

# 模糊决策树在高校师资管理中的应用

姚笑秋<sup>1,2</sup>, 何仁刚<sup>2</sup>, 陆秋<sup>3</sup>

(1. 华中科技大学 计算机科学与技术系, 湖北 武汉 430074;

2. 浙江海洋学院 计算机科学与技术系, 浙江 舟山 316004;

3. 广西桂林工学院 电子与计算机系, 广西 桂林 541004)

**摘要:**良好的教师资源是高校持续发展的重要保证。文中分析了当前高校教师资源管理中存在的问题,进而把融合了模糊理论和决策树优点的模糊决策树应用到这一领域;简要介绍了模糊决策树构造思想和 Fuzzy ID3 决策树学习算法,并将该思想和算法运用于高校师资管理之中,对引起高校教师资源流失的原因进行了初步分析。以若干高校提供的数据为基础的实验表明,本方法产生的分类结果基本符合实际情况,从而也为高校如何解决师资流失问题提供了理论依据。

**关键词:**模糊决策树;师资流失;Fuzzy ID3

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2007)05-0024-03

## Application of Fuzzy Decision Tree in Faculty Management of High Schools

YAO Xiao-qi<sup>1,2</sup>, HE Ren-gang<sup>2</sup>, LU Qiu<sup>3</sup>

(1. Dept. of Computer Science and Technology, Huazhong University of Science & Technology, Wuhan 430074, China;

2. Dept. of Computer Science and Technology, Zhejiang Ocean University, Zhoushan 316004, China;

3. Dept. of Electronics & Computer Science, Guilin University of Technology, Guilin 541004, China)

**Abstract:** Superior faculty ensures the sustainable development of high schools. By analyzing the problems existing in the faculty management of high schools, tries to apply into this field the fuzzy decision tree that incorporates the excels of both fuzzy theory and fuzzy tree. In this article, thought of fuzzy decision construction and algorithm of fuzzy ID3 decision tree are briefly introduced and then applied to the faculty management of high schools. Consequently, a tentative analysis of the reason causing the loss of teachers in high schools is conducted. The experiment based on the data from a couple of high schools shows that the results reached from this method comply to the facts, so that theoretical evidences are provided for high schools to tackle the problems in faculty management.

**Key words:** fuzzy decision tree; loss of teachers; fuzzy ID3

### 1 师资管理中存在的问题

目前高校的办学实力竞争越来越体现在师资力量这一重要环节上,师资薄弱的高校不仅在每年的招生大战中节节败退,更难以通过教育部的评估,无法立足于高校之林。如果没有一支业务能力和自身素质过硬的教师队伍,那么建设品牌学校并持续发展也就成了纸上谈兵。然而,现在很多学校的相关管理部门没能意识到教师资源危机,即使造成了大量的教师资源流失,也没有采取积极的解决方案和防范措施,使得学校

发展一再受阻。

在学校办学规模和招生规模不断发展壮大的同时,学校内部都积累了大量的与教师资源密切相关的数据。在这些海量的数据的背后隐藏着许多重要的模式和知识,对分析研究高校教师流失的原因具有重大意义。快速、准确、高效地将这些模式和知识挖掘出来,是学校提高教师资源管理水平和增强学校竞争力的一个重要手段。

传统使用的统计方法,基本上只能获得这些数据的表层信息,而不能挖掘出数据间潜在的相关性。这就需要有一种能够从这些海量数据中寻求有用信息的工具,数据挖掘就是在这种背景下应运而生的。

目前,决策树已成为一种重要的数据挖掘方法,而在决策树构造中, ID3 算法是最具有影响的一种决策

收稿日期:2006-07-16

基金项目:浙江省教育厅资助项目(20050113)

作者简介:姚笑秋(1980-),女,浙江衢州人,助教,硕士研究生,研究方向为人工智能、数据挖掘;导师:马光志,副教授,研究方向为软件工程、数据挖掘。

树生成算法,是1986年由Quinlan<sup>[1]</sup>提出的,但该算法不能很好地处理数据模糊性和不确定性问题,而模糊推理恰好补充了这一缺陷。模糊决策树作为一种数据挖掘方法,集成了模糊理论和决策树的优点,不仅具有很强的决策分析能力,而且能很好地处理模糊性和不确定性问题。

