知识工程中自然语义的模糊表达

陈国权 编著

哈姆雷特当时正慢慢走向门边

$\rightarrow$ 走[姓名＝哈姆雷特；$\pi_{速度}$＝慢；

$\pi_{时间}$＝过去；$\pi_{方向}$＝向[客体＝门]]

晏子不是非常高是非常真的

$\rightarrow \pi_{身高(晏子)}(u)=\mu^2_{真}(1-\mu^2_{高}(u))$

贾宝玉爱林黛玉 $\rightarrow \pi_X=\perp$

$X \triangleq_\mu$ 爱[姓名1＝贾宝玉；

姓名2＝林黛玉]

马力大的昂贵红色小汽车

$\rightarrow$ 小汽车[$\pi_{价格}$＝昂贵；$\pi_{颜色}$＝红色；

$\pi_{马力}$＝大]

科学出版社

知识工程中自然语义的模糊表达

陈国权 编著

科 学 出 版 社

1 9 8 9

内 容 简 介

本书是为关心知识工程这个新领域发展的读者写的，它系统地展示了模糊数学的应用对自然语言意义表达带来的重要突破．本书的出版将会引导读者朝着人工智能、专家系统和知识工程的研究方向深入发展，并促进其推广应用．

本书从基础理论开始，终结于实际应用，主要内容包括：为初学者准备一些必要的模糊数学知识，介绍以模糊数学为基础的可能性理论，阐述测分语义学的特性，介绍一种模糊意义表达语言，给出大量翻译实例，介绍自然语言意义表达法在近似推理中的应用，采用这种近似推理的计算机实现．附录中给出隶属函数的求取与自然语义量化的经验研究方法．

本书可供从事语言识别、人工智能、知识工程、自动控制、经济管理科学、语言学、心理学和模糊数学等领域的科研人员及高等院校有关专业的师生参考．

知识工程中自然语义
的模糊表达

陈国权 编著

责任编辑 鞠丽娜

科学出版社 出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1989年2月第一版 开本：787×1092 1/32
1989年2月第一次印刷 印张：6 5/8
印数：0001—2,660 字数：147.000

ISBN 7-03-000760-3/TP·48

定价：4.90元

前　　言

目前，机器系统和人文系统的控制都存在着智能化的倾向，计算机一体化加工系统（CIMS）则要求一系列的人工智能技术．这种系统，在生产的层面上要求把自动化的单机加以连接，把自动化技术向自动化程度低的前道工序和生产线末端的两个方面延伸，把生产线之间的连接建立起来，并且要求将生产过程与其技术和物质支持部门加以连接；在第二个层面上，企业人事、规划、计划、财务、营销、市场预测、新产品开发等也将逐渐成为整个系统的有机组成部分，通过计算机与生产过程合一，这样的目标，没有智能化是无法实现的．在单纯的社会系统的控制方面，由于这类系统中人的因素更占着主导地位，其理论基础本来就兼有数理学科和人文学科两种性质，因此对智能化的要求可以说更为直接和迫切．人们普遍认为，智能化的管理决策系统，即人工智能专家系统与决策支持系统（DSS）的结合应用是第二次电子革命的一个极为重要的方面，也是未来 10 年的主要挑战之一．

人类的知识可以分为两类，一种是已经由数学、物理、化学等学科所表达的结构性知识，另一种是由人类所掌握，但未能由硬科学各个部门所表述的以经验和直觉为基础的专家知识，后者或可称之为非结构性的知识．对智能化的要求，一方面表明人类要求把自身的非结构性的智力活动和专家知识的使用机器化，一方面表示现代硬科学未能覆盖人类知识的这个领域而需要进一步发展．

什么是人工智能？人工智能是人类学习、推理、计划、探

索、自适应、模式识别、自然语言理解、知识处理等能力的机器化。专家系统则是一种操作计算机的程序系统，这种程序系统能够在计算机中建立并改善表达专门知识的知识库，并模仿专家运用知识库中的知识进行演绎、归纳和解答问题。因此，专家系统是人工智能的组成部分。

