

## ФОРМИРОВАНИЕ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ВЕКТОРНОЙ ОПТИМИЗАЦИИ

Статья посвящена современным технологиям поиска знаний в данных. В ней приводится классификация основных методов и алгоритмов интеллектуального анализа данных по основным компонентам. Описан метод интеллектуального анализа, адаптированный к хранилищам данных, хранящим информацию о работе технических объектов. Эта технология базируется на методах векторной оптимизации. Процессом анализа управляет лицо, принимающее решение (ЛПР).

Современный уровень информационных технологий позволяет накапливать и обрабатывать базы данных, объемы которых выражаются экзабайтами. Это дало возможность в полном объеме сохранять информацию о различных процессах, происходящих в длительные промежутки времени. Человеческий мозг не способен воспринимать и анализировать такие объемы информации. В связи с этим возникает проблема интеллектуального анализа (Data mining) с помощью компьютера. Существуют многочисленные попытки создания алгоритмов Data mining, однако многообразие задач, которые требуется решить в ходе этих процедур, пока не позволило разработать универсальный алгоритм. В связи с этим представляется актуальной задача разработки новых подходов и средств для нахождения частных решений проблем Data mining'a [1].

Современные методы анализа данных возникли и развиваются на базе достижений прикладной статистики, распознавания образов, методов искусственного интеллекта, теории баз данных и др. Анализ различных систем Data Mining показал, что они основаны на методах и алгоритмах, которые можно классифицировать по ключевым компонентам, на которых они основаны [2].

**Предметно ориентированные аналитические системы** реализованы в методах технического анализа. Они представляют собой совокупность методов прогноза динамики показателей, основанных на различных эмпирических моделях динамики предметной области.

**Статистические методы** используют, в основном, корреляционный, регрессионный и факторный анализ. Недостатком методов этого класса считают то, что они опираются на статистическую парадигму, в которой главными объектами служат усредненные характеристики выборки, а эти характеристики при исследовании реальных объектов часто являются фиктивными величинами.

**Нейронные сети** имитируют архитектуру нервной ткани из нейронов. При разработке нейронной сети ее надо сначала «натренировать» на полученных ранее данных, для которых известны и значения входных параметров, и правильные ответы на них. Эта тренировка состоит в подборе весов межнейронных связей, обеспечивающих наибольшую близость ответов сети к известным правильным ответам.

Основным недостатком нейросетевой модели является необходимость иметь очень большой объем обучающей выборки. Другой недостаток заключается в том, что знания, зафиксированные как веса нескольких сотен межнейронных связей, совершенно не поддаются анализу и интерпретации.

**Системы рассуждений на основе аналогичных случаев** (case based reasoning), или метод «ближайшего соседа» (nearest neighbour), позволяют сделать прогноз на будущее или выбрать правильное решение, находя в прошлом близкие аналоги сложившейся ситуации и выбирая то решение, которое было для них правильным. Эти системы показывают очень хорошие результаты в самых разнообразных задачах, если правильно выбрана мера «близости» ситуаций, которая определяет размер множества прецедентов. Их основным недостатком является то, что они не создают моделей обобщения предыдущего опыта.

**Деревья решений** создают иерархическую структуру классифицирующих правил типа «ЕСЛИ..., ТО...». Для того чтобы решить, к какому классу отнести некоторый объект или ситуацию, требуется ответить на вопросы, стоящие в узлах этой структуры. Однако чем более разветвленной является структура, тем менее значимыми становятся вопросы и менее стабильной становится классификация данного объекта или ситуации. Кроме того, деревья решений дают полезные результаты только в случае независимых признаков. На настоящий момент данные проблемы не имеют решения.

**Эволюционное программирование** позволяет формулировать гипотезы о виде зависимости целевой переменной от других переменных, которые уточняются по определенному алгоритму. Система формулирует несколько гипотез, конкурирующих между собой в точности выражения искомой зависимости. Полученные результаты иллюстрируются средствами визуализации.

