

# 群体决策中多种研讨信息关系的自动识别方法<sup>①</sup>

李欣苗<sup>1</sup>, 张朋柱<sup>2</sup>, 李靖<sup>1</sup>

(1. 上海财经大学信息管理与工程学院, 上海 200433;

2. 上海交通大学安泰经济与管理学院, 上海 200052)

**摘要:** 决策是管理的核心, 决策贯穿管理的全过程. 群体决策中会产生海量的研讨信息, 研讨信息与决策方案之间存在多种关系. 论文研究了群体决策中多种研讨信息关系的自动识别方法, 构建了多种研讨信息关系自动识别模型, 并应用于实际的群体决策过程. 应用结果表明, 该模型较好地实现了研讨信息与决策方案之间强烈支持、一般支持、中立、一般反对和强烈反对关系的自动识别, 可以辅助人对群体研讨信息的整理和分析, 提高了群体决策过程信息组织的效率.

**关键词:** 群体决策; 研讨信息组织; 多种信息关系; 自动识别

**中图分类号:** C931.6   **文献标识码:** A   **文章编号:** 1007-9807(2012)09-0010-10

## 0 引言

决策是人类的基本活动之一, 随着社会的发展和技术的进步, 决策问题越来越具有动态性和复杂性. 面对快速多变的外部环境以及大量复杂的不确定性问题, 群体研讨在决策中的作用越来越重要<sup>[1]</sup>. 在群体决策中, 群体成员往往需要通过充分研讨, 才能综合形成合适的决策方案, 研讨过程是群体决策中群体智慧的结合和综合<sup>[2]</sup>, 群体研讨过程会产生海量的研讨信息, 群体决策中基本而重要的问题是对海量的研讨信息进行有效组织.

群体决策过程最主要的两个阶段是决策方案的产生阶段和评论阶段, 它们对群体决策的各个方面都起着关键性作用. 方案产生活动是群体决策过程的早期阶段, 研究表明, 在该阶段应当坚持当场不对任何方案作出评价的原则, 这样既可以防止评判约束群体成员的积极思维, 破坏自由畅

想的有利气氛, 也能够避免后续阶段的工作提前进行, 影响备选方案的产生<sup>[3]</sup>. 方案评论阶段的任务是对上一阶段形成的各种备选方案进行研讨和评价, 并判断各方案的可行性以及采用与否, 从而进行方案的优选. 这期间会产生海量的研讨评论信息, 评论信息与决策方案之间存在多种关系, 如强烈支持、一般支持、中立、一般反对、强烈反对等. 在传统的群体研讨活动中, 比如面对面的群体会议, 基本上是靠人工记录和整理群体研讨信息, 费时费力. 而在信息技术日益普及的今天, 以计算机为媒介的群体研讨活动产生的信息量远远超过传统的团队会议, 使人工方式难以胜任会议后的信息处理工作, 无法及时有效地进行分析和归纳, 导致所谓的“信息过载”, 群体成员因此很可能迷失在海量的信息中, 难以对信息进行有效的利用. 在这种情况下, 仍然依靠人工方式对信息进行组织已经无法满足群体决策的需要. 因此, 研究运用现代信息技术有效地组织群体决策中的研

① 收稿日期: 2010-09-21; 修订日期: 2011-03-22.

基金项目: 国家自然科学基金重大资助项目(79990580); 国家自然科学基金重点资助项目(70533030); 国家自然科学基金资助项目(71001059).

作者简介: 李欣苗(1975—), 女, 辽宁抚顺人, 副教授, 博士. Email: xinmiaoli@126.com

讨信息是非常必要和迫切的。

随着现代信息技术的飞速发展,各种基于计算机技术的信息关系辨识方法和信息分类方法的研究越来越受到广泛重视<sup>[4]</sup>。有些学者运用文本分析方法研究信息的特征<sup>[5]</sup>;有些学者研究利用人工神经网络、支持向量机、贝叶斯<sup>[6]</sup>等方法对电子会议产生的信息进行分类<sup>[7]</sup>,还有些学者通过挖掘电子头脑风暴评论信息的相似性识别信息之间的关系<sup>[8-9]</sup>。但是,结合本文的研究对象,以及群体决策支持的研究目的,上述研究存在如下不足:首先,研究没有关注群体决策中的方案评论阶段,而方案评论对最终决策至关重要,群体研讨尤其是复杂决策方案的评论和研讨,往往是痛苦而漫长的群体意见收敛过程。以往相关研究大多没有针对决策方案评论过程的特征和实际情况,没有考虑决策方案产生阶段与方案评论阶段的区别,其所研究的文本分类无法实现以决策方案为中心的研讨信息组织,难以有效支持群体决策;第二,上述研究多以文本中所涉及的概念或者主题作为类别<sup>[10-11]</sup>,或者通过比较其相似性而进行分类<sup>[8-9]</sup>,没有考虑决策方案评论阶段中,研讨评论信息对方案的支持和否定倾向性的需求,无法识别评论信息对决策方案的支持和否定关系;第三,上述工作多是基于西文文本,没有充分考虑中文文本的特殊性所带来的困难,难以有效地处理中文文本。

