АНАЛИЗ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ МЕТОДА ИНДЕКСАЦИИ ФАКТОВ

Осадчая И.А.

Национальный исследовательский Томский политехнический университет
Irishka_tomsk@mail.ru

Значение визуализации в слабоструктурированных областях постоянно возрастает. Управление информацией стало одной из ключевых задач и компании ищут способы сократить производственные расходы, позволяющие эффективно управлять электронным содержимым хранилищ данных.

Современные компьютерные технологии используют широкий спектр методов визуализации информации. Методы визуализации позволяют исследователю одним взглядом обнаружить особенности, выявить закономерности и аномалии в больших объемах информации. Однако не стоит забывать, что традиционные информационные системы независимо от доступных средств или применяемых технологий обеспечивают пользователя «сырой» информацией без каких-либо дополнительных пояснений. Они просто обрабатывают и распространяют данные, хранящиеся в базе данных. База данных – совокупность данных, хранимых в соответствии со схемой данных, манипулирование которыми выполняется в соответствии с правилами средств моделирования данных [1].

Системы поддержки принятия решений состоят не только из базы данных, но также из базы, содержащей техники, методы, прогнозы и статистику, что делает возможным комплексную обработку доступной информации. Они дают возможность «индивидуализированного» вывода информации, соответствующего потребностям пользователя.

Экспертные системы – это специализированные системы, которые могут заменить живого эксперта в определенных областях знаний. Они имитируют рассуждения живых экспертов по определенным предметам и иногда называются интеллектуальными помощниками. Экспертная система:

- полностью заменяет живого эксперта;
- охватывает узкую, специфичную область знаний;
- поясняет цепочку заключений (разъясняет, каким путем она пришла к какому-либо выводу);
- при работе со стохастической задачей для каждой альтернативы предоставляет степень ее вероятности.

Количество инструментов и приложений по экспертным системам удваивается каждый год, особенно для персональных компьютеров. Сегодня существует масса коммерческих прикладных

экспертных систем для различных типов производства, обслуживания и других видов организаций [1].

В основе современных экспертных систем лежат различные подходы, технологии и компьютерные конфигурации, поэтому не существует каких-либо стандартных решений. Большинство экспертных систем имеют три общих компонента:

- базу знаний;
- управляющую программу;
- механизм выработки решений.

Основным элементом экспертной системы является база знаний. Ее ценность зависит от качества ее содержимого и динамичности, с которой она может адаптироваться к изменениям в предметной области. Она состоит из декларативных (факты, понятия) и процедурных знаний (правила и действия), которые используются экспертами при решении задач. База знаний проектируется специалистами по разработке баз знаний и экспертами в предметной области таким образом, чтобы она отображала навыки экспертов, их знания и опыт. Знания в базе представляются посредством продукционных правил, семантических сетей, фреймов и других способов представления знаний.

Семантическая сеть является способом визуального описания связей между любыми видами объектов. Она состоит из узлов (например, людей, мест, предметов, понятий, событий или действий) и связей (взаимоотношении между узлами типа «является», «имеет», «сделай из» и т.п.). Семантические сети используются разработчиками систем с базой знаний для определения тех элементов базы знаний, которые связаны с декларативным описанием объектов, связей и взаимоотношений. Семантические сети наиболее широко используются при подтверждении правильности знаний экспертами в данной предметной области.

Управляющая программа (также называемая машиной вывода или интерпретатором правил) – это процедура, контролирующая процесс рассуждений, проводимый на базе знаний. Она направляет поиск по базе знаний, целью которого является получение какого-либо факта, решения или выбор следующего действия. Управляющая программа может взаимодействовать с пользователем системы, когда факт или правило не найдены в базе знаний. Также управляющая программа управляет работой системы путем

выбора правил, их оценки, исполнения и определения параметров изображения [1].