Другое направление эволюционного программирования связано с поиском зависимости целевых переменных от остальных в форме функций какого-либо определенного вида, например в форме полиномов.

**Генетические алгоритмы** представляют набор параметров, характеризующих объект в виде хромосом живого организма, к которым применимы методы эволюции: отбор наилучших (сильнейших) решений; структурный синтез (скрещивание) объекта из двух исходных; случайные изменения (мутация) одного или нескольких параметров. При этом предполагается достижение объектом состояния, которое не может быть улучшено по рассматриваемому критерию.

Недостатком генетических алгоритмов является эвристический характер отбора параметров и самой процедуры. То есть они не могут гарантировать глобального оптимального решения. Это сказывается при решении задач высокой размерности со сложными внутренними связями.

**Алгоритмы ограниченного перебора** вычисляют частоты комбинаций простых логических событий в подгруппах данных. При этом какой-либо параметр сравнивается с заранее определенными константами. Ограничением служит длина комбинации простых логических событий. На основании анализа вычисленных частот делается заключение о полезности той или иной комбинации для установления ассоциации в данных, их классификации, прогнозирования и т. д.

Метод демонстрирует более высокие по сравнению с другими алгоритмами показатели при решении практических задач.

В качестве такой задачи может быть рассмотрен *Data mining* в базе данных, описывающей динамику технического объекта, состоящего из множества однотипных элементов, например, динамику эксплуатации газового месторождения.

В настоящее время методы проектирования больших и сложных систем являются единствен-

ным доступным инструментом, который может быть использован в качестве базовой отрасли знаний, развивающей методы интеллектуального анализа баз данных технических объектов [3].

В общем случае метод системного анализа включает в себя:

- разработку общей схемы решения поставленной проблемы, в том числе построение обобщенной модели, и целевых программ;
- декомпозицию сложных систем и сложных объектов и операций их разработки и применения;
- методические основы формирования альтернатив;
- способы формирования целей и задач, назначение и согласование критериев;
- общую методологию подготовки и обоснования решений, включая неформальные процедуры.

Инструментом анализа являются математические модели.

Математические модели подразделяются на имитационные, оптимизационные и экспертные системы.

Имитационные модели позволяют осуществить сравнение альтернативных технических объектов воспроизведением в них течения технологического процесса. При этом альтернативные варианты объектов задаются извне, а не являются результатом решения оптимизационной задачи по выбору наилучшей конструкции. Таким образом, имитационные модели предназначены для сравнения конструкций путем вариантных просчетов. Оптимизационные модели в свою очередь направлены на отыскание оптимума конкретного объекта, они не предусматривают изменения исходных условий и корректировки конечных результатов.

Термины «имитационная» и «оптимизационная» характеризуют не саму модель, а способ ее использования. Одна и та же модель в зависимости от целей может быть имитационной и оптимизационной. Считается, что имитационные модели более сложные и включают большее число величин, задаваемых лицом, принимающим решения (ЛПР). Обычно имитационные модели используются в имитационных экспериментах, когда ЛПР взаимодействует с компьютером в диалоговом режиме, получая от него информацию о состоянии моделируемого процесса, анализируя ее и вводя в компьютер выработанные на основе анализа решения. То же самое происходит и при решении оптимизационных задач,

поскольку технология моделирования одинакова для всех видов математического моделирования, однако в имитационном моделировании ЛПР играет более важную роль, так как может накладывать условия и ограничения, которые трудно формализуются при включении в компьютерную программу.

Экспертные системы оперируют с логическими построениями на основе известного объема информации по указанной отрасли. С их помощью можно отыскивать технически осуществимые состояния рассматриваемого объекта. Они наиболее эффективны в тех случаях, когда оптимальное решение невозможно получить обычными вычислительными методами. Эти системы аналогичны вербальным моделям, но написаны языком математической логики.

