

# 基于概率图的三支决策模型研究

薛占熬 王朋函 刘 杰 朱泰隆 薛天宇

(河南师范大学计算机与信息工程学院 新乡 453007)

(“智慧商务与物联网技术”河南省工程实验室 新乡 453007)

**摘 要** 概率图模型是一类用图形模式表达基于概率关系的模型的总称,用该模型解决损失代价问题已成为当前的研究热点。结合概率图和三支决策理论,提出了基于概率图的三支决策模型。该模型通过对数据进行分析,构造其 Bayes 网络;并根据模型中节点的相互依赖关系,计算出条件概率分布函数;结合查询变量的先验概率和三支决策损失代价函数,建立了相应的决策规则,给出了概率推理决策中代价最小化问题的一种解决方法。最后通过教学评估实例验证了该模型的有效性。

**关键词** 三支决策,概率图模型,Bayes 网络,精确推理

**中图分类号** TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.1.007

## Three-way Decision Model Based on Probabilistic Graph

XUE Zhan-ao WANG Peng-han LIU Jie ZHU Tai-long XUE Tian-yu

(College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China)

(Engineering Lab of Henan Province for Intelligence Business & Internet of Things, Xinxiang 453007, China)

**Abstract** Probability graph model is a general term for a class of graph patterns that express the probabilistic model, and the problem of solving the loss cost by using the model has become a hot research hotspot. In this paper, the three-way decision model was put forward based on probability graph and three-way decision theory. By analyzing the information data set, the Bayesian network of the model is constructed. Then according to the interdependencies of the nodes in the model, the conditional probability distribution function is calculated. And combining with the prior condition of the data set and three decision-making cost loss function, its decision rules are established. The method to solve the problem of cost minimization is given in probability reasoning decision-making. Finally, a teaching evaluation illustration was presented to verify the effectiveness of model.

**Keywords** Three-way decision, Probabilistic graph model, Bayesian network, Precise reasoning

## 1 引言

加拿大里贾纳大学姚一豫教授通过扩展决策粗糙集,于 2010 年创立了三支决策理论,给出了粗糙集 3 个域应用的一种合理的语义解释<sup>[1,2]</sup>。传统的二支决策只考虑接受和拒绝两种状态,三支决策在此基础上增加了延迟决策,当信息不足以决定接受或者拒绝时,采取延迟决策实际问题更加符合人们的认知模式。近几年来,三支决策理论已成为国内外专家学者研究的热点。刘盾等对三支决策理论提出了新的研究方法和应用领域<sup>[3]</sup>;杜丽娜等将 Bayes 原理和基于决策粗糙集的三支决策规则引入到模糊综合评判中,提出了基于三支决策风险最小化的风险投资评判模型<sup>[4]</sup>;李华雄等采用代价敏感三支决策方法求得相应粒度下的最小代价决策<sup>[5]</sup>。

概率图模型是用图形模式表达概率关系的模型,是在概率论和图论的基础上发展而来的,为解决智能信息处理的

确定性和复杂性问题提供了直观而自然的解决方法,为计算机视觉、语音识别和知识发现等领域的发展提供了重要理论基础<sup>[6]</sup>。徐瑾等提出了一种基于概率图模型的研究生源质量评价方法,通过评分搜索的学习方法构建数学模型,实现推理过程<sup>[7]</sup>;王忆南等提出了基于概率图模型的态势估计,对战场态势估计中的相关理论进行了深入研究<sup>[8]</sup>。

三支决策在代价决策领域得到了广泛应用。但是已有的研究很少将概率图和三支决策相结合进行研究<sup>[9-14]</sup>。本文通过结合概率图和三支决策,构造基于概率图的三支决策模型,计算查询变量发生的概率,然后结合最小代价的三支决策,确定正域、负域和边界域的界定范围,从而得到最佳的决策;最后通过教学评估实例验证了该模型的有效性。

## 2 基础知识

**定义 1(联合分布)**<sup>[9]</sup> 设  $(X, Y)$  是一组随机变量,对任

到稿日期:2015-04-19 返修日期:2015-05-23 本文受国家自然科学基金计划项目(61273018),河南省基础与前沿技术研究计划项目(132300410174),河南省教育厅计划项目(14A520082),新乡市重点科技攻关计划项目(ZG14020)资助。