然而,在群体决策的解决方案评论阶段,对决策方案具有支持或者反对关系的评论信息具有结论性的作用,对于方案的优选具有很大帮助。因此,在大量的评论信息中提取出这样的信息,并自动识别其对方案的肯定或者否定关系,能够有效地挖掘原始评论的倾向,帮助决策群体对评论在总体上有较清楚的认识和把握,从而辅助最终决策。目前一些专家学者认识到群体研讨信息支持和反对关系对决策的重要性,设计并开发了一些群体决策支持系统帮助群体分析并组织研讨信息<sup>[12]</sup>,但是它们大都不能提供类似于人脑的智能分类手段,例如不能自动识别评论信息与决策方案之间的关系。目前信息关系识别一是依靠专门人员人工整理,二是由群体研讨成员自行标注信息关系,再借助信息技术按照发言人标明的关系

组织评论信息。但是这两种方式主要依靠人力,费时费力,并且容易受到发言人主观意识和工作态度的影响,导致信息关系标注的准确性不高,造成信息组织的混乱和错误。李欣苗,张朋柱<sup>[13-14]</sup>研究了团队创新研讨信息与创新问题解决方案之间的肯定和否定关系,提出了基于BP神经网络的团队创新信息关系自动识别方法,但是该方法只能识别肯定和否定两种信息关系。然而,在实际的决策方案评论过程中,绝对肯定和绝对否定的研讨信息并不多见,而是存在大量的一般支持、一般反对,以及没有明确立场和态度的中立性评论。评论信息对方案支持和反对的程度不同,对群体决策的影响也不同,如果将各种不同程度的研讨信息只简单归为支持或者反对两种关系,会导致决策方案评价不准确,甚至造成错误决策。因此,非常有必要进一步深入研究群体决策中多种信息关系的自动识别方法,从而更加准确和有效地支持群体决策。但是,多种信息关系识别涉及高维输入向量,且包含大量0分量。尽管支持向量机比神经网络等方法在处理高维输入向量时具有优势,但是过高的维数以及大量的0分量仍然会影响支持向量机的识别效果。因此,本文首先运用潜在语义分析方法(latent semantic analysis, LSA),将研讨信息的初始高维特征向量映射到潜在语义空间,在潜在语义空间中进行降维减噪得到新的特征矩阵;再运用支持向量机(support vector machine, SVM)方法,自动识别群体决策中多种研讨信息关系,最后提出并构建了群体决策中多种研讨信息关系的自动识别模型,并应用于实际的群体决策过程,取得了良好的效果。

## 1 群体研讨信息关系自动识别的概念模型

### 1.1 研讨信息关系自动识别的概念模型

本文研究了群体决策中研讨信息关系自动识别的概念模型,用于识别决策方案的研讨信息与方案之间的强烈支持关系、一般支持关系、中立关系、一般反对关系和强烈反对关系,如图1所示。这里假设在方案评论阶段,群体成员都针对目标

方案进行评论,不允许针对其他成员的评论信息进行二次评论.在图 1 中,首先从群体决策支持系统( group decision support system , GDSS) [12,15] 中提取出研讨信息作为待识别的目标信息,对目标信息进行特征词提取,生成初始特征向量;然后将研讨信息的初始特征向量在潜在语义空间中进行

奇异值分解,经过降维减噪得到新的特征矩阵后输入支持向量机,并运用支持向量机方法识别研讨信息与决策方案之间的关系,识别出来的结果即可作为该信息的属性用于信息组织.下面详细研究和阐述群体决策中研讨信息关系的自动识别方法.

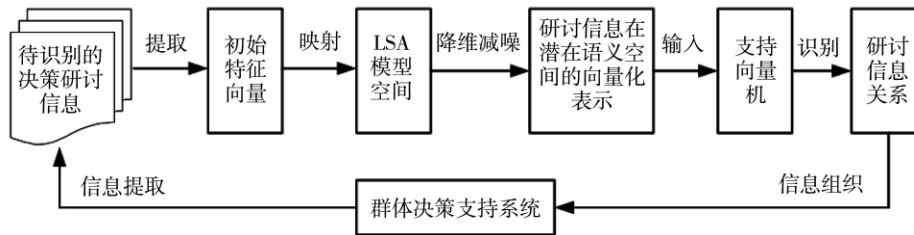


图 1 群体决策中研讨信息关系自动识别概念模型

Fig. 1 Conceptual model of automatic argumentation information relationship identification in group decision

## 1.2 基于潜在语义分析的研讨信息表示

### 1.2.1 研讨信息的特征提取

要实现群体决策中研讨信息的自动识别,首先需要研究研讨信息的表示,本文采用潜在语义分析法表示研讨信息.首先进行特征提取,特征提取是文本信息所属类别的共性与识别规则的归纳过程,是群体决策中研讨信息关系识别的重要环节,它直接影响到识别的效果.信息特征提取应以信息的内容为基础,要求所提取的特征能充分代表原始文本的内容.由于本文研究群体决策中研讨信息强烈支持、一般支持、中立、强烈反对和一般反对关系的识别,所以需要提取能够表征研讨者支持、反对和中立态度的关键词作为特征词来表示研讨信息.同时,还必须剔除那些在识别过程中会造成干扰的词.