人们往往对人工智能抱着过高的期望，这个学科发展30年来已经历了二至三个期望过高然后是失望的阶段。专家系统的瓶颈则是以收集和处理专家知识为任务的知识工程学。收集专家知识，弄清专家如何作判断显然是一个艰难的任务，它要求一套新的方法和技术。今天，开发一个专家系统一般要花几年的时间。

无论是人工智能、专家系统或是知识工程都需要自然语言的表达和理解技术。科学家们已经发现，这个任务比预料的更加艰难。美国国会技术评价办公室最近指出，要使计算机具备一个5岁小孩的自然语言理解能力说不定是20年之后的事。

自然语言的表达与理解的主要困难在于自然语言本身存在许多含混的因素，如定义不清，基准模糊，理解一个陈述句、一个问句或一种描述要求大量的背景材料。困难的内在原因是我们对人类如何贮存和处理模糊信息的机制尚不十分清楚，外在原因是我们还没有一种适合于处理这类信息的工具。

模糊数学方法对理解人类思维机制方面能否有所促进，作用如何，目前尚无足够的材料可以评价，但是大量文献表明，模糊数学方法是模糊信息处理的有效工具。由模糊集合论创始人L. A. 扎德亲自开拓的可能性理论、模糊语言方法以及由此而产生的模糊语言逻辑、自然语言意义表达和近似推理已构成一个知识分支，正在把克服上述计算机科学发展中的已经展示或正在展示出来的主要障碍当做自身的目

标。对于这样一个对软工程甚至软科学可能有深远影响的知识分支，目前尚无出版物对其作较系统的叙述，这本小册子是在作者为研究生授课的讲义的基础上补充整理而成的。它试图填补这方面的空白。

本书以自然语言意义表达为中心，先介绍作为自然语言意义表达的模糊数学方法的理论基础的可能性理论，进而阐述仿照人类理解自然语言的过程的一种测分语义实验。在这个基础上详细叙述自然语言的意义表达语言的基本结构及翻译规则，然后以大量实例说明翻译的方法；紧接着翻译举例的是近似推理方面的内容，它是作为自然语言意义表达法的应用而设置的，最后抱着使本书的方法尽可能早日付诸实用的愿望，介绍了模糊推理的计算机实现方法。

为了使这本小册子自成系统，使没有学过模糊数学的读者也能在不参考别的书籍的基础上读懂它，我们引入了模糊数学基本概念、基本运算方法以及与语言变量和模糊逻辑有关的内容。

这本小册子的内容可广泛地应用于自然语言处理、语言识别、人工智能、专家系统、知识工程、机器人学、信息检索、医疗诊断、决策支持系统等领域，可供从事计算机科学、自动控制、经济管理科学、语言学、心理学和模糊数学的研究工作者和高等院校的有关师生参考。

中国科学院心理学研究所的马谋超同志认真审阅了原稿，并提出了许多宝贵的意见。为保持本书的完整性，作者特邀请马谋超同志为本书撰写了附录，在此向他表示深切的谢意。

作者才疏学浅，书中不足之处在所难免，望广大读者不吝指正。

作者

1987年10月

目　　录

第一章　集合与模糊集合…… 1
1.1　集合 …… 1
1.2　模糊集合 …… 4
1.2.1　模糊集合的概念及术语 …… 4
1.2.2　模糊集合上的运算 …… 10
1.2.3　模糊关系 …… 14
1.2.4　投影和柱状模糊集合 …… 15
1.2.5　扩展原理 …… 19
1.2.6　有模糊隶属函数的模糊集合 …… 23
第二章　可能性理论…… 29
2.1　可能性分布 …… 30
2.1.1　模糊限制与可能性分布 …… 30
2.1.2　可能性测度 …… 37
2.1.3　可能性与信息 …… 40
2.2　多元可能性分布 …… 41
2.2.1　n 元可能性分布的概念 …… 41
2.2.2　边缘可能性分布 …… 44
2.2.3　条件可能性分布 …… 48
第三章　测分语义学的基本特性…… 52
3.1　语言实体的意义 …… 52
3.2　测分语义试验的基本特性 …… 54
3.2.1　数据库 …… 54
3.2.2　一些基本的试验 …… 58
3.2.3　基数试验 …… 62

