

УДК 519.68:007.52

Экспертная система разработки и оптимизации состава материалов для космической техники

В. Г. Ситало¹, Т. Н. Литвишко¹, А. М. Гупал², А. М. Цветков²

¹ Державне конструкторське бюро «Південне», Дніпропетровськ

² Науково-учбовий центр прикладної інформатики Національної академії наук України, Київ

Надійшла до редакції 05.05.95

Описывается экспертная система для создания новых материалов, применяемых в космической технике. Она автоматически строит базу знаний, используя методы построения деревьев и листьев решений.

На сегодняшний день в мире накоплен большой опыт создания специализированных компьютерных банков данных и экспертных систем в разных предметных областях. К примеру, известны системы DENDRAL, SECHS, SYNCHEM (Левин, Дранг, 1990; Лорьер, 1991), используемые для определения структуры химических веществ. Данные системы имеют широкую область применения, с их помощью был получен ряд серьезных научных результатов. Кроме того созданы программы для построения банков данных и экспертных систем в области металлургии, в связи с чем необходимо упомянуть разработки немецкой фирмы HEXAGON и Донецкого государственного университета. В области искусственного интеллекта создано большое количество методов индуктивного и дедуктивного вывода знаний. Процесс автоматизации научных исследований идет от развития банков данных к построению экспертных систем и повторяет исторический процесс познания: от сбора и обработки эмпирических данных к обобщению экспериментальных фактов и использованию их в качестве базиса научных теорий. Развитие систем искусственного интеллекта, таких как экспертные системы, указывает на появление нового типа моделирования познавательной активности — инженерии знаний, который будет играть важную роль

в различных областях науки и технологии, где математическое моделирование оказывается не достаточно эффективным (в химии, биологии, материаловедении и т. д.).

Количество веществ, которые используются в различных областях народного хозяйства, в том числе в космической отрасли, постоянно возрастает. Закономерности, которые управляют поведением химических элементов в этих материалах, часто остаются непостижимо сложными. Затраты на эксперименты при отсутствии надежных знаний об этих закономерностях становятся экономически неэффективными.

Специфические условия эксплуатации конструкций в открытом космосе ставят задачу создания металлических и композиционных материалов и их соединений, которые удовлетворяют конкретным требованиям. Для прогнозирования ресурса работы конструкций необходимо учитывать не только такие особенности самого материала, как прочность, устойчивость к коррозии и другие, но также физические и технологические факторы космического пространства. Сегодня не существует единого комплекса критериев для выбора материалов для создания конструкций и оценки их физико-механических и эксплуатационных характеристик, так как процессы деградации в большинстве случаев

развиваются по нелинейным законам и практически невозможно вычислить корреляционные зависимости между различными факторами и изменениями, которые они вызывают.

Достоверность прогноза возрастает с накоплением инженерного опыта испытаний конструкционных материалов в реальных условиях. Для обобщения и оптимального использования накопленных результатов при создании новых материалов и прогнозирования ресурса работы конструкций, эксплуатирующихся в конкретных условиях, идеально подходят методы искусственного интеллекта, в частности методология экспертных систем индуктивного вывода.

Предлагаемая статья посвящена разработке системы поддержки принятия решений для создания и оптимизации состава сплавов с заранее заданными характеристиками для использования в аэрокосмической технике. Данная работа выполняется совместно Научно-учебным центром прикладной информатики НАН Украины и КБ «Южное» в рамках государственной космической программы Украины.

ПОСТАНОВКА ЗАДАЧИ

Первым шагом на пути построения широкопрофильной экспертной системы для создания материалов с заданными свойствами является решение задачи прогнозирования механических свойств материала по его химическому составу и технологии изготовления. За основу была взята база данных об алюминиевых деформируемых сплавах, используемых в космической технике. База содержит количественную информацию о легирующих компонентах (Cu, Mg, Mn, Ni, Fe, Si, Zn, Ti, Cr, Zr, Be, Sc, Li), примесях (Fe, Si, Cu, Mn, Ni, Zr, Zn, Ti, Mg, Be, Cr), технологических параметрах (способ про-

изводства — прокатка, прессование, штамповка, состояние материала — без термообработки, отжиг, нагартовка, нагартовка повышенной прочности, закалка + естественное старение, закалка + искусственное старение), механических свойствах (предел прочности, предел текучести, удлинение, сужение, модуль упругости). База состоит из 250 записей, ее фрагмент приведен в табл. 1. Вопрос оптимального количества элементов в базе данных решается отдельно для каждого решаемого класса задач. Это число зависит от количества прогнозируемых параметров и характера вероятностных распределений на начальных данных. На основе этой информации одним из методов, описанных ниже, автоматически строится база знаний, которая содержит обобщенную информацию о закономерностях и зависимостях между химическим составом, технологией изготовления и механическими свойствами материалов. Детальное описание и обоснование методов можно найти в работах Гупала, Цветкова (1991, 1992) и Цветкова (1993). Затем эта база знаний используется для прогнозирования по заданным химическому составу и технологии изготовления механических свойств новых разрабатываемых материалов.

