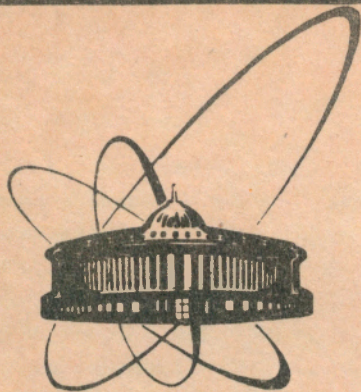


92-335



сообщения
объединенного
института
ядерных
исследований
дубна

P11-92-335

Ф. Фернандес Нодарсе

МОДЕЛЬ КОМПЬЮТЕРНОЙ ОБРАБОТКИ
ЭКСПЕРИМЕНТОВ В СИСТЕМЕ ПОДДЕРЖКИ
ПРИНЯТИЯ РЕШЕНИЙ COMEX

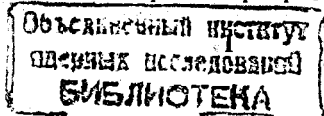
1992

В настоящее время жизнь ставит перед математическим моделированием новые задачи, связанные с исследованием взаимодействия сложнейших систем. Прогресс в информатике и вычислительной технике открывает перед математическим моделированием новые возможности. В этих условиях формируется "компьютерное моделирование" /1/ как новое направление, на стыке математического моделирования и системного программирования, с использованием в качестве цементирующего средства современных методов информационной технологии.

Целями разработки системы поддержки принятия решений (СППР) СОМЕХ являются: а) разработка СППР для оптимального или близкого к оптимальному выбору компьютерных средств для обеспечения экспериментов; б) моделирование предметной области с использованием алгоритмических и неалгоритмических знаний, т.е. разработка методики для оценки компьютерных ресурсов; в) использование созданной СППР для решения задач, связанных с планированием компьютерных ресурсов для проведения реальных экспериментов; г) исследование и создание СППР с новой архитектурой, базирующейся на знаниях, путем создания различных прототипов систем и проверки их практических возможностей. В этой работе подробно обсуждается задача компьютерного моделирования предметной области.

В отличие от традиционного математического моделирования, где под термином "формальное описание" понимается описание объекта на языке математики, в компьютерном моделировании под "формализованной информацией" понимают любую информацию, помещенную в память ЭВМ, если только в этой ЭВМ есть программы, способные "понять" ее, т.е. правильно интерпретировать эту информацию и использовать ее для порождения новой.

"Из подобной расширенной трактовки формализации непосредственно вытекает значительное расширение форм представления моделей - это и наборы экспериментальных данных, и тексты математических описаний, и экспертные высказывания, и графические образцы, и готовые программы. Как следствие, значительно расширяется область применимости. Так, если в математическом моделировании эта область естественным образом определяется наличием хорошо разработанного математического



аппарата, то в компьютерном моделировании - наличием соответствующей программы, например, содержащей некоторые эвристические правила" /1/.

Модель понимается как результат отображения одной абстрактной математической структуры на другую, тоже абстрактную, или как результат интерпретации первой в терминах и образах второй. Использование математической модели в современном, общепринятом в сфере научных исследований и техники смысле слова не связано с материальным воспроизведением подлежащих исследованию свойств и характеристик объекта и не предполагает экспериментальных процедур. Объект, описанный на языке математики, представляется некоторой математической структурой (дифференциальными уравнениями, передаточной функцией, графом и т.д.) с определенными параметрами, а процесс исследования (так называемое решение математической модели) заключается в применении к этой структуре совокупности математических преобразований и операций в соответствии с некоторым алгоритмом. Результатом вычислительного процесса является новая информация об исследуемом объекте, разумеется, в той части его свойств, которые нашли отражение в исходном математическом описании.