3.3 测分语义试验与意义表达 …………………… 66
第四章 模糊意义表达语言 …………………… 74
4.1 PRUF 翻译的基本概念 …………………… 74
4.1.1 模糊命题、模糊疑问句和模糊描述符 ……… 74
4.1.2 可能性赋值方程 …………………… 77
4.1.3 PRUF 的定义和表达式 …………………… 81
4.1.4 数据库、意义与信息 …………………… 83
4.1.5 语义等价与语义后承 …………………… 88
4.2 翻译规则 …………………… 90
4.2.1 几种标准的分布函数 …………………… 92
4.2.2 I 型翻译规则 …………………… 93
4.2.3 II 型翻译规则 …………………… 96
4.2.4 III 型翻译规则 …………………… 101
4.2.5 命题的修饰规则 …………………… 104
4.2.6 一致性、兼容性和真 …………………… 107
4.2.7 IV 型翻译规则 …………………… 111
第五章 翻译实例 …………………… 117
5.1 典型的翻译实例及其测分语义试验过程 …… 118
5.2 翻译实例续编 …………………… 130
第六章 近似推理 …………………… 141
6.1 语言变量和语言真变量 …………………… 141
6.1.1 语言变量的概念 …………………… 141
6.1.2 语言真值的计算 …………………… 144
6.2 近似推理 …………………… 149
6.2.1 推论规则 …………………… 149
6.2.2 推论规则的应用 …………………… 155
第七章 近似推理的计算机实现 …………………… 163
7.1 计算机近似推理语言 …………………… 163
7.2 解释型与编译/解释型求解方法 …………………… 167

7.2.1 解释型求解方法 …………………………………… 167
7.2.2 编译/解释型求解方法 ……………………………… 171
7.2.3 包含虚拟关系的解释型举例 ………………………… 173
7.2.4 编译/解释型举例 …………………………………… 176
7.3 指针及其他命令 …………………………………… 178
7.4 实例 ……………………………………………………… 183
附录 隶属函数的求取与自然语义量化的经验研究 …… 188
A.1 模糊统计试验与多级估量法 ……………………… 188
A.2 可能性分布函数的经验研究 ……………………… 191
A.3 随机集试验与区间估量法 ………………………… 193
A.4 语言值量化的经验研究 …………………………… 196
参考文献 ……………………………………………………… 201

第一章　集合与模糊集合

1.1　集　　合

一些事物的全体叫做一个集合．这些事物中每一个都称为这个集合的元素．集合有时简称为集．

如果某种事物只有一个，假定这个事物记作 a，那么称这种事物的全体是集合 $\{a\}$，a 是 $\{a\}$ 的唯一元素，只有一个元素的集合称为单元素集．

如果某种事物不存在，就称这种事物的全体是空集．规定任何空集都只是同一个集，记作 $\varnothing$，任何事物都不是空集的元素．

假定 a 是集合 A 的元素，记作

$$a \in A \text{或} A \ni a$$

“$\in$”读作“属于”，“$\ni$”读作“包含”．假定 a 不是 A 的元素，记作

$$a \notin A \text{或} A \not\ni a$$

“$\notin$”读作“不属于”，“$\not\ni$”读作“不包含”．

若有一些事物，全体写出来是 $a, b, c, \cdots$，那么，由定义，它们的全体是一个集，这个集可以记为 $\{a, b, c, \cdots\}$．这是用列举元素的办法来表示一个集合．

集合也可理解为“所有满足某条件的事物的全体”．若把“某个事物 x 满足某条件”这句话表示成一个逻辑公式 $p(x)$，那么按照这种理解，一个集合可以记作$\{x|p(x)\}$或$\{x:p(x)\}$．它所代表的意思是“所有使 $p(x)$ 成立的 x 的全体”．

