

# 团队创新信息关系的自动识别方法及其应用<sup>①</sup>

李欣苗<sup>1</sup>, 张朋柱<sup>2</sup>, 张兴学<sup>2</sup>

(1. 上海财经大学信息管理与工程学院, 上海 200433 2 上海交通大学安泰管理学院, 上海 200052)

**摘要:** 20 世纪 90 年代以来, 创新替代了效率和质量等因素成为企业最主要的核心竞争优势. 团队创新中会产生海量的信息, 研究提出了团队创新信息关系的自动识别方法以用于信息组织, 并应用于实际的团队创新过程中信息关系自动识别模型的建立. 在该模型中, 对词频特征提取方法进行了“特征词分级处理”、增加“小句数量”特征和增加“上一评论信息的态度”特征三方面改进, 以提高识别准确率. 应用结果表明, 该模型较好地实现了评论信息与方案肯定和否定关系的自动识别, 可以辅助人对团队评论信息的整理和分析, 提高了团队创新过程信息组织的效率.

**关键词:** 团队创新; 信息组织; 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(2007)05-0028-12

## 0 引言

在世界科技日益进步的今天, 创新能力的高低已经成为影响企业获得和保持竞争优势的重要因素. 特别是 20 世纪 90 年代以来, 创新被认为替代了效率和质量等因素, 而成为企业最主要的核心竞争优势<sup>[1]</sup>. 多年以来, 团队创新是组织进行创新的主要形式, 是企业在研发新产品中的重要方法<sup>[2~4]</sup>. 然而, 创新不是简单的任务, 它是存在创造性、挑战性和风险的复杂过程. 而且, 由于经济全球化和市场竞争的日趋激烈, 企业需要以更快的速度和更低的成本推出新的产品. 目前, 对于团队创新普遍缺乏有效的创新支持手段, 阻碍了团队创新任务快速有效地完成.

团队创新的过程在本质上是不断提出创新问题并解决之的过程<sup>[5]</sup>, 在这个过程中, 一个基本的问题是如何对团队创新过程中产生的海量信息进行有效的组织. 在传统的团队创新活动中, 比如面对面的团队会议, 基本上是靠会议秘书(或

者类似的角色)记录、整理团队信息, 以及与会成员的个人记忆和文本记录, 费时费力. 而在信息技术日益普及的今天, 以计算机为媒介的团队创新活动产生的信息量远远超过传统的团队会议<sup>[6]</sup>, 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”<sup>[7]</sup>, 团队成员因此很可能“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团队创新的需要. 因此, 研究运用现代信息技术有效地组织团队信息是非常必要和迫切的.

信息组织的本质在于分类<sup>[8,9]</sup>, 而其关键在于识别信息与目标对象, 以及信息与信息之间的关系. 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<sup>[10~13]</sup>的研究越来越受到广泛的重视. 目前, 一些学者研究利用人工神经网络等人工智能的方法对电子会议产生的信息进行分类, 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

① 收稿日期: 2005-11-21; 修订日期: 2007-07-09.

基金项目: 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(70271025).

作者简介: 李欣苗(1975-), 女, 辽宁抚顺人, 博士, 讲师. E-mail: ximmiao\_1@126.com.

系<sup>[14, 15]</sup>。但是, 结合本文的研究对象, 以及团队创新支持的研究目的, 上述研究存在如下不足: 一是没有考虑新产品开发 (new product development, NPD) 的实际背景, 无法实现以创新任务为中心的信息组织, 难以有效地支持 NPD 团队的工作; 二是没有考虑方案评论阶段, 判断评论信息对方案的支持和否定倾向性的需求, 上述研究一般以评论信息所涉及的概念作为电子会议评论信息的类别, 而没有考虑相关信息的评论倾向性; 三是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。目前有一些团队支持系统软件提供信息组织工具, 但是它们对团队信息组织的支持功能是非常有限的<sup>[12, 13, 16]</sup>, 大都不能提供类似于人脑的智能分类手段, 例如不能实现评论信息肯定和否定关系地自动识别。目前信息关系识别一是依靠专门人员分析整理; 二是由发表信息的人自行标注信息关系, 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 (包括发言人自己的标注), 前者费时费力, 后者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。针对上述情况, 本文研究提出了团队创新信息关系的自动识别方法, 并应用于实际的团队创新过程中信息关系自动识别模型的建立。

根据 Wallas Simon 和 Amabile 的创新过程模型以及 Alex Osborn 和 Sidney Parnes 提出的创新问题解决 (creative problem solving, CPS) 模型<sup>[17]</sup>, 本研究<sup>[18]</sup>针对具体的创新任务解决过程, 提出了面向创新任务的团队创新过程模型, 如图 1 所示。在图 1 中, 面向一个既定的创新任务, 该团队创新过程从子任务开始, 经过解决方案提出阶段和解决方案评论阶段, 形成阶段性的创新成果, 并经过评价反馈环节提供更多的改进需求, 从而推动下一个创新子任务的进行。这个创新过程不断循环, 直到产生最终的创新产品, 从而完成团队创新总任务。