薛占熬(1963—),男,博士,教授,主要研究方向为人工智能基础理论和粗糙集理论,E-mail:xuezhanao@163.com;王朋函(1986—),男,硕士生,主要研究方向为概率论和不确定推理;刘 杰(1989—),女,硕士生,主要研究方向为粗糙集和三支决策基本理论;朱泰隆(1990—),男,硕士生,主要研究方向为博弈论、决策论和粗糙集理论;薛天宇(1991—),男,硕士生,主要研究方向为机器学习、模糊集、代数。

意实数,  $\{X \leq x\}, \{Y \leq y\}$  同时发生的概率为:

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

称为二维随机变量  $(X, Y)$  的分布函数, 或者随机变量  $X$  和  $Y$  的联合分布。

定义 2(条件概率分布)<sup>[10]</sup>  $(X, Y)$  表示二维离散随机变量, 对满足条件  $P\{Y=y_j\} = p_j = \sum_{i=1}^{\infty} p_{ij} > 0$  的  $Y_j$ ,

$$p_{ij} = P\{X=x_i | Y=y_j\} = \frac{P\{X=x_i, Y=y_j\}}{P\{Y=y_j\}} = \frac{p_{ij}}{p_j}$$

是条件  $Y=y_j$  下的概率分布, 由条件概率分布的定义可得联合概率分布:  $P(X, Y) = P(X)P(Y|X)$ 。

定义 3(Bayes 公式)<sup>[11]</sup> 设  $H$  和  $E$  表示两个随机变量,  $H=h$  表示事件变量,  $E=e$  表示一组证据变量, 在加入证据  $E$  之前对事件  $H$  的概率估计  $P(H=h)$  称为先验概率, 加入证据变量之后,  $P(H=h|E=e)$  表示事件  $H=h$  的后验概率。Bayes 公式描述了先验概率和后验概率之间的关系:

$$P(H=h|E=e) = \frac{P(H=h)P(E=e|H=h)}{P(E=e)}$$

定义 4(Bayes 网络)<sup>[12,13]</sup> 定义随机变量集  $X$  的 Bayes 网络可以定义为  $BN=(B_S, B_P)$ 。其中  $B_S$  是 Bayes 网络的结构, 为定义在  $X$  上的有向无环图。每一个节点  $x_i$  都唯一对应着  $X$  中的一个随机变量, 并需要标注定量的概率信息; 每条有向边分别表示所连接的两个节点的条件依赖关系。若存在一条从节点  $x_j$  到  $x_i$  的有向边, 则称  $x_j$  是  $x_i$  的父节点。  $B_P$  为 Bayes 网络的条件概率集合,  $B_P = \{P(X_i | par(X_i))\}$ 。其中,  $par(X_i)$  表示  $x_i$  的所有父节点的相应取值,  $P(X_i | par(X_i))$  是节点  $x_i$  的一个条件概率分布函数, 它描述  $x_i$  的每个父节点对  $x_i$  的影响, 即节点  $x_i$  的条件概率表。

定义 5(三支决策)<sup>[14]</sup> 设  $\Omega = \{X, X^c\}$  表示两种状态集合, 分别表示对象属于集合  $X$  和对象属于集合  $X$  的补集。集合  $A = \{a_P, a_B, a_N\}$  对应 3 种决策动作: 接受、拒绝和延迟。

决策过程中往往存在不同的代价损失, 决策代价表如表 1 所列。设  $\lambda_{PP}, \lambda_{BP}$  和  $\lambda_{NP}$  表示当一个对象属于集合  $X$  时采取行动  $a_P, a_B$  和  $a_N$  所付出的代价,  $\lambda_{PN}, \lambda_{BN}$  和  $\lambda_{NN}$  分别表示属于  $X$  补集时采取行动  $a_P, a_B$  和  $a_N$  所付出的代价。

表 1 决策代价表

决策行动	代价函数	
	X	X <sup>c</sup>
$a_P$	$\lambda_{PP}$	$\lambda_{PN}$
$a_B$	$\lambda_{BP}$	$\lambda_{BN}$
$a_N$	$\lambda_{NP}$	$\lambda_{NN}$

对集合  $[x]$  中的对象采用不同的决策所产生的代价损失函数为:

$$R(a_P | [x]) = \lambda_{PP} \Pr(X|[x]) + \lambda_{PN} \Pr(X^c|[x]),$$

$$R(a_B | [x]) = \lambda_{BP} \Pr(X|[x]) + \lambda_{BN} \Pr(X^c|[x])$$

$$R(a_N | [x]) = \lambda_{NP} \Pr(X|[x]) + \lambda_{NN} \Pr(X^c|[x])$$

根据 Bayes 决策论, 提出了最小风险决策规则为:

(P) 如果  $R(a_P | [x]) \leq R(a_B | [x])$  且  $R(a_P | [x]) \leq R(a_N | [x])$ , 则  $x \in POS(X)$ ;

(B) 如果  $R(a_B | [x]) \leq R(a_P | [x])$  且  $R(a_B | [x]) \leq R(a_N | [x])$ , 则  $x \in BND(X)$ ;

(N) 如果  $R(a_N | [x]) \leq R(a_P | [x])$  且  $R(a_N | [x]) \leq R(a_B | [x])$ , 则  $x \in NEG(X)$ 。

通常情况下存在如下关系:

$$\lambda_{PP} < \lambda_{BP} < \lambda_{NP}, \lambda_{NN} < \lambda_{BP} < \lambda_{PP}$$

$$(\lambda_{PP} - \lambda_{BN})(\lambda_{NP} - \lambda_{BP}) > (\lambda_{BN} - \lambda_{NN})(\lambda_{BP} - \lambda_{PP})$$

于是简化可以得到:

$$\alpha = \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}$$

$$\beta = \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}$$

$$\lambda = \frac{(\lambda_{PN} - \lambda_{NN})}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}$$

### 3 基于概率图的三支决策模型及其应用

概率图模型主要用于推理和学习, 而精确推理就是利用 Bayes 网络构建数学模型, 通过联合概率分布计算变量的后验概率, 并最终实现推理过程的一种概率图模型。但在实际应用中, 推理的目的是做出最佳决策, 需要考虑到不同推理结果的误分类代价, 从而引入三支决策理论规则, 最终通过一个教学评估的实例对该模型的有效性进行验证。

概率图模型的构造利用了 Bayes 网络和联合概率分布, 以变量之间的条件依赖关系为起点, 其构造步骤如下:

(1) 首先建立一个独立的根节点, 并在图 1 中表示出来。

(2) 从每个根节点开始, 加入有依赖关系的对应子节点, 并在图 1 中用有向边表示其依赖关系。

(3) 重复步骤(2), 直到所有节点都在图 1 中表示出来。

(4) 标出根节点的先验概率, 以及中间节点和叶子节点的条件概率表。

在一个教学评估问题中, 学生的考试成绩一方面反映了教学的难度, 另一方面反映了学生的努力程度, 而四级考试的通过率也能反映学生是否努力学习, 最后学校在就业推荐时会优先考虑成绩比较好的学生, 于是可以构建概率图模型, 如图 1 所示。

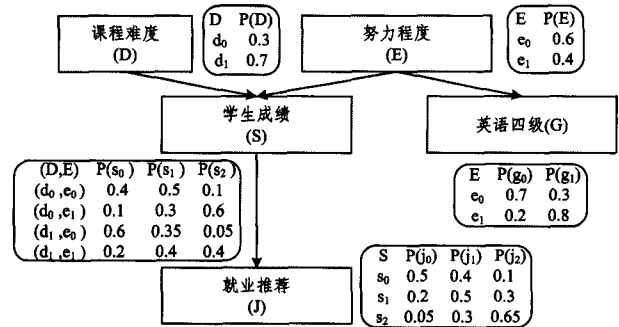


图 1 关于教学评估的概率图模型

在上述概率图模型中, 对河南师范大学计算机学院计算机科学与技术专业 2012 级的两个班级 172 名学生进行调查统计。首先让这些学生对当前的课程难度和自己平时的努力程度进行评估, 得到事件先验概率分布; 然后通过对期终考试的成绩进行样本统计, 将样本空间划分为优、良、差 3 个区域, 规定成绩大于或等于 85 分为优秀, 大于或等于 60 分但小于 85 分为良好, 小于 60 分为差。现在对 2014 年 7 月份该年级全体同学的期末考试成绩进行统计, 结果如表 2 所列。

表 2 期末考试成绩分布(单位: 人)

人数	成绩分布		
	优秀	良好	差
计科 1 班	32	46	8
计科 2 班	26	55	5

对 2014 年 6 月份参加四级考试的同学进行统计,根据考试结果对学生的努力程度进行评估,最终确定参加四级考试学生共 136 人,其中计科 1 班 71 人,计科 2 班 65 人,最终成绩在 425 以上的学生有 89 人,其人员分布情况如表 3 所列。