## 2 模糊决策树及其学习方法

### 2.1 决策树基本理论

归纳学习<sup>[2]</sup>着眼于一组无次序、无规则的事例中,找出所蕴涵的规律,事例一般是基于属性理论的,由特定的属性值得到问题的结论。决策树学习算法则是以实例为基础的归纳学习算法,它可以对数据进行分类或预测,是树形结构的知识表示。树中每个内部节点表示一个属性上的测试,每个分枝代表一个测试输出,而每个叶子节点代表类或类分布。树的最顶层节点是根节点。从根节点到叶子节点的一条路径形成一条分类规则。决策树可以很方便地转化为分类规则,是一种非常直观的分类模式表示形式。

决策树的构建是一种自上而下、分而治之的归纳过程,本质是贪心算法。决策树算法通常包括两部分:一是树的生成,开始时所有数据都在根节点;然后根据设定的标准选择测试属性,用不同的测试属性递归地进行数据分割,直至达到特定的终止条件。构造一棵决策树就是形成一个训练集的分类过程,人们可以在分类过程中完成对目标的获取策略与规则提取。其中,测试属性的选择和如何划分样本集是构建决策树的关键环节。不同的决策树算法在此使用的技术不尽相同<sup>[3]</sup>。

### 2.2 模糊集理论的基本概念

传统决策树是基于符号属性的,如ID3算法,在处理数值属性、属性选择等问题上存在许多不足,文献[4,5]都对ID3算法进行了优化。而在某些属性取值的模糊性与不确定性方面仍稍有不足,于是提出了模糊决策树算法。模糊决策树利用模糊划分的概念,解决了数值属性问题;通过引入模糊集理论中的模糊隶属函数,来处理模糊(不确定性)数据集的学习问题。在推理过程中,传统决策树每次只沿着满足条件的分支向下前进,最终到达仅仅一个叶结点(即结论);而模糊决策树以不同的信任度同时沿着多个分支向前搜索,最终到达多个叶结点。

模糊集理论是Zadeh于1965年提出的,随后推广应用到很多领域,包括人工智能领域。模糊集的关键概念就是对每一个集合中的元素赋予一个隶属度。与传统集合相似,模糊集也可以有许多算子操作计算,如

等于、包含、并和交等<sup>[6]</sup>。在归纳学习中要用到的两个算子是集合的并与交。令 $A, B$ 为2个模糊集, $\mu_A(x)$ ,  $\mu_B(x)$ 分别表示元素 $x$ 属于 $A, B$ 的隶属度。

由两个集合的并所形成的并集,可以表示为:

$$A \cup B \Leftrightarrow \forall x \in U, \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$$

由两个集合的交所形成的交集,可以表示为:

$$A \cap B \Leftrightarrow \forall x \in U, \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$$

### 2.3 模糊决策树算法 - Fuzzy ID3

目前存在许多模糊决策树学习算法,Fuzzy ID3算法就是其中之一,它将模糊理论引入到传统ID3算法中,以提高决策树处理数值属性和模糊性等问题的能力。

Fuzzy ID3算法的测试属性选择介绍如下:

设数据集(实例集合)为 $D = \{e_1, e_2, \dots, e_N\}$ 是定义于离散值论域 $X$ 上的例子集合,模糊属性集为 $\{A_1, A_2, \dots, A_M\}$ ,属性 $A_i$ 的属性值为 $T(A_i) = \{a_{i1}, a_{i2}, \dots, a_{iki}\} (1 \leq i \leq M)$ ,所要划分的类 $C = \{C_1, C_2, \dots, C_K\}$ 。第 $i$ 个例子 $e_i$ 关于第 $j$ 个属性的值由相应的隶属度 $\mu_{ij}$ 表示,它是一个定义在 $T(A_i)$ 上的模糊子集。若属性 $A_i$ 为符号值属性类分明情况,则 $\mu_{ij}$ 的取值为0或者1。另外设 $D_{c_k}$ 为类别 $c_k$ 的数据子集, $|D|$ 为 $D$ 中所有隶属值的总和,则数据集 $D$ 的信息熵为:

$$I(D) = - \sum_{k=1}^K P_k \cdot \log_2 P_k \quad (1)$$

其中 $K$ 为分类个数, $P_k = \frac{|D_{c_k}|}{|D|}$ ,而用属性 $A_i$ 划分数数据集 $D$ 后的模糊熵 $E(A_i, D)$ 为:

$$E(A_i, D) = \sum_{j=1}^{K_i} p_{ij} \cdot I(Da_{ij}) \quad (2)$$

其中 $Da_{ij}$ 为属性 $A_i$ 划分 $D$ 所得的子集,

$$P_{ij} = \frac{|Da_{ij}|}{\sum_{n=1}^{K_i} |Da_{in}|}$$

则相对于数据集 $D$ 的信息增益为 $G(A_i, D)$ :

$$G(A_i, D) = I(D) - E(A_i, D) \quad (3)$$

根据式(1)~(3),即可计算每个属性相对于数据集的信息增益。与传统ID3雷同,根据最大信息增益原理,每次选择信息增益最大的属性作为测试属性,实现对数据集的划分。在划分过程中,根据每个实例的某一测试属性的隶属度的取值,可将其归入多个子集(表示某一实例属于多个类型的可能性),以生成模糊决策树。