### 1.2.2 构建初始特征词—研讨信息矩阵

通过研讨信息的特征提取构建初始特征词—研讨信息矩阵( term-argumentation information matrix).假设针对群体决策方案,群体研讨过程中产生了  $n$  条研讨信息,提取了  $m$  个特征词,那么,可以构建特征词—研讨信息矩阵  $A_{m \times n}$ :

$$A_{m \times n} = [a_{ij}] = ( doc_1 \quad doc_2 \quad ; \dots \quad doc_n ) \\ = ( term_1 \quad term_2 \quad ; \dots \quad term_m )^T \quad (1)$$

式中:  $a_{ij}$  表示特征词  $i$  在研讨信息  $j$  中出现的频数;  $term_i$  和  $doc_j$  分别代表特征词和研讨信息的列

向量.表示每个特征词对应于矩阵  $A$  的一行,每条研讨信息对应于矩阵  $A$  的一列.矩阵  $A$  表示针对某决策方案的研讨信息集中特征词与研讨信息之间的相关关系.

由于对于任意一条研讨信息总提取有限个特征词,所以  $A$  是高维的稀疏矩阵,因此,本文将群体研讨中研讨信息的初始特征词—研讨信息矩阵映射到潜在语义空间中,经过降维减噪构建新的研讨信息特征矩阵.

### 1.2.3 初始特征词—研讨信息矩阵在潜在语义空间的映射

潜在语义分析法利用截断的奇异值分解方法( truncated singular value decomposition , TSVD)实现特征词和研讨信息在潜在语义空间( LSA space) 的映射[16].

设矩阵  $A_{m \times n}$  (设  $A$  的秩为  $r$ ),可以分解为两个正交矩阵和一个对角矩阵的乘积

$$A = TSD^T \quad (2)$$

式中:  $T$  是  $m \times m$  的正交矩阵,它的每一列是  $A$  的左奇异向量;  $D$  是  $n \times n$  的正交矩阵,它的每一列是  $A$  的右奇异向量;  $S$  是  $m \times n$  的矩形对角矩阵,对角元素  $s_1 \geq s_2 \geq s_3 \geq \dots \geq s_r > 0$ ,其中  $s_1, s_2, s_3, \dots, s_r$  是  $A$  的奇异值.在矩阵  $S$  的对角线中按降序排列,即

$$S = \begin{pmatrix} \left( \begin{matrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_r \end{matrix} \right) & 0 \\ 0 & 0 \end{pmatrix} \quad (3)$$

然后,选取前面  $k$  个最大的单值,其余的设置  
为0,即在  $S$  中取相应的  $k$  阶对角矩阵,  $T$  中取相应  
的  $k$  列,  $D^T$  中取相应的  $k$  行,然后按照式(4)进行  
奇异值分解反运算,得到矩阵  $A$  的相似矩阵

$$A \approx T_k S_k D_k^T \quad (4)$$

从而将初始特征词—研讨信息矩阵映射到  $k$  维的  
潜在语义空间中。

### 1.3 基于支持向量机的研讨信息关系自动识别 方法

文本分类算法主要包括简单向量距离分类  
法、贝叶斯算法、 $k$  近邻法、神经网络、支持向量机  
方法等。由于支持向量机方法应用 VC 维理论和  
结构风险最小化原则,能够克服维数灾难、局部最  
小以及过学习等问题,具有很好的泛化能力和出  
色的分类性能,可以迅速训练出较高性能的分类  
器。目前 SVM 已被推广到多类分类领域,取得了  
很好的分类效果。支持向量机分类方法的主要优  
点包括<sup>[17]</sup>:

1) 它是专门针对有限情况的,其目标是得到  
现有信息下的最优解而不仅仅是样本数趋于无穷  
大时的最优值;

2) 算法最终将转化成为二次型寻优问题,从  
理论上来说,得到的将是全局最优解,解决了在神  
经网络方法中无法避免的局部极值问题;

3) 算法将实际问题通过非线性变换到高维  
的特征框架,在高维空间中构造线性判别函数以  
替换原空间中的非线性判别函数,这样能保证学  
习机器有较好的推广能力。同时它巧妙地解决了  
维数问题,其算法复杂度与样本维数无关。在支持  
向量机方法中,只要定义不同的核函数,就可以实  
现多项式逼近、径向基函数方法、多层感知器网  
络等许多现有学习算法。支持向量机克服了神经  
网络分类和传统分类方法的许多缺点,具备较高  
的泛化能力,可以替代多层感知机、RBF 神经网络  
和多项式神经网络等已有的学习算法。

