

# 模糊集合的语言值近似

Approximating Linguistic Truth of Fussy Sets

91-93

欧灵

0159

(西南师范大学计算机科学系 重庆 400715)

**摘要** 本文在模糊集合的基础上,讨论了[0,1]区间上离散的模糊集间的相似问题,利用“重心”概念,可根据模糊集的语言近似特征,理解一类外延不明确的概念。

**关键词** 模糊集,语言值近似,重心 隶属函数

在模糊集合论基础上建立的可能性理论,特别适宜描述内涵不确定的概念<sup>[1]</sup>。在 Zadeh 引进了以 [0,1] 区间上的模糊子集代替其它模糊化的点值之后,用语言形式表示模糊集合便成了人们关心的问题,然而,当对模糊语言进行演算后,如何将得到的模糊子集翻译成语言元素的形式,即语言值近似还没有一种可供选择的“最好”近似概念,也没有计算“好的”近似概念的一般性技术<sup>[2]</sup>。Zadeh 在模糊演绎推理中利用求“模糊关系赋值方程”组的解的形式,刻画语言值近似,为度量近似程度,本文引入了模糊集合的“重心”概念,利用模糊集合的语言值特征,表达语言值近似关系。本文首先介绍模糊集合论中的模糊集运算,接着在实例基础上介绍似然推理,并根据似然推理的结果及模糊集表示的一般形式阐述了用“重心”概念来解决模糊集合语言值近似的问题。

## 1 模糊集合

Zadeh 在康托集合论的基础上进行推广,以处理没有清晰内涵和外延的集合,其推广主要体现在:

- ① 将值域由 {0,1} 变成 [0,1]。
- ② 引入隶属度  $\mu_A(x)$  来替代特征函数。
- ③ 模糊集合  $\bar{A}$  对应于论域 U 到 [0,1] 上的一个映射。

推广之后的模糊集合,  $\mu_{\bar{A}}(x)$  只取 0 和 1 两个值时,即论域 U 上的元素 x 属于  $\bar{A}$  ( $\mu_{\bar{A}}(x)=1$ ) 或不属于  $\bar{A}$  ( $\mu_{\bar{A}}(x)=0$ ), 则  $\bar{A}$  就成为一普通子集。一般情况下,  $0 < \mu_{\bar{A}}(x) < 1$ , 即可属于  $\bar{A}$  也可不属于  $\bar{A}$ , 或表示在一定程度上属于  $\bar{A}$ 。

在模糊集合上可以定义很多运算<sup>[1]</sup>, 下面列出常用到的包含、交、并、补运算, 引用 Zadeh 的定义, 设  $\bar{A}, \bar{B}$  是 U 上的模糊子集。

- (1) 包含关系, 记为“ $\supseteq$ ”
- $\bar{A}$  包含  $\bar{B}$  或  $\bar{B}$  被  $\bar{A}$  包含, 记为  $\bar{A} \supseteq \bar{B}$  或  $\bar{B} \subseteq \bar{A}$

$$\bar{A} \subseteq \bar{B} =_{def} (\forall x \in U, \mu_{\bar{A}}(x) \leq \mu_{\bar{B}}(x))$$

(2) 相等关系, 记为“=”

$$\bar{A} = \bar{B} =_{def} \bar{A} \supseteq \bar{B} \text{ 且 } \bar{B} \supseteq \bar{A}$$

(3) 交集记为“ $\cap$ ”

定义  $\bar{A}$  和  $\bar{B}$  的交集 ( $\bar{A} \cap \bar{B}$ ) 的隶属函数

$$\mu_{\bar{A} \cap \bar{B}}(x) = \min(\mu_{\bar{A}}(x), \mu_{\bar{B}}(x))$$

(4) 并集, 记为“ $\cup$ ”

定义  $\bar{A}$  和  $\bar{B}$  的并集 ( $\bar{A} \cup \bar{B}$ ) 的隶属函数

$$\mu_{\bar{A} \cup \bar{B}}(x) = \max(\mu_{\bar{A}}(x), \mu_{\bar{B}}(x))$$

(5) 补集, 记为“ $\bar{\bar{A}}$ ”

$$\forall x \in U, \mu_{\bar{A}}(x) = 1 - \mu_{\bar{\bar{A}}}(x) \Leftrightarrow \bar{\bar{A}} = \bar{A}, \text{ 称 } \bar{A} \text{ 和 } \bar{\bar{A}} \text{ 互补。}$$