Модель компьютерной обработки эксперимента есть создаваемый с целью получения и (или) хранения информации специфический объект, отражающий свойства, характеристики и связи объекта - оригинала произвольной природы, существенные для задачи, решаемой субъектом. Модель представляет собой четырехместную конструкцию, компонентами которой являются субъект, задача, решаемая субъектом, объект - оригинал и язык описания или способ материального воспроизведения модели. В компьютерном моделировании различные средства описания объекта при моделировании, включая математическое, качественное /2/, имитационное, экспертное, информационное, теряют свою прежнюю автономию, превращаясь в компоненты более общего направления - представления знаний в ЭВМ, развивающегося на стыке искусственного интеллекта и системного программирования. "Компьютерное моделирование может заполнить брешь между математическим моделированием и искусственным интеллектом, включить математическое моделирование в процесс формирования средств представления знаний" /1/.

Основные этапы процесса компьютерного моделирования, и в том числе нашей задачи : 1) предмодельный анализ, включая сбор и обработку исходной информации, уточнение постановки задачи,

проектирование состава и структуры модели; 2) построение спецификации модели, т.е. разработка и отладка представления отдельных подмоделей, сборка модели в целом, идентификация параметров; 3) исследование модели, включая исследование ее качественных особенностей, адекватности, прогностических возможностей; 4) исследование моделей в различных режимах и в различных задачах; 5) интерпретация результатов моделирования; 6) генерация отчетов.

Для решения этой задачи автор обсуждает проблему соединения математической модели с экспертной СППР путем единого представления модели объекта. Проблема заключается в создании представления модели, учитывающей все особенности. В нашей работе ее роль играла модель классификации объектов или признаков.

1. О необходимости экспертных СППР для решения задачи выбора компьютерных средств при обработке экспериментальных данных

Неопределенность во всех оценках вычислительных ресурсов физических экспериментов, и в том числе в ИЕР, основывается на следующих факторах: а) в экспериментальной физике невозможно детально планировать вычислительные ресурсы физических экспериментов, т.к. многое может меняться в процессе эксперимента (в его начале и конце м. б. различны), включая идеи о data rate, эффективности детектора, распределение пучка (план) ускорителя и направления в физических исследованиях; б) скорость аккумуляции событий может значительно увеличиваться между началом и концом периода, тогда будет наблюдаться неуклонный рост числа аккумулярованных событий (которое определяет объем master Data Summary Files (DST)); в) место, где различные вычислительные задачи должны выполняться, может меняться в ту или иную сторону; г) некоторые из оценок могут оказаться такими высокими, что исследователи могут постараться уменьшить некоторые факторы, которые, как подсказывает опыт, будут реалистичны; д) системы обработки - это системы с обратной связью. Так как исследователи начинают понимать некоторые проблемы работы с данными, они могут использовать различные подходы.

Ключевые факторы, которые определили, что разработка экспертной СППР COMEX разумна, - это характер (задача требует оперирования символами и эвристических решений), сложность и широта постановки нашей задачи (задача представляет практический интерес и имеет решение).

Экспертная СППР, и в том числе COMEX, может функционировать в

качестве теории обработки информации или модели решения задачи в заданной области, давая ожидаемые ответы в конкретной ситуации и показывая, как изменяются эти ответы в новых ситуациях. Экспертная система может подробно объяснить, каким образом новая ситуация привела к изменениям. Это позволяет пользователю оценить возможное влияние новых фактов или информации и понять, как они связаны с решением. Аналогично пользователь может оценить влияние новых стратегий или процедур на решение, добавляя новые правила или изменяя уже существующие.

Если имеющаяся информация надежна и четко задана, то использование экспертной системы не рекомендуется. Если же доступные данные "замусорены", то применять экспертные системы рекомендуется. Тогда начнет играть нечеткая, неточная или какая-нибудь "нечистая" логическая схема.

Способность оценить собственные результаты - одна из важных особенностей, присущая практически любой экспертной СППР. Необходимость в такой оценке связана прежде всего с неполнотой, неточностью, а иногда и противоречивостью имеющейся информации о предметной области и с использованием вследствие этого правил правдоподобных рассуждений, не гарантирующих логической истинности заключения.

Исходя из типичных категорий способов применения экспертных систем, СППР СОМЕХ является планированной системой (определение последовательности действий, т.е. оптимальный или близкий к оптимальному план использования необходимых технических средств на единицу времени (например, на год - единицу срока эксперимента) для обеспечения требований к обработке эксперимента) и проектированной экспертной системой, т.е. выбирается оптимальный или близкий к оптимальному компьютерный комплекс.