根据此方法产生的模糊决策树的叶节点并不是一个唯一的类别,而是以信任度(隶属值)标定的类。而且不同的决策树,其叶节点的产生也有所不同,即学习结束条件有多种情形:



- a)某个分类的信任度大于给定的阈值  $\beta$ ;
- b) 数据集的隶属值的总和小于给定的阈值  $\alpha$  ;
- c)所有测试属性使用完毕。

### 3 模糊决策树在高校师资管理中的应用

#### 3.1 师资管理的任务

任何高校的发展都渴求拥有良好、稳定的师资队伍。教师的流动是影响教师队伍建设的方面,“流水不腐,户枢不蠹”,合理的流动,可以解决学非所用,用非所学;可以使教师水平不断上升。然而目前,很多高校都面临着大量的教师流失问题。

师资管理的任务,不仅要解决怎样使在校的教师能够发挥其最大的作用,怎样安排其合适的岗位等问题,更要分析大量流失的教师当中,不同类型的教师的流失原因(或者称为离职原因)。只有深入分析离职者的身份、特点和离职原因之间的关系,学校才能有针对性地制定相关政策,以各种方法或措施来挽留学校所需人才,并对新的求职者做出预测。

#### 3.2 师资管理中模糊决策树方法应用

分析教师的流失原因及其类型,势必要讨论评测教师的各项指标,如科研测评、教学测评、职称、学历、年龄、流失原因等等。其中,科研测评和教学测评数据是根据专家和师生的打分所得,和年龄一样,为离散性数据。而职称、学历为符号类值分明属性。

对于离散性数据的处理,若通过划分区间法将其转换为符号类属性,容易在边界处带来突变。如,规定 1.78m 以上的身高为高个子,则 1.779m 的也不是高个子,但其实两种非常接近,强行分区容易带来不少问题,如增加误差。鉴于数据的模糊性,因此,必须引入能处理模糊性问题的方法。模糊决策树则提供了一个解决问题的方法。

这里用到的原始数据是若干个高校的流失(离职)人员基本信息,包括:姓名、性别、年龄、教学情况、科研能力测评等等。将这些信息收集之后,如表 1 所示。

表 1 离职教师基本信息原始数据

ID	姓名	职称	性别	民族	年龄	学历	...
1	王林	讲师	女	汉	34	硕士	...
2	马云	教授	男	汉	39	博士	...
...	...	...	...	...	...	...	...

在这些原始数据中抽取一小部分数据作为训练集,同时根据用户兴趣度提取出数据挖掘所需要的相关字段(去掉一些与离职原因没多大关系的属性,如:联系方式、民族、籍贯等),并对离散性属性进行模糊化处理。最终得到如表 2 所示训练集。

其中“离职原因”为分类属性,鉴于篇幅问题,一些相关属性如学位、性别等未列入本例当中。

利用上述 2.3 节介绍的 Fuzzy ID3 算法,检验所有信息增益的特征属性,选择信息增益最大的属性作为测试属性,由于计算过程十分复杂,仅给出最终生成的模糊决策树,如图 1 所示。