下面根据需要定义  $\lambda$  水平截集。

(6)  $\lambda$  水平截集

设  $\bar{A}$  为模糊集, 定义分明基  $\# \bar{A}$  为:

$$\# \bar{A} =_{def} \{x \mid \exists a, (x, a) \in \bar{A}\}$$

则模糊集  $\bar{A}$  的  $\lambda$  水平截集  $\bar{A}_\lambda$  定义为:

$$\bar{A}_\lambda =_{def} \{(x, \mu_{\bar{A}}(x)) \mid x \in \# \bar{A}, \mu_{\bar{A}}(x) \geq \lambda\},$$
$$0 < \lambda \leq 1$$

## 2 似然推理

在数理逻辑中, 肯定前提, 则有:

$$\frac{P, P \rightarrow Q}{Q}$$

在模糊前提下:

$$\frac{\bar{P}_1, \bar{P}_1 \rightarrow \bar{Q}_2}{\bar{Q}_2}$$

$$\bar{Q}_2 = \bar{P}_1 \cdot (\bar{P}_1 \rightarrow \bar{Q}_2) = \bar{P}_1 \cdot ((\bar{P}_1 \wedge \bar{Q}_2) \vee \bar{P}_2) \quad (1)$$

$\bar{P}_1, \bar{P}_2, \bar{Q}_2, \bar{Q}_1$  表示模糊子集,  $\bar{P}_2$  表示  $\bar{P}_1$  的补集, 运算“ $\cdot$ ”表示模糊关系的复合。通过转换, 可化为矩阵运算, 现举例说明。

设  $U = \{5, 10, 15, 20, 25, 30\}$ ,  $\bar{P}_1: x$  是小的, 表示为  $1/5 + 0.6/10 + 0.2/15 + 0.1/20 + 0/25 + 0/30$

$\bar{P}_2 \rightarrow \bar{Q}_2: x$  与  $y$  近似相等,  $x$  与  $y$  之间的二元关系可用矩阵 M 表示:

$$M = \begin{bmatrix} 1 & 0.5 & 0.2 & 0.1 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0.2 & 0.1 & 0 \\ 0.2 & 0.5 & 1 & 0.5 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.5 & 1 & 0.5 & 0.2 \\ 0 & 0.1 & 0.2 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0.1 & 0.2 & 0.5 & 1 \end{bmatrix}$$

矩阵为对称矩阵,即  $x$  与  $y$  近似相等也有  $y$  与  $x$  近似相等。令

$N = (1 \ 0.6 \ 0.2 \ 0.1 \ 0 \ 0)$ , 作运算  $N \cdot M = (1 \ 0.6 \ 0.5 \ 0.2 \ 0.2 \ 0.1)$ , 比较  $\tilde{P}_1(1 \ 0.6 \ 0.2 \ 0.1 \ 0 \ 0)$  与  $\tilde{Q}_3(1 \ 0.6 \ 0.5 \ 0.2 \ 0.2 \ 0.1)$ , 可以认为  $\tilde{P}_1$  与  $\tilde{Q}_3$  所表示的模糊命题大致相当<sup>[5]</sup>。用语言描述可理解为:“ $x$  是小的,  $x$  与  $y$  近似相等, 则有  $y$  大致也是小的”。对于这类模糊语言演算所得的模糊集, 因为它们的论域一致, 而且论域上的覆盖也一致, 只需比较各自的隶属度。然而, 也存在这样的问题, 虽由同一基元集扩展生成, 但可以有不同的覆盖, 如图1所示。它们均以年龄为基元, 但

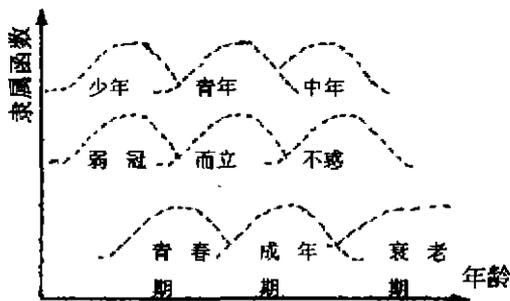


图1 同一基元集上的不同覆盖

描述的是不同的含义或者语言值。考察人类自然语言, 对于外延不确定的概念有很多, 比如“年轻”、“年老”、“高”、“矮”等。对这些概念还可加修饰词, 比如语气修饰词“非常”、“相当”, 组成短语“非常年轻”、“相当矮”, 也可用模糊数量词“很多”、“多数”、“少数”, 组成短语“多数年轻人”、“几个高”个子。就人类语言的不确定性而言, 可以用数量、程度、概率和频率等几种方式来描述<sup>[6]</sup>。当模糊集做某种运算后, 比如精确逻辑中的“与”、“或”、“非”、“异或”运算, 对结果的语义解释显得尤其重要。定义相应的运算, 然后加以证明, 这种形式化的方法当然可以。考察文[7]给出的三种模型: 相容度模型、包含度模型和贴近度模型, 它们都采用了比较两模糊集  $\tilde{A}, \tilde{B}$  的方法, 计算  $\tilde{A}$  相对于  $\tilde{B}$  的真值  $\lambda \in [0, 1], \lambda = f(\tilde{A}, \tilde{B})$ , 并且就认为  $\tilde{A}$  在推理过程中化作了真值  $\lambda$ , 而将真值  $\lambda$  沿着推理链传递。上述三种模型就是根据  $f$  的不同选择而得。函数  $f$  的确是文[2]中近似类比推理的关键之一。综合上述, 本文利用函数逼近, 引入“重心”概念来描述模糊集之间的近似程度。

