计算机科学2003Vol. 30№. 10

输出为可能性分布的模糊决策树

袁修久[1,2] 张文修[1]

(西安交通大学理学院信息与系统科学研究所 西安710049)[1] (空军工程大学文理学院 西安710051)[2]

Fuzzy Decision Trees with Possibility Distributions as Output

YUAN Xiu-Jiu[1,2] ZHANG Wen-Xiu[1]

(Institute of Information and System Science, Science College of Xi'an Jiaotong University, Xi'an 710049)[1]

(College of Arts and Science, Air Force Engineering University, Xi'an 710051)[2]

Abstract More than one possible classifications for a given instance is supposed. A possibility distribution is assigned at a terminal node of a fuzzy decision tree. The possibility distribution of given instance with known value of attributes is determined by using simple fuzzy reasoning. The inconsistency in determining a single class for a given instance diminishes here.

Keywords Fuzzy decision trees, Possibility distribution, Uncertainty reasoning, Similarity degree

1 引言

建立决策树的方法已有多种。从决策树可以提取知识，提取的知识用 IF-THEN 规则表示。决策树的一个优点是提取的知识易于理解和解释。由于噪声、测量误差使得决策树处理连续属性时出现了困难和在识别判断的过程中客观地存在着模糊性，很多文献已将决策树的方法推广成了模糊决策树，如文[1,2,5,6]等。已有的模糊决策树的文献都是假设一个对象是属于一个类的，是有监督的学习。但是模糊决策树涉及的属性变量和决策变量中有模糊变量，其取值是模糊集，从而使得一个对象可以属于这个类也可以属于另一个类，只不过属于不同类的程度可能不一样，因此假设一个对象对应于一个可能性分布(即一个对象可以属于多个类，属于不同类的程度可以不一样)则更合理，解释起来更加自然。另外由于假设一个对象属于一个类，在模糊决策树的终端结点只赋给一个类，在利用从模糊决策树提取的规则确定属性值已知的对象对应的类时，常会出现一个对象对应多个类的不一致的情况。如果假设一个对象对应于一个可能性分布，则可以避免这个问题。如果假设一个对象对应于一个可能性分布，归纳学习由有监督学习变成了无监督学习。

设论域 U 由 n 个对象 $e_i, i=1,\cdots,n$ 组成，e_i 可以用变量 $(X_1, X_2, \cdots, X_p, Y)$ 的取值描述，其中 $X_1, X_2, \cdots, X_p$ 是属性变量，Y 是决策变量。它们中至少有一个是模糊变量，其余的为符号变量。设 X_k 取 L_k 个值 $\tilde{X}_{kj}, j=1,\cdots,L_k$，它们都是模糊集(取符号值可以看成模糊集的特殊情况)，第 i 个对象 e_i 属于 $\tilde{X}_{kj}$ 的隶属度为 x_{ikj}。Y 取 m 个值 $\tilde{Y}_1, \tilde{Y}_2, \cdots, \tilde{Y}_m$(即 m 个类)，且第 i 个对象属于第 j 个类 $\tilde{Y}_j$ 的隶属度为 $\tilde{y}_{ij}$，即每个 e_i 对应于一个可能性分布 $(y_{i1}, y_{i2}, \cdots, y_{im})$，称为该对象的可能性分布。本文利用这些数据建立一个模糊决策树，从模糊决策树提取规则，利用简单住是方法对每个属性已知的对象确定其可能性分布。

2 模糊决策树的生成

模糊决策树的扩展是在一定的准则下，选择属性变量不断地划分当前树的终端结点完成的。开始将所有的数据全部放在根结点 t_1，每个对象属于根结点的隶属度为1，根结点的可能性分布就是落入根结点的所有对象的可能性分布的平均，即 $\bar{W}(t_1)=(\bar{W}_1(t_1), \bar{W}_2(t_1), \cdots, \bar{W}_m(t_1))$，其中 $\bar{W}_j(t_1)=\frac{1}{n}\sum_{i=1}^{n} y_{ij}, j=1,\cdots,m$。判断 $\bar{W}(t_1)$ 做为属于 t_1 的所有对象的可能性分布的估计的好坏的一个标准，是看模糊离差平方和 $Q(t_1)=\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{W}_j(t_1))^2$ 是否较小。