因此,本文运用支持向量机方法实现研讨信  
息关系的自动识别。在支持向量机模型中,一对一  
算法(1-a-1)往往具有较高的分类精度,但是  
相比于一对余算法(1-a-r)需要构造的支持向  
量机数目较多,对于类别数目较多的分类问题,训  
练速度较低。由于本文分类数为5,若采用1-a-  
1算法所需要的支持向量机数目为10,并不算多,  
并且目前普遍使用的SVM工具箱 libsvm 也采用  
了该算法,因此本文运用一对一算法研究研讨信  
息关系的自动识别。支持向量机核函数的选择决  
定了特征空间的结构,本文选用高斯径向基核函  
数(radial basis function, RBF)作为支持向量机  
核函数。基于RBF的支持向量机需要确定惩罚参  
数  $c$  和核参数  $g$ 。SVM 参数选择与其性能有密切  
联系,核函数的核参数和惩罚参数是影响 SVM 性  
能的关键因素。本文采用网格搜索算法进行参数  
选择,首先确定参数搜索范围,然后设定固定的搜  
索步长,这样在由核参数和惩罚参数构成的坐标  
系上就得到个二维网络,对应网格上的每一组参  
数值,按照交叉检验方法计算出平均准确率,最后  
将各组参数值对应的准确率用等高线绘出,得到  
相应的等高线图,进而选出最优的参数值。

## 2 研讨信息关系自动识别方法在群 体决策中的应用

本文将上述信息关系自动识别方法应用于某  
保险公司核保核赔决策过程中,基于“开放式群  
体决策研讨平台”(该平台由上海交通大学安泰  
经济与管理学院设计开发),保险公司的决策群  
体通过“开放式群体决策研讨平台”研讨公司的  
核保核赔决策方案。本文随机选取针对某核保核  
赔决策方案的373条研讨信息进行研究。

### 2.1 研讨信息的特征提取

按照上述研究思路和方法,在研讨信息关系  
的自动识别中,首先提取研讨信息的特征向量。特  
征项的选取应该使得支持向量机分类器能够根据  
特征项集,将待识别的目标信息同其他类别的信  
息相区分,即确定研讨信息所属的强烈支持、一般  
支持、中立、强烈反对或者一般反对类别。本文选

取那些能够表征支持和反对意义的词和短语作为特征项,并列出了影响识别效果的剔除词,详见表1.

表1 特征词及剔除词表

Table 1 Feature words and rejection words

强烈支持	完全同意、不得不、强烈支持、非常支持、赞成、赞成!、支持、支持!、对、对!、好啊!、好啊!、太棒了!、太妙了!、同意!、同意、很想、认同!、认同……
一般支持	必要、比较合理、合理、可以接受、同意、支持、赞成、还是可以的、基本同意、基本上同意、基本合理、有可能、喜欢、适合、不错、还行……
中立	不确定、不清楚、不一定、不晓得、不知道、没想好……
强烈反对	坚决反对、完全错误、根本不可能、不赞成、不赞成!、不对、不对!、不同意、不同意!……
一般反对	不公平、不合理、没有、没必要、不赞成、不是、不好、不能、不同意、不应、不应该、错误、不可行、不太好、不太可行、不宜、不想、不适合、不喜欢、喜欢…但、不错…但、不错…可是、还行…但……
剔除词	是不是、肯定、不可避免、对、好……

### 2.2 构建特征词—研讨信息矩阵

根据表1中的特征词,构建初始特征词—研讨信息矩阵 $A_{198 \times 373}$ . 矩阵的行向量代表特征词,共有198个特征词;列向量代表研讨信息,共有373条研讨信息;矩阵中的每一个元素代表该行的特征词在该列研讨信息中出现的频数. 根据式(2),本文运用Matlab软件对初始特征词—研讨信息矩阵进行奇异值分解,构建出新的特征词—研讨信息矩阵. 新特征词—研讨信息矩阵中的对角矩阵为 $S_{198 \times 373}$ ,相应的研讨信息矩阵为 $D_{373 \times 373}$ . 潜在语义空间转换实质上是降维过程, $k$ 是语义空间的维数, $k$ 值的选择非常重要,其大小直接关系到信息关系识别的质量和效率. 如果 $k$ 值过小会丢失有用信息,达不到区分研讨信息关系的目的;过大则捕获不了特征词之间的关联性,存在噪音,无法有效地识别研讨信息关系,同时也增加了计算量和计算复杂度. 因此,本文从对角矩阵中选取元素奇异值发生较大变化的6个 $k$ 值,将这些 $k$ 值对应的研讨信息矩阵分别输入支持向量机,对识别效果进行测试,并与不降维的原始研讨信息矩阵

的识别效果进行对比,通过比较支持向量机对不同 $k$ 值特征矩阵的识别效果,从而最终确定最优 $k$ 值. 表2列出了本文所选的 $k$ 值及其对应的对角矩阵元素的奇异值与研讨信息矩阵.

表2 所选 $k$ 值和对应的奇异值及对应的研讨信息矩阵

Table 2 Selected  $k$ , the corresponding singular value and argumentation information matrix

$k$ 值	对应的奇异值 $S_k$	对应的研讨信息矩阵 $D$
132	4.138 469 314 202 880e - 229	$D_{373 \times 132}$
125	0.618 0	$D_{373 \times 125}$
95	1.000 0	$D_{373 \times 95}$
66	1.589 0	$D_{373 \times 66}$
35	2.000 0	$D_{373 \times 35}$
13	3.243 4	$D_{373 \times 13}$

### 2.3 基于支持向量机的研讨信息关系识别

运用支持向量机方法识别研讨信息与决策方案之间的关系,支持向量机的输入为经过LSA降维的新研讨信息特征向量. 经过训练后支持向量机输出研讨信息所属的类别. 如果:研讨信息与决策方案之间是强烈支持关系,则输出类别取值为1;研讨信息与决策方案之间是一般支持关系,则输出类别取值为2;研讨信息与决策方案之间是中立关系,则输出类别取值为3;研讨信息与决策方案之间是一般反对关系,则输出类别取值为4;研讨信息与决策方案之间是强烈反对关系,则输出类别取值为5.

