

ЭКСПЕРТНЫЕ СИСТЕМЫ И РАСПОЗНАВАНИЕ ОБРАЗОВ

Н.Г.Загоруйко

Термин "экспертная система" отражает как способ создания такой системы, так и ее основную характеристику. В процессе создания экспертная система наполняется информацией (знаниями) о предметной области, и эти знания берутся у специалистов данной области, обладающих высокой квалификацией, т.е. у экспертов. После того как с помощью этих знаний система построит достаточно полную модель предметной области, она получает возможность отвечать на нетривиальные вопросы так же квалифицировано, как лучшие эксперты в данной области. Таким образом, система строится экспертами и сама становится экспертом, помогающим принимать решения.

Разработчик экспертной системы должен создать средства, которые помогали бы экспертам передавать системе свой опыт и знания, и средства, которые позволяли бы системе использовать обобщенные знания экспертов для выработки квалифицированных ответов на сложные вопросы.

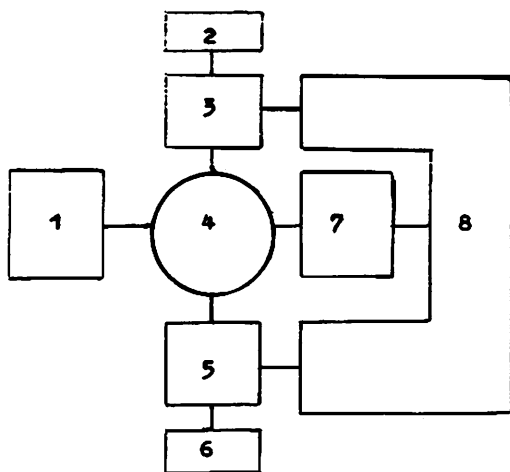
Прагматические требования к экспертной системе. Будем придерживаться следующего определения системы: экспертной системой называется система, предназначенная для удобной^{х)}, квалифицированной поддержки информационной деятельности человека в конкретной прикладной области. Имеются в виду такие виды деятельности, как познание (изучение объектов или явлений), проектирование объектов или процессов, анализ и прогнозирование ситуаций, принятие решений, планирование действий, направленных на выполнение принятого решения или на достижение определенной цели.

х) Подчеркнутые слова определяются ниже.

Понятие "удобная" поддержка деятельности связывается с наличием диалога на проблемно-ориентированном языке, близком к естественному. Диалог желательно иметь в привычной для человека графической или звуковой форме и в приемлемом масштабе времени. Поддержка деятельности может считаться "квалифицированной", если система будет обеспечивать правильное понимание входных сообщений, выполнение нужных процедур и выдачу разумных выходных сообщений.

Понять сообщение - это значит распознать его форму, т.е. узнать, что это: информация для запоминания, вопрос для поиска ответа или приказ для выполнения действия. Кроме того, нужно определить субъекта (кто говорит), объект (о ком говорит), предмет (что говорит) и степень достоверности сообщения. Если этих данных нет в сообщении, система должна их запросить. Процедуры, которые должна выполнять система, включают в себя такие технические действия, как стирание, запись, замена информации, а также и более сложные: такие, как поиск ответа на заданный вопрос, проверка на непротиворечивость новых данных и знаний старым данным и знаниям, заполнение пробелов в данных и знаниях, получение новых знаний из данных, адаптация к особенностям конкретного пользователя.

"Разумные" сообщения должны заслуживать доверие и быть прозрачными. Мы склонны больше доверять решениям, выработанным по правилам, которые в прошлом часто позволяли получать оправдывающиеся решения или прогнозы. Сообщение можно считать "прозрачным",



если оно легко интерпретируется пользователем или если система может объяснить все этапы получения решения на языке, понятном пользователю.

Структура экспертной системы. Прагматические требования к системе определяют как ее структуру, так и характеристики отдельных блоков. В состав системы должны входить блоки, показан-

ные на рисунке. Подсистема "Диалог" (1) обеспечивает удобную для человека связь с системой. "Диспетчер" (4) осуществляет планирование действий системы в ответ на полученное входное сообщение и управляет взаимодействием подсистем в ходе выполнения запланированных действий.