Основными средствами представления знаний и получения решения в СППР являются распространяющие таблицы и специальные встроенные функции. СППР подобны пакетам прикладных программ (ППП), в которых для описания модели прикладной области используются уравнения, неравенства и другие математические конструкции с соответствующими методами получения оптимального или близкого к оптимальному решения, а также применяются средства представления знаний в виде правил и объяснения получаемых с их помощью решений, а также более качественное описание модели прикладной области. Помимо средств, типичных для систем, основанных на знаниях (СОЗ), или систем, базирующихся на знаниях (СБЗ), СППР СОМЕХ содержит реляционную

СУБД, подсистему работы с электронными таблицами, деловую графику, подсистему обработки текстов и другие средства интерфейса пользователя.

Выделенные знания о предметной области называются базой знаний, а общие знания о нахождении решений задач называются механизмом вывода или решения. Программа, которая работает со знаниями, организованными подобным образом, называется системой, основанной на знаниях. База знаний ЭС содержит факты (данные) и правила (или другие представления знаний), использующие эти факты как основу для принятия решений задач. Механизм вывода содержит интерпретатор, определяющий, каким образом применять предметные знания, и диспетчер, устанавливающий порядок применения различных кусков предметных знаний. В.Н. Агафонов и М.И. Дехтярь /3/ приходят к следующей практически полезной и применяемой в системе баз знаний (СБЗ) классификации знаний: 1) понятия (математические и нематематические), 2) факты, 3) правила, зависимости, законы, связи, 4) алгоритмы, процедуры.

База знаний состоит из правил или источников знаний, определяющих преобразования, которые нужно провести в БД в том или ином ее состоянии. Правила включают такие элементы, как область определения, область действия, условия применения, условия выбора, действие, приоритет и др. Для задания условий применения, условий выбора и действий определены специальные базисные операторы, обеспечивающие доступ и поиск по образцу в БД и списке выполнимых правил, а также возможности изменения БД, загрузки модулей БЗ и взаимодействия с пользователем.

Анализируя схемы систем баз знаний (СБЗ), можно выделить в них три основные части: базы знаний (БЗ), механизм получения решений (МР) и интерфейс (ИФ). Каждая из этих частей может быть устроена по-разному в различных системах, причем отличия могут быть как в деталях, так и в принципах. Следует также отметить, что между этими частями нет абсолютной границы, их размежевание довольно условно, они могут "пересекаться". В этом контексте полезно различать алгоритмические и неалгоритмические знания.

Принципы работы МР тесно связаны со способами представления знаний в БЗ. Для знаний, представленных в БЗ уравнениями, МР является процедурой решения уравнений, для знаний, представленных логическими формулами или правилами специального вида, это некоторый механизм вывода и т.д. В любом случае МР содержит алгоритм получения решения, т.е. специальное алгоритмическое

знание. Так как в МР содержится по крайней мере часть семантики БЗ, определенная в процедурной форме, это подтверждает относительность границы между МР и БЗ.

2. Модель компьютерной обработки эксперимента

Основные проблемы обработки данных могут быть сформулированы в виде общей оптимизационной задачи при соответствующем выборе оптимизируемого критерия качества метода. Покажем, как при надлежащем выборе оптимизируемого критерия качества метода и класса допустимых решений можно в рамках единого унифицированного подхода получать известные методы и модели в задачах: а) статистического исследования зависимостей; б) классификации объектов или признаков; в) сжатого представления данных.

Когда говорится о классификации совокупности объектов, подразумевается, что каждый из них задан соответствующим столбцом матрицы $X(t)$ [1], где каждый элемент $x_i^{(k)}(t)$ - значение k -го признака, характеризующего состояние i -го объекта в момент времени t , либо, что геометрическая структура их попарных расстояний (связей) задана матрицей. Аналогично интерпретируется исходная информация в задаче классификации совокупности признаков, что каждый из признаков задается соответствующей строкой матрицы.

