

经管
交大

群体研讨支持系统 及其文本分析方法

Group Argumentation Support Systems
and Related Text Analysis Methods

李 嘉 张朋柱 著



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

卓越管理论丛

国家自然科学基金项目(71171131, 71001038, 71371005, 71471064)和国家自然科学基金委创新研究群体(71421002)研究专著

群体研讨支持系统 及其文本分析方法

李 嘉 张朋柱 著

上海交通大学出版社

内容提要

群体研讨支持系统是一种支持群体研讨的信息系统,对于节约会议成本、缩短会议时间和提高决策效果具有重要意义。全书分上下两部分,上半部分重点介绍了群体研讨支持系统研究的最新进展,主要涉及群体决策中相关的任务类型,群体支持系统的设计,外部信息刺激对群体创新绩效的影响,以及在线群体研讨的信息组织结构设计。下半部分重点介绍了以言语行为分析为代表的群体研讨文本分析方法,主要涉及言语行为自动分类研究综述,言语行为分类体系设计框架,网络舆情冲突类言语行为分析,以及基于多阶段和分层方法的言语行为分类方法。

本书对于高等学校和科研机构从事群体研讨、群决策支持系统和研讨文本分析的研究人员具有理论参考价值,也可作为开发群件的系统分析员和软件工程师的参考书。

图书在版编目(CIP)数据

群体研讨支持系统及其文本分析方法/李嘉,张朋柱著. —上海:上海交通大学出版社,2015

ISBN 978-7-313-14290-0

I. ①群… II. ①李…②张… III. ①言语行为—决策支持系统 IV. ①H0-39

中国版本图书馆 CIP 数据核字(2015)第 309337 号

群体研讨支持系统及其文本分析方法

著 者: 李 嘉 张朋柱

出版发行: 上海交通大学出版社

邮政编码: 200030

出 版 人: 韩建民

印 制: 上海颀辉印刷厂

开 本: 787mm×960mm 1/16

字 数: 254 千字

版 次: 2015 年 12 月第 1 版

书 号: ISBN 978-7-313-14290-0/H

定 价: 45.00 元

地 址: 上海市番禺路 951 号

电 话: 021-64071208

经 销: 全国新华书店

印 张: 14.5

印 次: 2015 年 12 月第 1 次印刷

版权所有 侵权必究

告读者: 如发现本书有印装质量问题请与印刷厂质量科联系

联系电话: 021-57602918

前 言

随着信息化的发展和互联网的普及,在线研讨已经成为一种普遍现象,并且正在变得越来越流行。人们通过在线平台进行沟通,协作完成各种各样的群体任务。这些典型的在线研讨平台包括网络论坛、在线社区、BBS、即时通讯工具、邮件列表、新闻组、微博、Mooc等。这些不同形式的在线研讨平台,已经渗入了社会的方方面面,极大地改变了传统的工作模式。因此针对在线群体研讨的研究,具有重要的应用价值。

群体研讨支持系统是一种特殊的群体支持系统(GSS),重点支持与群体研讨相关的活动。在信息系统领域,关于群体支持系统的研究已经持续了30余年,并且在超过15年的时间内成为信息系统领域的研究热点。直到21世纪的第2个10年,群体研讨的相关研究仍然吸引了信息系统领域大量的学者,并得到国家自然科学基金委的持续资助。在国家自然科学基金项目(71171131,71001038,7137105,71471064)和国家自然科学基金委创新研究群体(71421002)的支持下,本书的作者对群体研讨支持系统及其文本分析方法进行了一系列较为深入的研究。本书的内容正是最近5年来相关研究的最新进展。

本书收录的研究大致遵循设计科学的思路展开,试图通过设计新的群体研讨支持系统或文本分析方法,实现系统的自动化和智能化。当然,这些研究在评估所设计的人造物时,也采用了一些行为研究的方法,如测量用户对系统的感知有用性、感知易用性、满意度等。

项目的研究凝聚了许多人的劳动和智慧。除了本书的著作者外,还有很多科研人员参与了课题的讨论和研究。其中,上海万慧科技有限公司的夏侯锋负责了Web 2.0环境下一种新型群体支持系统的开发工作。硕士生蒋御柱同学与作者一起开发了GASS 3原型系统,在研究思路给予作者很大帮助,并完成了第12章的研究内容。因篇幅有限,恕不一一赘述。

