基于粒关联规则的冷启动推荐方法

巫文佳 何 旭
(闽南师范大学粒计算及其应用重点实验室 漳州 363000)

摘 要 推荐系统已被广泛应用于电子商务等多个领域。冷启动问题是推荐系统的一个难点。基于粒关联规则的冷启动推荐方法,运用粒来描述用户和产品,通过满足粒关联规则的4个指标,挖掘出用户和产品之间的关联规则,匹配合适的规则,最后根据这些规则向用户做出相应的推荐。在公开有效的数据集MovieLens上进行了实验,结果表明,用粒关联规则所挖掘出的规则可以有效地用于训练集和测试集上的推荐,并且具有较好的准确性。

关键词 粒计算,关联规则,推荐系统,冷启动问题,数据挖掘
中图法分类号 TP18 **文献标识码** A

Cold-start Recommendation Based on Granular Association Rules

WU Wen-jia HE Xu
(Laboratory of Granular Computing, Minnan Normal University, Zhangzhou 363000, China)

Abstract Recommendation systems have been widely used in many fields such as e-commerce. The cold-start problem is one of difficulties on recommendation systems. This paper designed a cold-start recommendation approach based on granular association rules. First, we used granules to describe users and items. Then we generated rules between users and items through satisfying four measures of granular association rules. Finally, we matched the suitable rules to recommend items to users. Experiments were undertaken on a publicly available dataset MovieLens. Results show that granular association mining rule can be used for the recommendation on training and testing sets effectively and accurately.

Keywords Granular computing, Association rule, Recommendation system, Cold-start problem, Data mining

1 引言

推荐系统是向用户推荐感兴趣的产品,随着数据量的逐渐增加,它已成为电子商务的一个重要组成部分。目前,越来越多的推荐方法被提出,主要有以下几种类型:基于内容(Content-based)推荐[1,2],即知道用户和产品的属性信息以及用户选择产品的历史数据,向当前用户推荐其过去比较偏好的产品;协同过滤(Collaborative Filtering)推荐[3],即结合当前用户和其他兴趣相仿用户所选择产品的历史数据,根据总体的偏好程度向用户推荐产品;混合方法(Hybrid-based)推荐[4,5],即综合了基于内容推荐和协同过滤推荐两种方法的优点,向用户做出推荐。

冷启动问题[6]是推荐系统的一个难点。主要分为3类:第一类是对新用户推荐已有的产品;第二类是对老用户推荐新的产品;第三类是对新用户推荐新的产品。第一类和第二类冷启动问题是对称的,所以可以用类似的方法来解决,目前已经采用了许多不同的方法,例如上述提到的基于内容推荐和协同过滤推荐等方法。然而,第三类冷启动问题却很少被考虑,因为新用户选择产品和新产品被选择的历史数据都是未知的,研究这种情况下的问题更具挑战性。

粒关联规则[7,8]是一种新的多关系数据挖掘方法,用于寻找隐藏在多关系数据表中的模式。目前,粒度关联规则已经开展了一些工作,比如粒关联规则概念的提出[7,8]、基于粗糙集设计的一种更有效挖掘规则的算法[9]、对数值型数据的离散化处理研究[10]以及属性多值问题[11]的研究。

本文提出了一种基于粒关联规则的冷启动推荐方法来处理上述3类冷启动问题,特别是对新用户推荐新产品。首先,通过信息粒来描述用户和产品。例如,"男生"、"喜剧电影"和"2013年上映的冒险电影",这些都是信息粒。然后建立多对多实体关系系统,通过信息粒来挖掘用户和产品之间的关联规则。例如,在用户与电影构成的多对多实体关系系统中,可以挖掘到下面这条关联规则:"60%的男生观看了40%上映于2013年的动作电影;用户信息表中20%的对象是男生,电影信息表中5%的对象是上映于2013年的动作电影。"这里20%、5%、60%和40%分别是源覆盖、目标覆盖、源置信度和目标置信度,它们是粒关联规则的4个度量指标。通过恰当地设定这4个指标的阈值,可以得到一些有意义的关联规则。接着,针对不同用户进行合适的规则匹配。最后,根据匹配的