### 3 模糊集合相似

为了比较模糊集合的相似度, 考虑论域  $U$  上的两模糊集  $\tilde{A}, \tilde{B}$ 。

定义1 假设  $\tilde{A}$  表示为  $\sum_{x=1}^n \mu_{\tilde{A}}(x_i)/x_i$ , 若

$\int_U \mu_{\tilde{A}}(x) dx$  收敛, 则  $\tilde{A}$  的重心  $g(\tilde{A})$  定义为:

$$g(\tilde{A}) =_{\text{def}} \frac{\int_U \mu_{\tilde{A}}(x) dx}{\sum_{x=1}^n \mu_{\tilde{A}}(x_i) \cdot x_i} \quad (2)$$

其中,  $\tilde{A}_\lambda$  表示  $\tilde{A}$  的  $\lambda (0 < \lambda \leq 1)$  水平截集。

定义2 若  $g(\tilde{A}) > g(\tilde{B})$  称模糊集  $\tilde{A}$  大于  $\tilde{B}$ , 反之, 称  $\tilde{A}$  小于  $\tilde{B}$ 。

定义3 若模糊集  $\tilde{A}$  和  $\tilde{B}$  的中心分别为  $U_{a_1}$  和  $U_{a_2}$ , 称  $D_{AB} = |U_{a_1} - U_{a_2}|$  为两模糊集的重心之差。

定义4 模糊集  $\tilde{A}, \tilde{B}, \# \tilde{A} \cap \# \tilde{B} \neq \emptyset$ , 若满足  $D_{AB} \leq \delta, \delta$  为一任意小的正数, 则称  $\tilde{A}$  相似  $\tilde{B}$ , 记为  $\tilde{A} \approx \tilde{B}$ 。当  $D_{AB} = 0$  时称  $\tilde{A}$  极相似  $\tilde{B}$ , 记为  $\tilde{A} \simeq \tilde{B}$ 。

定义5 称  $\# \tilde{A}$  与  $\# \tilde{B}$  有最大公共元个数, 若对于  $u_{\tilde{A}}(x) \neq 0$ , 不存在  $\tilde{B}_i (\tilde{B}_i \in U), \# \tilde{A} \cap \# \tilde{B}_i$  的公共元个数大于  $\# \tilde{A} \cap \# \tilde{B}$  的公共元个数, 记为  $\max | \# \tilde{A} \cap \# \tilde{B} |$ 。

定义6 称  $\tilde{A}$  与  $\tilde{B}$  语言值近似, 若  $\tilde{A} \simeq \tilde{B}$ , 或  $\tilde{A} \approx \tilde{B}$  且  $\max | \# \tilde{A} \cap \# \tilde{B} |$ , 并且不存在其他的  $\tilde{B}_i \in U$ , 满足  $D_{AB} > D_{AB_i}$ 。

求模糊集  $\tilde{A}$  的语言值近似算法描述如下:

(1) 令模糊集  $P_A$  和  $P_U$  为空。

(2) 在论域  $U$  上求出  $\# \tilde{A} \cap \# \tilde{B}_i \neq \emptyset$  的子集  $\tilde{B}_i, \tilde{B}_i \in U, \tilde{B}_i$  应该是语言值明晰的, 或在第(5)步之后是明晰的。将这样的  $\tilde{B}_i$  置入  $P_A$  中,  $P_A$  代表已解释的集合。若所有的  $\# \tilde{A} \cap \# \tilde{B}_i = \emptyset$ , 则在  $U$  上不存在  $\tilde{A}$  的语言值近似翻译, 转(4)。

(3) 设  $\# \tilde{A} \cap \# \tilde{B}_i = \emptyset$  的  $\tilde{B}_i$  共有  $K$  个, 求出  $\max | \# \tilde{A} \cap \# \tilde{B}_i |$ , 假设  $\max | \# \tilde{A} \cap \# \tilde{B}_i | = | \# \tilde{A} \cap \# \tilde{B}_j |$ , 然后求出  $D_{AB_j}$  和  $D_{AB_i} (j \neq i)$ , 将  $K$  个  $D$  值排序, 若  $D_{AB_j} < D_{AB_i} (j \neq i)$ , 则根据定义6, 有  $\tilde{A}$  与  $\tilde{B}_j$  两模糊集的语言值近似。将  $\tilde{A}$  加入集合  $P_A$  中, 否则, 可因两模糊集的隶属函数不同而导致不存在  $\tilde{A}$  的语言值近似翻译。