МЕТОД ДЕРЕВЬЕВ РЕШЕНИЙ

Дерево решения является наглядным представлением процедуры определения принадлежности объекта к тому или иному классу. Каждый узел дерева определяет или имя класса, или тест, в соответствии с которым производится дальнейшее разбиение на поддерева. Формально дерево решения определяется как структура, состоящая из:

- а) листовых узлов (или узлов ответа), содержащих название класса;

Таблица 1. Фрагмент базы данных

Марка сплава	Легирующие компоненты						Технология			Механические свойства		
	Cu	Mg	Mn	Si	Ti	Zr	отжиг	естественное старение	искусственное старение	Sv	S02	S
AMГ2	—	1.8—2.6	0.2—0.6	—	—	—	+	—	—	17	8	15
01570	0.1	5.3—5.6	0.2—0.6	—	—	0.05—0.15	—	—	—	43	31	5
B93	0.8—1.2	1.7—2.8	0.1	—	—	—	—	—	+	48	44	6
AMГ6	—	5.8—6.9	0.5—0.8	—	0.02—0.1	—	—	—	—	38	29	6
AD32	0.15—0.4	0.8—1.2	—	0.4—0.8	—	—	—	+	—	18	11	15
1911	0.1—0.25	1.6—2.1	0.1—0.5	—	—	0.13—0.22	—	—	+	38	28	10

б) нелистовых узлов (или узлов решения), содержащих атрибутный тест, соединенных с другими узлами в соответствии со значениями тестируемого атрибута.

Каждый пример описывается списком пар атрибут — значение и приписывается к тому или иному классу. Для простоты будем считать, что имеется всего два класса; примеры, принадлежащие первому считаются положительными, второму — отрицательными. Представленный ниже алгоритм естественным образом может быть обобщен на несколько классов.

Множество атрибутов, используемых для описания примеров обозначим через A , a — конкретный атрибут, $a \in A$, $1 \leq i \leq |A|$, где $|A|$ — число атрибутов. Для каждого атрибута a множество возможных значений обозначим через V_i , v_{ij} — конкретное значение атрибута, где $1 \leq j \leq |V_i|$, и $|V_i|$ — количество значений, которое может принимать атрибут a_i .

Алгоритм построения дерева решений состоит из двух шагов:

1. Если все примеры принадлежат одному классу, тогда дерево решений — листовая узел, содержащий имя класса.

2. В противном случае:

а) определяется a_{best} , как атрибут с наименьшей E -мерой;

б) для каждого значения $v_{\text{best}, i}$ атрибута a рекурсивно строятся деревья на основе примеров, имеющих значение $v_{\text{best}, i}$ атрибута a_{best} .

E -мера вычисляется следующим образом. Пусть для данного узла p — число положительных примеров; n — число отрицательных примеров; p_{ij} — число положительных примеров со значением v_{ij} атрибута a_i ; n_{ij} — число отрицательных примеров со значением v_{ij} атрибута a_i . Тогда

$$E(a_i) = \sum_{j=1}^{|V_i|} \frac{p_{ij} + n_{ij}}{p + n} I(p_{ij}, n_{ij}),$$

где

$$I(x, y) = \begin{cases} 0, & \text{если } x = 0, \\ 0, & \text{если } y = 0, \\ -\frac{x}{x+y} \lg \frac{x}{x+y} - \\ -\frac{y}{x+y} \lg \frac{y}{x+y} & \text{иначе.} \end{cases}$$

Можно показать, что выбор атрибута с наименьшей E -мерой эквивалентен выбору атрибута с наиболь-

шей информационной мерой, определяемой как $I(p, n) - E(a_i)$.

E -функция — это теоретико-информационная мера, основанная на энтропии. Эта функция оценивает меру неопределенности в классификации при использовании выбранного атрибута в узле решения. Считается, что атрибут, имеющий наименьшую E -меру, обладает наибольшей классифицирующей силой.