到稿日期:2013-05-15 返修日期:2013-07-28 本文受国家自然科学基金面上项目(61170128),福建省自然科学基金项目(2012J01294),福建省自然科学基金省属高校专项(JK2012028),福建省计算机应用技术和信号与信息系统研究生教育创新基地(闽高教[2008]114号)资助。
巫文佳(1980—),男,硕士生,讲师,主要研究方向为体育数据挖掘,E-mail:wuwenjiayo@163.com;何 旭(1990—),男,硕士生,主要研究方向为粒计算与粗糙集。

规则向用户做出相应的推荐，并通过计算推荐准确率来评价该推荐系统的性能。

本文是粒关联规则的一个很好的应用方向，同时也进一步推广了粒计算的相关研究工作[12-15]。我们设计了一个公开的软件 Grale[16] 进行实验，在 MovieLens 数据集上的实验结果表明，运用粒关联规则方法挖掘到的规则在训练集和测试集上表现相似，都很稳定，通过合适的阈值设置，该方法可以向用户做出良好的推荐。

2 基本概念

本节介绍一些基本概念，包括信息系统及粒的相关概念、多对多实体关系系统以及粒关联规则等方面的内容。

2.1 信息系统及粒的相关概念

定义 1[7] 信息系统 $S=(U, A)$，其中，$U=\{x_1, x_2, \cdots, x_n\}$ 表示所有的对象集合，$A=\{a_1, a_2, \cdots, a_m\}$ 表示所有属性集合，$a_j(x_i)$ 表示对象 x_i 在属性 a_j 下的值（$i=1,2,\cdots,n$ 以及 $j=1,2,\cdots,m$）。

在信息系统中，任意一个属性集合的子集 $A'\subseteq A$ 都满足下面的等价关系[17,18]：

$$E_{A'}=\{(x, y)\in U\times U \mid \forall a\in A', a(x)=a(y)\} \tag{1}$$

将对象集合 U 划分成若干个子集合，我们称之为块或者粒。每个粒 x 满足：

$$E_{A'}(x)=\{y\in U \mid \forall a\in A', a(y)=a(x)\} \tag{2}$$

定义 2[14] 粒是一个三元组

$$G=(g, i(g), e(g)) \tag{3}$$

其中，粒标记为 g，该粒的特征记为 $i(g)$，即粒的内涵，该粒所对应的对象集合记为 $e(g)$，即粒的外延。

在信息系统中，(A', x) 决定一个粒，$i(g)$ 表示该粒所包含的各属性值对的结合，即：

$$i(g(A', x))=\bigwedge_{a\in A'}\langle a:a(x)\rangle \tag{4}$$

$e(g)$ 表示粒所对应的对象集合，所表达的含义和式(2)相同，但是可以被更加形象地刻画，故有：

$$e(g(A', x))=E_{A'}(x) \tag{5}$$

粒的置信度是 $e(g)$ 所含的对象个数和全部的对象个数的比值，定义式如下：

$$supp(g(A', x))=supp\left(\bigwedge_{a\in A'}\langle a:a(x)\rangle\right)=supp(E_{A'}(x))=\frac{|ET(A', x)|}{|U|} \tag{6}$$

2.2 多对多实体关系系统

定义 3[7] 两个对象集合 $U=\{x_1, x_2, \cdots, x_n\}$ 和 $V=\{y_1, y_2, \cdots, y_k\}$，任意的 $R\subseteq U\times V$，且 R 是 U 到 V 的二元关系，$x\in U$ 的领域满足：

$$R(x)=\{y\in V \mid (x, y)\in R\} \tag{7}$$

当 $U=V$ 时，R 就表示等价关系，$R(x)$ 就表示对象 x 的等价类。同样地，我们也可以得到 $y\in V$ 的领域满足：

$$R^{-1}(x)=\{x\in U \mid (x, y)\in R\} \tag{8}$$

定义 4[7] 多对多实体关系系统（MMER）包含 5 个元素，$ES=(U, A, V, B, R)$，其中，U 和 V 是对象集合，A 和 B 是属性集合，(U, A) 和 (V, B) 构成两个信息系统，$R\subseteq U\times V$ 是 U 到 V 的二元关系。