В своей общей (нестрогой) постановке проблема классификации объектов заключается в том, чтобы всю анализируемую совокупность объектов $O = \{O_i\}$, $i=1, n$ разбить на сравнительно небольшое число однородных групп или классов. Для формализации этой проблемы удобно интерпретировать анализируемые объекты в качестве точек в соответствующем факторном пространстве; если исходные данные представлены в виде [1], то эти точки являются непосредственным геометрическим изображением многомерных наблюдений x_1, x_2, \dots, x_n в p -мерном пространстве X с координатными осями $Ox(1), Ox(2), \dots, Ox(p)$. Таким образом, в общем случае задачу классификации исследуемой совокупности объектов O , статистически представленной в виде [1], можно сформулировать как задачу поиска такого разбиения (правила классификации) z^* заданной совокупности O на непересекающиеся классы $z_1^*, z_2^*, \dots, z_k^*$:

$\bigcup_{j=1}^k z_j^* = O$, $z_i^* \cap z_j^* = \emptyset$ при $i \neq j$, при котором функционал качества

$Q(z)$ достигает своего экстремального значения на A , т.е.

$$Q(z^*) = \min_{z \in A} Q(z) \quad \text{или} \quad Q(z^*) = \max_{z \in A} Q(z);$$

при этом число классов k может быть как заранее заданным, так и неизвестным.

В общем виде проблему классификации применительно к модели компьютерной обработки эксперимента можно представить так:

а) входные и промежуточные данные:

$$S_{ijk}, \quad i=1,9; \quad j=1,10$$

i - элементы временного ряда < 10

j - тип объекта (требования)

$j=1$ - требование к CPU

2 - к суммарному файлу данных

3 - к среднему размеру файлов

4 - к минимальному сегменту файлов

5 - к дисковой памяти

6 - к оперативной памяти

7 - к передаче данных

8 - число пользователей

9 - число групп

10 - средняя ежедневная загрузка

$k=1, \dots, 6$ - задачи $\{F_k\}$, которые описывают область применения

Часть массива S_{ijk} можно определить, исходя из числа событий и "RUNS".

M - число уровней компьютерного комплекса ($1 \leq M \leq 3$)

G_{km} - процент задачи F_k , которая должна выполняться на уровне M .

U - выводы, полученные на основе использования качественной модели, когда это было необходимо на этапе u в процессе решения задачи.

R - процент ошибок в данных массива S

На основе (S, M, U) определяется X_m и Y_m

- X_m - число терминалов на уровне m

- Y_m - число рабочих станций на уровне m

Q_{ijk} - массив распределения требований на уровнях и по годам получен на основе (S, G, U)

- i - тип объекта

$i=1$ - мощность процессора

2 - размер суммарного файла данных

3 - средний размер файлов

4 - размер минимального сегмента файлов

5 - объем дисковой памяти

6 - объем оперативной памяти

7 - манипуляции с носителями данных

- j - уровень

- k - год

H_{hjk} - массив требований на уровне и по годам для каждого типа устройства, который определяется на основе (Q, X, Y, U, R) . Это называется системой требований.

- h - требования к устройствам = $\{A, B, C, D\}$

- j - уровень

- k - год (временного ряда),
где

$a_{i,w}$ - элемент массива A i-го требования к процессору w на уровне j и года k,
где i = 1- длина слова,
2- есть сопроцессор или нет
3- есть многопроцессорная система или нет
4- интерфейс
...

$b_{i,w}$ - элемент массива B i-го требования к внешнему запоминающему устройству w на уровне j и года k,
где i = 1- емкость
2- интерфейс
...

$c_{i,w}$ - элемент массива C i-го требования к устройству ввода / вывода w на уровне j и года k,
где i = 1- графический или нет
2- ввод / вывод
3- разрешающая способность
5- скорость передачи данных

d_i - элемент массива D i-го требования к линиям связи или сетям на уровне j и года k,
где i = 1- скорость передачи данных
2- тип линии

Массив H меняется в пределах:

$H_{h,j,k} \pm \epsilon$, где ϵ обозначает ошибки входных данных,

P (p1, p2, ...) - технические характеристики процессора из множества процессоров в базе данных

p1- скорость
p2- оперативная память
p3- стоимость
p4- интерфейс
p5- ежегодное изменение цены
...