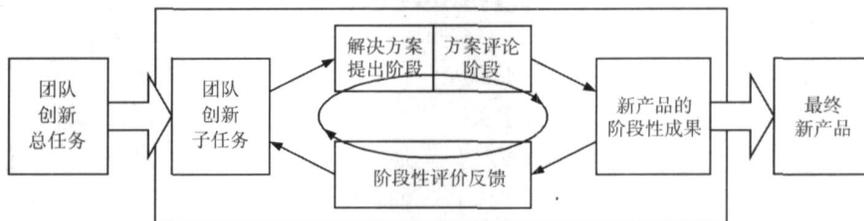


图 1 面向任务的团队创新过程

Fig. 1 Task-oriented team creation process

在团队创新过程中, 最主要的两个阶段是创新问题的解决方案产生阶段和方案评论阶段, 它们至关重要而又缺一不可, 对团队创新的各个方面都起着关键性的作用<sup>[19]</sup>。方案产生活动是团队创新过程的早期阶段, 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团队成员的积极思维, 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, 影响解决方案的产生<sup>[6, 7, 19]</sup>。而对该阶段产生的各种方案的评价则放在后续的方案评论阶段进行。方案评论阶段的任务是对上一阶段形成的各种创新问题的解决方案进行论证和评价, 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。这期间团队会产生海量的评论信息。

在团队创新的解决方案评论阶段, 对方案具有明确肯定或者否定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, 帮助团队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。本文研究提出了团队信息中的评论信息与方案之间肯定和否定关系的自动识别方法, 并应用于实际的团队创新过程中信息关系自动识别模型的构建, 用于信息组织。

# 1 团队创新信息关系的自动识别概念模型

## 1.1 信息关系的自动识别概念模型

在实际的团队创新过程中, 进行方案评论的时候, 各条信息之间具有多种关系, 如肯定、否定、质疑等. 但是, 就方案的优选问题而言, 团队最关心的是那些对评论对象 (即评论方案) 具有肯定和否定态度倾向性的评论信息. 这些信息反映了

评论人 (团队成员) 对当前方案的态度, 如果能够在众多的评论中识别出对方案具有明确的肯定关系或否定关系的信息, 就可以帮助团队从总体上把握对该方案的态度倾向性, 从而更好地支持方案优选. 为此, 本文研究提出了评论信息对方案的肯定或者否定关系的自动识别方法, 构建了基于神经网络的团队信息关系自动识别概念模型. 这里假设在方案评论阶段, 团队成员都针对目标方案进行评论, 不允许针对其它成员的评论信息进行二次评论.

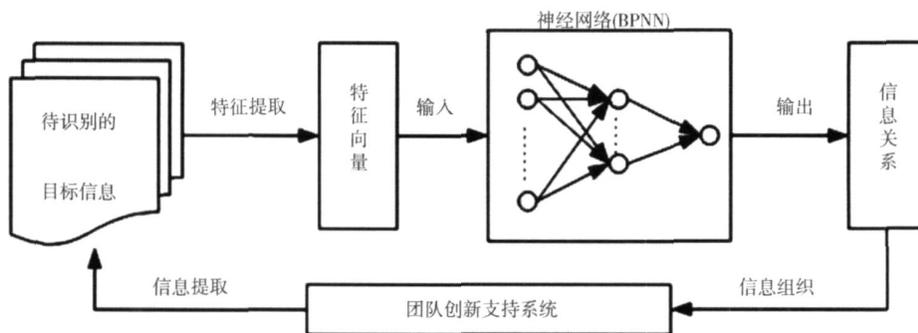


图 2 基于神经网络的信息关系自动识别概念模型

Fig 2 Conceptual model of automatic information relationship identification based on NN

团队信息关系自动识别的概念模型, 用于识别方案评论信息与方案之间的肯定关系和否定关系, 如图 2 所示. 在图 2 中, 首先从团队创新支持系统 (team creation support system, TCSS)<sup>[18 20 21]</sup> 中提取出团队评论信息作为待识别的目标信息, 然后对评论信息进行特征提取, 接下来将提取的特征输入神经网络分类器, 运用人工神经网络的方法识别评论信息与方案的关系, 识别出来的结果即可作为该信息的属性用于信息组织. 如果发现原始评论信息经过特征提取后, 特征向量中不包含任何肯定和否定意义的特征词, 则可将该评论信息直接归入中立关系 (本文将不具有明确的肯定或否定关系的评论均归入中立关系), 无须经由神经网络识别. 团队创新支持系统 (TCSS) 的研究与开发受国家自然科学基金“团队创新的支持模式及促进方法研究”资助, 本文的工作为该系统的重要研究内容之一. 团队信息关系自动识别的概念模型包括信息的特征提取方法和基于人工神经网络的信息关系自动识别方法两个关键环节, 其具体内容在下面进行详细地研究和阐述.

## 1.2 信息的特征提取方法

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, 是方案评论阶段信息关系识别的重要环节, 特征提取算法的优劣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. 目前在信息处理方向上, 文本的表示主要采用向量空间模型 (vector space model, VSM)<sup>[22 23]</sup>. 在向量空间模型中, 文本信息用相应的特征向量表示, 信息被看作是一系列相互独立的特征词组的集合  $T = \{T_1, T_2, \dots, T_n\}$ <sup>[24]</sup>, 对每个特征项  $T_k$ , 有其相应的权重,  $W = \{W_1, W_2, W_3, \dots, W_n\}$ , 其中  $W_k$  为第  $k$  个特征项的权重. 因为词、词组和短语是组成文本的基本元素, 它们在文本中出现的频率与原始文本的内容之间具有较强的相关性, 因此提取文本的词频特征作为特征项<sup>[25]</sup>, 以下本文将从文本信息中提取的特征项称为特征词. 这样, 在向量空间模型中, 每条信息 (用其特征向量代表) 就可以看作是  $N$  维空间中的一个点 (或向量), 该空间的维数由特征词组集合  $T$  决

定, 该点的位置由权重集合  $W$  决定. 信息分类问题即等效于按照某种空间距离测度对众多信息点进行分群 (划分) 的过程. 这里权重的计算采用基于词频的方法, 即权重等于特征词在文本中出现的频率.