由定义可知，构建多对多实体关系系统需要两个信息系统和一个关系表，如表 1—表 3 所列，其中表 1 和表 2 分别表示用户和电影的信息，表 3 是一个关系表，即用户评价了哪些电影。

表 1 用户信息

用户编号	年龄	性别	职业
1	[22,27)	男	教师
2	[48,55)	女	其他
3	[22,27)	男	作家
…	…	…	…
943	[22,27)	男	学生

表 2 电影信息

电影编号	上映年份	动作类	冒险类	…	欧美类
1	[1994,1995)	否	否	…	否
2	[1994,1995)	否	是	…	否
3	[1994,1995)	否	否	…	否
…	…	…	…	…	
1682	(1922,1980]	否	否	…	否

表 3 评价记录

用户编号\电影编号	1	2	3	…	1682
1	否	是	否	…	否
2	是	否	否	…	是
3	否	否	否	…	是
…	…	…	…	…	…
943	否	否	是		是

2.3 粒关联规则

现在我们讨论联系用户和产品关系的方法。文献[7]提出了粒关联规则的基本形式：

$$\bigwedge_{a\in A'}\langle a:a(x)\rangle\Rightarrow\bigwedge_{b\in B'}\langle b:b(y)\rangle \tag{9}$$

其中，$A'\subseteq A, B'\subseteq B$。

在介绍评价指标之前，我们观察这样一条粒关联规则："60%的男生观看了 40%上映于 2013 年的动作电影；用户信息表中 20%的对象是男生，电影信息表中 5%的对象是上映于 2013 年的动作电影。"这里 20%、5%、60%和 40%分别是源覆盖、目标覆盖、源置信度和目标置信度，下面分别给出它们的定义。

首先根据式(4)、式(5)和式(9)，给出对象集所满足粒关联规则的左手边和右手边的定义：

$$LH(GR)=E_{A'}(x) \tag{10}$$

$$RH(GR)=E_{B'}(y) \tag{11}$$

接着，给出粒关联规则的 4 个度量指标的定义式，即源覆盖、目标覆盖、源置信度和目标置信度。

源覆盖定义式如下：

$$scov(GR)=\frac{|LH(GR)|}{|U|} \tag{12}$$

目标覆盖定义式如下：

$$tcov(GR)=\frac{|RH(GR)|}{|V|} \tag{13}$$

因为源置信度和目标置信度是相互影响和相互制约的，所以二者的值都不能单独地通过一个式子来得到，计算任何一个，都需要确定另一个的阈值。在这里，我们设定了目标置信度的阈值 tc，则粒关联规则的源置信度可以定义如下：

$$sconf(GR, tc)=\frac{|\{x\in LH(GR) \mid \frac{|R(x)\cap RH(GR)|}{|RH(GR)|}\geqslant tc\}|}{LH(GR)} \tag{14}$$

实际上，源覆盖和目标覆盖表明所挖掘规则的一般性，即这部分规则占所有规则的比重，而源置信度和目标置信度表示的是所挖掘规则的强度。

3 基于粒关联规则的冷启动推荐方法

冷启动问题[5]是推荐系统的一个难点，主要分为3类：第一类是对新用户推荐已有的产品；第二类是对老用户推荐新的产品；第三类是对新用户推荐新的产品。第一类和第二类是对称的，所以可以用相同的方法来解决，目前已有一些比较经典的方法，例如基于内容推荐[1,2]和协同过滤推荐[3]等方法。而对于第三类，由于所能掌握的信息比较少，因此目前研究很少，显然它具有一定的挑战性。

基于粒关联规则的冷启动推荐方法不仅可以用来处理只有新用户或者新产品的冷启动问题，也可以用来解决同时存在新用户和新产品的情况。通过粒关联规则挖掘到的许多规则集，我们需要在一组相对小的规则集上进行测试，现在有一个关键的问题：在规则集上，哪些规则可以被用于推荐？其解决方法就是首先设置合适的4个指标阈值，然后挖掘出所有满足这些阈值的规则用于推荐。

问题1 粒关联规则挖掘问题

输入：一个多对多实体关系系统 $ES=(U,A,V,B,R)$，最小源覆盖阈值 ms，最小目标覆盖阈值 mt，最小源置信度阈值 sc 和最小目标置信度阈值 tc。