(4) 将模糊集  $\tilde{A}$  及其上的隶属度作为  $[0, 1]$  上的一个未解释子集, 加入  $P_U$  集合中。

(5) 选择适当的时候, 对  $P_A$  和  $P_U$  集合中的子集重复(1)、(2)两步, 以逐步增加  $P_A$  中的元素, 减少  $P_U$  中的元素。

结束语 定义“重心”概念, 使得在  $[0, 1]$  上离散的模糊集可依此度量, 并进而理解模糊集之间的

93-94

遗传程序设计

赋值函数

节点分类 (25)

计算机科学 1998 Vol. 25 No. 3

## 遗传程序设计中的赋值函数和节点分类

The Assignment Function and Node Classification in Genetic Programming

齐越胜 王保中 康立山

(武汉大学软件工程国家重点实验室 武汉 430072)

TP311 0242.23

**摘要** To make sure that the crossover operator produces meaningful offsprings, it is always a must that every function node of the tree has the property of closeness. This paper points out the problems of this closeness requirement, especially when assignment function is needed in the function set. In this paper, we propose a node classification method based on the return data type of the node so as to cancel this closeness requirement. Using this method we evolved a program to model a multioutput problem. The computational result is quite satisfactory.

**关键词** Genetic programming, Assignment function, Node classification

## 一、遗传程序设计

遗传程序设计(GP)是用演化方法让计算机自己发现解决问题的算法,生成可以解决实际问题的程序。一个程序可以用树结构很好地表示出来,因此,演化一个程序也就是演化一个合适的树的过程,通常,树的节点分为两类:函数节点和终端节点<sup>[1]</sup>,函数节点取自函数集(根据具体问题进行定义),用向下的分支表示其参数,终端节点取自终端集(不同的问题也有不同的终端集),不再有分支。比如,在 $+(a, b)$ (即 $a+b$ )这个表达式中(也是一个简单的程序), $+$ 是函数节点,有两个参数,而 $a, b$ 是终端节点。遗传程序设计的演化过程是这样的<sup>[2]</sup>(同遗传算

法):

1. 设定最大代数 MaxGeneration, 随机生成一群个体, 记当前代 Generation 为 0;
  2. 判定当前代的每一个体的性能(计算适应值);
  3. 若已经得到一合符要求的程序, 则此次演化成功, 结束运行, 否则做 4;
  4. 若当前代已经达到最大代数, 则此次演化失败, 结束运行, 否则做 5;
  5. 通过杂交, 变异和复制, 产生新一代个体, 当前代 Generation 增加 1, 返回 2。
- 杂交是两个树通过交换它们的子树而得到两个

相似和语言近似。在这里隶属函数显得很重要。Zadeh 在定义“年龄”这一概念时, 曾简单地给出了几个隶属函数。根据人类的思维习惯, 在考虑隶属函数时, 一般倾向于逼近或拟合二次函数, 如果大部分隶属函数能基于此考虑的话, 对模糊的演算将变得容易些。然而对于人类来说, 一个概念不仅与其自身的含义有关, 也与对象所处的环境、人们的心理等相关因素有联系。因而, 与其说概念本身模糊还不如说人们用以建立该概念的标准是模糊的<sup>[3]</sup>。如上所述, 要解决带有模糊的一系列问题, 比如模糊知识表示、模糊推理、模糊集合的语言值等问题, 由于模糊逻辑的理论基础还不完善, 也还只处于探索应用阶段, 其根本的解决办法只能从揭示人类思维的奥秘, 了解其本质入手, 才能解决机器理解概念、命题以及进行演绎、归纳推理的非清晰性和不明确性。

## 参考文献

- [1] L. N. Kanal, and J. F. Lemmer, Uncertainty in Artificial Intelligence, North-Holland, 1986
- [2] P. Z. Wang and H. M. Zhang, Truth-Valued Flow Inference and its Dynamic Analysis, Jour. of Beijing Normal Univ., No. 1, 1989
- [3] 赵沁平译, 李未等校, 人工智能中的逻辑, 北京大学出版社, 1990年
- [4] 陆汝钫编著, 人工智能, 上、下册, 科学出版社, 1996年
- [5] 王雨田, 现代逻辑科学导引, 下册, 中国人民大学出版社, 1987年
- [6] 魏长华, 人类自然语言不确定性及粗集理论的表示, 计算机科学, 24(5), 1997
- [7] 吴望名, 模糊推理的原理和方法, 贵州科技出版社, 1994年
- [8] 冯嘉礼著, 思维与智能科学中的性质论方法, 原子能出版社, 1991年