我们感谢美国南加州大学信息科学研究所高级教育技术研究中心的 Jihie

Kim 博士为本研究提供的语料和相应支持。

本著作的出版得到了国家自然科学基金项目(71171131,71001038,71371005,71471064)和国家自然科学基金委创新研究群体(71421002)的资助,在此表示衷心的感谢。

李 嘉 张朋柱

2015 年 11 月于上海

目 录

第 1 章 绪论	1
1.1 背景	1
1.2 本书主要内容	2
第 2 章 群体决策中任务类型的相关综述	8
2.1 群体决策任务的分类体系	8
2.1.1 McGrath 的分类体系	8
2.1.2 Campbell 的分类体系	11
2.2 群体决策任务的实现	14
2.3 群体决策任务的绩效指标及测量方法	16
2.4 与群体任务类型相关的理论	18
2.4.1 任务—媒介匹配理论	18
2.4.2 任务—技术匹配理论	19
2.5 总结与讨论	20
第 3 章 Web 2.0 环境下一种新型群体支持系统的设计研究	26
3.1 引言	26
3.2 文献综述	27
3.3 Web 2.0 环境下群体支持系统的新需求	29
3.4 一种新型群体支持系统的设计	30
3.4.1 系统概述和用户界面	30
3.4.2 适合海量数据的信息组织结构的设计	32
3.4.3 统计分析和自动摘要工具的设计	34
3.5 系统评估	37

3.5.1	系统功能有用性评价	38
3.5.2	决策效率	38
3.5.3	决策效果	39
3.6	结论	39
第4章	外部信息刺激对群体创新绩效的影响研究	43
4.1	引言	43
4.2	文献综述	44
4.2.1	GSS 环境下刺激群体创新的技术	44
4.2.2	决策引导	45
4.2.3	外部信息支持	46
4.3	研究模型和研究假设	47
4.4	研究设计	50
4.4.1	研讨平台系统	51
4.4.2	测量方法	52
4.4.3	被试	53
4.4.4	实验任务	53
4.4.5	实验过程	53
4.5	结果分析和讨论	54
4.5.1	变量操纵检测	54
4.5.2	研讨气氛	55
4.5.3	主意创新性	56
4.5.4	感知易用性	57
4.6	讨论	58
4.6.1	主要发现	58
4.6.2	理论贡献	58
4.7	结论	59
第5章	在线群体研讨的信息组织结构研究:基于任务—技术匹配的观点	63
5.1	引言	63
5.2	文献综述	65
5.2.1	研讨的线性和树状信息组织结构	65

5.2.2	线性和树状信息组织结构的比较研究	65
5.2.3	在线研讨系统中的任务—技术匹配研究	66
5.3	研究模型和研究假设	67
5.3.1	任务需求描述	69
5.3.2	技术能力描述	70
5.3.3	任务—技术匹配描述	73
5.4	研究设计	74
5.4.1	研讨平台系统	74
5.4.2	测量方法	75
5.4.3	被试	78
5.4.4	实验任务	79
5.4.5	实验过程	79
5.5	结果分析和讨论	80
5.5.1	决策效率	80
5.5.2	决策效果	80
5.5.3	决策过程满意度	81
5.5.4	决策结果满意度	81
5.6	讨论	82
5.6.1	主要发现	82
5.6.2	理论贡献	83
5.6.3	现实意义	84
5.7	结论	84
第 6 章	支持模糊型任务的信息组织结构设计研究	91
6.1	引言	91
6.2	文献综述	93
6.2.1	在线研讨的信息组织结构研究	93
6.2.2	群体支持系统中关于模糊型任务的研究	93
6.3	线性+树状信息组织结构的设计	95
6.3.1	基于任务—技术匹配的分析方法	95
6.3.2	线性+树状结构的设计	97
6.4	研究模型和研究假设	100

6.5	研究设计	102
6.5.1	基准系统	102
6.5.2	测量方法	102
6.5.3	被试	103
6.5.4	实验任务	103
6.5.5	实验过程	104
6.6	结果分析和讨论	104
6.6.1	决策效率	104
6.6.2	决策效果	105
6.6.3	决策过程满意度	105
6.6.4	决策结果满意度	106
6.7	讨论	106
6.8	结论	107
第7章	言语行为自动分类研究综述	111
7.1	引言	111
7.2	言语行为的基本理论	112
7.3	言语行为的分类体系	113
7.4	言语行为分类特征	115
7.4.1	语言学资源	117
7.4.2	处理资源	117
7.4.3	领域资源	118
7.5	特征选择方法	119
7.6	言语行为分类算法	120
7.6.1	有监督学习方法	121
7.6.2	无监督学习方法	125
7.7	言语行为分类的评价方法	127
7.7.1	有监督学习	127
7.7.2	无监督学习	127
7.8	总结与讨论	128