输出：一组规则集，其中每条规则都满足：源覆盖值$\geq ms$，目标覆盖值$\geq mt$，源置信度$\geq sc$，目标置信度$\geq tc$。

问题1通过算法1来实现，算法首先挖掘出两个信息系统(U,A)和(U,B)中满足源覆盖值$\geq ms$、目标覆盖值$\geq mt$的粒的集合，然后由二元关系 R 得到候补规则，最后输出满足源置信度$\geq sc$、目标置信度$\geq tc$的规则集。由于整个过程从式(9)的两端开始，再向中间运行，因此该算法称为Sandwich算法。在文献[7]中还做了相应的改进，提出了Forward算法和Backward算法，进一步提高了运算的效率。在本文，研究的是粒关联规则产生推荐的准确性，所以在此不对改进的算法做深入的探讨。

通过算法1获得了一组粒关联规则集。针对每个用户，我们通过这个规则集向他们推荐感兴趣的产品。由于只向用户推荐相匹配的规则，因此一些用户可能会有许多推荐，而另一些可能会很少。该推荐系统的性能主要是由推荐的准确性来评价的。假设向用户推荐的产品个数为 M 个，其中成功的推荐个数为 N 个，则推荐的准确率为 N/M。

算法1 粒度关联规则挖掘的Sandwich算法

输入：ES=(U, A,V,B, R),ms,mt, sc, tc

输出：一组规则集

方法：Sandwich算法

1. SG(ms)={(A′,x)∈2^A×U||$E_{A'}$(x)|/|U|≥ms};
   //信息系统(U, A)中满足最小源覆盖阈值 ms 的粒的集合。
2. TG(mt)={(B′,y)∈2^B×V||$E_{B'}$(x)|/|V|≥ms};
   //信息系统(U, B)中满足最小源覆盖阈值 mt 的粒的集合。
3. for each g∈SG(ms) do
4. for each g′∈TG(mt) do
5. GR=(i(g)=>i(g′));
6. if sconf(GR,tc)≥sc then
7. add GR to the rule set;
8. end if
9. end for
10. end for

在本文实验中，我们把用户选择过的产品按一定比例随机划分为训练集和测试集，通过训练集的数据预测用户的喜好，再利用测试集对预测结果进行评估。我们比较以下4组训练集和测试集的方案：

(1)随机推荐，即向不同用户随机推荐产品。在这里不需要训练的步骤，随机推荐的结果可以用于初步判断推荐的好坏。如果某个推荐方法的准确率比随机推荐的准确率还低，那么这种方法就失去研究的价值。

(2)将用户划分成训练集和测试集。这种方案适用于处理第一类冷启动问题，即对新用户推荐已有的产品。

(3)将产品划分成训练集和测试集。这种方案适用于处理第二类冷启动问题，即对老用户推荐新的产品。

(4)将用户和产品同时划分成训练集和测试集。这种方案适用于处理第三类冷启动问题，即对新用户推荐新的产品。这是推荐系统的一个难点，也是本文研究的重点。

4 实验

在本节，我们进行多组实验，主要回答了以下几个问题：

(1)运用基于粒关联规则的冷启动推荐方法，处理3类不同的冷启动问题，哪种的推荐性能最好？

(2)在保持随机划分比例不变的情况下，随着指标阈值的变化，训练集和测试集推荐的准确率是如何变化的？

(3)在保持指标阈值不变的情况下，随着随机划分比例变化，训练集和测试集推荐的准确率是如何变化的？

4.1 数据集

我们的实验在MovieLens数据集上进行，它来自于一个大型的电影推荐系统网站(www. movielens. org)，目前已被广泛应用于推荐系统。该网站由美国明尼苏达大学计算机科学与工程学院的GroupLens产品组创建，是一个非赢利的、以科学研究为目的的实验性站点。我们在MovieLens数据集上运用粒关联规则挖掘的技术，向用户推荐他们感兴趣的电影。数据表及包含的属性如下：

- 用户(用户编号，年龄，性别，职业)
- 电影(电影编号，上映年份，电影类型)
- 评分关系(用户编号，电影编号)