表 3 四级通过情况分布(单位:人)

人数	四级通过情况	
	通过	未通过
计科 1 班	52	19
计科 2 班	37	28

可以得到样本的成绩分布函数,用大写字母  $D, E, S, G$  和  $J$  分别表示节点所选课的“课程难度”、学生平时学习的“努力程度”、“学生成绩”、“四级”考试是否通过,以及毕业时学校“就业推荐”工作岗位的差别。 $d_0, d_1$  分别表示课程难度“难”和“容易”两种情况; $e_0, e_1$  分别表述学生在学习过程中“努力”和“不努力”两种情况;对学生的成绩同样进行划分, $s_0, s_1, s_2$  对应成绩“优秀”、“良好”和“差”;而四级考试的结果只有“通过”和“没有通过”,对应的用  $g_0$  和  $g_1$  表示。学校在就业推荐时,通常有 3 种决策:好的职位推荐给成绩好的学生、一般的岗位推荐给成绩一般的学生、对于表现特别差的学生则不予推荐,分别对应着  $j_0, j_1$  和  $j_2$ 。以上面的概率图模型为例,通过收集毕业生的信息,在知道学生期末学习成绩、四级成绩以及就业推荐情况下,推断教学质量的优劣。由于数据量比较大,先建立数学模型对整体进行评估,用三支决策的规则进行分类:

定义  $L(x)$  表示条件  $x$  下的学生总人数,如  $L(j_0)$  表示就业推荐岗位比较好的总人数。

(P') 当  $L(j_0) > L(j_1)$  且  $L(j_0) > L(j_2)$  时,判定整体就业推荐情况“很好”,用  $j_0$  表示;

(B') 当  $L(j_1) > L(j_0)$  且  $L(j_1) > L(j_2)$  时,判定整体就业推荐情况“一般”,用  $j_1$  表示;

(N') 当  $L(j_2) > L(j_0)$  且  $L(j_2) > L(j_1)$  时,判定整体就业推荐情况“较差”,用  $j_2$  表示。

对于学生成绩,当平均成绩大于或等于 80 时划分为“优秀”,大于或等于 60 但小于 80 时划分为“良好”,平均分小于 60 时视为“差”。同样利用上述规则, $L(s_0) > L(s_1)$  且  $L(s_0) > L(s_2)$  时,整体视为“优秀”; $L(s_1) > L(s_0)$  且  $L(s_1) > L(s_2)$  时,整体视为“良好”;当  $L(s_2) > L(s_1)$  且  $L(s_2) > L(s_0)$  时,整体视为“差”。

在概率图模型的推理结果中,对推理结果的两个状态集采取 3 种不同的决策动作会带来不同的代价损失,推理结果的三支决策代价函数如表 4 所列。

表 4 推理结果的三支决策代价函数

决策动作	课程难度	
	困难	容易
降低难度	提高整体学习水平产生的代价较小	不利于能力的提高,对教学方案造成误导,会产生较大的代价
保持不变	提高少数人的能力,打击多数人学习积极性,产生的代价一般	普遍提高学生成绩,不利于能力的提高,产生的代价一般
提高难度	打击学生学习积极性,会产生较大的代价	有助于提高学习积极性,产生的代价较小

在上述的概率图模型中,得到数据  $U = \{j_1, g_0\}$ ,在该条件下评估该院计算机科学与技术专业 2012 级“课程难度”的概率  $P(D|j_1, g_0)$ 。则由精确推理可得:

$$P(D|j_1, g_0) = \theta P(D, j_1, g_0) = \theta \sum_E \sum_S P(D, E, S, j_1, g_0)$$

其中,  $\theta$  表示归一化常数,确保条件概率总和为 1;已知  $D$  有两种取值即  $d_0$  和  $d_1$ ,当  $D$  取不同值时,分别用 Bayes 网络的概率分布公式进行计算。