设当前的树为 $T, t\in\tilde{T}$($\tilde{T}$ 为 T 的终端结点的集合)。假若用属性变量 X_v 划分结点 t，此时可以产生 L_v 个子结点，由于 X_v 是模糊变量，每个对象可以属于这个子结点也可以属于另一个子结点。第 i 个对象属于 t 的第 k 个子结点 $t(\tilde{X}_{vk})$ 的隶属度这样计算：设 $x^i(t)$ 是第 i 个对象 e_i 属于结点 t 的隶属度，则 e_i 属于子结点 $t(\tilde{X}_{vk})$ 的隶属度为 $T(x^i(t), x_{ivk})$，T 为三角模，常用的三角模是"取小"和"乘积"。落入子结点 $t(\tilde{X}_{vk})$ 的对象 e_i 属于第 j 个类的隶属度为 $T(T(x^i(t), x_{ivk}), y_{ij})$。$\bar{W}(t)=(\bar{W}_1(t), \bar{W}_2(t), \cdots, \bar{W}_m(t))$，称为结点 t 的可能性分布，其中 $\bar{W}_j(t)=\sum_{i=1}^{n}T(x^i(t), y_{ij})/\sum_{i=1}^{n}x^i(t), j=1,\cdots,m$。用 $\bar{W}(t)$ 估计落入结点 t 的对象的可能性分布的模糊离差平方和为 $Q(t)=\sum_{i=1}^{n}\sum_{j=1}^{m}x^i(t)(y_{ij}-\bar{W}_j(t))^2$。用属性变量 X_v 划分结点 t 得到 L_v 个子结点，第 k 个子结点 $t(\tilde{X}_{vk})$ 的可能性分布为 $\bar{W}(t(\tilde{X}_{vk}))=(\bar{W}_1(t(\tilde{X}_{vk})), \bar{W}_2(t(\tilde{X}_{vk})), \cdots, \bar{W}_m(t(\tilde{X}_{vk})))$，其中 $\bar{W}_j(t(\tilde{X}_{vk}))=\sum_{i=1}^{n}T(T(x^i(t), x_{ivk}), y_{ij})/\sum_{i=1}^{n}T(x^i(t), x_{ivk}), j=1,\cdots,m$。用属性变量 X_v 划分结点 t 后总的模糊离差平方和为

$$Q(t, X_v)=\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{L_v}T(x^i(t), x_{ivk})(y_{ij}-\bar{W}_j(t(\tilde{X}_{vk})))^2$$

$Q(t, X_v)$ 从总体上反映了结点 t 的各子结点的对象的可能性分布同各子结点的可能性分布的差别。若要用 X_v 划分结点 t，只有 $Q(t, X_v)\leqslant Q(t)$ 才有必要，并且应选择达到结点 t 的路径上没有出现过的，且使得 $Q(t, X_v)$ 达到最小的属性变量 X_v 划分结点 t。

在得到决策树后，从根结点到终端结点的每一条路径提取一条模糊规则。这条模糊规则的可信度是我们要关心的。由

袁修久 博士研究生，主要研究方向为知识发现与数据挖掘。张文修 教授，博士生导师，主要研究方向为人工智能、模糊集、随机集等。

于落入一个终端结点的对象的可能性分布可能是不完全相同的，用该终端结点的可能性分布估计这些分布，终端结点的可能性分布同落入它的对象的可能性分布一般是不一样的。如果这些可能性分布与终端结点的可能性分布的相似度越大，这个估计就越好，反之则估计就比较差。估计的好坏同对象属于终端结点的隶属度的大小也是有关系的，如果一个对象其可能性分布同该终端结点的可能性分布的相似度较小，但是该对象隶属于该终端结点的隶属度也小，则认为这个估计仍然是一个好的估计。用终端结点 t 的加权相似度

$$r=\frac{\sum_{i=1}^{n}x^i(t)l(y_i,\overline{W}(t))}{\sum_{i=1}^{n}x^i(t)},t\in\overline{T}$$

表示模糊规则的可信度，也称该终端结点的可信度。其中 $l(y_i,\overline{W}(t))$ 表示 y_i 同 $\overline{W}(t)$ 的相似度，$y_i=(y_{i1},y_{i2},\cdots,y_{im})$。相似度可取下列的形式，

$$l(y_i,\overline{W}(t))=\sqrt{1-\frac{1}{m}\sum_{j=1}^{m}(y_{ij}-\overline{W}_j(t))^2},$$

$$l(y_i,\overline{W}(t))=\frac{\sum_{j=1}^{m}(y_{ij}\wedge\overline{W}_j(t))}{\sum_{j=1}^{m}(y_{ij}\vee\overline{W}_j(t))}$$