В "Базе данных" (БД) (3) накапливаются факты типа: свойство x_j у объекта a_i в момент времени t_k имело значение Q_{ijk} с достоверностью P_{ijk} : $\langle a[i], x[j], t[k], [Q_{ijk}], [P_{ijk}] \rangle$. Такие "атомарные" факты могут быть организованы в трехмерные таблицы типа "объект-свойство-время".

Знания в "Базе знаний" (БЗ) (5) удобно накапливать в виде логических конструкций (импликаций и продукций) типа:

Если $(x_1 \geq 3) \wedge (x_4 \neq 0) \wedge (8 \leq x_{11} \leq 15)$, то $(x_0 > 28)$.

В "Базе знаний" могут храниться также формулы для численных решений, тексты на естественном языке.

Блок "Предобработка БД" (2) должен делать слияние массивов, выделение части массива, нормировку данных, удаление отдельных частей массива данных. Блок "Предобработка БЗ" (6) должен выполнять те же процедуры на массиве знаний.

Подсистема "Аналитик" (7) включает в себя процедуры (программы) работы с массивами данных и знаний в процессе поиска ответа на заданный вопрос.

Подсистема "Гомеостат" (8) должна обеспечивать сохранение и даже улучшение свойств экспертной системы в ходе ее эксплуатации. Она должна обнаруживать противоречия между информацией, хранящейся в "Базе данных" и "Базе знаний" и новой информацией, заполнять пробелы в данных и знаниях, получать новые знания на основе старых знаний и новых данных, обнаруживать изменение закономерностей, убирать в архив устаревшие данные и знания, адаптировать систему к особенностям работы с конкретным пользователем.

Существующие экспертные системы не имеют многих из указанных функций, чем и объясняется тот факт, что 90% систем, называемых сейчас экспертными, работают на уровне специалистов ниже среднего уровня.

Методы распознавания в экспертной системе. Экспертные системы могут с успехом применяться для решения задач распознавания образов. Так, при распознавании слитной речи требуется использовать знания о физической природе речевого сигнала, о процессах

речеобразования, о фонетической, слоговой или морфемной структуре слов, о синтаксических, семантических и прагматических ограничениях [1]. Обращение к таким знаниям, полученным из экспериментальных данных или от специалистов-экспертов в области речи, языка и прикладной области, позволяет повысить надежность понимания речи. Однако в данной работе будет рассмотрена другая связь экспертных систем и распознавания образов, а именно использование методов распознавания в экспертной системе любого назначения.

Методы и средства распознавания образов должны использоваться прежде всего в подсистеме "Диалог". С их помощью можно облегчить ввод данных и знаний в экспертную систему и организовать удобный для человека диалог с ней в ходе эксплуатации. Для этого требуются читающие автоматы для ввода письменных (в том числе рукописных) текстов, схем, графиков, снимков. Важное значение имеет устный диалог человека с системой, при котором, во-первых, человек получает в свое распоряжение дополнительный и удобный способ общения с экспертной системой и, во-вторых, возникает возможность диалога с системой на расстоянии по телефону.

Достигнутые результаты в области распознавания и синтеза речи позволяют уже сейчас включать в блок "Диалог" устройства для устного общения человека с экспертной системой.

Система распознавания речи [2,3], разработанная в Институте математики Сибирского отделения Академии наук СССР совместно с Новосибирским государственным университетом, позволяет распознавать изолированные команды при словаре 200 слов и подстройке под диктора с надежностью не хуже 98%. Такую же надежность имеет и система распознавания дискретной речи, т.е. фраз, произносимых словами или короткими словосочетаниями, разделенными друг от друга короткими паузами. Устройство реализовано на микро-ЭЕМ. В качестве признаков используются интенсивности на выходе шести полосовых фильтров, измеряемые каждые 16 мсек. Распознавание слов ведется с помощью метода динамического программирования с адаптивным коридором [2,4]. При распознавании фраз используются лингвистические ограничения: синтаксические, семантические и прагматические, которые представлены в виде дерева сопрягаемости слов во фразах, каждая ветвь которого отображает допустимое продолжение фразы. При словаре 120 слов для 140 фраз длиной 3-7 слов коэффициент ветвления в вершинах дерева был в среднем равен 13.