V (v1, v2, ...) - технические характеристики внешнего запоминающего устройства прямого доступа из множества внешних запоминающих устройств в базе данных

v1- объем
v2- интерфейс
v3- стоимость
...

T (t1, t2, ...) - технические характеристики внешнего запоминающего устройства последовательного доступа из множества внешних запоминающих устройств в базе данных

t1- объем
t2- интерфейс

t3- стоимость
...

W (w1, w2, ...) - технические характеристики устройства ввода / вывода из множества устройств ввода / вывода в базе данных

w1- скорость записи / чтения
w2- интерфейс
w3- стоимость
...

L (l1, l2, ...) - технические характеристики линии связи и сетей из множества в базе данных

l1- скорость передачи данных
l2- тип линии
...

КЛАСС - это тип конфигурации вычислительного комплекса, который получается на основе матриц P, V, T, E, L, и набор правил (принципов) проектирования компьютерного комплекса, и в том числе набор правил, связанных с совместимостью между компонентами комплекса.

б) постановка задачи:

Выбор оптимального класса в зависимости от множества требований H и характеристик {P, V, T, E, L}, обладающего минимальной стоимостью.

При этом решаемая задача характеризуется нечеткостью и неполнотой используемых исходных данных.

Особенный интерес имеет понятие "стоимость". Стоимость каждого типа устройства или процессора - это функция от времени (ежегодный фактор обесценивания каждого типа устройства или процессора). Стоимость используется как взвешивающий фактор в оптимизационном уравнении для выбора компьютерного комплекса. Все значения каждой переменной хранятся в базах данных СППР СОМЕХ.

Из матриц технических характеристик (хранятся в базах данных процессоров и устройств СППР СОМЕХ) можно временно выбирать для каждого выполнения СППР СОМЕХ ("RUN") подмножество этой базы и обновлять все его значения. Это можно представить как трансформацию или проекцию данных матриц для каждого выполнения ("RUN") СППР СОМЕХ.

При определении компьютерного комплекса эксперимента, часть которого (комплекса) уже существует или будет использоваться для других целей, оказывается полезным ввести еще одну задачу (фиктивную), представляющую все задачи, которые выполняются в компьютерном комплексе, но не принадлежат этому эксперименту. Тогда переменную "стоимость" можно рассматривать как стоимость за час.

Пусть n категорий объектов, описанных замерами на определенном наборе переменных. Затем предъявляется еще один объект и набор замеров для этого объекта на том же наборе переменных. Проблема классификации состоит в определении, к какой из n категорий принадлежит объект.

Наиболее просто поставленная проблема решается в том случае, если указанные категории линейно сепарабельны. Это значит, что между каждой из категорий можно поместить разделяющую поверхность или ряд поверхностей.

Общее уравнение поверхности имеет вид

$$Y = \sum_{i=0}^{i=n} b_i X_i, \text{ где } b_i - \text{соответствующие константы или коэффициенты; } X_i - \text{переменная или расстояние, измеренное по } p\text{-осям.}$$

В результате получается значение y , однозначно определяющее, к какой группе принадлежит тот или иной объект.

Переменные X_i , $i=1, n$ могут быть независимы друг от друга или некоррелированы, и значения этих переменных, соответствующие конкретному объекту, определяют его положение в пространстве описания.

Можно предположить, что категории являются взаимоисключающими, т. е. объект в состоянии попасть только в одну из них. В этом случае поместим его в наиболее вероятную категорию. Другими словами, поместим его в такую категорию, для которой значение $P(H:E)$ максимально. Здесь гипотеза H определяет одну категорию, а событие E - все свидетельства, относящие объект к членам этой категории.

$P(H)$ - вероятность того, что событие H истинно.
 $P(E)$ - вероятность того, что событие E произошло.