等。

生成模糊决策树的过程可以归纳成以下几步。

1）计算根结点 t_1 的可能性分布、可信度和 $Q(t_1)$。如果可信度大于 α（α 是提前给的阈值），根结点就是所要的树。

2）如果根结点的可信度小于 α，对每个属性变量 X_v 计算 $Q(t_1,X_v)$，选择使得 $Q(t_1,X_v)$ 达到最小且 $Q(t_1,X_v)\leqslant Q(t)$ 的 X_v 划分结点 t_1。

3）删除当前树的空的终端结点，对每个非空的终端结点 t 计算其可能性分布、可信度。如果可信度超过 α，则停止划分该终端结点；如果可信度小于 α，则计算 $Q(t)$ 和达到该终端结点的路径没有出现过的属性变量 X_v 的模糊离差平方和 $Q(t_1,X_v)$。选择使 $Q(t_1,X_v)$ 达到最小且 $Q(t)\geqslant Q(t,X_v)$ 的 X_v 划分结点 t。

4）重复3）直到当前树的所有终端结点的可信度超过 α，或已没有属性变量划分终端结点为止。在每个终端结点赋给它自己的可能性分布就得到所要的模糊决策树。

利用事先给定的可信度的阈值 α 可以控制树的大小。如果 α 小，则树就比较小，如果 α 大，则树就比较大。但是当 α 超过一定的值，树的大小就不再变化。

从根结点到每个终端结点的一条路径都可以表示成一个模糊规则，模糊规则的条件是路径上所有属性值的合取，结论是该终端结点的可能性分布。该终端结点的可信度就是该条模糊规则的可信度。

3 应用提取的模糊规则确定属性值已知的对象的可能性分布

在提取了模糊规则后，可以利用简单模糊推理的方法确定属性值已知的对象的可能性分布，即通过已知的属性值计算出该对象对每条模糊规则的条件的满足程度，然后把它作为该模糊规则后件对应的可能性分布的权重，把权重归一化后，对模糊规则结论的可能性分布加权平均得到该对象的可能性分布。

如果要求输出是一确定值，可以利用非模糊化的方法，如最大隶属函数法，重力重心法得到一个确定的输出值。

例 假设有4个属性变量 X_1＝天气，X_2＝温度，X_3＝湿度，X_4＝是否有风。X_1 取三个值：晴天、阴天、下雨；X_2 取三个值：热、暖和、凉；X_3 取两个值：湿、正常；X_4 取两个值：有风、无风。决策变量 Y 取三个值：游泳、排球、举重，依次被解释为“适宜于游泳”、“适宜于打排球”、“适宜于举重”。某个星期六的天气适宜于三种运动中的哪一种不是绝对的。如果天热适宜于游泳的程度大一些，而另两种运动也不是绝对不适宜的，只不过是适宜的程度小一些，因此在某种天气状况下给出一个适宜于三种运动的一个可能性分布则解释更加自然，表1是把16个星期天作为对象的一个训练集（见文[2]）。

表1

对象	X_1			X_2			X_3		X_4		Y		
	晴天	阴天	下雨	热	暖和	凉	湿	正常	有风	无风	游泳	排球	举重
1	0.9	0.1	0.0	1.0	0.0	0.0	0.8	0.2	0.4	0.6	0.0	0.8	0.2
2	0.8	0.2	0.0	0.6	0.4	0.0	0.0	1.0	0.0	1.0	1.0	0.7	0.0
3	0.0	0.7	0.3	0.8	0.2	0.0	0.1	0.9	0.2	0.8	0.3	0.6	0.1
4	0.2	0.7	0.1	0.3	0.7	0.0	0.2	0.8	0.3	0.7	0.9	0.1	0.0
5	0.0	0.1	0.9	0.7	0.3	0.0	0.5	0.5	0.5	0.5	0.0	0.0	1.0
6	0.0	0.7	0.3	0.0	0.3	0.7	0.7	0.3	0.4	0.6	0.2	0.0	0.8
7	0.0	0.3	0.7	0.0	0.0	1.0	0.0	1.0	0.1	0.9	0.0	0.0	1.0
8	0.0	1.0	0.0	0.0	0.2	0.8	0.2	0.8	0.0	1.0	0.7	0.0	0.3
9	1.0	0.0	0.0	1.0	0.0	0.0	0.6	0.4	0.7	0.3	0.2	0.8	0.0
10	0.9	0.1	0.0	0.0	0.3	0.7	0.0	1.0	0.9	0.1	0.0	0.3	0.7
11	0.7	0.3	0.0	1.0	0.0	0.0	1.0	0.0	0.2	0.8	0.4	0.7	0.0
12	0.2	0.6	0.2	0.0	1.0	0.0	0.3	0.7	0.3	0.7	0.7	0.2	0.1
13	0.9	0.1	0.0	0.2	0.8	0.0	0.1	0.9	1.0	0.0	0.0	0.0	1.0
14	0.0	0.9	0.1	0.0	0.9	0.1	0.1	0.9	0.7	0.3	0.0	0.0	1.0
15	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.8	0.2	0.0	0.0	1.0
16	1.0	0.0	0.0	0.5	0.5	0.0	0.0	1.0	0.0	1.0	0.8	0.6	0.0