本文提出的多种研讨信息关系自动识别方法已应用于某保险公司核保核赔决策过程中的方案评论阶段. 将随机选取的373条研讨信息随机分成两份,随机选取其中的187条研讨信息用来确定支持向量机的参数,剩下的186条研讨信息用来测试本文所构建的支持向量机模型的效果. 测试时采用五折交叉检验法,即将样本大致平均分为5组,每次让其中的1组作为检验集,其它4组作为训练集,如此反复5次,记录下准确率、查全率及查准率. 采用五折交叉检验法可以更有效地检验支持向量机模型的效果.

#### 2.3.1 支持向量机参数的确定

选取高斯径向基函数作为支持向量机的核函数,运用网格法确定惩罚参数 $c$ 和核参数 $g$ . 为了

更准确地确定参数值,分为参数粗略选择和参数精细选择两步来确定支持向量机的惩罚参数  $c$  和核参数  $g$ ,参数粗略选择是为了缩小并确定最优参数的合理范围,参数精细选择是为了最终确定最优参数.下面以  $k = 66$  时的参数确定过程为例进行阐述,参数确定过程采用 matlab 编程实现.

1) 参数的粗略选择 首先根据本文的研究问题选取比较大的参数取值范围.

设  $2^{-10} < c < 2^{10}$   $2^{-10} < g < 2^{10}$   $cstep = 0.8$ ,  $gstep = 0.8$ . 其中  $cstep$  为  $c$  的变化步长  $gstep$  为  $g$  的变化步长.

由此可得由  $c$  和  $g$  构成的二维网络. 在支持向量机模型中取网格上相应的每一组  $c$  和  $g$ ,按照五折交叉检验法计算出平均准确率(本文使用 libsvm-mat-3.0-1 作为 SVM 软件包) 并将各组参数值对应的平均准确率用等高线绘出,即得到  $k = 66$  的等高线图(图 2) 和相应 3D 视图(图 3),从图 2 和图 3 可以确定最优参数的范围,并可以粗略选出一组最优参数值,此时最优  $c$  为 21.112 1,最优  $g$  为 0.435 28,对应的交叉检验平均准确率为 74.331 6%.

2) 参数的精细选择 经过参数的粗略选择,将  $c$  范围缩小到  $2^{-2} \sim 2^8$   $g$  范围缩小到  $2^{-4} \sim 2^4$ ,并将  $cstep$  和  $gstep$  设置为 0.5,由此得出精细选择后  $k = 66$  的等高线图(图 4) 和相应的 3D 视图(图 5).至此,经过参数的粗略选择和精细选择可以得到  $k = 66$  的最优参数  $c = 11.313 7$   $g = 0.707 11$ ,对应的交叉检验平均准确率为 81.818 2%.