Для экспертной системы профессионального назначения подстройка под диктора, которая в нашей системе состоит из однократного произнесения слов словаря, трудностей не представляет. Если экспертная система предназначена для общения с произвольным диктором, то высокой надежности удастся достичь при использовании признаков, имитирующих известный из психоакустики "эффект маскировки" [5,6]. Нерешенной пока следует признать проблему распознавания слитной речи произвольного диктора.

Успехи автоматического чтения машинописных текстов хорошо известны. Проблема чтения рукописных текстов находится пока в стадии лабораторных экспериментов. Что касается чтения рукописных цифр, то имеются методы, которые позволяют решать эту задачу достаточно хорошо. Так, разработанные в нашем институте алгоритмы [7] позволяют машине читать рукописные цифры с такой же надежностью, с какой это делает человек. Судя по публикациям, можно ожидать, что в недалеком будущем машина с достаточной надежностью и за приемлемое время сможет читать схемы, чертежи и фотоснимки.

Большую роль методы распознавания должны играть в подсистеме "Аналитик". Многие существующие экспертные системы используют только те знания, которые в готовом виде вводятся в них экспертами. Однако знания можно получать и автоматически из информации, накапливающейся в базе данных. В распознавании образов существуют эффективные методы построения логических решающих правил [8] и именно в том виде, который удобен для накопления в базе знаний. В медицинских экспертных системах наряду с правилами диагностики, сформулированными опытными врачами, следует накапливать информацию из фактических историй болезни с верифицированным диагнозом. Анализ этих данных позволит проверить врачебные правила и получить новые диагностические правила. При накоплении достаточно больших массивов данных возможно и целесообразно решать в системе все типичные задачи распознавания образов [1]: делать таксономию (кластерный анализ), выбирать подмножество наиболее информативных характеристик, строить решающие правила и предсказывать по ним имя образа для нового объекта. Обобщение последней задачи приводит к решению проблемы предсказания не только характеристик, измеренных в шкале наименований, но и в более сильных шкалах - шкале порядка, интервалов, отношений, абсолютной шкале [9]. Хорошо зарекомендовал себя при решении широкого круга прикладных задач алгоритм ZET для заполнения пробелов в эмпирических таблицах [9,

10]. При достаточно богатой базе данных его можно использовать для поиска ответа на вопрос: "Каково значение характеристики x_j у объекта a_i , если другие его характеристики имеют такие-то значения?"

В экспертной системе, предназначенной для управления экономикой сельского хозяйства, алгоритм ZET используется нами для прогнозирования, например, надоев молока по России на год вперед. Ошибка прогноза при этом не превышает 1,5%. Решаются и другие задачи прогнозирования сельскохозяйственного производства.

Если говорить о нерешенных задачах блока "Аналитик", то они связаны, в основном, с обработкой информации, хранящейся в "Базе знаний". В большинстве существующих экспертных систем знания анализируются только техникой, заложенной в языке ПРОЛОГ. При этом в процессе поиска ответа на вопрос используется метод резолюций, известный большой трудоемкостью, особенно если объем информации в "Базе знаний" достаточно велик.

Для ускорения многих процедур над знаниями было бы целесообразно сделать их таксономию, чтобы приближенные решения искать на множестве типичных представителей таксонов, а затем для более точного решения анализировать содержимое соответствующего таксона. Имеют смысл аналогии и других задач распознавания образов на массиве знаний: выбор системы информативных предикатов при той или иной фиксированной классификации имплицитивных связей или продукций; построение обобщенных логических закономерностей на базе импликаций и продукций; заполнение пробелов в знаниях.

Для этого необходимо иметь возможность измерять степень "близости", "похожести" между любыми двумя знаниями, сформулированными в виде импликаций. Вопросам построения метрики пространства знаний посвящена работа [11].