这样，"小于 10 的自然数"是一个集合，按第一种写法写为

$$A=\{1,2,3,4,5,6,7,8,9\}$$

按第二种写法写为

$$A=\{x \mid x\in N \text{ 且 } x<10\}$$

式中N是自然数集。

一个集合又可以用其特征函数来表示。这个函数定义于论域U上，但只取 0,1 两值。因此，集合的特征函数是从论域U到两元素集合$\{0,1\}$上的一个映射，写为

$$c_A(u):U\to\{0,1\} \tag{1.1.1}$$

$c_A(u)=1$ 意味着 $u\in A$，$c_A(u)=0$ 意味着 $u\notin A$，不存在其他情况。

假定A和B都是集合，B的每个元素都是A的元素，那么称B为A的一个子集，记作$B\subset A$或 $A\supset B$，"$\subset$"读作"包含于"，而"$\supset$"读作"包含"。

集合的最基本运算是并(或称联)、交、补和笛卡儿积(或称直接积)。

属于集合A与属于集合B的元素的总体构成集合A和B的并集，记为

$$A\cup B$$

例如，若$A=\{1,3,6,9,11\}$，$B=\{2,3,5,6,7,8,11\}$，则

$$A\cup B=\{1,2,3,5,6,7,8,9,11\} \tag{1.1.2}$$

若A的特征函数为 $c_A(u)$，B的特征函数为 $c_B(u)$，那么，并集的特征函数为

$$c_{A\cup B}(u)=c_A(u)\vee c_B(u) \tag{1.1.3}$$

式中$\vee$是布尔代数中的"或"运算，也称布尔和。$\vee$有时可写为 Max，因为实际上这个运算是在诸被运算量中取其大者。

同时属于A和B两个集合的元素的总体构成集合A和B的交集，记为

$$A \cap B$$

例如，若 $A = \{1, 3, 6, 9, 11\}$，$B = \{2, 3, 5, 6, 7, 8, 9, 11\}$，则

$$A \cap B = \{3, 6, 11\}$$

交集的特征函数为

$$c_{A \cap B}(u) = c_A(u) \wedge c_B(u) \tag{1.1.4}$$

式中$\wedge$是布尔代数中的“与”运算，也称布尔积，$\wedge$有时可写为Min，因为它在诸被运算量中取其小者。

论域中不属于集合A的元素的总体构成集合A的补集，记为

$$\neg A \text{ 或 } A'$$

例如，设论域为

$$U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

集合 $A = \{1, 3, 6, 9, 11\}$，则论域U中A的补集为

$$A' = \{2, 4, 5, 7, 8, 10, 12\}$$

补集的特征函数为

$$c_{A'}(u) = 1 - c_A(u) \tag{1.1.5}$$

这个运算和我们熟知的布尔代数中的一元运算“非”一致。

集合的并、交、和补运算可从图 1.1 直观理解。

设A,B是两个集合。$A \times B$为A与B的笛卡儿积集，其定义为

$$A \times B = \{(a, b) | a \in A, b \in B\} \tag{1.1.6}$$

例如，设 $A = \{1, 2, 3\}$，$B = \{4, 5\}$，则

$$A \times B = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$$

我们看到$A \times B$的元素是有序数对。

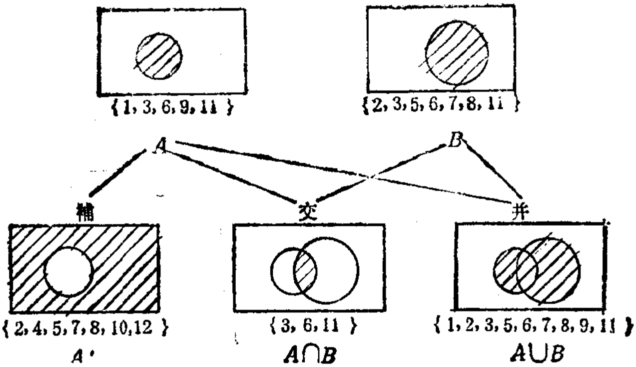

图 1.1 普通集合的基本运算

在并和交运算中，集合 A 和 B 都是同一论域的集合．在笛卡儿积运算中，情况不同，在这里 A，B 分别属于论域 U_1 和 U_2．积集的特征函数可写为