Известные алгоритмы ускорения логического вывода, в частности, RETE [2], ускоряют вывод приблизительно на три порядка, не устраняя, однако, экспоненциальной сложности задачи поиска решений в продукционной модели знаний. Радикально ускорить извлечение фактов можно при условии запоминания результатов предыдущих обращений к базе знаний в виде вторичных фактов (прецедентов). Первичными будем называть факты, которые не являются результатами работы правил. В работе [3] показано, как механизм прецедентов позволяет не только устранить необходимость углубления по дереву поиска при повторном решении задачи, но существенно сократить даже время первой резолюции цели, поскольку дерево решений может содержать большое число повторяющихся фрагментов, спуск в которые может быть заменен обращением к прецеденту.

Однако создание базы прецедентов влечет за собой проблему контроля актуальности прецедентов в условиях изменчивости базы фактов. В данной работе была исследована концептуальная модель базы знаний, предложенная И.А. Бессмертным [4].

Метод индексации фактов реализован в программе SemanticV1.5beta. Данная программа, разработанная в Санкт-Петербургском НИИ университета информационных технологий и оптики, предназначена для представления и визуализации знаний в виде семантических сетей, а также для доступа к базам знаний с помощью графического интерфейса и языка запросов. Данный программный продукт может использоваться в режиме обучающей системы, в котором база знаний содержит в неявном виде некоторую гипотезу, например, диагноз, а пользователь, задавая вопросы, должен выявить данную гипотезу, затратив на это минимум вопросов. Кроме того, программа может работать в режиме экспертной системы, сопоставляя имеющиеся факты с правилами в базе знаний.

Поскольку прямой логический вывод предполагает независимую обработку каждого правила, длительность вывода имеет линейную зависимость от числа правил. Следовательно, оценку быстродействия можно оценивать на одном правиле. Тестирование данного метода проводилось на базе знаний, описывающей родственные отношения. Эксперименты на тестовом наборе фактов показали, что эффективность индексации фактов зависит также от результативности правил.

Если количество успешных применений правил невелико, то время вывода с использованием индексов стремится к нулю. Если каждая комбинация исходных фактов приводит к успешной резолюции (ситуация, на практике, маловероятная), то использование индексов только увеличивает время за счет издержек на индексацию. В использованном примере успешными были приблизительно 15% всех применений правил.

Кроме того в работе было проведено сравнение метода предложенного И.А. Бессмертным и алгоритмом *Rete*. Главное отличие заключается в том, что только индексация фактов выполняется заблаговременно, а операция предварительного отбора фактов выполняется непосредственно перед резолюцией. Это означает, что факты и правила могут существовать отдельно, в т.ч. на разных сетевых ресурсах, что соответствует концепции Глобальной Семантической Сети [5]. В ходе предварительного отбора фактов для каждого условия правила создаются множества кортежей значений переменных, используя которые можно в отдельных случаях отказаться от резолюции правил, а, используя операции реляционной алгебры над кортежами, сразу получить множество решений правила. Тестирование метода индексации фактов показало его работоспособность и ускорение приблизительно на порядок по сравнению с «наивным» логическим выводом.

Литература

1. Информационные технологии в бизнесе / Под ред. М.Желены. – СПб: Питер, 2002.
2. Forgy C. L. RETE: A fast algorithm for the many pattern / many object pattern match problem // *Artificial Intelligence*, 1982. Vol. 19, pp. 17-37.
3. Бессмертный И.А. Методы поиска информации с использованием интеллектуального агента // *Известия вузов. Приборостроение*. 2009. № 12. С. 26-31. - ISSN 0021-3454.
4. Бессмертный И.А. Семантическая паутина и искусственный интеллект // *Научно-технический вестник СПбГУИТМО*. - Санкт-Петербург: СПбГУИТМО, 2009. - Т. 64, вып. 6. - С. 77-83. - 122 с. - ISSN 1819-222X.
5. Информационные технологии в бизнесе / Под ред. М.Желены. – СПб: Питер, 2002.
Бессмертный И.А. Управление базами знаний с использованием прецедентов // *Научно-технический вестник СПбГУИТМО*. - Санкт-Петербург: СПбГУИТМО, 2011.