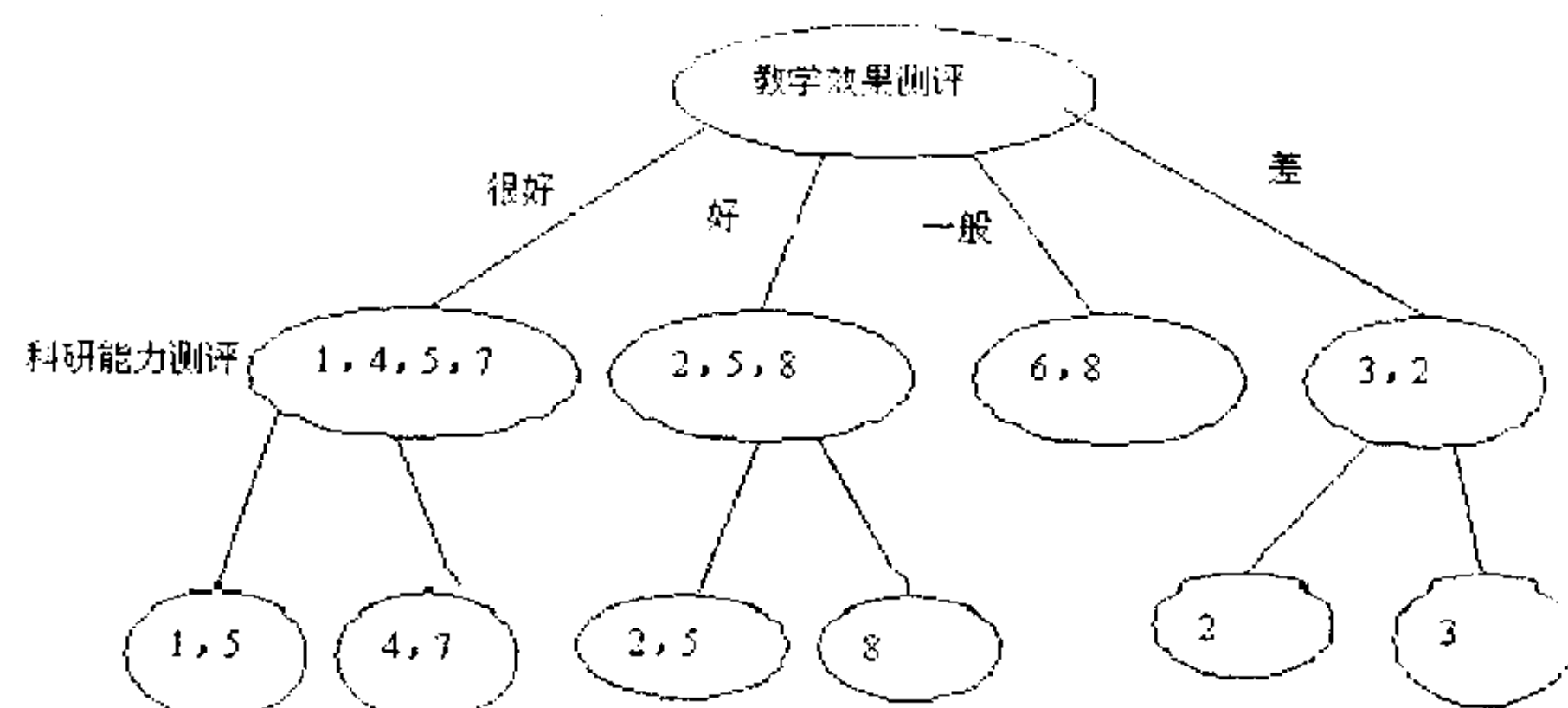


图 1 最终决策树

#### 3.3 知识规则描述

决策树推理是从根节点开始,通过重复地进行属性及其取值的比较,沿着决策树的分支向下搜索,一直到达叶节点,得到所要的结论。而模糊决策树总是以某个信任度(隶属度)沿着多个分支向下的搜索,最终到达多个叶节点(同一实例属于多个叶节点)。生成决策树之后,接下来要做的就是将模糊决策树以规则形式表示出来。

和传统决策树一样,模糊决策树也可表示成 if-then 形式的分类规则。以图 1 表示的决策树为例,得

表 2 模糊化处理后的训练集(离职教师信息表)

ID	教学效果测评				职称				科研能力测评			年龄	离职原因
	很好	好	一般	差	教授	副教授	讲师	助教	强	中	弱		
1	0.6	0.2	0.1	0.1	0	1	0	0	0.5	0.3	0.2	45	薪资不满意
2	0.1	0.3	0.2	0.4	1	0	0	0	0.7	0.2	0.1	39	薪资不满意
3	0	0.2	0.2	0.6	0	1	0	0	0.6	0.3	0.1	35	不能发挥作用
4	0.7	0.1	0.1	0.1	0	0	0	1	0.1	0.3	0.6	28	不能发挥作用
5	0.5	0.4	0	0.1	0	0	1	0	0.4	0.4	0.2	31	薪资不满意
6	0.1	0.2	0.6	0.1	0	0	1	0	0.1	0.3	0.6	29	夫妻两地分居
7	0.6	0.2	0.1	0.1	0	0	1	0	0.2	0.5	0.3	31	不能发挥作用
8	0.1	0.4	0.3	0.2	1	0	0	0	0.2	0.3	0.5	40	夫妻两地分居

这些方案,找到需要调用的服务类型,然后由服务类型找到组成此类服务类型的具有最佳性能的基本服务,给其起始服务以输入,按动态组合方案执行,调用其后服务,生成期望输出。