$$c_{A\times B}(u_1, u_2) = c_A(u_1) \wedge c_B(u_2) \tag{1.1.7}$$

积集的特征函数在这里是一个二元函数．

当然，并、交、笛卡儿积都很容易扩展到多于两个集合的情况．

普通集合还有许多其他的基本性质，在以后用到时再加介绍．

1.2 模 糊 集 合

1.2.1 模糊集合的概念及术语

上面 1.1 节的叙述表明，普通集合是一种边界明确的集合，这是由于我们只用属于和不属于两种状态对论域中的元素进行分类．普通集合的特征函数只有 0 和 1 两个值，也说明一个元素与一个集合之间只有完全隶属或完全不隶属两种关系，不存在中间状态．

但是，实际世界中存在许多边界不明确的分类。例如，“远大于1的实数”就是论域——实数轴——上的一个没有明确边界的分类。对于这样一个分类，可以找到许多与其隶属关系不明确的元素。例如，我们不能肯定5这个数是不是远大于1。显然，说5这个数对于“远大于1的实数”这个分类的隶属程度是0.2，比起肯定地说5属于或不属于这个分类要合理得多。这一修改意味着把普通集合特征函数的值域从$\{0, 1\}$扩展到区间$[0, 1]$之中并因此产生了模糊集合这个新概念。

严格地讲，论域U中的一个模糊集合A由一个隶属函数[1] $\mu_A(u): U \to [0, 1]$所表征；隶属函数把区间$[0,1]$中的一个数$\mu_A(u)$与U中每一个元素u对应起来，说明u对A的隶属程度。

两个模糊集合A，B的并集记为$A \cup B$。利用隶属函数，并运算的定义是

$$\mu_{A \cup B}(u) = \mathrm{Max}(\mu_A(u),\ \mu_B(u)) = \mu_A(u) \vee \mu_B(u) \tag{1.2.1}$$

式中Max和$\vee$都表示在被运算量中取大者。

两个模糊集合A，B的交集记为$A \cap B$。利用隶属函数，交运算的定义是

$$\mu_{A \cap B}(u) = \mathrm{Min}(\mu_A(u),\ \mu_B(u)) = \mu_A(u) \wedge \mu_B(u) \tag{1.2.2}$$

式中Min和$\wedge$都表示在被运算量中取小者。

模糊集合A的补记为$\neg A$或A'。利用隶属函数，补运算的定义是

---

1) 隶属函数（membership function）在《模糊集合、语言变量及模糊逻辑》（陈国权，科学出版社，1982年）一书中曾称为资格函数，因为当时尚未有统一译名。

$$\mu_{A'}(u) = 1 - \mu_A(u) \tag{1.2.3}$$

式中一代表普通的减法.

一个模糊集合是空的当且仅当其隶属函数在论域 U 上恒为 0，空模糊集合记为 $\varnothing$.

两个模糊集合相等，写为 $A = B$，当且仅当对 U 中所有元素 u 都有 $\mu_A(u) = \mu_B(u)$. 有时为了简化，用 $\mu_A = \mu_B$ 代替 $\mu_A(u) = \mu_B(u)$ 的写法.

模糊集合 A 包含于 B 中，或者说 A 是 B 的一个子集的充要条件是 $\mu_A \leqslant \mu_B$. 用符号表示为

$$A \subset B \Longleftrightarrow \mu_A \leqslant \mu_B \tag{1.2.4}$$

式中 $\Longleftrightarrow$ 表示“互为等效”，或“等价”.

模糊集合 A 的支集是 U 中 $\mu_A(u)$ 为正的点的集合. 模糊集合的支集是普通集合. 模糊集合 A 的高度是遍及 U 上 $\mu_A(u)$ 的最大值. 模糊集合 A 的过渡点是 U 上的一个点，这个点在 A 中的隶属程度是 0.5.