В технологии формирования объекта его облик формируется на множестве альтернативных вариантов комплекса в замкнутом итерационном цикле с последовательным приближением к наиболее рациональному решению.

Для ускорения этого процесса используются человеко-машинные процедуры по выбору наилучших альтернатив. При векторном показателе эффективности выбор осуществляется на паретовом множестве, при этом могут иметь место два подхода:

- рациональный, построенный на оптимизационных расчетах и требующий формулировки принципа оптимальности [4];
- психологический, базирующийся на теоретических и экспериментальных психологических исследованиях, изучающих процессы принятия решений с позиции того, как это делает ЛПР, даже если его выбор и не всегда оптимален [5].

Оба подхода объединяются на базе диалоговых процедур ЛПР с компьютером, в которых математические модели несут в себе рациональное начало, а предпочтения ЛПР – психологическое.

ЛПР при сравнении объектов по векторному показателю эффективности может использовать две группы стратегий:

- стратегии компенсации, в которых человек стремится сопоставить оценки одной альтернативы с оценками другой;
- стратегии исключения, в которых исключаются из рассмотрения альтернативы, не удовлетворяющие требованиям по одному или нескольким показателям.

Если альтернативных вариантов не много, то обычно используются стратегии компенса-

ции. При большом числе альтернатив и показателей используется смешанная стратегия: сначала стратегия исключения, приводящая к небольшому числу альтернатив, а затем стратегия компенсации.

Принятие решения путем диалога ЛПР с компьютером позволяет выявлять предпочтения одновременно с исследованием допустимого множества действий. Процесс принятия решения при этом состоит из этапов синтеза, оптимизации и анализа. Общая схема процесса принятия решения приведена на рисунке 1.

В таком процессе происходит адаптация модели к системе предпочтения, вырабатываемой ЛПР, а ЛПР адаптируется к задаче. Адаптация модели к ЛПР связана с учетом информации, получаемой от ЛПР в виде решающего правила. Адаптация ЛПР к задаче происходит в результате осмысливания соотношения меж-

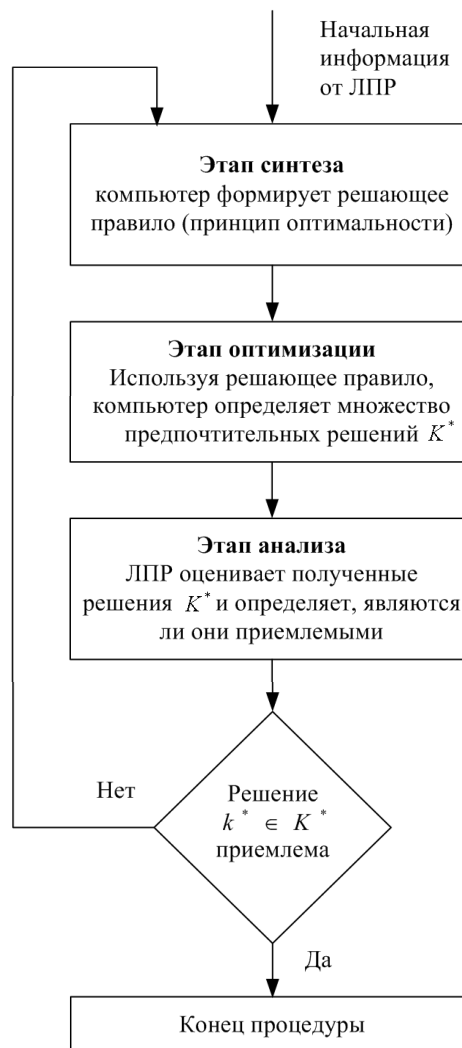


Рисунок 1. Человеко-машинная схема принятия решений в диалоговом режиме

ду своими потребностями и возможностями их достижения.