我们采用的数据集包含了943个用户对1682部电影的100000个评分记录，每个用户至少评价了20部电影。原来的关系表包括用户对电影的评分(评分值：1～5分)，但是MMER系统考虑两个信息系统间的二元关系，即用户是否对某部电影进行评分，所以暂时不考虑评分体制。由于用户的年龄分布不均匀，同时只有少数电影上映于1980年之前，绝大多数的电影上映于1990年之后，因此为了能够挖掘到更多更好的规则，这里先对数据预处理，即通过对数值型数据的离散化处理[10]，将用户的年龄划分为5个区间：[7,22)，[22,27)，[27,31)，[31,39)，[39,48)和[48,73]；将电影的上映年份划分为5个区间：[1922,1980)，[1980,1993)，[1993,1994)，[1994,1995)，[1995,1996)，[1996,1997)和[1997,

1998]。因为电影的类型是多值属性，所以我们分别列出 18 种不同的布尔属性，然后通过多值属性方法[11]进行处理。

4.2 实验结果

我们通过以下设置来回答本节开头所提出的问题。在实验中，我们把用户评分过的产品按一定比例随机地划分为训练集和测试集，为了使实验的结果更加精确，每个实验都进行 30 次，这样可以得到 30 组不同的随机产生的训练集和测试集下的实验结果，最后求出这 30 次实验的平均实验结果。阈值取得不同，产生的结果也会不一样，但是反映出的总体趋势一般情况下是一致的。我们通过以下几组不同的设置来进行实验并做出相应的分析。

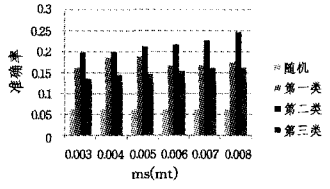

图 1　3 种冷启动问题的推荐准确率

当随机划分比例为 0.6、$sc=tc=0.4$ 时，我们比较了对新用户在已有产品上的推荐、对老用户在新产品下的推荐以及对新用户在新产品下的推荐这 3 类冷启动问题的推荐准确率，结果如图 1 所示。随机推荐的准确率为 0.062，我们观察到这三者都比随机推荐的效果好，说明它们的推荐都有意义，其中对老用户，在新产品下的平均准确率最好。虽然用户和产品都是新的，没有过多的信息可以参考，但是通过我们的推荐方法，还是可以做出较好的推荐，只是准确率相对会比较低一些。推荐的个数随着 ms 和 mt 的增加而减少，如图 2 所示。

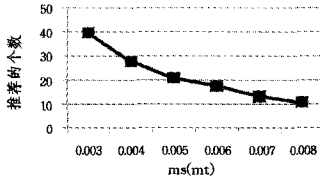

图 2　推荐个数

当随机划分比例为 0.6、$sc=tc=0.4$ 时，我们对新用户在新产品下的训练集和测试集推荐进行分析，如图 3 所示。我们观察到训练集比测试集的推荐效果好，推荐的平均准确率随着 ms 和 mt 的增加刚开始平稳增加，最后有下降的趋势。由于粒关联规则是从训练集上得到的，而测试集是新的用户和产品，因此综合考虑，测试集的推荐效果会差一些，但总体来说还是比较稳定。

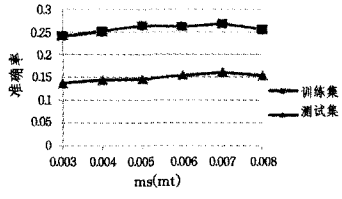

图 3　源覆盖率和目标覆盖率变化下的训练集和测试集的准确率

当 $ms=mt=0.005$、$sc=tc=0.4$ 时，我们继续对新用户在新产品下的训练集和测试集推荐进行分析，如图 4 所示。我们观察到在训练集上的推荐比测试集上的强，随着划分比例的增加，训练集的比重逐渐增大，有更多的已有用户和产品及历史评价记录，我们可以挖掘出更好的用户喜爱的规则，从而根据规则为用户推荐合适的电影，因此，训练集和测试集的推荐准确率逐渐提高。