但是, 传统的词频特征提取方法用于解决本文的研究问题, 效果却不是很好. 因此, 本文对传统的词频特征提取方法进行了改进, 改进方法在第 2 节的“信息关系识别方法在实际团队创新信息关系识别建模中的应用”部分将结合具体实例进行详细的分析和阐述.

### 1.3 基于人工神经网络的信息关系自动识别方法

基于 VSM 的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、神经网络方法等<sup>[22]</sup>. 由于神经网络具有与人脑相似的高度并行性、良好的容错性和联想记忆功能, 以及自适应和自学习等特点, 可以很好地解决传统文本分类方法在实现和应用过程中遇到的一些问题, 例如:

(1) 准确性问题: 以 BP 神经网络 (back propagation neural network, BPNN) 为基础的有监督人工神经网络 (artificial neural network, ANN) 分类器可以非常好地学习人的分类经验, 并将这种经验转化为网络形式的隐性知识表达, 使分类的准确性较无监督学习更高;

(2) 自学习问题: 在系统输出结果与实际结果相差太大时, 它的学习算法会自动调整系统本身, 改变知识的存储. 同时由于采用了神经网络技术, 系统能自然地实现模糊推理功能;

(3) 自适应问题: 神经网络技术能根据所提供的数据, 通过学习, 找出和输出结果之间的内在联系, 从而求得问题的解答, 而不仅仅依靠对问题的先验知识和规则, 因而它具有很好的适应性;

(4) 容错性问题: 具有很强的鲁棒性和容错性, 善于联想、概括、类比和推广, 任何局部的操作不会影响整体效果;

(5) 运行速度问题: 人工神经网络具有并行处理的特点, 运行速度快, 因而一方面可存储大量的知识, 另一方面可保持高的运行速度, 从而很好地解决这个矛盾.

因此, 本文采用神经网络的方法识别评论信

息与方案之间的肯定关系和否定关系. 在各种人工神经网络模型中, 在识别分类中应用最多的也是最成功的当数多层前馈网络, 其中又以多层感知器为代表<sup>[26]</sup>. BPNN 是由于其权值采用反向传播学习算法而得名, 它的结构简单, 工作状态稳定, 仅具有一个隐层的 BPNN 即足以实现任意分类问题, 因此本文采用三层 BPNN 识别团队信息的关系.

## 2 信息关系识别方法在实际团队创新信息关系识别建模中的应用

上述方案评论阶段的信息关系自动识别方法已应用于“华东电网规划系统的研究与开发”项目. 在该项目的方案评论阶段, 选取若干子方案进行研究, 训练样本从原始方案中随机抽取, 其中随机抽取某子任务的 4 种解决方案, 其中前两个解决方案的 100 条评论信息作为训练集, 另外两个解决方案的 100 条评论信息作为检验集. 下面详细阐述应用过程和结果.

### 2.1 信息关系自动识别模型的构建

#### 2.1.1 信息的特征提取

按照上述研究思路和方法, 在方案评论信息关系的自动识别中, 首先提取方案评论信息的特征向量. 特征项的选取应该使得神经网络分类器能够根据特征项集, 将待识别的目标信息同其它类别的信息相区分, 即确定信息所属的肯定或者否定类别.

首先, 本研究选取那些能够表征肯定和否定意义的词和短语作为特征项. 对于本文研究的团队方案评论信息, 其文本较小, 这导致按照传统词频方法得到的特征向量质量不高, 其中会包含大量 0 分量. 为了降低特征向量空间的维数, 但是又不至于丢失太多的信息, 同时避免特征向量中出现大量的 0 分量, 本研究进一步根据实验的效果, 将选取的特征词按照语气的强弱分为六级, 分别表示不同的肯定和否定程度: 十分肯定、一般肯定、轻微肯定、十分否定、一般否定和轻微否定的特征词. 在实验过程中还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

十分肯定	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!
一般肯定	必要、比较合理、合理、可以接受、同意、支持、赞成
轻微肯定	还是可以的、基本同意、基本上同意、基本合理、有可能
十分否定	坚决反对、完全错误、根本不可能、不赞成、不赞成!、不对、不对!
一般否定	不公平、不合理、没有、没必要、不赞成、不是、不好、不能、不同意、不想、不应、不应该、错误、不可行
轻微否定	遥不可及、不太好、不一定、不太可行、不宜
剔除词	是不是、肯定、不可避免、不确定性