Подсистема "Гомеостат" и методы распознавания. Одна из функций подсистемы "Гомеостат" состоит в получении новых знаний путем анализа информации из "Базы данных". Для этого удобно пользоваться аппаратом логических решающих правил [7], который в таблице "объект-свойство" может находить логические закономерности имплицитивного вида. Для обнаружения быстрых изменений в состоянии динамических объектов можно применять сравнение логических решающих правил, полученных в последовательные моменты времени, используя для этого описанные выше меры расстояния между импликациями.

Алгоритм ЗЕТ в одном из своих режимов дает возможность определить, с какими данными из имеющихся в "Базе данных" новые данные связаны определенными закономерностями. Если обнаружится, что некоторая часть "Базы данных" перестает быть связанной с новыми данными, то это значит, что эта часть данных "устарела" и ее можно переслать в архив, оставляя в "Базе данных" только те данные, которые отражают текущие закономерности. Аналогичным способом можно выделять устаревшую часть и в "Базе знаний", обеспечивая хранение в указанных блоках только актуальной информации, что будет сохранять высокую реактивность экспертной системы.

Алгоритм ЗЕТ позволяет обнаруживать в массиве данных грубые ошибки, что предохранит экспертную систему от нарастания "склеротического шума". Располагая этими методами распознавания образов и анализа данных и их аналогами для анализа знаний, экспертная система будет улучшать свои характеристики в ходе работы.

Заключение. Была показана роль методов распознавания образов для построения наиболее важных блоков экспертных систем. Наряду с методами, хорошо изученными и давно используемыми, указывались задачи, требующие своей доработки или даже первоначального исследования. Дополнительно следует отметить, что экспертные системы, в которых данные и знания накапливаются во времени, делают весьма актуальными разработку методов анализа трехмерных таблиц типа "объект-свойство-время". При этом возникают новые задачи, которых не было при анализе плоских таблиц "объект-свойство", и новые возможности для обнаружения более глубоких закономерностей изучаемой предметной области.

Л и т е р а т у р а

1. ЗАГОРУЙКО Н.Г. Методы распознавания и их применение. -М.: Сов.радио, 1972. - 206 с.
2. БЕЛИЧКО Б.М., ЗАГОРУЙКО Н.Г. Автоматическое распознавание ограниченного набора устных команд. -В кн.: Вычислительные системы. Вып. 36. -Новосибирск, 1969, с.101-110.
3. БЕЛИЧКО Б.М., ЗАГОРУЙКО Н.Г. Автоматическое распознавание 200 устных команд. -В кн.: Вычислительные системы. Вып. 37. - Новосибирск, 1969, с. 75-76
4. SACOE H., CHIBA S. Dynamic Programming Algorithm optimization for Spoken word Recognition.- IEEE Trans.Acoust.Speech and Signal Processing, 1978, v.ASSP-26, N 1, p.43-49.

5. ЗАГОРУЙКО Н.Г., ЛЕБЕДЕВ В.Г. Эффект маскировки и автоматический анализ речевых сигналов. - В кн.: Вычислительные системы, вып. 61. Эмпирическое предсказание и распознавание образов. Новосибирск, 1975, с.103-111.
6. ЛЕБЕДЕВ V.G., ZAGORUIKO N.G. Auditory perception and speech recognition.-Speech Communication,1985, N 4, p.97-103.
7. НУДЕЛЬМАН А.С. Язык для описания плоских рукописных черно-белых фигур. - В кн.: Вычислительные системы. Вып. 44. Новосибирск, 1971, с. 141-154.
8. ЛБОВ Г.С. Обработка разнотипных экспериментальных данных. - Новосибирск: Наука, 1981. - 290 с.
9. PFANZAGLE J. Theory of Measurement.-Wüburd:Physica.-1971.
10. ZAGORUIKO N.G., YOLKINA V.N. Inference and Data Tables with Missing Values.-In: Handbook of Statistics, Volumer, North-Holland Publ. Comp.,1982, p.493-500.
11. БУШУЕВ М.В., ЗАГОРУЙКО Н.Г. Метрика пространства знаний. - Настоящий сборник, с. 24-35.

Поступила в ред.-изд.отд.
10 июля 1986 года