E -метрика обладает определенным недостатком — она дает при построении дерева преимущество атрибутам с большим числом возможных значений. Улучшенный вариант метрики, устраняющий этот недостаток, называется метрикой отношения и имеет вид:

$$E^r(a_i) = \begin{cases} \frac{I(p, n) - E(a_i)}{IV(a_i)}, & \text{если } I(p, n) - E(a_i) \geq g; \\ -\infty & \text{в противном случае.} \end{cases}$$

$$IV(a_i) = - \sum_{j=1}^{|V_i|} \frac{p_{ij} + n_{ij}}{p + n} \lg \frac{p_{ij} + n_{ij}}{p + n},$$

$$g = \frac{1}{|A|} \sum_{i=1}^{|A|} [I(p, n) - E(a_i)].$$

Данный алгоритм позволяет построить для непротиворечивого множества примеров полное дерево, т. е. дерево, на основе которого каждый пример из исходного множества будет классифицироваться правильно.

МЕТОД ЛИСТЬЕВ РЕШЕНИЙ

Пусть понятие S определено на множестве объектов, каждый из которых описывается набором пар атрибут — значение. Пусть N — множество атрибутов, a_i — i -й атрибут, $1 \leq i \leq N$, A_i — множество значений i -го атрибута, $|A_i|$ — количество различных значений атрибута a_i , a_{ij} — значение i -го атрибута j -го объекта, где $1 \leq j \leq m$, m — число объектов, описывающих понятие. Каждый объект приписывается к одному из классов $P_k \in P$, $1 \leq k \leq l$, где l — число классов. Очевидно, $l \leq m$. Каждый объект с соответствующим классом называется примером S . Далее для простоты будем считать, что $|A_i| = q$ для всех i . Таким образом, число всевозможных термов длины меньшей или равной k составляет

$$W = \sum_{i=0}^k C_i^{2nq} = O(n^k).$$

Лист решений — это список L пар

$$(f_1, v_1), \dots, (f_r, v_r), \quad (1)$$

где каждый f_i — терм из числа W , каждое v_i — значение из P и последняя функция f_r — постоянная функция «истина». В случае, если $v_r \in \{0, 1\}$, выражение (1) естественным способом определяет булеву функцию: для каждого набора $x \in X_n$, $L(x)$ равен v_j , где j — наименьший индекс, такой, что $f_j(x) = 1$.

Лист решений формализует правило типа if-then-else-if...else...

В качестве меры информативности атрибутов при построении листа решений используется мера, основанная на энтропии. Пусть Y — дискретная переменная со значениями из R_y . Для каждого $y \in R$ пусть $p(y) = P\{Y = y\}$. Энтропия Y определяется формулой:

$$H(Y) = \sum_{y \in R_y} p(y) \log_2 \frac{1}{p(y)}.$$

Энтропия $H(y)$ отражает степень неопределенности для Y . Теперь для двух дискретных случайных переменных X и Y определим меру неопределенности для X при условии, что наблюдается Y . Пусть x — значение из X и y — значение из Y .

Обозначим

$$p(y|x) = P\{Y = y | X = x\},$$

$$p(y, x) = P\{Y = y, X = x\}.$$

Тогда условная энтропия $H(Y|X)$ определяется формулой:

$$H(Y|X) = \sum_{y \in R_y, x \in R_x} p(x, y) \log_2 \frac{1}{p(y|x)}.$$

Условная энтропия отражает количество неопределенности для Y при условии, что перед этим наблюдалась X . Следовательно, величина

$$I(Y, X) = H(Y) - H(Y|X)$$

представляет количество информации о Y , даваемое X .

Покажем теперь, как вычислять $I(Y, X)$ на множестве примеров S . Пусть

$$S_y = \{ \langle x, c \rangle | c = y \},$$

$S_{x,y} = \{ \langle x, c \rangle | c = y, x \text{ имеет значение } x \text{ для атрибута } X \},$

$$\pi_y = \frac{|S_y|}{|S|}.$$

Условная вероятность определяется формулой

$$p(x|y) = \frac{p(x, y)}{p(y)}.$$

Так как

$$p(x|y) = \frac{|S_{x,y}|}{|S_y|},$$

то

$$p(x, y) = \pi_y \frac{|S_{x,y}|}{|S_y|}.$$

С другой стороны,

$$p(x) = \sum_{y \in R_y} \pi_y \frac{|S_{x,y}|}{|S_y|}.$$

Окончательно

$$p(y|x) = \frac{p(x, y)}{p(x)}.$$

Алгоритм построения листа решений

Вход: непустое множество примеров S ,
DL — пустой лист решений.