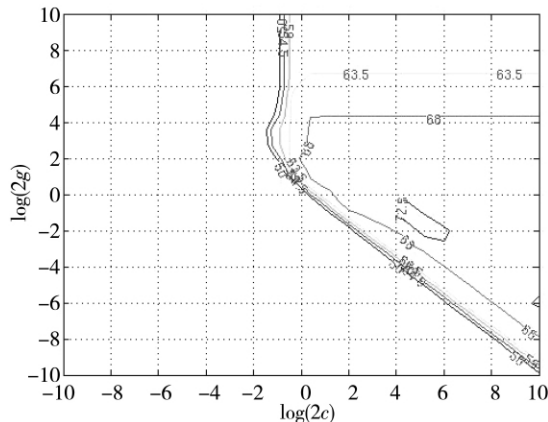


图 2  $k = 66$  时的参数粗略选择等高线图

Fig. 2 Contour map on searching for  $c$  &  $g$  roughly when  $k = 66$

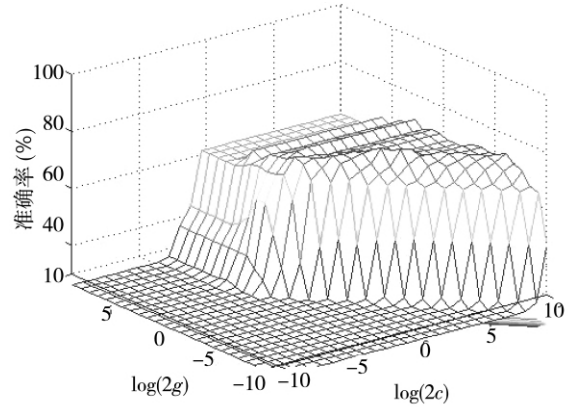


图 3  $k = 66$  时的参数粗略选择 3D 视图

Fig. 3 Contour 3 on searching for  $c$  &  $g$  roughly when  $k = 66$

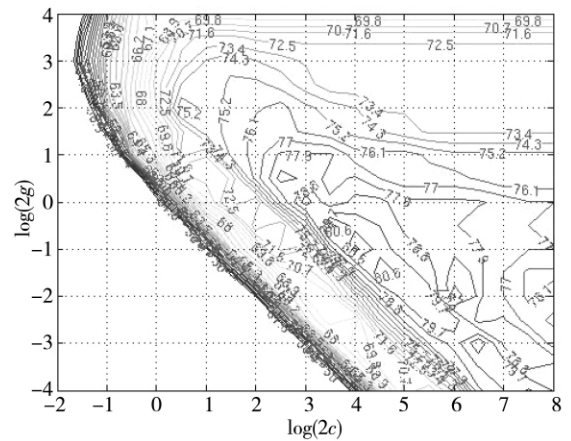


图 4  $k = 66$  时的参数精细选择等高线图

Fig. 4 Contour map on searching for  $c$  &  $g$  accurately when  $k = 66$

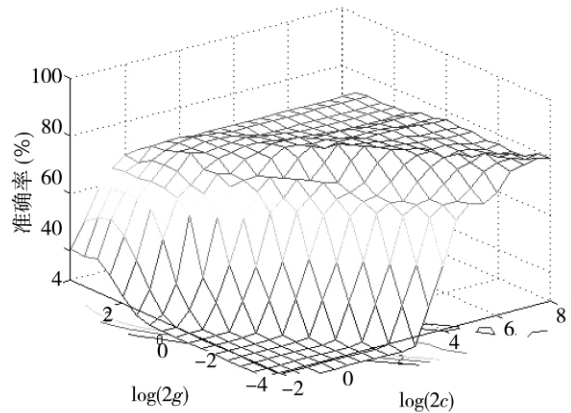


图 5  $k = 66$  时的参数精细选择 3D 视图

Fig. 5 Contour 3 on searching for  $c$  &  $g$  accurately when  $k = 66$

同样可以得到前面所取各维度的最佳 SVM 参数以及对应的平均准确率(如表 3).

### 2.3.2 支持向量机模型的测

在不同输入维度的最优参数确定以后,用剩下的 186 条研讨信息来测试支持向量机模型的有

效性. 这里同样采用五折交叉检验法,即将样本大致平均分为 5 组,每次让其中的 1 组作为检验集,其他 4 组作为训练集,如此循环 5 次.并通过比较支持向量机的平均准确率、各类别平均查全率和查准率,以及 5 类平均查全率和查准率,选出最优的  $k$  值(表 4—表 10 给出了  $k$  取不同值时,支持向量机的准确率、查全率及查准率).

本文中查全率和查准率可如下计算

$$R_i = \frac{\text{第 } i \text{ 类分类正确的研讨信息}}{\text{未分类前属于该类的研讨信息}} \times 100\% \quad (5)$$

$$P_i = \frac{\text{第 } i \text{ 类分类正确的研讨信息}}{\text{分类以后被分为该类的研讨信息}} \times 100\% \quad (6)$$

$$\bar{R} = \frac{\sum_{i=1}^n R_i}{N} \times 100\% \quad (7)$$

$$\bar{P} = \frac{\sum_{i=1}^n P_i}{N} \times 100\% \quad (8)$$

其中  $R_i$  为第  $i$  类研讨信息的查全率,  $P_i$  为第  $i$  类研讨信息的查准率,  $\bar{R}$  为平均查全率,  $\bar{P}$  为平均查准率,  $N$  为类别数,本文中为 5 类,故  $N = 5$ .

表 3 选定维度对应的 SVM 最优参数以及相应的交叉检验平均准确率

Table 3 Selected  $k$ , the corresponding best  $c$  &  $g$  and CV accuracy

维度 ( $k$ 值)	最优 $c$ 值 ( Best $c$ )	最优 $g$ 值 ( Best $g$ )	交叉检验准确率 ( CV accuracy)
198	5.656 85	0.707 107	51.336 9%
132	11.313 7	0.707 107	67.379 7%
125	16	0.5	79.679 1%
95	16	0.5	75.935 8%
66	11.313 7	0.707 11	81.818 2%
35	5.656 85	1.414 21	79.679 1%
13	11.313 7	5.656 85	63.636 4%

表 4 研讨信息取 198 维特征的自动分类结果

Table 4 Results of automatic classification when  $k = 198$

类别	查准率	查全率	平均查准率	平均查全率	平均准确率
强烈支持类别	52.60%	64.99%	79.50%	72.10%	72.43%
一般支持类别	61.27%	77.32%			
中立类别	100.00%	88.00%			
一般反对类别	88.61%	77.08%			
强烈反对类别	95.00%	53.10%			

表 5 研讨信息取 132 维特征的自动分类结果

Table 5 Results of automatic classification when  $k = 132$

类别	查准率	查全率	平均查准率	平均查全率	平均准确率
强烈支持类别	63.02%	65.97%	81.21%	70.39%	67.03%
一般支持类别	43.02%	65.94%			
中立类别	100.00%	95.00%			
一般反对类别	100.00%	76.36%			
强烈反对类别	100.00%	48.67%			

表 6 研讨信息取 125 维特征的自动分类结果

Table 6 Results of automatic classification when  $k = 125$

类别	查准率	查全率	平均查准率	平均查全率	平均准确率
强烈支持类别	47.60%	66.08%	78.98%	68.41%	63.78%
一般支持类别	47.29%	60.60%			
中立类别	100.00%	94.29%			
一般反对类别	100.00%	72.51%			
强烈反对类别	100.00%	48.57%			