在表 1 中, 从十分肯定、一般肯定, 到轻微肯定, 不仅特征词的肯定语气从强到弱, 而且在识别分类的实验中, 三类特征词对肯定关系的识别效果也是逐渐减弱; 同样, 从十分否定、一般否定, 到轻微否定, 特征词的否定语气从强到弱, 在识别分类的实验中, 三类特征词对否定关系的识别效果也是逐渐减弱。比如, “对” 单独一个词不宜作为评论信息的特征词, 因为“对” 是个常用词, 对肯定和否定关系的识别效果很差, 还常常干扰正确的识别, 但是“对, ” 和“对!” 就是对肯定关系识别效果很好的特征词, 因此本研究将其列在十分肯定这类里面。同样的情况, “不对” 单独一个词也不宜作为评论信息的特征词, 因为它常常干扰对否定关系的正确识别, 但是“不对, ” 和“不对!” 就是对否定关系识别效果很好的特征词, 因此本章将“不对, ” 和“不对!” 列在十分否定这类里面。再比如, “合理” 和“比较合理” 这两个词虽然在肯定语气的强弱上有所区别, 但是他们在对肯定关系的识别分类效果上基本一样, 因此本文将它们都列在一般肯定类里面。还有一些词, 比如“不” 不能作为特征词, 而是将“不是”、“不能”、“不赞成”、“不合理” 等作为特征词, 这样会大大提高识别分类的正确率。剔除词在识别分类的实验中干扰肯定与否定关系的正确识别, 本研究将这样的词作为剔除词, 在对评论信息进行特征提取时不予考虑。

对于每条文本信息, 根据 VSM, 确定与每一条信息所对应的特征向量, 在这里, 特征向量的每个分量即等于相应级别的特征词在文本中出现的频率之和。例如: 评论文本信息“对于图层加载问题, 我认为实例化 SeConnection 类的方法比较合理。

我同意用该方法。” 其中, 包含两个特征词“比较合理” 和“同意”, 这两个特征词都属于“一般肯定” 级别的特征词 (见表 1), 所以, 该评论文本信息的特征向量为 020000

考虑到单纯依赖词频特征, 在表达评论信息的态度倾向性方面精确度难以进一步提高, 因此, 有必要挖掘评论文本的其它特征以更好的反映文本内容, 从而提高后续分类的效果。本文在上述“肯定否定六级分类” 的基础上, 进一步增加了“上一评论文本的态度” 和“小句数量” 作为后续分类器的输入特征:

(1) 上一评论文本的态度: 反映了在连续评论文本中, 相邻评论文本之间在语义上的隐含关联, 主要用于提高连续评论情况下的识别效果。尽管本文在实际应用中规定评论的对象应该是当前需要解决的方案, 而不是其他人的评论, 但是, 在多个团队成员对同一方案进行连续评论时, 前后相继的文本在一定程度上总会存在一定的语义关联, 如果能够将这种关联关系作为特征提取出来, 就应该可以达到比孤立评论文本更高的识别准确率。因此, 本文进一步提取了“上一评论文本的态度” 这一特征, 辅助对当前评论文本的态度倾向性判别。在当前评论紧密承接上一评论并且上一评论的态度倾向性比较明确的情况下, 可以向后续分类器中输入该特征, 其它情况下该特征变量输入恒为 0 表示当前评论文本的态度与上一评论无关。

(2) 小句数量: 是指评论文本中所包含的分句的数量, 本文以标点符号为界, 大多数标点符号如逗号、叹号、分号、句号等都可以在实际中用来分隔小句 (顿号、破折号、引号、书名号例外)。引入

“小句数量”这个特征可以更好的表征肯定否定特征词在评论文本中的分布,而这种分布通常可以反映当前文本的态度倾向.例如,评论文本信息“我认为在没有真正掌握 AddDataSet方法的情况下,不应该选用该方法进行图层的加载”,该文本包含两个否定词“没有”和“不应该”,这两个否定词分别分布于两个小句中,经过特征提取以后,其特征向量的前 6 项表示为 000020 显然,该文本与方案“用 AddDataSet方法加载图层”之间是否定关系;而另一条评论文本信息,“采用 AddDataSet的方法加载图层不是不可行”,该文本也包含两个否定词“不是”和“不可行”,不同的是,这两个否定词分布于同一个小句中,经过特征提取以后,其特征向量的前 6 项也表示为 000020 但是,该文本中的两个否定词却构成了双重否定,而表示肯定意义,从而后者与方案“用 AddDataSet方法加载图层”之间是肯定关系.上述情况,如果单纯依赖词频的方法而不引入特征词的分布特征,就难以区分,因此,本文引入了小句数量这个特征,以表征特征词在评论文本中的分布,从而提高识别的

准确率.

### 2 1 2 基于 BP神经网络的信息分类器的构建

本节采用三层 BP神经网络作为信息分类器.该神经网络包括三层:以特征提取阶段提取的特征向量作为网络的输入,共包括 8 个分量,其中 6 个用于输入肯定词和否定词特征,1 个用于输入上一评论文本的态度,1 个用于输入小句数量;输出为一个神经元,用 1 和 -1 来表示当前信息属于哪一类.隐含层神经元的数目是根据实验的效果来确定的,在本章中采用 4 个神经元可以得到很好的效果.该神经网络的逻辑结构如图 3 所示.其中  $X$  表示 8 个输入向量,分别对应 8 个特征项; $IW$  为从输入层到隐层的  $8 \times 4$  的权值矩阵; $b_1$  为隐层 4 个神经元的阈值;隐层用  $tansig$  函数作为神经元的激活函数; $LW$  为隐层与输出层之间的  $4 \times 1$  的权值矩阵; $b_2$  为输出层 1 个神经元的阈值;输出层用  $purelin$  函数作为神经元的激活函数; $Y$  为网络的输出向量,表示评论信息与方案的肯定或否定的关系.系统允许误差  $10^{-2}$ .