图9表示了按照图8生成的动态服务组合方案在服务组合执行过程中的应用,通过对图9的描述可以了解文中关于基于性能的动态服务组合的关键所在。

图9中,在组合服务引擎工作前,由基本服务所组成的服务类型已经被生成并记录,然后给服务组合引擎以一个 Input,动态服务组合以此 Input 和动态服务组合策略,生成最佳的组合方案,服务组合引擎依照此方案寻找相应的服务类型组成动态服务组合执行路径,然后挑选服务类型中所记录的此服务类型的性能最佳的服务来依次执行,最终产生期望输出 Output。

### 3 结 论

伴随着 Web 服务的发展,动态服务组合方案的生成一直被广泛注意和研究。文中以改变传统的动态服务组合结构为基础,借助 QoS 中有关服务发现的基本知识,较为合理地解决了动态服务组合的生成和性能问题,并且开阔了动态服务组合的应用前景。

#### 参考文献:

- [1] Booth D. Web Services Architecture[EB/OL]. 2004-02.  
http://www.w3.org/TR/ws-arch, W3C Working Group

(上接第26页)

出的部分有价值的挖掘结果如下:

教学效果测评和科研能力测评,有一项表现不太好的,往往是学校强加于教师身上的定位使得教师不能发挥自己的作用;

教学效果测评高的,并且科研能力强的,一般对薪资要求也比较高;

不管年龄如何,多年未解决夫妻两地分居问题者多有离职的倾向。

### 4 结束语

当然,高校中还存在着其它的“离职原因”,如:高校所处位置、单位裁员下岗、年轻教师因为要继续深造而脱离原单位等等。在此,仅说明模糊决策树在分析离职教师基本信息的过程中的应用,所以属性取值考虑并不全面。

由以上分类挖掘产生的结果数据表明,数据挖掘得出的结果与现实情况还是基本符合。一些事先并没

Note.

- [2] Liang Q A. Composite Service Discovery, Description and Invocation[J]. International Journal of Web Service Research, 2004,1(4):64-89.
- [3] Zou Zhile, Duan Zhenhuan, Wang Jianli. A Comprehensive Framework for Dynamic Web Services Integration[C]//E-COWS2006, 4th IEEE European Conference on Web Services. [s.l.]:[s.n.],2006.
- [4] Deora V, Shao J, Shercliff G, et al. Incorporating QoS Specifications in Service Discovery[C]//Web Information Systems - WISE2004. [s.l.]:[s.n.],2004:252-263.
- [5] Shen Derong, Yu Ge, Nie Tiezheng, et al. Modeling QoS for Semantic Equivalent Web Services[C]//Advances in Web - Age Information Management: 5th International Conference. WAIM2004. [s.l.]: Springer,2004:478-488.
- [6] Hwang San - Yih, Wang Haojun, Srivastava J, et al. A Probabilistic QoS Model and Computation Framework for Web Services - Based Workflows[C]//Conceptual Modeling - ER 2004,23(1) International Conference on Conceptual Modeling. Xi'an:[s.n.], 2004:596-609.
- [7] Vahid S, Mavaddat H F. A Graph - Based Approach to Web Services Composition[J]. Computer Society,2005,50(2):2-3.
- [8] L Jing - Shan, L Hua - Ming, H Zi - Feng, et al. A dynamic service composition method based on semantic interface description in pervasive computing[J]. Journal of Computer Research and Development,2004,41(7):1124-1134.

有预料的结果,也客观地反映了实际情况。将模糊决策树应用于分析离职教师基本信息中,可以使得师资管理的相关工作人员,甚至整个学校的领导能够引起重视,并提出相应的解决措施,尽最大可能挽留一切可以挽留的人才。

#### 参考文献:

- [1] Quinlan J R. Induction of decision tree[J]. Machine Learning, 1986,1(1):81-106.
- [2] Wu Xindong. Inductive Learning[J]. Journal of Computer Sci & Tech,1993,8(2):118-132.
- [3] 邵峰晶,于忠清.数据挖掘原理与算法[M].北京:中国水利水电出版社,2003:126-170.
- [4] 王静红,王熙照,邵艳华,等.决策树算法的研究及优化[J].微机发展,2004,14(9):30-32.
- [5] 杨宝华.基于粗集的决策树构建的探讨[J].计算机技术与发展,2006,16(8):83-87.
- [6] Kantardzic M. 数据挖掘——概念、模型、方法和算法[M].北京:清华大学出版社,2003:219-239.