

ПРЕДСТАВЛЕНИЕ ЗНАНИЙ В ДЕДУКТИВНЫХ БАЗАХ ДАННЫХ

Н.Н. Дацун

Донецкий Государственный Технический Университет, Донецк, Украина

Под дедуктивными базами данных (ДБД) понимают системы нового типа, которые позволяют проводить дедуктивные рассуждения на реляционных базах данных больших объемов. Формально ДБД есть тройка: $EDB = \langle EDB, IDB, IC \rangle$, где EDB - экстенциональная БД (множество фактов); IDB - интенциональная БД (множество правил вида $A \leftarrow B_1, B_2, \dots, B_n$); IC – множество ограничений целостности. Такое сочетание возможностей систем управления базами данных (СУБД) и логического программирования (ЛП) придает дедуктивным базам данных новые возможности, которые “усиливают” достоинства каждой из названных компонент. Использование логического программирования на реляционных БД позволяет использовать единый формализм - язык ЛП - для описания запросов и проверки ограничений целостности. Использование баз данных во внешней памяти позволяет создавать при логическом программировании большие экстенциональные базы данных, в том числе многопользовательские, что позволяет в итоге повысить эффективность систем основанных на знаниях, например, экспертных систем. Поэтому ДБД также называют системами управления базами знаний (что иллюстрирует современную тенденцию слияния данных и знаний).

Классическим примером связывания реляционных СУБД и логического программирования является язык Datalog (подкласс ДБД, использующих Datalog как язык ЛП, называют Datalog-базами данных). Как и его предшественник Пролог, Datalog является языком ЛП, основанным на логике предикатов

первого порядка. Однако синтаксис Datalog'a предельно упрощен: в качестве термов предиката могут быть использованы только константы или переменные; нет встроенных предикатов сравнения, арифметики и т.п.; отсутствуют предикаты нулевой аргументности. В отличие от Пролога язык Datalog имеет несколько вычислительных моделей, корректность и полнота этих моделей доказаны. Как язык запросов к реляционным БД классический язык Datalog (без операции отрицания) имеет вычислительную мощность положительной реляционной алгебры. Однако возможности Datalog'a гораздо шире, потому что как язык ЛП, Datalog предоставляет возможности создания рекурсивных запросов, а также проверки произвольных по сложности ограничений целостности БД.

Рассматривая дедуктивные БД как хранилища данных и знаний, необходимо исследовать возможности реализации в ДБД различных методов представления знаний. Для каждого из методов в докладе анализируется представление пассивной (П) компоненты (собственно знаний) и активной (А) компоненты (механизма вывода) при реализации Datalog-баз данных.

Логические подходы. Это наиболее "естественный" для ДБД способ представления знаний. EDB (П) и IDB (А) описывают множество объектов и отношений между ними.

Произведения. (П) Интенциональная часть ДБД представляет собой продукционные правила $A \leftarrow B_1, B_2, \dots, B_n$, причем атомами B_i могут быть предикатные символы EDB и IDB. Результатом применения произведения является кортеж реляционной БД с предикатным символом А из левой части произведения. Поэтому размер экстенциональной и интенциональной частей системы произведений ограничивается только объемом доступного дискового пространства. (А) Вычислительные модели Datalog'a основаны на "поиске в ширину" в дереве И-ИЛИ; если необходимо применить иной механизм поиска, его программируют на Datalog'e.

Семантические Сети. (П) Если EDB представляет собой прямое отображение реляционной модели, то множество правил IDB позволяет реализовать сетевую модель, используя косвенную рекурсию правил. (А) Встроенный и программируемый механизмы вывода Datalog'a позволяют реализовать обход семантической сети (СС). Свойства, распределенные по узлам СС, наследуются либо терминами предикатов, либо предикатами правил.

Фреймы. (П) Предикат Datalog'a синтаксически является структурой, причем предикатный символ может быть трактован как имя фрейма, его аргументы - как слоты, значения этих аргументов - как значения слотов. (А) Если в качестве фрейма рассматривать заголовок правила, то в теле этого правила может быть инициирована произвольная "присоединенная процедура" в виде подцели. Аналогично СС, в фреймах реализуется наследование свойств.

Нечеткие Множества. (П) Для реализации этого метода представления знаний в классическом Datalog'e нет арифметических операторов. Расширив Datalog названной группой операторов, мы получим языковые средства для представления нечетких множеств. Поэтому Datalog-базы данных могут быть

использованы для представления информации в условиях неопределенности и неточности.

Сочетание возможностей реляционных СУБД (в качестве EDB) и логического программирования (IDB) позволяет использовать ДБД как хранилище данных и знаний.