例 1.2.1 令论域 U 是区间 $[0, 100]$，U 的元素 u 代表人的年纪. 这时，老年人的概念可表达为 U 的一个模糊集合 A，其隶属函数可定义为

$$\mu_A(u) = \begin{cases} 0, & \text{对于 } 0 \leqslant u \leqslant 50 \\ \left[1 + \left(\dfrac{u - 50}{5}\right)^{-2}\right]^{-1}, & \text{对于 } 50 < u \leqslant 100 \end{cases} \tag{1.2.5}$$

在此情况下，A 的支集是区间 $(50, 100]$，A 的高度是 1，A 的过渡点是 55.

从这个例子可以看到，年老这个词的意义可由**年纪**的论域中的一个模糊集合来表示.

以后，当论域中的一个模糊集合 A 代表语言中一个词的意义时，为突出其特点，我们将用这个词本身代替 A 来标记这

个模糊集合。为区别起见，英文词用大写，中文词用楷体，这样，式(1.2.5)的隶属函数可写为 $\mu_{老}(u)$；这时“模糊集合 A”与“模糊集合*老*”两种说法被认为是一致的。这种标记方法同样用于模糊集合的图示上，例子可见图 1.2。

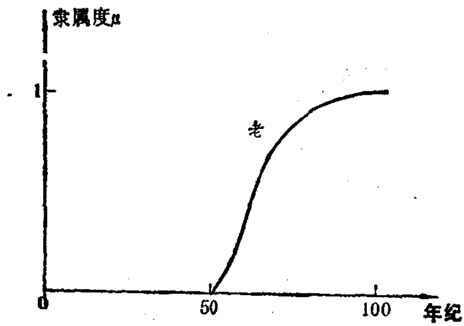

图 1.2 用文字标记模糊集合

为简化模糊集合的表达方法，将使用下面的概念。

一个非模糊(普通)有限集合

$$U=\{u_1, \cdots, u_n\} \tag{1.2.6}$$

将记为

$$U=u_1+u_2+\cdots+u_n \tag{1.2.7}$$

或

$$U=\sum_{i=1}^{n} u_i \tag{1.2.8}$$

这里，约定用+号代表并运算而不代表算术和。这样做实际上是把一个集合看成由这个集合的每一元素构成的单元素集的并集。

把式(1.2.7)推广到U的模糊子集 A，有

$$A=\mu_1 u_1+\cdots+\mu_n u_n \tag{1.2.9}$$

或

$$A = \sum_{i=1}^{n} \mu_i u_i \tag{1.2.10}$$

式中 μ_i, $i=1,\cdots,n$, 是 u_i 在 A 中的隶属程度。 当 u_i 是数字时,字串 $\mu_i u_i$ 中 μ_i 和 u_i 可能辨别不清。 在这种情况下,我们在中间加上一个分离符来消除含混,这时式 (1.2.9) 可写为

$$A = \mu_1/u_1 + \cdots + \mu_n/u_n \tag{1.2.11}$$

而式 (1.2.10) 则变为

$$A = \sum_{i=1}^{n} \mu_i/u_i \tag{1.2.12}$$

例 1.2.2 令 $U=\{a, b, c, d\}$ 或等价地写为

$$U = a + b + c + d \tag{1.2.13}$$

在这种情况下,U 的某一模糊集合 A 可以不含混地写为

$$A = 0.3\,a + b + 0.9\,c + 0.5\,d \tag{1.2.14}$$

另一种情况,若

$$U = 1 + 2 + \cdots + 100 \tag{1.2.15}$$

则 U 的某一模糊集合 A 写为

$$A = 0.3/25 + 0.9/30 \tag{1.2.16}$$

以避免含混。

和式 (1.2.8) 一样, 式 (1.2.10) 是模糊集合的一种表达法。在这种表达法中, 我们把一个模糊集合看成由它的元素构成的单元素集 $\{\mu_i u_i\}$ 或 $\{\mu_i/u_i\}$ 的并集。 从并的定义出发,若 A 的表达式中出现 $u_i = u_j$, 则有

$$\mu_i u_i + \mu_j u_i = (\mu_i \vee \mu_j) u_i$$

例如

$$A = 0.3\,a + 0.8\,a + 0.5\,b$$

可以写成为

$$A = (0.3 \vee 0.8)\,a + 0.5\,b$$