第 8 章 面向在线群体研讨的言语行为分类体系设计框架研究	134
8.1 引言	134
8.2 言语行为分类体系的研究框架	135
8.2.1 核心理论	136
8.2.2 基本需求	136
8.2.3 基本设计	136
8.2.4 可测试的假设	137
8.3 言语行为分类体系设计案例:以 E-learning 语料为例	138
8.3.1 语料简介	138
8.3.2 基础言语行为分类体系的选择	138
8.3.3 语料标注方法	139
8.3.4 言语行为分类体系	140
8.3.5 辨别能力分析	143
8.3.6 泛化能力分析	144
8.4 结论	144
 第 9 章 网络舆情冲突类言语行为分析研究	 148
9.1 引言	148
9.2 文献综述	149
9.2.1 言语行为分类体系	149
9.2.2 言语行为分类	151
9.2.3 冲突类言语行为类别构建	153
9.3 研究问题和假设	154
9.4 研究设计	154
9.4.1 系统总述	154
9.4.2 特征抽取	155
9.4.3 特征选择	156
9.4.4 分类器	156
9.5 系统评估	157
9.5.1 评估指标	157
9.5.2 实验结果	158
9.6 结论	159

第 10 章 基于多阶段转换学习的群体研讨文本言语行为分类	163
10.1 背景	163
10.2 言语行为类别	164
10.3 基于转换学习的言语行为分类	166
10.3.1 特征选取	166
10.3.2 算法实现	167
10.4 实验结果	172
10.5 结论	174
第 11 章 基于多阶段和分层方法的言语行为分类研究	176
11.1 背景	176
11.1.1 分层的言语行为分类体系	177
11.1.2 基于转换的学习	177
11.1.3 支持向量机	178
11.2 言语行为分类体系	179
11.3 特征选取	181
11.4 基于分阶段和分层方法的言语行为分类方法	183
11.4.1 参数选择的实验设计	184
11.4.2 各类别分类的难易程度	185
11.4.3 前指发言的层次	187
11.4.4 确定分类器运行顺序	188
11.5 系统评估	190
11.6 结论	191
第 12 章 群体研讨支持系统中研讨主题自动聚类及其可视化研究	195
12.1 背景	195
12.2 GASS 中自动化研讨主题聚类的挑战	196
12.3 相关研究	197
12.4 自动化研讨主题聚类的解决方案	198
12.4.1 自动索引	198
12.4.2 聚类分析	200
12.5 系统评估	203

12.5.1 实验设计	203
12.5.2 实验结果与讨论	205
12.6 结论及进一步的研究方向	206
第 13 章 总结与展望	209
13.1 主要工作和贡献	209
13.2 进一步研究展望	214
索引	216

第 1 章

绪 论

1.1 背景

群体支持系统(Group Support Systems, GSS)是用以支持智力型协作活动的计算机系统^[1]。例如,群体决策支持系统(GDSS)、电子会议系统(EMS)、计算机支持的协作活动(CSCW)等均属 GSS 的研究范畴。群体支持系统的诞生和发展是组织变化和技术进步的结果。电子会议室、局域网络、电话会议及决策支持软件等促进了对该领域的研究。此外,组织外部环境的根本变化也促进了群体支持系统的研究和应用。GSS 的研究和应用从 20 世纪 70 年代末至今先后经历了奠基阶段、探索阶段、实验研究阶段、实地研究阶段及目前正在开展的深入研究阶段。GSS 支持的群体活动包括:群体沟通、计划、主意产生、问题求解、论点讨论、谈判、冲突解决、系统分析与设计及协作的群体活动(如文稿准备)等。

群体研讨支持系统(Group Argumentation Support System, GASS)是一种特殊的群体支持系统(GSS),重点支持与群体研讨相关的活动。GASS 的概念提出和系统实现主要是在上海交通大学管理信息系统研究中心完成的。在过去 10 年的时间内,上海交通大学管理系统研究中心先后设计、开发、部署和实验了 3 个版本的 GASS,试图让 GASS 有更完善的功能和更广泛的应用领域。GASS 3 的基础功能开发在 2006 年 9 月完成,以国际主流群体支持系统 GroupSystems 为主要参考,同时其设计思路也受到了国内的综合集成研讨厅体系^[12-18]、基于互联网的开放式群体决策研讨过程^[19, 20]、研讨任务分解、电子公共大脑^[21, 22]、研讨信息组织模型^[23]、研讨分析中的共识与评价^[24]、研讨意见可视化分布与智能可视化^[25, 26]等理论的影响。

