну брать на рыбалку? Мы будем решать эту задачу по алгоритму «снизу-вверх».

1) В условной части заданных правил ищем условие, совпадающее с исходной ситуацией «лето влажное» (правило 3). Делаем вывод: «вода в реке поднимается».

2) Принимаем этот вывод как новое условие, следствие которого следует определить, и ищем соответствующее правило (правило 1). Делаем вывод: «рыба выходит на мель».

3) В условной части правил ищем совпадение для ситуации «рыба выходит на мель» (правило 5). Окончательный вывод: блесну следует брать легкую.

Если представить отдельное правило в виде прямоугольника, темная часть которого означает условие, а светлая - заключение, то прямая стратегия отобразится в виде рис. 1.

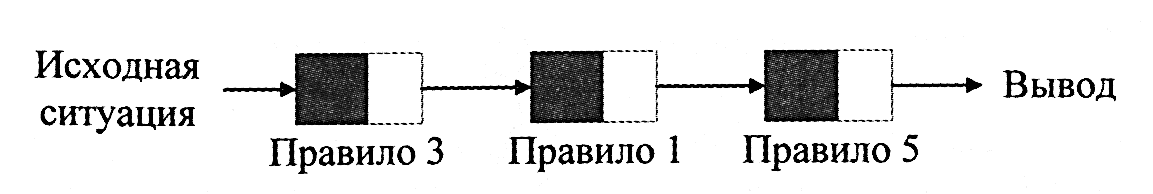


Рис. 1. Стратегия «снизу-вверх» для приведенного примера

В общем случае правило удобно изображать в виде графа-дерева. Если оно имеет несколько условий (Рi) и одно заключение (A), то граф будет иметь вид рис. 2. Но правило может содержать и несколько заключений, и несколько действий.

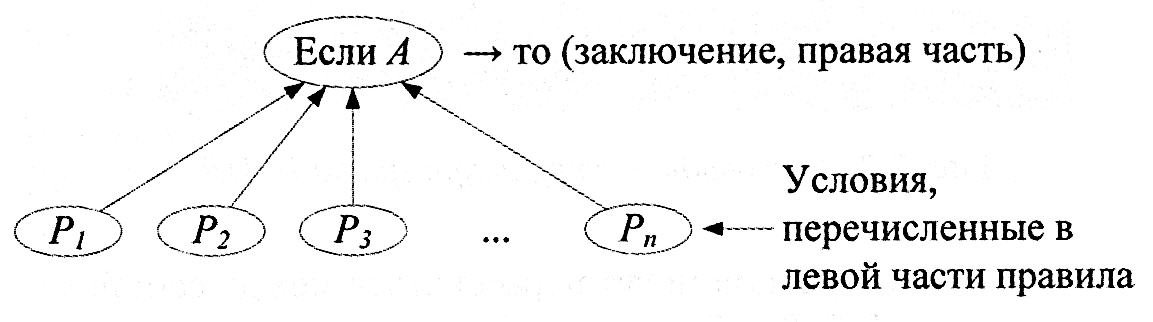


Рис. 2. Структура правила

Очевидно, что правила могут пересекаться между собой:

• по условиям: ;

• по заключениям: ;

• по действиям: и т.д.

С учетом возможных пересечений база знаний может быть представлена древовидной структурой типа И/ИЛИ-графа, которая хорошо показывает связи между результатами вывода по отдельным правилам.

Вывод происходит «снизу-вверх» от состояния к цели. В случае подтверждений условий фактами БД осуществляется подъем к некоторому заключению, которое является условием вместе с заключениями по другим правилам для получения другого заключения более высокого уровня. В случае альтернативного развития вывода («ИЛИ») выполняются генерация конфликтного набора правил, его разрешение с помощью метаправил и «подъем» далее. Данная стратегия вывода называется прямой или стратегией «от состояния к цели». Могут быть варианты этой стратегии:

а) возвратная, когда возврат осуществляется опять к первому правилу, то есть одна и та же продукция применяется до тех пор, пока это требуется;

б) безвозвратная, когда каждая продукция применяется один раз.