В распоряжении ЛПР имеет место БД, которую можно рассматривать как численное решение математической модели, описывающей состояние системы при некоторых известных граничных и начальных условиях. Подобное решение позволяет построить экспертную систему, которую можно подвергнуть анализу соответствующими методами, которые применяются для экспертных систем. Однако полученные результаты будут ограничены сложившимся комплексом возмущений и управлений. Для того чтобы иметь возможность решать имитационные и оптимизационные задачи на основании имеющейся БД, необходимо преобразовать ее в БД, содержащую меньшее количество элементов, являющихся критериями оценки состояния системы.

На основе проведенного анализа синтезируется технический объект с заданными свойствами. В этом случае управление объектом осуществляется изменениями однотипных управляющих параметров элементов, содержащихся в анализируемом объекте.

Формирование управлений (управляющих параметров) осуществляется на основе анализа ограниченного комплекса критериев, обеспечивающих необходимую достоверность состояний системы. В качестве критериев состояния системы – технического объекта удобно принять параметры эффекта. Под параметрами эффекта будем понимать промежуточное числовое значение между целевой функцией и характеристиками объекта. Характеристики объекта – это физические параметры, которые непосредственно не выражают пользу от применения объекта, хотя могут входить в состав параметров эффекта. Основным признаком параметра эффекта является то, что он определен через действие, выполняемое объектом, и поэтому является объективным показателем сравнительной оценки одних и тех же процессов.

Таким образом, Data mining удобнее проводить на комплексе параметров эффекта – массиве, составленном из параметров эффекта  $W_1$  в виде:

$$W_1 = W_1(r), 1 \leq l \leq L,$$

где  $r$  – множество параметров внутреннего состояния объекта;

$l$  – номер элемента массива параметров эффекта;

$L$  – размерность массива.

Для отыскания оптимальной области удобно применить алгоритмы ограниченного перебора [2]. Тогда ограничения параметров эффекта имеют вид

$$\begin{aligned} W_1 &\leq A_1, \\ B_1 &\leq W_1 \leq C_1, \\ W_1 &\leq D_1, \end{aligned}$$

где  $A_1, B_1, C_1, D_1$  – действительные числа.

Необходимое количество элементов комплекса параметров эффекта определяется исходя из точности описания моделью реального процесса. Эта точность не может превосходить точности вычислений параметров эффекта. Обычно точность вычислений для технических объектов, по мнению В.А. Панфилова, не хуже 20% [6]. Если использовать предложенную В.Г. Гмошинским [7] для метода инженерного прогнозирования нормирующую функцию весомостей характеристик технического объекта, требуемая точность требует использования пяти параметров эффекта.

Можно выделить три группы параметров эффекта технологического объекта [8]:

- параметры эффекта, характеризующие масштабы технологического процесса (например, производительность);
- параметры эффекта, определяющие ресурсосберегающее ведение процесса (например, энергоемкость);
- параметры эффекта, обеспечивающие необходимое качество получаемого продукта.

Параметры эффекта образуют  $n$ -мерную поверхность, которую можно подвергнуть анализу методом векторной оптимизации, т. е. выделяя участки поверхности, соответствующие введенным ограничениям параметров эффекта, и объявляя их оптимальными.

Метод векторной оптимизации на дискретном множестве [3] заключается в том, что в оптимальной области выбираются крайние точки, т. е. такие, в которых достигается максимум по какому-либо параметру эффекта. Эти точки будут эффективными. Из оставшихся точек выявляются те, для которых из числа найденных точек окажется хотя бы одна лучшая, у которой все параметры эффекта не хуже и хотя бы один лучше. Выявленные точки неэффективны и исключаются из рассмотрения. Из оставшихся точек опять выбираются крайние, оставшиеся снова сравниваются с найденными крайними точками и выявляют неэффективные. И так продолжается до тех

пор, пока не будут исследованы все исходные точки.