$P(H:E)$ или $P(H:\text{не}E)$ - апостериорная вероятность гипотезы H , т. е. вероятность H при условии, что нам известен факт существования E . Согласно теореме Байеса, после некоторой перегруппировки получаем $P(H:E) = P(E:H)P(H)/P(E)$ и $P(E) = P(E:H)P(H) + P(E:\text{не}H)P(\text{не}H)$. В случае нескольких взаимоисключающих гипотез H_i , $i=1, n$ можно использовать общую формулу для вероятности некоторого свидетельства:

$$P(E) = \sum_{i=1}^n P(E:H_i)P(H_i).$$

Другой вариант, который использован в системе PROSPECTOR, $P(H:E) = LS * P(H) / P(\text{не}H) + LS * P(H)$, $LS = P(E:H) / P(E:\text{не}H)$.

Все приведенные формулы в действительности эквивалентны, так как LS - это отношение правдоподобия.

Для решения этой типовой задачи К. Нейлор /4/ предлагал более удобный алгоритм: работать поэтапно, суммируя отдельные свидетельства и их влияние на условную вероятность по мере наступления отдельных E_i . Это можно сделать, используя априорные и апостериорные вероятности.

Заметим, что математически $P(H:E)$ является уравнением поверхности. Зная $P(H:E)$, определяют константы b_i для уравнения поверхности, а на основании свидетельства E по каждой переменной вычисляют значения X_i . Если же категории не являются взаимоисключающими, то можно поместить объект в несколько категорий. В этом случае не просто выбираем наиболее вероятную категорию, а имеем дело с пороговым критерием.

Если разделяющая поверхность имеет вид $Y = \sum_{i=0}^{i=n} b_i X_i$, то можно отнести объект к какой-либо категории, если $Y > Y_c$.

в) стратегии поиска решений

Обоснуем выбор последовательных процедур для решения задач. По сравнению с параллельной процедурой, в которой используется выражение

$$Y_k = \sum_{i=0}^{i=n} b_{ik} X_i \quad (\text{для всех категорий } k),$$

в последовательной процедуре, которая в основном используется в СППР СОМЕХ, осуществляется переход от одной переменной к другой с учетом информации, полученной на каждом шаге, последовательно по всем i для любых категорий k , т. е.

$$Y_{ik} = Y_{i-1,k} + b_{ik} X_i.$$

Поскольку какой-либо вывод может быть получен на основе использования лишь части всей возможной информации, то целесообразнее применить последовательную процедуру, так как результат можно получить гораздо быстрее.

С использованием метода узлов для решения нашей задачи (математически такой вход описывается единичной ступенчатой функцией, которая и задает нелинейное преобразование данных) появился вопрос о промежуточных выводах.

Промежуточный вывод иногда используется в качестве входной переменной для другой ступени экспертной системы. Все зависит от

такого каверзного вопроса, как предположение о независимости различных свидетельств E_i . В большинстве статистических методов это часто ничем не подтверждается. Если существует n коррелированных свидетельств, подтверждающих справедливость гипотезы H , а вычисления проделаны в предположении, что такой корреляции не существует, то гипотеза H получит большую поддержку, чем этот факт того заслуживает. Введение промежуточных выводов в процессе рассуждений может помочь исключить подобный эффект. Присутствие промежуточных выводов позволяет в значительной степени избавиться от накопленных ошибок, но наличие таких выводов усложняет процесс написания системы. Последовательность вычислений, при меньшем числе свидетельств, снижает поэтому риск формирования новой неопределенности.

Применяется две наиболее эффективные и возможные стратегии поиска решения: 1- основанная на наличии цели и 2- основанная на наличии данных. Независимо от стратегии поиска решения существует некоторая свобода, с учетом которой должно быть проверено следующее значение X_i .