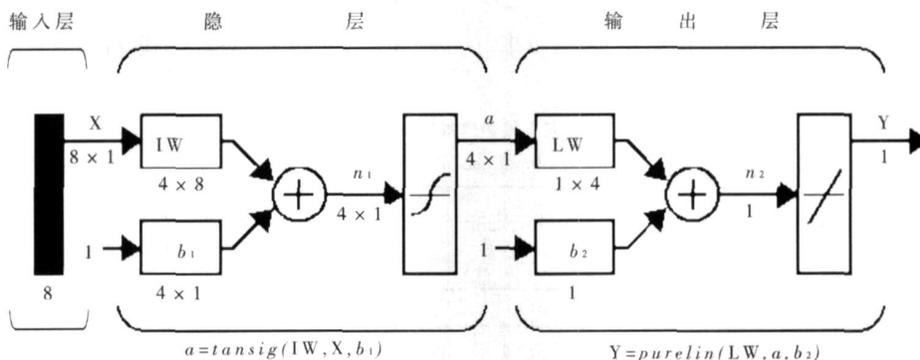


图 3 BP神经网络的结构图

Fig 3 Framework of BPNN

该网络的输入层有 8 个神经元:

\* 表示十分肯定意义的特征词权重:见表 1 特征词的权重等于特征词在文本信息中出现的频率,表示十分肯定意义的特征词权重等于十分肯定类的所有特征词在文本信息中出现的频率之和.由于本研究处理的是团队方案评论阶段的评论信息,这些信息包含的文字量一般较少,通常不

超过 100 字,特征词出现的频率之和一般也不会太大,所以,根据训练集文本信息的特征提取情况,本研究将表示十分肯定意义的特征词的权重取值范围定为  $\{0, 1, 2, 3, 4, 5\}$ .

\* 表示一般肯定意义的特征词权重:同理,表示一般肯定意义的特征词权重等于一般肯定类的所有特征词在文本信息中出现的频率之和,取值

范围为 {0 1 2 3 4 5}.

\* 表示轻微肯定意义的特征词权重: 同理, 表示轻微肯定意义的特征词权重等于轻微肯定类的所有特征词在文本信息中出现的频率之和, 取值范围为 {0 1 2 3 4 5}.

\* 表示十分否定意义的特征词权重: 同理, 表示十分否定意义的特征词权重等于十分否定类的所有特征词在文本信息中出现的频率之和, 取值范围为 {0 1 2 3 4 5}.

\* 表示一般否定意义的特征词权重: 同理, 表示一般否定意义的特征词权重等于一般否定类的所有特征词在文本信息中出现的频率之和, 取值范围为 {0 1 2 3 4 5}.

\* 表示轻微否定意义的特征词权重: 同理, 表示轻微否定意义的特征词权重等于轻微否定类的所有特征词在文本信息中出现的频率之和, 取值范围为 {0 1 2 3 4 5}.

\* 上一评论文本的态度: 该特征反映了在连续的评论文本中, 相邻评论文本之间在语义上的隐含关联, 更加丰富准确的特征信息有助于提高分类的准确性, 因此该特征的引入可以提高连续文本的识别效果, 达到比孤立评论文本更高的识别准确率. 在当前评论紧密承接上一评论并且上一评论的态度倾向性比较明确的情况下, 向分类器

中输入该特征. 输入变量范围是 {- 1 0 1}, - 1 表示上一评论文本对方案持否定态度, 1 表示上一评论对方案持肯定态度, 0 表示当前评论文本的态度与上一评论无关或者上一评论态度倾向不明显.

\* 当前评论信息中所包含的小句数: 指评论文本中所包含的分句数量. 取值范围为 {0 1 2 3 4 5}.

例如: 评论信息“我赞成用实例化 SeConnection 类的方法创建数据库链接加载图层.”, 按照上述特征和方法经过特征提取以后, 特征表示为 01000001

输出层有 1 个神经元, 如果评论信息与方案之间是肯定关系, 该神经元的输出取值为 1; 反之, 如果评论信息与方案之间是否定关系, 该神经元的输出取值为 - 1

## 2.2 信息关系自动识别模型的检验

基于神经网络的信息关系识别模型经过训练, 能否进行可靠的应用, 还需要经过检验样本的检验以验证其可行性和有效性, 本文选用 100 个样本作为训练结果的检验样本, 样本来源前面已经介绍过, 不再赘述. 表 2 列出了部分检验样本的输入, 表 3 列出了检验样本的期待输出和模型输出.

表 2 检验样本的输入

Table 2 Input data of test samples

样本	十分肯定特征词权重	一般肯定特征词权重	轻微肯定特征词权重	十分否定特征词权重	一般否定特征词权重	轻微否定特征词权重	上一评论文本的态度	评论信息包含的小句数
1	0	1	0	1	0	1	0	3
2	0	1	0	1	1	0	0	3
3	0	0	0	1	0	2	0	3
4	1	0	0	0	0	1	0	3
5	2	0	0	0	1	0	0	3
6	0	0	2	1	0	0	0	3
7	0	0	0	2	0	1	0	3
8	0	0	1	2	0	0	0	3
9	0	2	0	0	0	0	1	2
10	0	0	1	0	0	0	0	1
...	...	...	...	...	...	...	...	...

表 3 检验样本的输出

Table 3 Output data of test samples

样 本	期待输出	模型输出
1	- 1	- 1 000 0
2	- 1	- 1 000 0
3	- 1	- 1 000 0
4	1	1 000 0
5	1	0 999 9
6	- 1	- 1 000 0
7	- 1	- 1 000 0
8	- 1	- 1 000 0
9	1	- 1 000 0
10	1	0 950 8
...	...	...

### 2 3 信息关系自动识别模型的应用效果

比较训练样本的期待输出和模型输出, 以及表 3 检验样本的期待输出和模型输出, 可以得到如表 4 所示的实验结果.