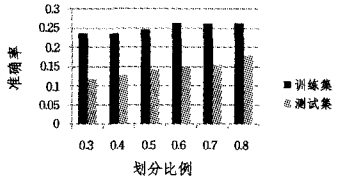

图 4　划分比例变化下的训练集和测试集的准确率

结束语　本文研究了一种基于粒关联规则的冷启动推荐方法，用于对新用户的推荐、老用户在新产品下的推荐以及新用户在新产品下的推荐，并分析了它们的性能。在 MovieLens数据集上进行的实验表明，通过合适的阈值设置可以挖掘出有意义的规则，将其应用于对用户的推荐取得了较好的效果。在未来的工作里，我们会继续完善粒关联规则的冷启动推荐方法，然后与其他的推荐方法进行比较。

参 考 文 献

[1] Mooney R J, Roy L. Content-based book recommending using learning for text categorization [C]// Proceedings of the fifth ACM conference on Digital libraries. ACM, 2000: 195-204

[2] Pazzani M J, Billsus D. Content-based recommendation systems [M]. The adaptive web. Springer Berlin Heidelberg, 2007: 325-341

[3] Goldberg D, Nichols D, Oki B M, et al. Using collaborative filtering to weave an information tapestry [J]. Communications of the ACM, 1992, 35(12): 61-70

[4] Cremonesi P, Turrin R, Airoldi F. Hybrid algorithms for recommending new items [C]// Proceedings of the second International Workshop on Information Heterogeneity and Fusion in Recommender Systems. ACM, 2011: 33-40

[5] Balabanović M, Shoham Y. Fab: content-based, collaborative recommendation [J]. Communications of the ACM, 1997, 40(3): 66-72

[6] Schein A I, Popescul A, Ungar L H, et al. Methods and metrics for cold-start recommendations [C]// Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval. ACM, 2002: 253-260

[7] Min F, Hu Q, Zhu W. Granular association rules with four subtypes [C]// Granular Computing (GrC), 2012 IEEE International Conference on. IEEE, 2012: 353-358

[8] Min F, Hu Q, Zhu W. Granular association rules on two universes with four measures [J]. arXiv preprint arXiv: 1209. 5598, 2012

[9] Min F, Zhu W. Granular association rule mining through parametric rough sets [M]. Brain Informatics. Springer Berlin Heidelberg, 2012: 320-331

[10] He X, Min F, Zhu W. A comparative study of discretization approaches for granular association rule mining [C]// Proceedings of the 2013 Canadian Conference on Electrical and Computer

Engineering. 2013

[11] Min F, Zhu W. Granular association rules for multi-valued data [C]//Proceedings of the 2013 Canadian Conference on Electrical and Computer Engineering. 2013

[12] Lin T Y. Granular computing on binary relations I: data mining and neighborhood systems [J]. Rough sets in knowledge discovery, 1998, 1: 107-121

[13] Zhu W, Wang F Y. Reduction and axiomization of covering generalized rough sets [J]. Information sciences, 2003, 152: 217-230

[14] Yao Y, Deng X. A granular computing paradigm for concept learning [M]. Emerging Paradigms in Machine Learning. Springer Berlin Heidelberg, 2013: 307-326

[15] 苗夺谦. 粒计算: 过去, 现在与展望[M]. 北京: 科学出版社, 2007

[16] Min F, Zhu W, He X. Grale: Granular association rules[OL]. http://grc.fjzs.edu.cn/~fmin/grale/, 2013

[17] Pawlak Z. Rough sets [J]. International Journal of Computer & Information Sciences, 1982, 11(5): 341-356

[18] Skowron A, Stepaniuk J. Approximation of relations [M]. Rough Sets, Fuzzy Sets and Knowledge Discovery. London: Springer, 1994: 161-166

(上接第 54 页)