Пусть переменные X_i , которые должны вводиться в систему, имеют либо соответствующие минимальные и максимальные значения, либо справедливо $\bar{X}_i \pm \sigma_i$, тогда эта информация позволит процедуре принятия решения (MP) быстрее сформировать необходимое заключение. Здесь \bar{X}_i - математическое ожидание (часто используется и другой термин - среднее значение) и σ_i^2 - дисперсия. Допустим, все $\max(X_i)$ всегда приводят к определенной категории при классификации, тогда как все $\min(X_i)$ оспаривают подобную классификацию. Тогда для неизвестных еще нам переменных X_i вычисляются два возможных исхода, основанных на $\max(X_i)$ и $\min(X_i)$ (или $\bar{X}_i \pm \sigma_i$). Если ни один из них не опровергнет наиболее вероятный текущий исход, формируемый системой, то этот наиболее вероятный возможный исход и будет взят в качестве корректного возможного исхода. Тогда такие значения X_i фактически оказываются ненужными. Для выбора последующих переменных используется следующая процедура /4/:
1- в случае известных $\max(X_i)$ и $\min(X_i)$ (или $\bar{X}_i \pm \sigma_i$) для каждой неизвестной на текущий момент переменной X_i можно вычислить разность $[Y_{ik} \text{ (используя } \max(X_i)) - Y_{ik} \text{ (используя } \min(X_i))]$, чтобы определить ее максимально возможное изменение, 2- подсчитать значение $[P(H:E) - P(H:\text{не}E)]$, где E - наблюдение X_i , чтобы выяснить максимальное изменение переменной, вызванное каждым новым свидетельством.

Чтобы уменьшить затраты времени на обработку, вычисляется:
а) либо $\max(\max(X_i) - \min(X_i))$, либо справедливо $\max \sigma_i^2$ для того, чтобы судить о степени важности вопроса.

б) $\max(\text{дисперсия}(Y_{ik}))$, для того чтобы судить, до какой степени та или иная переменная используется при установлении категорий.

Недостаток метода "а" состоит в том, что значения X_i могут изменяться в широком диапазоне, но не быть важными в каком-то отношении; и недостаток метода "б" состоит в том, что переменная, которая, возможно, будет часто использоваться, имеет небольшую разницу ($\max(X_i) - \min(X_i)$) и большой диапазон изменения Y_{ik} , что и приводит к такому результату.

Другой подход состоит в использовании знаний о конкретной области применения и, в том числе, набора правил, которые позволяют скорректировать значения переменных и, в том числе, выбор последующих переменных, основанных на формуле полной вероятности ($P(E) = \sum P(E:Q_i) * P(Q_i)$), где условия Q_i , $i=1, n$, при которых известные условные вероятности события Q образуют полную систему событий для подсчета вероятности события $P(E)$).

Посмотрим проблему о неопределенности в ответах пользователя на основе решения К. Нейлора /4/. Пользователь, возможно, хотел ответить, ориентируясь на определенную шкалу, например, от $-r$ до $+r$, где $-r$ означает <нет>; 0 <не знаю>; $+r$ <да> и $r \in N$.

Пусть априорная вероятность любой реакции обозначается через $P(E)$. Пусть P - это отклик пользователя, тогда /4/

$$P(E) = \begin{cases} P(E) + (1 - P(E))P/r & \text{если } P \geq 0 \\ P(E) + P(E)P/r & \text{если } P \leq 0. \end{cases}$$

При $P=0$ параметр $P(E)$ остается неизменным. Тогда новое значение $P(H:E)$; $P(H:E)P(E) + P(H:\text{не}E)P(\text{не}E)$, т.е. вычисляются оба исхода и взвешивается окончательный исход с учетом определенности, которую пользователь выразил как в пользу данного свидетельства, так и против него.

Использование качественного рассуждения более предпочтительно, чем количественный подход, когда необходимо сделать выводы на базе небольшого количества информации (в таких случаях количественный подход не может быть осуществлен); выразить подтверждение в требуемых терминах, и также, где полная информация, необходимая для количественного моделирования, является непригодной (недоступной), где она является слишком сложной, чтобы быть

использованной успешно, или где она является практически незначимой для результата моделирования. Факторы, перечисленные выше, присутствуют в нашей задаче, поэтому следует использовать качественный подход к ее моделированию. Качественное рассуждение работает на структурном описании системы (компьютерного комплекса) обработки экспериментальных данных, которая моделируется. Структурные описания состоят из описаний системных компонент, их режимов и связей между компонентами. Автор комбинирует классическую математическую и качественную модель /5/ для того, чтобы более эффективно использовать их свойства для решения поставленной задачи /2/. Качественная модель, соединенная с экспертной СППР СОМЕХ, которая также включает эвристическое знание, была использована в качестве источника для рассуждения о ситуациях, в которых эвристические правила, полезные для решения общих проблемных случаев, были недостаточны.