当  $D$  为  $d_0$  时,

$$\begin{aligned} P(d_0 | j_1, g_0) &= \theta \sum_E \sum_S P(d_0, E, S, j_1, g_0) \\ &= \theta \sum_E \sum_S P(d_0) P(E) P(S | d_0, E) P(j_1 | S) P(g_0 | E) \\ &= \theta P(d_0) \sum_E P(E) P(g_0 | E) \sum_S P(S | d_0, E) P(j_1 | S) \\ &= \theta P(d_0) \{ P(e_0) P(g_0 | e_0) [ P(s_0 | d_0, e_0) P(j_1 | s_0) + P(s_1 | d_0, e_0) P(j_1 | s_1) + P(s_2 | d_0, e_0) P(j_1 | s_2) ] + P(e_1) P(g_0 | e_1) [ P(s_0 | d_0, e_1) P(j_1 | s_0) + P(s_1 | d_0, e_1) P(j_1 | s_1) + P(s_2 | d_0, e_1) P(j_1 | s_2) ] \} \end{aligned}$$

将图 1 概率图模型中的概率分布函数代入上式可得:

$$\begin{aligned} P(d_0 | j_1, g_0) &= \theta \times 0.3 \{ 0.6 \times 0.7 [ 0.4 \times 0.4 + 0.5 \times 0.5 + 0.1 \times 0.3 ] + 0.4 \times 0.2 [ 0.1 \times 0.4 + 0.3 \times 0.5 + 0.6 \times 0.3 ] \} \\ &= \theta \times 0.3 \times ( 0.6 \times 0.7 \times 0.44 + 0.4 \times 0.2 \times 0.37 ) \\ &= \theta \times 0.3 \times 0.2144 \\ &= 0.06432\theta \end{aligned}$$

当  $D$  为  $d_1$  时,

$$\begin{aligned} P(d_1 | j_1, g_0) &= \theta \sum_E \sum_S P(d_1, E, S, j_1, g_0) \\ &= \theta \sum_E \sum_S P(d_1) P(E) P(S | d_1, E) P(j_1 | S) P(g_0 | E) \\ &= \theta P(d_1) \sum_E P(E) P(g_0 | E) \sum_S P(S | d_1, E) P(j_1 | S) \\ &= \theta P(d_1) \{ P(e_0) P(g_0 | e_0) [ P(s_0 | d_1, e_0) P(j_1 | s_0) + P(s_1 | d_1, e_0) P(j_1 | s_1) + P(s_2 | d_1, e_0) P(j_1 | s_2) ] + P(e_1) P(g_0 | e_1) [ P(s_0 | d_1, e_1) P(j_1 | s_0) + P(s_1 | d_1, e_1) P(j_1 | s_1) + P(s_2 | d_1, e_1) P(j_1 | s_2) ] \} \\ &= 0.7\theta \{ 0.6 \times 0.7 [ 0.6 \times 0.4 + 0.35 \times 0.5 + 0.05 \times 0.3 ] + 0.4 \times 0.2 [ 0.2 \times 0.4 + 0.4 \times 0.5 + 0.4 \times 0.3 ] \} \\ &= 0.7 \times \theta ( 0.6 \times 0.7 \times 0.43 + 0.4 \times 0.2 \times 0.4 ) \\ &= 0.7 \times \theta \times 0.2126 \\ &= 0.14882\theta \end{aligned}$$

最后对结果归一化处理,取  $\theta = 1 / ( 0.06432 + 0.14882 )$  得:

$$P(D|j_1, g_0) = \theta ( 0.06432, 0.14882 ) = ( 0.3018, 0.6982 )$$

因此,在已知的统计数据下,可以推断出现有课程“难”的概率是 0.3018,“容易”的概率是 0.6982。

在表 1 所列代价函数的基础上,构建一组带数学模型的三支决策代价函数。设状态集  $\Omega = \{X, X^c\}$  分别表示课程的难易,决策集  $A = \{a_P, a_B, a_N\}$  分别表示接受、延迟和拒绝决策, $\lambda_{PP}, \lambda_{BP}$  和  $\lambda_{NP}$  分别对应  $x$  属于  $X$  (困难)时采取动作  $a_P, a_B, a_N$  带来的代价损失; $\lambda_{PN}, \lambda_{BN}$  和  $\lambda_{NN}$  则对应  $x$  属于  $X$  补集 (容易)时采取动作  $a_P, a_B, a_N$  带来的代价损失。根据实际情况,令各损失函数如表 5 所列。

表 5 不同决策带来的代价函数

决策动作	代价损失	
	X	X <sup>c</sup>
$a_P$	$\lambda_{PP} = 0.1$	$\lambda_{PN} = 0.7$
$a_B$	$\lambda_{BP} = 0.3$	$\lambda_{BN} = 0.4$
$a_N$	$\lambda_{NP} = 0.8$	$\lambda_{NN} = 0.15$