右端点分别记为 $\alpha_1^L=0.517$, $\alpha_1^U=0.705$, $\beta_1^L=0.226$, $\beta_1^U=0.420$。据此, 可以得出 1-置信度下的风险偏好者以及风险厌恶者的决策阈值; 同时也给出 α、β 的阈值下界和阈值上界。根据不同风险倾向者的决策阈值, 可以分别得出风险偏好者以及风险厌恶者的决策规则。更一般地, 可以取 α、β 的 η-截集, $\eta\in(0,1]$。得到置信度 η 下的不同风险倾向者的决策阈值, 从而得出相应的决策规则。可以看出, 在损失函数不确定的条件下, 本文通过一系列模糊运算, 将这些不确定的信息保留到模糊决策阈值 α、β 中, 在信息损失最少的情况下帮助决策者在实际决策中做出合理的决策。

结束语 本文从决策粗糙集出发, 利用模糊数来刻画损失函数, 首先提出用模糊量来处理损失函数不确定性特征的问题; 其次, 通过模糊统计得到损失函数的梯形分布; 再次, 介绍通过模糊运算得出 α、β 模糊分布的方法; 然后, 介绍一种通过比较模糊数 α、β 与条件概率大小而生成决策规则的方法; 最后, 用一个石油投资的例子来说明模型的应用过程。然而, 本文对引入模糊量进行粗糙集决策的研究尚在初级阶段, 其相关数学性质以及如何充分利用模糊数的分布来进一步分析决策过程有待进一步挖掘。

参 考 文 献

[1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Science, 1982, 11(5): 341-356

[2] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data[M]. Boston: Kluwer Academic Publishers Press, 1991: 90-166

[3] Yao Y Y. Three-way decision: an interpretation of rules in rough set theory[J]. LNAI, 2009(5589): 642-649

[4] Yao Y Y. Three-way decisions with probabilistic rough sets[J]. Information Sciences, 2010, 180: 341-353

[5] Yao Y. The superiority of three-way decisions in probabilistic rough set models[J]. Information Sciences, 2011, 181(6): 1080-1096

[6] Pawlak Z, Wong S K M, Ziarko W. Rough sets: probabilistic versus deterministic approach[J]. Inter. Journal of Man-Machine Studies, 1988, 29: 81-95

[7] Yao Y Y, Wong S K M. A decision theoretic framework for approximation concepts[J]. Inter. Journal of Man-machine Studies, 1992, 37: 793-809

[8] Yao Y Y, Wong S K M. A decision theoretic framework for approximating concepts[J]. International Journal of Man-machine Studies, 1992, 37(6): 793-809

[9] Ziarko W. Variable precision rough set model [J]. Journal of computer and system sciences, 1993, 46(1): 39-59

[10] Slezak D. Rough sets and Bayes factor[J]. LNCS Transactions on Rough Sets, 2005, 111: 202-229

[11] Slezak D, Ziarko W. The investigation of the Bayesian rough set model[J]. International Journal of Approximate Reasoning, 2005, 40: 81-91

[12] 刘盾, 李天瑞, 李华雄. 区间决策粗糙集[J]. 计算机科学, 2012, 39(7): 178-181, 215

[13] 谢季坚, 刘承平. 模糊数学方法及其应用[M]. 武汉: 华中科技大学出版社, 2000: 29-32

[14] Liu X. Measuring the satisfaction of constraints in fuzzy linear programming[J]. Fuzzy Sets and Systems, 2001, 122(2): 263-275

[15] Yager R R. On choosing between fuzzy subsets[J]. Kybernetes, 1980, 9(2): 151-154

[16] Yager R R. A procedure for ordering fuzzy subsets of the unit interval[J]. Information Sciences, 1981, 24(2): 143-161

[17] Adamo J M. Fuzzy decision trees[J]. Fuzzy sets and systems, 1980, 4(3): 207-219

[18] Chang W. Ranking of fuzzy utilities with triangular membership functions[C]//Proceedings of International Conference on Policy Analysis and Systems. 1981: 272

[19] de Campos Ibáñez L M, Muñoz A G. A subjective approach for ranking fuzzy numbers[J]. Fuzzy sets and systems, 1989, 29(2): 145-153

[20] Xie G, Yue W, Wang S, et al. Dynamic risk management in petroleum project investment based on a variable precision rough set model[J]. Technological Forecasting and Social Change, 2010, 77(6): 891-901

[21] Yusgiantoro P, Hsiao F S T. Production-sharing contracts and decision-making in oil production: The case of Indonesia[J]. Energy economics, 1993, 15(4): 245-256