表 7 研讨信息取 95 维特征的自动分类结果

Table 7 Results of automatic classification when  $k = 95$

类别	查准率	查全率	平均查准率	平均查全率	平均准确率
强烈支持类别	59.60%	87.32%	85.29%	75.95%	75.68%
一般支持类别	78.83%	62.10%			
中立类别	100.00%	83.33%			
一般反对类别	96.00%	80.98%			
强烈反对类别	92.00%	66.00%			

表 8 研讨信息取 66 维特征的自动分类结果

Table 8 Results of automatic classification when  $k = 66$

类别	查准率	查全率	平均查准率	平均查全率	平均准确率
强烈支持类别	78.69%	91.51%	91.43%	88.89%	88.65%
一般支持类别	84.50%	83.21%			
中立类别	100.00%	100.00%			
一般反对类别	93.99%	88.21%			
强烈反对类别	100.00%	81.5%			

表 9 研讨信息取 35 维特征的自动分类结果

Table 9 Results of automatic classification when  $k = 35$

类别	查准率	查全率	平均查准率	平均查全率	平均准确率
强烈支持类别	74.96%	69.95%	83.22%	72.68%	70.27%
一般支持类别	41.15%	82.21%			
中立类别	100.00%	100.00%			
一般反对类别	100.00%	68.05%			
强烈反对类别	100.00%	43.21%			

表 10 研讨信息取 13 维特征的自动分类结果

Table 10 Results of automatic classification when  $k = 13$

类别	查准率	查全率	平均查准率	平均查全率	平均准确率
强烈支持类别	91.94%	63.38%	74.47%	66.54%	65.95%
一般支持类别	39.51%	92.29%			
中立类别	100%	94.29%			
一般反对类别	93.06%	67.73%			
强烈反对类别	47.84%	15.00%			

比较表 4 至表 10 可以得到  $k = 66$  时支持向量机分类准确率为 88.65% ,研讨信息关系识别模型在应用中的平均查准率为 91.43% ,平均查全率为 88.89% ,且各个类别的查全率和查准率都达到了较好的效果 ,这表明当  $k = 66$ 、 $c = 11.3137$ 、 $g = 0.70711$  时 ,本文所建立的多种研讨信息关系自动识别模型能够较好地支持群体决策中研讨信息与决策方案关系的自动识别 ,从而支持群体决策方案的分析和归纳 ,提高

了群体决策过程研讨信息组织的效率.

再次结合相应维度所对应的奇异值进行分析 ,当  $k > 124$  时 ,对应的奇异值都很小 ( $\leq 1.242342208025828e-16$ ) ,若保留这些维度的研讨信息 ,则会产生较多“噪音” ,不利于研讨信息关系识别. 而当  $k$  取较小值 ,例如  $k = 13$  ,尽管所对应的奇异值为 3.3412 ,但是由表 10 可以看出 ,自动分类结果仍然不理想. 可见选择过小维度会导致重要语义信息缺失 ,同样不利于研讨信息关系



识别. 图6为各维度与其相应平均识别准确率的函数曲线图, 从图中可以直观地看出不同维度对于研讨信息识别的影响.

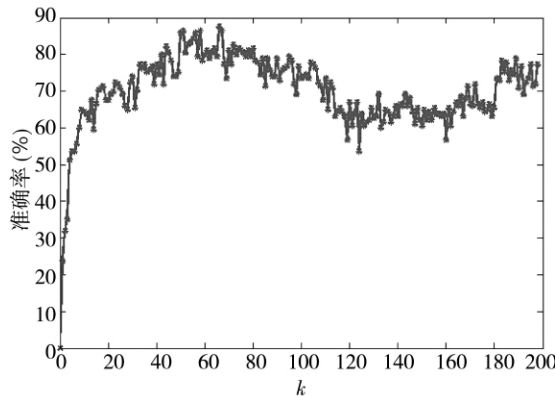


图6 各维度与相应平均准确率的函数曲线图

Fig. 6 The relationship between the value of  $k$  and average accuracy

### 3 结束语

决策中的许多复杂问题往往需要通过群体研讨才能得到有效解决, 群体研讨过程会产生海量的研讨信息. 群体决策过程中的基本问题是如何对产生的海量研讨信息进行有效组织. 本文提出了群体决策中研讨信息与决策方案之间的强烈支持、一般支持、中立、一般反对和强烈反对关系的自动识别方法, 并应用于实际的群体决策过程中. 应用结果表明, 该模型较好地实现了研讨信息与决策方案之间的强烈支持、一般支持、中立、一般反对和强烈反对关系的自动识别, 可以辅助人对群体研讨信息的整理和分析, 提高了群体决策过程信息组织的效率.