代入阈值计算公式:

$$\alpha = \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \beta = \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}$$

得:

$$\alpha = \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})} = \frac{0.7 - 0.4}{(0.7 - 0.4) + (0.3 - 0.1)} = 0.6$$

$$\beta = \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})} = \frac{0.4 - 0.15}{(0.4 - 0.15) + (0.8 - 0.1)} = 0.263$$

根据三支决策规则,可以得到以下决策:

- (P) 若  $P(X|[x]) \geq \alpha$ , 则  $x \in POS(X)$ ;
- (B) 若  $\beta < P(X|[x]) < \alpha$ , 则  $x \in BND(X)$ ;
- (N) 若  $P(X|[x]) \leq \beta$ , 则  $x \in NEG(X)$ 。

结合表 5 中的代价函数分析,现将针对课程难易程度  $d_0$  和  $d_1$  两种状态分别采取 3 种决策动作(即肯定决策、拒绝决策和延迟决策)进行分析。又因  $d_0$  和  $d_1$  概率之和为 1,故  $d_0$  和  $d_1$  的概率不可能同时小于  $\beta$ ,也不可能同时大于  $\alpha$ ,即  $d_0$  和  $d_1$  的决策动作不可能同时为肯定决策和拒绝决策。若  $d_0$  划分到负域中,无论  $d_1$  划分到正域中还是边界域中,采取“提高难度”这一决策所带来的代价损失最小;同理,若  $d_1$  划分到负域中,无论  $d_0$  划分到正域中还是边界域中,采取“降低难度”这一决策所带来的代价损失最小;若  $d_0$  和  $d_1$  同时划分到边界域中,采取“保持不变”这一决策是最佳的。最终所有可能的决策结果如表 6 所列。

表 6 最终所有可能的决策结果

决策结果	课程容易程度( $d_1$ )		
	$a_P$	$a_B$	$a_N$
课程困难程度( $d_0$ )	$a_P$	—	降低难度
	$a_B$	提高难度	保持不变
	$a_N$	提高难度	—

由三支决策规则得: $d_0$  划分到边界域,而  $d_1$  划分到正域中。由表 6 知,此时最佳决策结果应为“提高难度”。故推理结果是:现有课程偏容易,在以后的教学安排中,需要适当提高难度,以便于提高整体教学水平。

**结束语** 本文以概率图模型为基础,通过 Bayes 推理方法,结合三支决策基本理论对代价决策问题进行研究,为解决现实问题提供了理论依据。在推理阶段,先从事件的因果关系中构建概率图模型,然后计算出推理结果发生的条件概率,再从代价分析的角度对三支决策理论进行扩展应用。在教学评估的问题中,根据不同的观测条件,对现实问题进行推理分析,最终做出最佳的决策方案。本研究基于以代价函数为确定值的三支决策模型,而在一些实际应用中,可能无法得到一组确定的代价函数,所以代价函数为模糊值情况下的应用研究将是我们下一步要研究和学习的要点。

### 参考文献

[1] Li Hua-xiong, Liu Dun, Zhou Xian-zhong. Review on decision-theoretic rough set model[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition),

2010,22(5):624-630(in Chinese)

李华雄,刘盾,周献中. 决策粗糙集模型研究综述[J]. 重庆邮电大学学报(自然科学版),2010,22(5):624-630

[2] Yao Yi-yu. Three-way decision: an interpretation of rule in rough set theory[M]//Rough Sets and Knowledge Technology: Lecture Notes in Computer Science,2009:642-649

[3] Liu Dun, Li Tian-rui, Miao Duo-qian, et al. Three-way decisions and granular computing [M]. Beijing: Science Press, 2013 (in Chinese)

刘盾,李天瑞,苗夺谦,等. 三支决策与粒计算[M]. 北京:科学出版社,2013

[4] Du Li-na, Xu Jiu-cheng, Liu Yang-yang, et al. Research on the evaluation of venture investment based on the risk minimization of three-way decision[J]. Journal of Shandong University(Natural Science),2014,49(8):66-72(in Chinese)

杜丽娜,徐久成,刘洋洋,等. 基于三支决策风险最小化的风险投资评估应用研究[J]. 山东大学学报(理学版),2014,49(8):66-72

[5] Zhang Li-bo, Li Hua-xiong, Zhou Xian-zhong, et al. Multi-granularity cost-sensitive three-way decision for face recognition[J]. Journal of Shandong University (Natural Science),2014(8):48-57(in Chinese)