表 4 实验结果

Table 4 Experimental results

识别结果	训练集	检验集
样本个数	100	100
正确率	100%	73. 33%

从表 4 的实验结果可知, 神经网络分类器训练集的准确率为 100%, 检验集的准确率为 73. 33%. 可见, 本研究所构建的模型的识别结果是较为满意的, 可应用在实际的信息自动分类中. 这种实际应用的效果可进一步通过查全率和查准率两个指标描述. 对于本研究而言, 查全率和查准率可如下计算

$$R_i = \frac{\text{第 } i \text{ 分类正确的评论信息}}{\text{未分类前属于该类的评论信息}} \times 100\% \quad (1)$$

$$P_i = \frac{\text{第 } i \text{ 分类正确的评论信息}}{\text{分类以后被分为该类的评论信息}} \times 100\% \quad (2)$$

$$\bar{R} = \frac{\sum_{i=1}^n R_i}{n} \times 100\% \quad (3)$$

$$\bar{P} = \frac{\sum_{i=1}^n P_i}{n} \times 100\% \quad (4)$$

其中,  $R_i$  为第  $i$  类评论信息的查全率,  $P_i$  为第  $i$  类评

论信息的查准率,  $\bar{R}$  为平均查全率,  $\bar{P}$  为平均查准率,  $n$  为类别数, 本研究中为两类, 故  $n = 2$

经过应用, 得到肯定关系类和否定关系类的查全率和查准率的结果如表 5

表 5 自动分类的结果

Table 5 Results of automatic classification

分类结果	查准率	查全率	平均查准率	平均查全率
肯定类别	76. 47%	72. 22%	75. 74%	75. 59%
否定类别	75%	78. 95%		

从表 5 的结果可见, 本文所构建的方案评论信息的关系识别模型在应用中的平均查准率为 75. 74%, 平均查全率为 75. 59%, 这表明该模型能够较好地支持团队创新过程中方案评论信息关系的自动识别, 从而支持团队创新任务解决方案的分析和归纳, 提高了团队创新过程信息组织的效率.

### 2 4 信息关系自动识别模型的使用过程

本文结合上述团队创新信息关系的自动识别方法, 设计并实现了团队创新信息关系的自动识别子系统. 该子系统是团队创新支持系统 (TCSS) 的重要组成部分之一, 它基于已经训练好的神经网络分类器, 可用于自动识别创新问题解决方案与评论信息的关系. 其识别后的结果会作为评论信息的重要属性, 用于团队信息的组织. 结合图 2, 信息关系自动识别模型用于团队信息组织的过程如下:

(1) 在进行团队信息组织之前, 首先是团队评论信息的产生, 产生的团队评论信息存储在团队创新支持系统 (TCSS) 中, 团队评论信息的产生及其存储过程如下:

在团队创新过程中, 团队成员在进行方案评论活动时所发表的评论信息存储在团队创新支持系统 (TCSS) 的团队记忆子系统中. 团队记忆子系统中的评论信息定义了“信息隶属的方案、信息名、信息提出人、信息内容”等要素, 匿名的团队活动中“信息提出人”可以缺省, 每条信息由计算机

自动编码确定其标识符,“信息提出时间”由团队记忆子系统自动记录并定义;

(2) 在对团队信息进行组织时,团队信息关系自动识别子系统从团队创新支持系统(TCSS)的团队记忆子系统中提取评论信息,作为待识别的目标信息;

(3) 运用该自动识别子系统的核心部分——基于神经网络的信息关系自动识别模型,识别评论信息与评论方案之间的关系.其使用方法如下:

(3.1) 利用词频统计程序对信息进行逐遍扫描.在此过程中,需对照剔除词表去除剔除词,并对照本章建立的肯定否定特征词表提取相应的词和短语,对提取的特征词进行计数,然后利用本文改进的特征提取方法构造原始文本信息的特征向量.

(3.2) 将提取的特征向量输入神经网络分类器,运用神经网络的方法识别评论信息与方案的关系,根据 ANN 的输出识别当前信息与评论方案之间的关系.

(4) 识别后的结果作为评论信息的重要属性用于信息组织,并重新更新进入到团队创新支持系统(TCSS).方案评论阶段的信息组织如图 4 所示.在图 4 中,首先以任务为中心,按照属于该任务的各个方案分别组织评论信息.图中首层为任务层;第二层为子任务层;第三层为方案层,方案与任务之间是隶属关系;最下面是评论信息层,评论信息与方案之间是肯定关系、否定关系和中立关系(本文将不具有明确的肯定或否定关系的评论均归入中立关系),其中,肯定关系和否定关系由本文所构建的信息关系自动识别模型进行识别.