Вторая стратегия вывода - обратная - «сверху-вниз» (или от цели к состоянию). Обратный вывод аналогичен прямому, но возникают дополнительные проблемы ограничения конфликтных наборов правил и выбора алгоритма разрешения конфликтов. Есть и другие проблемы, например, оценки условий в условной части. Все они решаются по-разному в рамках соответствующих инструментальных систем продукционного типа.

Согласно алгоритму вывода «сверху-вниз» вначале ищут совпадения имени цели с частью то какого-либо правила. Соответствующая ей часть если определяет следующее условие, по которому ищут совпадения с частью то другого правила и т.д. Процесс, таким образом, идет от цели к ситуации. Для простоты понимания рассмотрим процедуру обратного вывода на примере правил, записанных на языке исчисления высказываний.

Пример. Допустим, что в базе знаний имеются правила:

П1 ,

П2 ,

ПЗ ,

П4 ,

П5 ,

П6 ,

П7 .

Дано: и - истинны. Доказать, что - истинно.

Алгоритм:

1. Ищем правило, где в правой части - П7.

2. будет истинным, если - истинно; ищем продукцию, где в правой части, их две - П2 и ПЗ. , таким образом, будет истинным, если истинна хотя бы одна из конъюнкций в П2 и ПЗ. В П2: по условию, но неясно ; в ПЗ: по условию, неясно .

3. Ищем продукцию, в которой является целью. Это П1. Здесь истинно по условию, и неизвестны.

4. Из П5: , так как по условию.

5. Из П4: так как , то и .

6. Обратимся к П1. Все высказывания в левой части здесь истинны, следовательно, .

7. Так как , то из П2: .

8. Из П7: , что и требовалось доказать.

Теперь можно обобщить обе стратегии вывода на продукциях. В обоих случаях алгоритм начинается с сопоставления исходной ситуации: в первом случае с условием (Если), во втором - с целью (то). В результате могут быть активизированы сразу несколько правил (в нашем примере П2 и ПЗ). Образуется «конфликтное множество», оно разрешается различными способами по заданному критерию выбора. Выбранное правило выполняется, что позволяет осуществить необходимые действия (о действиях - ниже). Обобщенная стратегия показана на рис. 4.

Процесс интерпретации очень хорошо может быть проиллюстрирован как поиск решения в пространстве состояний. Поскольку невозможно построить сразу все пути решения задачи, движение в пространстве состояний осуществляется в соответствии с правилами-эвристиками, составляющими содержание критериев.

Каждое правило легко представить в виде отдельного «кирпичика» (модуля, блока и т.д.), управляемого данными. Правила не зависят друг от друга, что позволяет наращивать и изменять БЗ без изменения самого механизма функционирования и БЗ.