张里博,李华雄,周献中,等. 人脸识别中的多粒度代价敏感三支决策[J]. 山东大学学报(理学版),2014(8):48-57

[6] Zhao Yue. Probability graph model learning theory and its application[M]. Beijing: Tsinghua University Press, 2012 (in Chinese)

赵悦. 概率图模型学习理论及其应用[M]. 北京:清华大学出版社,2012

[7] Xu Jin, Yue Kun, Qian Wen-hua, et al. A probabilistic-graphical-model based approach for quality evaluation of graduate student pools[J]. Journal of Yunnan University (Natural Sciences Edition),2011,33(S2):299-304(in Chinese)

徐瑾,岳昆,钱文华,等. 一种基于概率图模型的研究研究生源质量评价方法[J]. 云南大学学报(自然科学版),2011,33(S2):299-304

[8] Wang Yi-nan. Sitaiton assessment using Probabilistic Graphical Models[D]. Beijing: Beijing Institute of Technology, 2011 (in Chinese)

王忆南. 基于概率图模型的态势估计[D]. 北京:北京理工大学,2011

[9] Miao Duo-qian, Wang Yu. On the relationships between information entropy and roughness of knowledge rough set theory [J]. Pattern Recognition and Artificial Intelligence, 1998, 11 (1):34-40(in Chinese)

苗夺谦,王珏. 粗糙集理论中知识粗糙性与信息熵关系的讨论[J]. 模式识别与人工智能,1998,11(1):34-40

[10] Wang Wan-sen. The principle of artificial intelligence and its application[M]. Beijing: Electronic Industry Press, 2000 (in Chinese)

王万森. 人工智能原理及其应用[M]. 北京:电子工业出版社,2000

[11] Cheng Qiang, Chen Feng, Dong Jian-wu, et al. Variational Approximate Inference Methods for Graphical Models[J]. Acta