如果可信度的阈值 $\alpha=0.7$，则得到决策树如图1。从该树提取的模糊规则如下

（1）若晴天天热则适宜于三种运动的可能性分布 $d_1=\{\frac{0.37}{\text{排球}},\frac{0.69}{\text{游泳}},\frac{0.12}{\text{举重}}\}$，$(r=0.78)$。

（2）若阴天天热则适宜于三种运动的可能性分布 $d_2=\{\frac{0.43}{\text{排球}},\frac{0.53}{\text{游泳}},\frac{0.12}{\text{举重}}\}$，$(r=0.79)$。

（3）若天热下雨则适宜于三种运动的可能性分布 $d_3=\{\frac{0.17}{\text{排球}},\frac{0.07}{\text{游泳}},\frac{0.75}{\text{举重}}\}$，$(r=0.75)$。

（4）若天气暖和有风则适宜于三种运动的可能性分布 $d_4=\{\frac{0.17}{\text{排球}},\frac{0.07}{\text{游泳}},\frac{0.75}{\text{举重}}\}$，$(r=0.79)$。

（5）若天气暖和无风则适宜于三种运动的可能性分布 d_5

$=\{\frac{0.64}{排球},\frac{0.28}{游泳},\frac{0.24}{举重}\}$,$(r=0.71)$。

(6)若晴天天凉则适宜于三种运动的可能性分布 $d_6=\{\frac{0.00}{排球},\frac{0.30}{游泳},\frac{0.70}{举重}\}$,$(r=1.00)$。

(7)若天凉阴天则适宜于三种运动的可能性分布 $d_7=\{\frac{0.38}{排球},\frac{0.01}{游泳},\frac{0.61}{举重}\}$,$(r=0.76)$。

(8)若天凉下雨则适宜于三种运动的可能性分布 $d_8=\{\frac{0.02}{排球},\frac{0.00}{游泳},\frac{0.98}{举重}\}$,$(r=0.97)$。

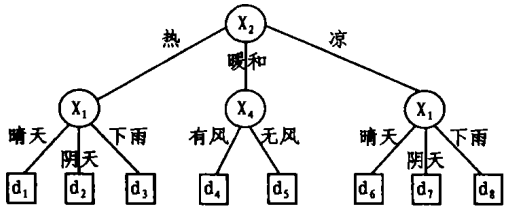

图1 $d_i, i=1,\cdots,8$为对应终端结点的可能性分布(见规则的结论)

在可信度大于0.7的条件下,上述各条模糊规则同湿度没有关系。假设已知一个对象属于属性值的隶属度为(0.0,1.0,0.0;0.5,0.5,0.0;0.7,0.3;1.0,0.0)。则该对象对8条规则的条件的满足程度分别为0.0,0.5,0.0,0.5,0.0,0.0,0.0,0.0。由于满足第二条规则和第四条规则的程度都达到最大,大多数模糊决策树都是在适宜于游泳和适宜于举重中任选一个,作为该对象所属的类,这带有很大的不确定性。若假定输出对应的是可能性分布,利用简单模糊推理可以得到该对象的可能性分布为$\{\frac{0.30}{排球},\frac{0.30}{游泳},\frac{0.44}{举重}\}$。此时适宜于三种运动的程度相差不远,而适宜于举重的程度略大。

结论 本文在假设一个对象可以属于多个类的情况下,给出了模糊决策树的算法,所给算法得出的模糊决策树的终端结点赋给的是一个可能性分布。简单模糊推理被用于确定一个属性值已知对象所属类的可能性分布。本文所得的模糊树能够避免在确定属性值已知的对象时所属类出现不一致的情况,对所建的模糊决策树提取的规则解释更加自然。