На рисунке 2 изображена схема управления техническим объектом. В процессе эксплуатации технического объекта в течение длительного времени его технико-экономические показатели заносятся в хранилище данных. Для обеспечения достоверности содержимого хранилища данных необходимо произвести его верификацию. Верифицированная информация формирует исходную БД. Ситуация, сложившаяся в процессе эксплуатации объекта, диктует проведение оптимизационного или имитационного моделирования ситуации. В зависимости от решаемой задачи информация из исходной базы данных трансформируются в базу данных параметров эффекта непосредственно – при решении оптимизационной задачи либо через построение корреляционных зависимостей – при решении имитационных задач. При этом мы исходим из того, что имитационные задачи требуют экстраполяции состояния исходной базы данных. Сформированная в процессе решения оптимизационной или имитационной задачи база данных параметров эффекта подвергается векторной оптимизации по комплексу ограничений параметров эффекта, задаваемых ЛПР. В результате векторной оптимизации формируется оценка состояния технического объекта, сообщаемая ЛПР, которое задает управляющее воздействие на технический объект. При помощи описанного алгоритма могут быть решены

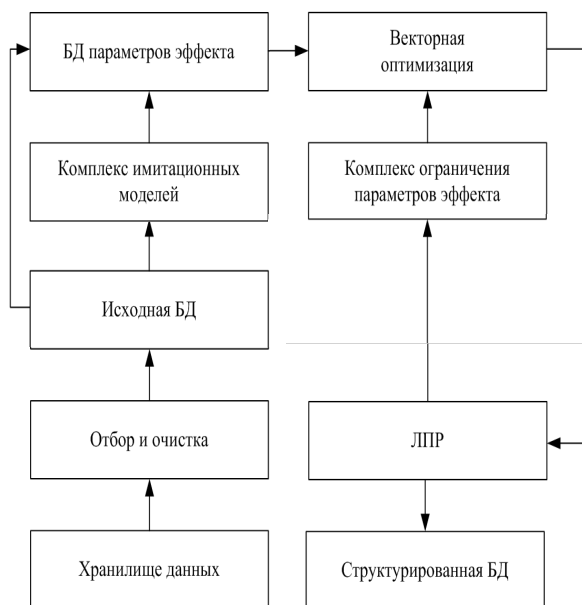


Рисунок 2. Схема интеллектуального анализа базы данных технического объекта

задачи мониторинга технического объекта и прогнозирования его состояния.

Таким образом, существует перспектива использования в задачах интеллектуального анализа баз данных технических объектов человеко-машинных систем алгоритмов, разработанных с использованием элементов методологии САПР. Реализация такого подхода позволяет организовать целенаправленный поиск ответов на поставленные вопросы, что может быть интерпретировано как «проектирование» знаний.

**Список использованной литературы:**

1. Дюк В., Самойленко А. Data mining: учебный курс. – СПб.: Питер, 2001. – 386 с.
2. Филиппов В.А. Интеллектуальный анализ данных: методы и средства. – М.: Эдиториал УРСС, 2001. – 52 с.
3. Формирование технических объектов на основе системного анализа / В.Е. Руднев, В.В. Володин, К.М. Лучанский, В.Б. Петров. – М.: Машиностроение, 1991. – 320 с.
4. Макаров И.М., Виноградская Т.М., Рубчинский А.А., Соколов В.В. Теория выбора принятия решений. М.: Наука, 1982. 327 с.
5. Проблемы принятия решения / Под ред. П.К. Анохина и В.Ф. Рубахина. М.: Наука, 1976. 319 с.
6. Панфилов В.А. Технологические линии пищевых производств (теория технологического потока). – М.: Колос, 1993. – 288 с.
7. Гмошинский В.Г. Инженерное прогнозирование. – М.: Энергоиздат, 1982. – 208 с.
8. Карташов Л.П., Полищук В.Ю. Системный синтез технологических объектов АПК. Екатеринбург: УрО РАН, 1998. – 185 с.