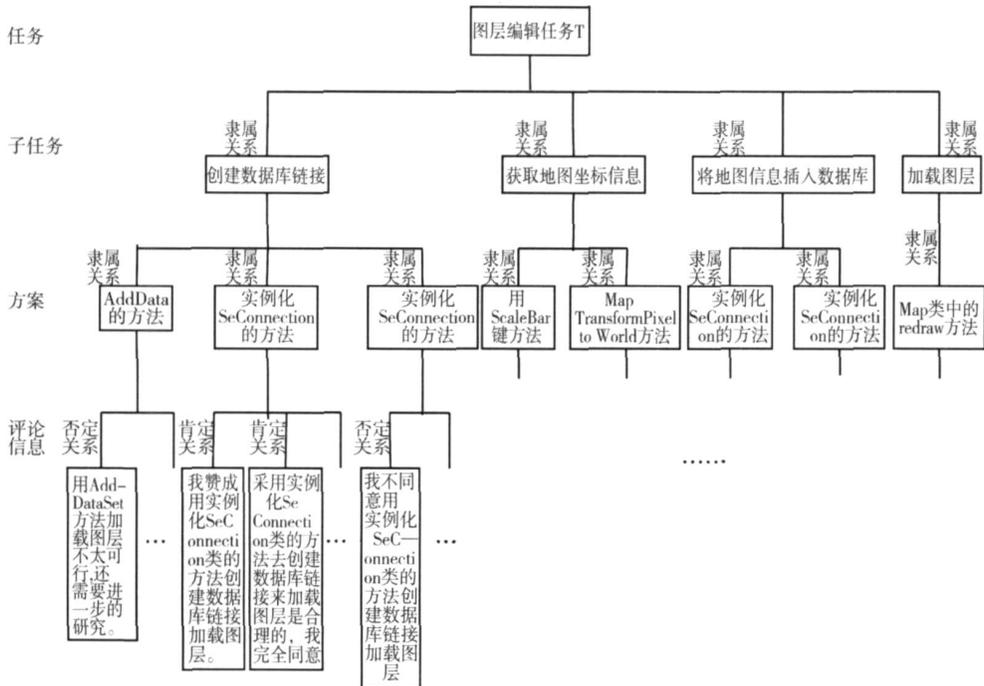


图 4 方案评论阶段的信息组织

Fig 4 Information organizing in the stage of solution comment

### 3 结 论

团队创新过程中一个基本的问题是如何对产生的海量信息进行有效的组织. 本文研究提出了团队创新信息中的评论信息与创新问题解决方案之间肯定和否定关系的自动识别方法, 并应用于实际的团队创新过程中信息关系自动识别模型的建立. 在该模型中, 评论文本的特征提取是信息关系自动识别的重要环节, 特征提取算法的优劣直接影响到识别的效果, 传统的词频特征提取方法用于解决本文的研究问题效果不是很好. 本文对该基本方法进行了改进, 体现在如下三方面: 首先, 按照语气轻重将肯定和否定特征词进行分级

处理, 既减少了特征向量维数, 从而提高了分类效果, 又不至于丢失太多的信息; 第二, 考虑到单纯依赖词频特征, 在表达评论信息的态度倾向性方面, 其精确度难以进一步提高, 因此有必要挖掘评论文本的其它特征以更好反映文本内容, 从而提高分类效果, 本文在上述“特征词六级分类”的基础上, 进一步增加“上一评论文本的态度”这个特征, 用于提高连续评论文本情况下的识别效果; 第三, 进一步增加“小句数量”这个特征, 可以更好的表征肯定和否定特征词在评论文本中的分布. 应用结果表明, 该模型较好地实现了评论信息与方案肯定和否定关系的自动识别, 可以辅助人对团队评论信息的整理和分析, 提高了团队创新过程信息组织的效率.

### 参 考 文 献:

- [1] Bolwin P T, Kumpe T. Manufacturing in the 1990s-productivity flexibility and innovation [ J]. Long Range Planning 1990, 23: 44-57
- [2] Michael P, Gallupe R B. Knowledge creation in focus groups: Can group technologies help? [ J]. Information & Management 2000, 38: 47-58
- [3] Schilling M A, Hill C. Managing the new product development process: Strategic imperatives [ J]. Academy of Management Executive 1998, 12(3): 67-81
- [4] Morgan D L. Focus Groups as Qualitative Research, Qualitative Research Methods [ M ]. Newbury Park: Sage Publications Series 16, 1988
- [5] Nunamaker J F, Briggs R O, Romano N C. Meeting Environments of the Future: Meeting to Plan Work or Meeting to Do Work? [ C ]. Proceedings of GroupWare'94, Europe 1994, 6: 521-545
- [6] Trauth E M, Jessup L M. Understanding computer mediated discussions: positives and interpretive analyses of group support system use [ J]. MIS Quarterly 2000, 24(1): 43-79
- [7] Grise M, Gallupe R B. Information overload: addressing the productivity paradox in face-to-face electronic meetings [ J]. Journal of Management Information Systems 1999-2000, 16(3): 157-185
- [8] 程少川, 王刊良, 席西民. 基于行为的思维分析和 DSS 中信息组织方式的研究 [ J]. 系统工程理论与实践, 1995, 11(11): 25-30  
Cheng Shaochuan, Wang Kanliang, Xi Youmin. On behavior-based thought analysis and information organization in DSS [ J]. Systems Engineering Theory & Practice, 1995, 11(11): 25-30 (in Chinese)
- [9] 程少川, 程向前, 张朋柱. 基于本体系统的知识管理人工智能研究 [ J]. 管理科学学报, 2003, 6(2): 17-22  
Cheng Shaochuan, Cheng Xiangqian, Zhang Pengzhu. Study on artificial intelligence of knowledge management based on ontology [ J]. Journal of Management Sciences in China 2003, 6(2): 17-22 (in Chinese)
- [10] Chen H, Schuffels C, Owig R. Internet categorization and search: A self-organizing approach [ J]. Journal of Visual Communication and Image Representation 1996, 7(1): 88-102