- Automatica Sinica, 2012, 38(11): 1721-1734 (in Chinese)
- 程强, 陈峰, 董建武, 等. 概率图模型中的变分近似推理方法[J]. 自动化学报, 2012, 38(11): 1721-1734
- [12] Xiao Qin-kun, Gao Song. Application of Bayesian Network in Intelligent Information Processing[M]. Beijing: National Defense Industry Press, 2012 (in Chinese)
- 肖秦琨, 高嵩. 贝叶斯网络在智能信息处理中的应用[M]. 北京: 国防工业出版社, 2012
- [13] Liu Dun, Li Tian-rui, Liang De-cui. Based on rough set three policy decision analysis[M]//Jia Xiu-yi, Shang Lin, Zhou Xian-zhong, et al. Three-way Decision theory and application. Nanjing: Nanjing University Press, 2012; 49-60 (in Chinese)
- 刘盾, 李天瑞, 梁德翠. 基于决策粗糙集的三支政策决策分析[M]//贾修一, 商琳, 周献中, 等. 三支决策理论与应用. 南京: 南京大学出版社, 2012; 49-60
- [14] Jia Xiu-yi, Shang Lin, Zhou Xian-zhong, et al. Three-way Decision theory and application [M]. Nanjing: Nanjing University Press, 2012 (in Chinese)
- 贾修一, 商琳, 周献中, 等. 三支决策理论与应用[M]. 南京: 南京大学出版社, 2012
- 
- (上接第 13 页)
- [14] Gonzalez G C, Sharma P N, Galletta D F. The antecedents of the use of continuous auditing in the internal auditing context [J]. International Journal of Accounting Information Systems, 2012, 13(3): 248-262
- [15] Rutgers Accounting Web[OL]. 2014. <http://raw.rutgers.edu>
- [16] Overpeck J T, Meehl G A, Beny S, et al. Dealing with data [J]. Science, 2011, 331(6018): 639-806
- [17] Gartner E S. 10 Critical Tech Trends for the Next Five Years [EB/OL]. <http://www.forbes.com/sites/ericavitz/2012/10/22/gartner-10-critical-tech-trends-for-the-next-five-years>
- [18] Gong Xue-qing, Jin Che-qing, Wang Xiao-ling, et al. Data-Intensive Science and Engineering: Requirements and Challenges[J]. Chinese Journal of Computers, 2012, 35(8): 1563-1578 (in Chinese)
- 宫学庆, 金澈清, 王晓玲, 等. 数据密集型科学与工程: 需求和挑战[J]. 计算机学报, 2012, 35(8): 1563-1578
- [19] Chen C L P, Zhang C Y. Data-intensive applications, challenges, techniques and technologies A survey on Big Data[J]. Information Sciences, 2014, 275: 314-347
- [20] Manyika J, Chui M, Brown B, et al. Big data: The Next Frontier for Innovation, Competition, and Productivity [R]. McKinsey Global Institute, 2011
- [21] Kelly J. Apache drill brings sql-like, ad hoc query capabilities to big data [EB/OL]. <http://wikibon.org/wiki/v/Apache-Drill-Brings-SQL-Like-Ad-Hoc-Query-Capabilities-to-Big-Data>
- [22] Melnik S, Gubarev A, Long J J, et al. Dremel; interactive analysis of webscale datasets[J]. Proceeding of the 36th International Conference on Very Large Data Bases, 2010, 3(1): 330-339
- [23] Gulisano V, Ricardo J P, Marta P M, et al. Streamcloud: an elastic and scalable data streaming system[J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(12): 2351-2365
- [24] Sqlstream [OL]. <http://www.sqlstream.com/products/server>
- [25] Bell G, Hey T, Szalay A. Beyond the data deluge[J]. Science, 2009, 323 (5919): 1297-1298
- [26] Hey T, Tansley S, Tolle K. The fourth paradigm: data-intensive scientific discovery[R]. Microsoft Research, 2009
- [27] James P A, Bruce H, Gabrielle L, et al. Data-intensive science in the us doe: case studies and future challenges[J]. Computing in Science and Engineering, 2011, 13(6): 14-24
- [28] Lynch C. Big data: how do your data grow? [J]. Nature, 2008, 455(7209): 28-29
- [29] Divyakant A, Philip B, Elisa B, et al. Challenges and Opportunities with Big Data[R]. Cyber Center Technical Reports, Purdue University, 2011
- [30] Deam J, Ghemawat S. Mapreduce: simplified data processing on large clusters[J]. Communications of the ACM, 2008, 51(1): 107-113
- [31] Tabealu [OL]. <http://www.tableausoftware.com>, 2014
- [32] CCF 大数据专家委员会. 中国大数据技术与产业发展白皮书 [R]. <http://www.ccf.org.cn>, 2013
- [33] Chen Wei, Zhang Jin-cheng, Qiu R. A Survey on Computer-assisted Audit Techniques(CAATs)[J]. Computer Science, 2007, 34(10): 290-294 (in Chinese)
- 陈伟, 张金城, Qiu R. 计算机辅助审计技术(CAATs)研究综述[J]. 计算机科学, 2007, 34(10): 290-294
- [34] Simeon S, Michael H B, Arturas M. Visual Data Mining: Theory, Techniques and Tools for Visual Analytics [M]. Springer, 2008
- [35] Geng B, Li Y, Tao D C, et al. Parallel lasso for large-scale video concept detection[J]. IEEE Transactions on Multimedia, 2012, 14(1): 55-65
- [36] Heer J, Mackinlay J D, Stolte C, et al. Graphical histories for visualization: supporting analysis, communication, and evaluation [J]. IEEE Transactions on Visualization and Computer Graphics, 2008, 14(6): 1189-1196
- [37] Thompson D, Levine J A, Bennett J C, et al. Analysis of large-scale scalar data using hixels[C]//IEEE Symposium on Large Data Analysis and Visualization (LDAV). 2011: 23-30
- [38] Chen Wei, Smieliauskas W. Study on Online Auditing Methods in Cloud Computing Environments [J]. Audit Research, 2012 (3): 37-44 (in Chinese)
- 陈伟, Smieliauskas W. 云计算环境下的联网审计实现方法探析[J]. 审计研究, 2012(3): 37-44
- [39] Armbrust M, Fox A, Griffith R, et al. A view of cloud computing [J]. Communications of the ACM, 2010, 53(4): 50-58
- [40] Sakr S, Liu A, Batista D M, et al. A survey of large scale data management approaches in cloud environments[J]. IEEE Communications Surveys & Tutorials, 2011, 13(3): 311-336
- [41] Lee G. Using in-memory analytics to quickly crunch big data [J]. IEEE Computer Society, 2012, 45(10): 16-18
- [42] Nathan M, James W. Big data: principles and best practices of scalable realtime data systems[M]. Manning, 2012