While не все примеры из S принадлежат к одному классу K **do**

P = пустое множество и term = true

While не все примеры из S принадлежат к одному классу K **do**

x = выбор литерала (S , term);

образовать конъюнкцию x и term;

удалить из S все примеры со

значением 0 для x и

поместить их в P ;

end;

Добавить к листу решений пару (term, K);

$S = P$;

end.

Выбор литерала предлагается производить одним из двух способов. Исходное множество примеров S преобразуется в множество S^* следующим образом.

Для каждого атрибута a_i со значениями из A_i создаются булевы переменные $b_{a_i}^{(c)}$ вида $a_i = c$,

$c \in A_i$, и исходная таблица соответствующим образом преобразовывается в бинарную.

Итак, первый алгоритм выбора литерала имеет вид:

Итак, первый алгоритм выбора литерала имеет вид:

Вход: множество примеров S^* .

Для каждого атрибута $b^{(c)}$ определить $I(b^{(c)}|P)$;

Выбрать $b_{a_i}^{(c)}$ с максимальным значением I ;

Для $i = 0, 1$

пусть S^* — множество примеров в S со значением i атрибута $b_{a_i}^{(c)}$.

Если $H_{s_0}(P) \leq H_{s_1}(P)$, тогда

выбранный атрибут — ($a_i = c$),

иначе

выбранный атрибут — ($a_i \neq c$).

Таблица 2. Результаты тестирования

Сплав	Химический состав, %	Примеси	Технология	Прогнозируемые механические свойства	Реальные значения механических свойств
АМГ2	Mg: 1.8—2.6 (4) Mn: 0.2—0.6 (3)	Fe: 0.4 (2) Si: 0.4 (5) Cu: 0.1 (2) Zn: 0.2 (3) Ti: 0.1 (3) Cr: 0.05 (2)	отжиг	S _v : 14—27 S ₀₂ : 1—15 S: 5—8	19 10 6
АМГ6	Mg: 5.8—6.8 (6) Mn: 0.5—0.8 (3) Ti: 0.02—0.1 (3)	Fe: 0.4 (2) Si: 0.4 (5) Cu: 0.1 (2) Zn: 0.2 (3)	отжиг	S _v : 28—39 S ₀₂ : 1—15 S: 15—16	34 15 20

Второй способ не требует преобразования S в S^* :

Вход: множество примеров S .

Выбирается атрибут a_i с наибольшим значением $I(a_i | P)$;

Для каждого значения $a_i = c_n$, $c_n \in A_i$ строятся подмножества S_{c_1}, \dots, S_{c_n} , где S_i — подмножество примеров S со значением f атрибута a_i ;

Выбирается подмножество с минимальным значением $H(P)$ и выбранный атрибут — ($a_i = f$).

В отличие от первого способа генерируются только литералы типа ($a_i = f$).

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ЭКСПЕРТНОЙ СИСТЕМЫ

Для решения поставленной задачи создана оболочка экспертной системы индуктивного вывода АВС, в которой реализованы описанные выше методы. Она позволяет обрабатывать базы данных, созданные в формате DBASE, выбирать метод решения задачи, просматривать и выводить на печать получаемые результаты прогнозирования.

Система тестировалась на Al-деформируемых сплавах АМГ2 и АМГ6, которые были изъяты из исходной базы данных, и следовательно, были неизвестны системе. По химическому составу и технологии изготовления прогнозировались механические свойства данных материалов (табл. 2). В скобках после процентного состава указано значение, соответствующее принятому в системе шкалированию.

В результате тестирования получена достоверность прогноза на уровне 70 %, что позволяет сделать вывод о целесообразности применения разработанной системы для решения различных классов задач из области материаловедения.

Гупал А. М., Цветков А. М. Разработка и реализация алгоритмов индуктивного вывода // Использование математических методов и ЭВМ в системах управления и проектирования. — Киев: Ин-т кибернетики им. В. М. Глушкова АН Украины, 1991.—С. 33—37.

Гупал А. М., Цветков А. М. Об одном методе индуктивного вывода, основанном на комитетных конструкциях // Кибернетика и системный анализ.—1992.—№ 5.—С. 159—161.

Левин Р., Дранг Д. Практическое введение в технологию искусственного интеллекта. — М.: Финансы и статистика, 1990.—238 с.

Лорьер Ж. Л. Системы искусственного интеллекта. — М.: Мир, 1991.—568 с.

Цветков А. М. Разработка алгоритмов индуктивного вывода, основанных на построении деревьев решений // Кибернетика и системный анализ.—1993.—№ 3.—С. 174—178.