- [ 11] Chen H, Lynch K J. Automatic construction of networks of concepts characterizing document databases[ J]. *IEEE Transactions on Systems Man, and Cybernetics*, 1992, 22( 5): 885—902
- [ 12] Ellis C A, Gibbs S J, Rein G L. Groupware: Some issues and experiences[ J]. *Communications of the ACM*, 1991, 34( 1): 38—58
- [ 13] George J F, Easton G K, Nunamaker J F, *et al*. A study of collaborative group work with and without computer-based support[ J]. *Information Systems Research*, 1990, 1( 4): 394—415
- [ 14] Chen H, Hsu P, Orwig R, *et al*. Automatic concept classification of text from electronic meetings[ J]. *Communications of the ACM*, 1994, 37(10): 56—73
- [ 15] Roussinov D, Zhao J L. Automatic discovery of similarity relationships through web mining[ J]. *Decision Support Systems*, 2002, 1—18
- [ 16] 张朋柱, 汪应洛. DSS信息组织技术的现状与未来[ J]. *系统工程理论与实践*, 1995, ( 10): 18—24  
Zhang Pengzhu, Wang Yinglu. Present and future technique of information organization in DSS[ J]. *Systems Engineering Theory & Practice*, 1995, ( 10): 18—24 ( in Chinese)
- [ 17] Boston R P, Nagasundaram M. Research in Creativity and GSS[ C]. *IEEE Proc. 31st Annual Hawaii International Conference on System Sciences*, 1998, 391—404
- [ 18] 李欣苗, 张朋柱. 面向任务的团队创新支持理论及其系统研究[ D]. 上海: 上海交通大学, 2005  
Li Xinmiao, Zhang Pengzhu. Task-Oriented Team Creation Support Theory and System[ D]. Shanghai: Shanghai Jiaotong University, 2005. 96—130 ( in Chinese)
- [ 19] Briggs R O, Nunamaker J F, Sprague R H. 1001 unanswered research questions in GSS[ J]. *Journal of Management Information Systems*, 1998, 14( 3): 3—21.
- [ 20] 李欣苗, 张朋柱, 唐协平. 基于机械产品的团队创新支持系统研究[ J]. *中国机械工程*, 2004, 15( 16): 1479—1482  
Li Xinmiao, Zhang Pengzhu, Tang Xieping. Study on team creation support system for mechanical products[ J]. *China Mechanical Engineering*, 2004, 15(16): 1479—1482 ( in Chinese)
- [ 21] 李欣苗, 张朋柱. 综合集成过程中研讨信息记录模式的识别方法研究[ J]. *系统工程理论与实践*, 2005, 25( 2): 24—29  
Li Xinmiao, Zhang Pengzhu. Study on argumentation information record methods in metarsynthesis[ J]. *Systems Engineering Theory & Practice*, 2005, 25( 2): 24—29 ( in Chinese)
- [ 22] 庞剑锋, 卜东波, 白 硕. 基于向量空间模型的文本自动分类系统的研究与实现[ J]. *计算机应用研究*, 2001, ( 9): 23—26  
Pang Jianfeng, Bu Dongbo, Bai Shuo. Research and implementation of text categorization system based on VSM[ J]. *Application Research of Computers*, 2001, ( 9): 23—26 ( in Chinese)
- [ 23] 乔智勇, 刘志镜. Web数据挖掘系统的设计及实现[ J]. *计算机工程与设计*, 2002, 23( 7): 36—45  
Qiao Zhiyong, Liu Zhijing. Research on the design and implementation of web mining system[ J]. *Computer Engineering and Design*, 2002, 23( 7): 36—45 ( in Chinese)
- [ 24] 汪晓岩, 胡庆生, 李 斌, 等. 面向 Internet的个性化智能信息检索[ J]. *计算机研究与发展*, 1999, 36( 9): 1039—1046  
Wang Xiaoyan, Hu Qingsheng, Li Bin, *et al*. A system of personalized intelligent information retrieval for internet[ J]. *Journal of Computer Research and Development*, 1999, 36( 9): 1039—1046 ( in Chinese)
- [ 25] 牛忠兰, 陈跃新, 徐正同, 等. 网络文本自动分类系统的研究与设计[ J]. *微机处理*, 2002, ( 2): 41—43  
Niu Zhonglan, Chen Yuxin, Xu Zhengtong, *et al*. The study and design of the automatic documents classification system[ J]. *Microprocessors*, 2002, ( 2): 41—43 ( in Chinese)
- [ 26] 边肇祺, 张学工. 模式识别[ M]. 北京: 清华大学出版社, 2000  
Bian Zhaoqi, Zhang Xuegong. *Pattern Recognition*[ M]. Beijing: Tsinghua University Press, 2000 ( in Chinese)

## Automatic identification method of team creation information relationship and application

*LIX in-miao*<sup>1</sup>, *ZHANG Peng-zhu*<sup>2</sup>, *ZHANG Xing-xue*<sup>2</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** In the 1990s, creation was regarded as the main resource of competitive advantage for firms replacing efficiency and quality. A lot of information is produced in team creation. In this paper, the automatic identification method of the information relationship in team creation is researched and put forward. Furthermore, the method is applied to setting up the automatic identification model of information relationship in actual team creation process. In the model, the primary word frequency method is improved from the following three aspects: classifying feature words into six levels according to their tone; introducing the new feature of clause count and the new feature of the last comment's attitude. The results of application show that the method realizes the automatic identification of the relationship between solution and comment effectively. It can help team members to organize the large amount of team comments effectively and increase the efficiency of information organizing.

**Key words** team creation; information organization; information relationship; automatic identification