Заключение

Применение теории экспертных систем, в том числе СППР, создание компьютерной модели являются новым решением проблемы выборки компьютерного комплекса, обеспечивающего научные эксперименты. В итоге:

1.-Были разработаны более изощренные средства, включаемые в БЗ (база знаний) и МР (механизм получения решения), ориентированные на более тонкое и глубокое моделирование. При этом фактически осуществлен переход от анализа типа "что если" к анализу типа "каков наилучший вариант", т.е. к системам, позволяющим вырабатывать наилучшие решения, а не только более или менее предпочтительные. Такие СППР подобны ППП, в которых для описания модели прикладной области используются уравнения, неравенства и другие математические конструкции с соответствующими методами получения оптимального или близкого к оптимальному решения. Развитие СППР происходит также в сторону применения средств представления знаний в виде правил и объяснения получаемых с их помощью решений.

2.-Была создана компьютерная модель области применения. Она представлена в виде модели классификации. Существуют обоснование ее решения, характеристика процесса решения задачи и ее особенностей и соединение математической модели с экспертной СППР путем единого представления модели объекта.

3.-Использована методика, позволяющая описывать не только

количественные, но и качественные параметры задачи и ее модель.

4.-Применена СППР для решения данных задач, что позволяет получать более обоснованные результаты, близкие к оптимальному решению, накапливать и хранить их результаты, входные данные и относительную информацию общего характера в виде текстов в машинной памяти и в большей степени "готовых к употреблению" для анализа и решения последующих задач.

5.-Базы знаний и данных СППР СОМЕХ открыты для будущего развития и дополнения, что является большим преимуществом, поскольку можно обогащать пользовательские знания об области применения. БЗ и БД позволяют накапливать информацию разных видов, что может быть полезным.

Автор приносит благодарность Г.А. Ососкову за критические замечания и полезные обсуждения.

Литература

- 1-В.А. Геловани, В.В. Юрченко, Компьютерное моделирование, Математическое моделирование, М.: Академия Наук СССР, 1989, т.1, N.1, 1989, с.3-12
- 2-Fernandez Nodarse, Ivanov V.G., Blackboard architecture and qualitative model in a computer aided assistant designed to define computers for NEP computing, PREPRINT JINR E11-91-560, Dubna, 1991
- 3-Системы управления базами данных и знаний, Справ. изд./ А.Н. Наумов, А. М. Вендров, В.К. Иванов и др. -М: Финансы и статистика, 1991
- 4-Chris Naylor, Build your own expert system, John Wiley&Sons Ltd., Chichester, 1987
- 5-J.MC Donell, Domain modelling in second generation expert system, Comput. Phys. Commun 61 (1990), 13-21, North - Holland

Рукопись поступила в издательский отдел
31 июля 1992 года.

Ф.Фернандес Нодарсе

P11-92-335

Модель компьютерной обработки экспериментов в системе поддержки принятия решений COMEX

В работе обосновывается положение о необходимости экспертной системы поддержки принятия решений (СППР) COMEX для оптимального или близкого к оптимальному выбора компьютерного комплекса для обработки экспериментальных данных. Обсуждается проблема представления знаний, и описывается компьютерная модель области применения, представленная в виде модели классификации. Приведены обоснование ее решения, характеристика процесса решения задачи и ее особенностей и соединение математической модели с экспертной СППР путем единого представления модели объекта.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1992

Перевод авторов

F.Fernandez Nodarse

P11-92-335

Model of the Computer Processing of the Experiments on the Computer Aided Assistant COMEX

In this paper the necessity to use an expert Computer Aided Assistant (CAA) COMEX to select the multicomputer system required to process the experimental data is found. The knowledge representation is analysed and the applied domain computer model using the classification model is described. The characteristics and peculiarities of the task solution and the combination of the mathematical model with the expert CAA using a united representation of the object model are found.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1992