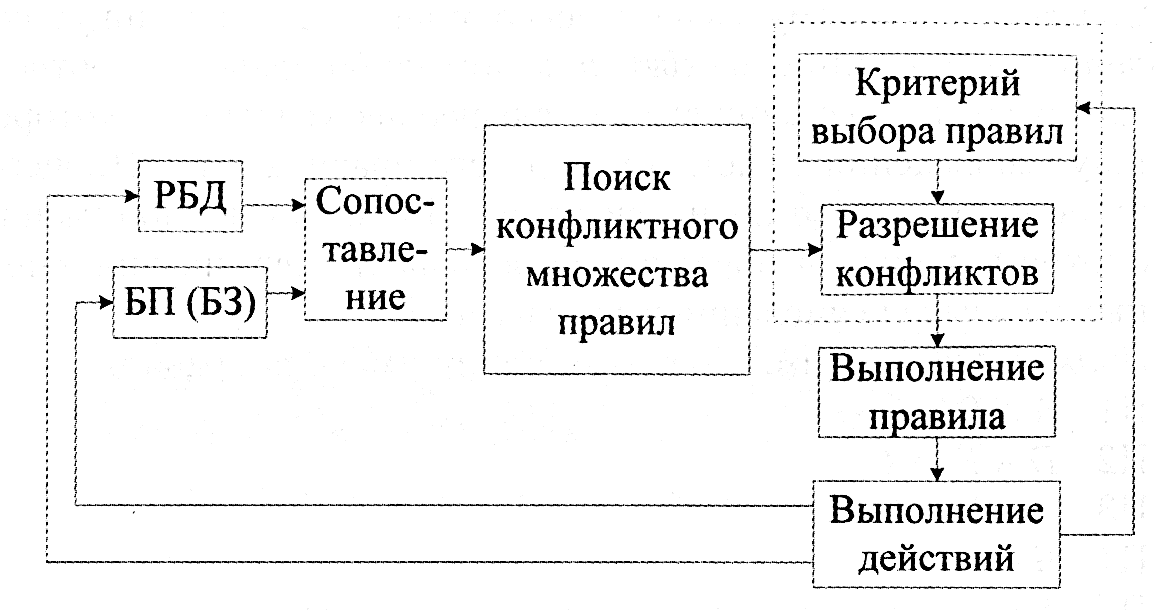


Рис. 4. Стратегия управления выводом на продукциях

**Эвристические принципы управления правилами**

Несмотря на «прозрачность» механизма вывода, в продукционных системах (ПС) возникает проблема выбора, то есть проблема активизации правил. Этим занимается система управления, являющаяся частью всей машины вывода. Подходящих продукций в каждый момент времени может быть несколько, и в принципе здесь возможен параллельный вывод. Эта возможность зависит от имеющегося ресурса и архитектуры вычислительной техники. Используется несколько эвристических принципов «упаковки» продукций в БЗ. Перечислим наиболее популярные из них.

1. Принцип «стопки книг». Наиболее часто используемые продукции актуализируются первыми. Этот принцип целесообразно применять, если продукции относительно независимы друг от друга.

2. Принцип наиболее длинного условия. Первой выбирается продукция, у которой наиболее «длинное» условие выполнимости ядра. Этот принцип целесообразен, когда продукции хорошо структурированы по отношению «частное-общее».

3. Принцип метапродукций. В систему продукций вводятся специальные правила-метапродукции, задачей которых является разрешение ситуаций в условиях конфликтного набора правил.

4. Принцип приоритетного выбора. На продукции вводятся статические или динамические приоритеты. Статические формируются априори, чаще всего на основе экспертных оценок. Динамические приоритеты вырабатываются в процессе функционирования системы продукций и могут отражать, например, такой параметр, как время нахождения продукции во фронте готовых продукций. К данному типу управления относится задание последовательности приоритетов с помощью специальной каузальной семантической сети. В этом .случае задается некоторый каузальный сценарий, движение по которому определяется складывающимися ситуациями и в каждой вершине которого задана функция выбора очередной продукции из фронта готовых продукций. Если продукции работают над семантической сетью, приоритеты могут формироваться не в системе продукций, а в самой сети.

5. Принцип «классной доски». Основан на идее спусковых функций. При реализации принципа «классной доски» выделяется специальное рабочее поле памяти - аналог классной доски, на которой мелом пишут объявления и стирают их при необхо-димости. На этой «доске» параллельно выполняющиеся процессы находят информацию, инициирующую их запуск, на нее же они и выносят информацию о своей работе, которая может оказаться полезной для других процессов. Как правило, на «классной доске» выделяются специальные поля для формирования условий применимости ядер продукций, различные для разных сфер применения продукций, специальные поля для записи результатов срабатывания продукций и для записи постусловий, если они адресованы другим продукциям. С принципом «классной доски» может комбинироваться принцип управления с помощью метапродукций, т.к. он требует проверки некоторых условий, которые фиксируются в рабочем поле памяти, а также другие принципы управления.