参考文献

1 Janikon C Z. Fuzzy Decision Trees: Issues and Methods. IEEE Trans. Syst. ,Man,Cybern,B,1988,28:1～14
2 Yuan Y,Shaw M J. Induction of Fuzzy decision Trees. Fuzzy sets and Systems,1995,69:125～139
3 Quinlan J R. Induction of Decision Trees. Machine learning,1986,1:71～99
4 Quinlan J R. C4.5,Programs for Machine Learning. San Mateo,CA:Morgan Kuauffman,1993
5 Chang I, Hsu J. Fuzzy Classification Trees for Data Analysis. Fuzzy Sets and Systems,2002,130:87～99
6 Myung L K,Mi K L,Hyong L J,Kwang H L. A Fuzzy Decision Tree Induction Method for Fuzzy Data. In:Proc. IEEE Int. Fuzzy Systems Conf. 1999. 1～16
7 张文修,梁广锡. 模糊控制与系统. 西安交通大学出版社,1998

(上接第57页)

和基于限制容差关系的广义 Rough 集下集合之间相似性度量的方法和性质,实际上这两种度量方法是有联系的,而且是统一的。下面将对这两种集合相似度量方法进行一些比较。

(1)经典 Rough 集和广义 Rough 集中集合之间相似性度量的方法具有一致性,可以为它们建立一个统一模型。这两种度量方法实际上就是一种度量方法,即 Rough 集之间的相似性度量方法,它可用于不确定推理和不确定信息的处理。只是当信息系统完备时,用经典 Rough 集之间的相似性度量方法;当信息系统不完备时,用广义 Rough 集之间的相似性度量方法。

(2)经典 Rough 集和广义 Rough 集中集合之间相似性度量方法满足一些共同的性质。如经典 Rough 集中定理1～3分别与广义 Rough 集中定理4～6相对应。

(3)经典 Rough 集下集合之间相似性度量是基于完备信息系统下的不分明关系的;而广义 Rough 集下集合之间相似性度量是基于不完备信息系统下的限制容差关系的。

(4)不管是经典 Rough 集还是广义 Rough 集之间的相似程度,可根据实际要求确定一个阈值,当集合之间的相似程度落在限定的阈值范围内时,就认为两个集合之间具有相似性。

(5)经典 Rough 集中集合之间相似性度量方法是广义 Rough 集中集合之间相似性度量方法的特殊情况。当基于限制容差关系的广义 Rough 集中的信息从不完备状态变为完备状态时,则基于限制容差关系的广义 Rough 集下集合之间相似性度量的方法即为基于不分明关系的经典 Rough 集下集合之间相似性度量的方法。

(6)除基于限制容差关系的广义 Rough 集之外,对于基于容差关系、非对称相似关系的广义 Rough 集,也可以定义类似的广义相似性度量,限于篇幅,不再详细讨论,有兴趣的读者可以自行设计。

结论 Rough 集理论在不完备信息系统中的应用,是将 Rough 集理论进一步推向实用的关键之一,因为现实需要处理的数据基本都是在一定程度上是不完备的,所以就有必要研究能够直接处理不完备数据处理的方法。本文在分析研究经典 Rough 集与广义 Rough 集的一些基本概念理论的基础上,分别提出并讨论了基于不分明关系的经典 Rough 集之间的相似度量及性质和基于限制容差关系的广义 Rough 集之间的相似度量及性质。最后对它们的度量方法进行了一些比较。对于本文提出的度量方法,经过仿真实验,效果良好,可用于完备信息系统或不完备信息系统中 Rough 集之间的相似度量,从而更好地应用于不确定信息的数据处理。本文所提出的 Rough 集之间的相似度量,也可用于不确定性推理中。基于不确定集合相似度量的匹配与推理系统,是我们下一步的研究工作。

参考文献

1 Kryszkiewicz M. Rough set approach to incomplete information systems. Information Sciences, 1998,112:39～49
2 Liu Qing. λ-Level Rough equality relation and the inference of rough paramodulation. In:Proc. of Intl. Workshop on Rough Set Theory and Granular Computing (RSTGC-2001),2001,5:424～431
3 江娟,刘清. 基于程度不分明关系的广义 Rough 集定义. 计算机科学,2001,28(5.专刊):8～9
4 王国胤. Rough 集理论在不完备信息系统中的扩充. 计算机研究与发展,2002,39(10)
5 李凡,徐章艳. Vague 集之间的相似度量. 软件学报,2001,12(6):922～927
6 王国胤. Rough 集理论与知识获取. 西安:西安交通大学出版社,2001
7 Chen S M. Measures of similarity between vague sets. Fuzzy Sets and Systems,1995,74(2):217～223
8 Hong D H,Kim C. A note on similarity measures between vague sets and between elements. Information Sciences, 1999,115(1):83～96