#### 参考文献:

- [1] Lu Jing-yan, Lajoie S P. Supporting medical decision making with argumentation tools [J]. *Contemporary Educational Psychology*, 2008, 33(3): 425 - 442.
- [2] 刘怡君, 唐锡晋. 一种支持协作与知识创造的“场” [J]. *管理科学学报*, 2006, 9(1): 79 - 85.  
Liu Yijun, Tang Xijin. Ba for collaboration and knowledge creation [J]. *Journal of Management Sciences in China*, 2006, 9(1): 79 - 85. (in Chinese)
- [3] Paul S, Nazareth D L. Input information complexity, perceived time pressure, and information processing in GSS-based work groups: An experimental investigation using a decision schema to alleviate information overload conditions [J]. *Decision Support Systems*, 2010, 49(1): 31 - 40.
- [4] Zhang Wen, Yoshida Taketoshi, Tang Xi-jin. A comparative study of TF\* IDF, LSI and multi-words for text classification [J]. *Expert Systems with Applications*, 2011, 38(3): 2758 - 2765.
- [5] Altıncay H, Erenel Z. Analytical evaluation of term weighting schemes for text categorization [J]. *Pattern Recognition Letters*, 2010, 31(11): 1310 - 1323.
- [6] Lu Shing-Hwa, Chiang Ding-An, Keh Huan-Chao, et al. Chinese text classification by the Naïve Bayes Classifier and the associative classifier with multiple confidence threshold values [J]. *Knowledge-Based Systems*, 2010, 23(6): 598 - 604.
- [7] Song W, Choi L C, Park S C, et al. Fuzzy evolutionary optimization modeling and its applications to unsupervised categorization and extractive summarization [J]. *Expert Systems with Applications*, 2011, 38(8): 9112 - 9121.
- [8] Chen Yao-tsung, Chen Meng-chang. Using chi-square statistics to measure similarities for text categorization [J]. *Expert Systems with Applications*, 2011, 38(4): 3085 - 3090.
- [9] Roussinov D, Zhao J L. Automatic discovery of similarity relationships through web mining [J]. *Decision Support Systems*, 2003, 35(1): 149 - 186.
- [10] Frunza O, Inkpen D, Matwin S, et al. Exploiting the systematic review protocol for classification of medical abstracts [J]. *Artificial Intelligence in Medicine*, 2011, 51(1): 17 - 25.
- [11] Song Wan-peng, Liu Wen-yin, Gu Nai-jie, et al. Automatic categorization of questions for user-interactive question answering [J]. *Information Processing & Management*, 2011, 47(2): 147 - 156.
- [12] 蒋御柱, 张朋柱, 张兴学. 群体研讨支持系统中的智能可视化研究 [J]. *管理科学学报*, 2009, 12(3): 1 - 12.  
Jiang Yuzhu, Zhang Pengzhu, Zhang Xingxue. Research on intelligence visualization in group argument support system [J]. *Journal of Management Sciences in China*, 2009, 12(3): 1 - 12. (in Chinese)

- [13]李欣苗,张朋柱,张兴学. 团队创新信息关系的自动识别方法及应用[J]. 管理科学学报,2007,10(5): 28-39.  
Li Xinmiao, Zhang Pengzhu, Zhang Xingxue. Automatic identification method of team creation information relationship and application[J]. Journal of Management Sciences in China, 2007, 10(5): 28-39. (in Chinese)
- [14]Li Xin-miao, Zhang Peng-zhu. A system to support team creation[J]. International Journal of Information Technology & Decision Making, 2007, 6(3): 523-539.
- [15]Zhang Peng-zhu, Sun Jing-le, Chen Hsin-chun. Frame-based argumentation for group decision task generation and identification[J]. Decision Support Systems, 2005, 39(4): 643-659.
- [16]余正涛,樊孝忠,郭剑毅,等. 基于潜在语义分析的汉语问答系统答案提取[J]. 计算机学报,2006,(29): 1893-1898.  
Yu Zhengtao, Fan Xiaozhong, Guo Jianyi, et al. Answer extracting for Chinese question-answering system based on latent semantic analysis[J]. Chinese Journal of Computers, 2006, (29): 1893-1898. (in Chinese)
- [17]过蓓蓓,方兆本. 基于SVM的Web日志挖掘及潜在客户发现[J]. 管理工程学报,2010,24(1): 129-133.  
Guo Beibei, Fang Zhaoben. Application of SVM in mining potential customers from web log[J]. Journal of Industrial Engineering and Engineering Management, 2010, 24(1): 129-133. (in Chinese)

## Automatic identification method of multiple argumentation information relationship in group decision-making

LI Xin-miao<sup>1</sup>, ZHANG Peng-zhu<sup>2</sup>, LI Jing<sup>1</sup>

1. School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China;
2. Antai Management School, Shanghai Jiaotong University, Shanghai 200052, China

**Abstract:** Decision making is essential to management. A lot of argumentation information is produced in group decision-making. There are multiple relationships between the argumentation information and decision solution. In this paper, a automatic identification method of multiple argumentation information relationship in group decision-making is researched and put forward. A automatic identification model of the argumentation information relationship in group decision-making is built. Furthermore, the method is applied to actual group decision process. The results of the application show that the method realizes the automatic identification of the strongly supportive, supportive, neutral, strongly against, and against relationship between decision solution and argumentation information effectively. It can help group members to organize the large amount of argumentation information effectively and increase the efficiency of information organizing in group decision-making.

**Key words:** group decision-making; argumentation information organization; multiple information relationship; automatic identification