Если система продукций работает над некоторой сетевой моделью в базе знаний, то необходимо (особенно при альтернативных продукциях) принимать специальные меры защиты от «порчи знаний» работающими продукциями. Для этого также используется рабочее поле памяти, куда временно переносится фрагмент знаний, которым оперируют продукции. Этот фрагмент фиксируется на специальном поле «классной доски».

**Достоинства и недостатки продукционной модели**

Продукционным моделям не хватает строгой теории. Пока на практике при работе с продукционными моделями в них используют эвристики. При задании модели проблемной области в виде совокупности продукций нельзя быть уверенным в ее полноте и непротиворечивости. Причина неудач создания теории кроется в расплывчатости понятия продукции из той интерпретации, которая приписывается ядру, а также в различных способах управления системой продукций. Переход к алгоритмической схеме мало что дает, так как при этом исключается основное свойство продукционных систем - их модульность и проистекающие из этого асинхронность и параллельность выполнения продукций в системе.

Популярность продукционных моделей определяется несколькими факторами.

1) Подавляющая часть человеческих знаний может быть записана в виде продукций.

2) Системы продукций являются модульными. За небольшим исключением удаление и добавление продукций не приводит к изменениям в остальных продукциях.

3) При необходимости системы продукций могут реализовывать любые алгоритмы и, следовательно, способны отражать любое процедурное знание, доступное ЭВМ.

4) Наличие в продукциях указателей на сферу применений продукции позволяет наиболее эффективно организовать память, сократив время поиска в ней необходимой информации. Классификация сфер может быть многоуровневой, что еще более повышает эффективность поиска знаний, так как позволяет наследовать информацию в базе знаний.

5) При объединении систем продукций и сетевых представлений получаются средства, обладающие большой вычислительной мощностью.

6) Естественный параллелизм в системе продукций, асинхронность их реализации делают продукционные системы удобной моделью вычислений для ЭВМ новой архитектуры, в которой идея параллельности и асинхронности является центральной.

Продукционные модели имеют, по крайней мере, два недостатка.

1) При большом числе продукций становится сложной проверка непротиворечивости системы продукций. Это заставляет при добавлении новых продукций тратить много времени на проверку непротиворечивости новой системы.

2) Из-за присущей системе недетерминированности (неоднозначного выбора выполняемой продукции из множества активизированных продукций) возникают принципиальные трудности при проверке корректности работы системы.

Считается, что если в ИС число продукций достигает тысячи, то мало шансов, что система продукций во всех случаях будет правильно функционировать. Именно поэтому число продукций, с которым, как правило, работают современные ИС, не превышает тысячи.

ВОПРОСЫ.

1. Крестьянину необходимо перевести через реку волка, козу и капусту. Рядом с крестьянином в лодке может поместиться либо только волк, либо только коза, либо только капуста. Очевидно, что на берегу нельзя оставлять волка с козой, а козу с капустой. Построить БЗ и РБД для случая, когда крестьянин все-таки перевез все, что хотел, на другой берег. Найти решение задачи.

2. К реке подошли три рыцаря с оруженосцами. У берега оказалась одна двухместная лодка. Как организовать переправу при условии, что ни один оруженосец нигде не должен оставаться в обществе других рыцарей без хозяина? Построить БЗ и РБД успешного решения задачи. Найти решение задачи.

3. Тот же случай, но рыцарей с оруженосцами - четыре.

4. Восьмиведерный бочонок полон кваса. Требуется разделить этот квас пополам при помощи двух других бочонков объемом в 3 и 5 ведер. Построить БЗ и РБД. Найти решение задачи.

5. Имеется равнобедренный треугольник. Доказать, что углы при основании равны. Построить БЗ и РБД.

6. Задача о вождении автомобиля. Вождение (увеличение/уменьшение скорости или останов, пропуск сзади идущего транспорта) осуществляются в зависимости от качества дороги (главная, второстепенная), указателей на дороге (ограничение на скорость, переход). Смоделировать процесс управления автомобилем.