关于群体研讨的相关研究,从大的方面可以分成行为研究和设计科学两种。

行为研究关注在群体研讨、协作或决策过程中体现出来的个体或群体行为,或者在某种技术环境下用户体现出来的行为规律。设计科学关注如何设计一种人造物(系统、流程或算法),来更好地支持群体研讨、协作或决策。本书收录的研究大致遵循设计科学的思路展开,试图通过设计新的群体研讨支持系统或文本分析方法,实现系统的自动化和智能化。当然,这些研究在评估所设计的人造物时,也采用了一些行为研究的方法,如测量用户对系统的感知有用性、感知易用性、满意度等。

1.2 本书主要内容

本书可以分为上下两个部分。其中上半部分主要介绍群体研讨支持系统的最新进展,主要涉及群体决策中相关的任务类型,群体支持系统的设计,外部信息刺激对群体创新绩效的影响,以及在线群体研讨的信息组织结构设计。本书的下半部分重点介绍了以言语行为分析和聚类分析为代表的群体研讨文本分析方法,主要涉及言语行为自动分类研究综述,言语行为分类体系设计框架,网络舆情冲突类言语行为分析,基于多阶段和分层方法的言语行为分类方法,以及研讨主题自动聚类和可视化方法。

第2章介绍了群体决策中的任务类型。任务类型是群体决策中一个非常重要的变量,往往可以解释多达50%的群体绩效。然而,缺乏关于群体任务类型的相关综述,在一定程度上妨碍了国内学者理解群体任务的本质,并以此来解释群体绩效的变异或设计新的过程和技术来支持群体活动。为了填补这一空白,本章综述了与群体决策任务相关的几个问题:群体决策任务的分类体系,群体决策任务的实现,每类决策任务的绩效指标和测量方法,以及与群体任务类型相关的理论。

第3章介绍了Web 2.0环境下一一种新型群体支持系统的设计研究。传统的群体支持系统都是在Web 1.0时代开发的,无法满足Web 2.0环境的需求。本研究分析了Web 2.0的特点,将Web 2.0环境下群体支持系统的新需求归纳为4个方面:适合海量数据的信息组织结构、提供研讨统计分析和摘要的工具、支持多平台多应用访问和增强的用户体验。本章按照设计科学的方法,研究设计了一种新型群体支持系统,来满足这4个方面的需求。实验结果表明,用户对其中3项功能(信息组织结构、方案气泡图和研讨报告)都给予了较高的有用性评价。同时,与传统群体支持系统GASS相比,使用新型群体支持系统可以提高感知的决策效果。

第4章介绍了外部信息刺激对群体创新绩效的影响研究。外部信息刺激对群体创新造成的影响,既包括内容干涉的作用,也包括过程干涉的作用。这是因为向

群体提供外部信息的时候,必然会影响群体正常的研讨过程。但是,过去关于群体研讨外部信息支持的研究,只关注了内容干涉的影响,忽略了过程干涉的影响。本章以认知网络理论和解释水平理论为基础,提出了外部信息刺激的相关性(内容干涉)和刺激时机(过程干涉)如何影响群体研讨气氛、主意创新性以及感知易用性的研究模型。实验结果表明,提供相关刺激比提供无关刺激更能活跃研讨气氛,也更有利于提高用户的感知易用性。但是主意的创新性与外部刺激的相关性无关,只与外部刺激的提供方式有关。研讨过程中的任何刺激,不管刺激的相关性如何,对于提高群体的创新性都有正向作用。

第5章介绍了在线群体研讨的信息组织结构研究。虽然在线研讨平台已经广泛应用并受到学术界的大量关注,它的信息组织结构却还是停留在非常原始的阶段。关于哪一类信息组织结构更适合在线研讨的研究一直很少,仅有的一些研究也很零散,并且研究结果难以统一,因此迫切需要一个可以系统化地解释在线群体研讨中最优信息组织结构的理论。本研究从任务—技术匹配的观点出发,从沟通、协调组织和信息处理三个维度阐述了任务和信组织结构之间的匹配关系,并将其用于预测不同任务—技术匹配环境下的研讨绩效。我们设计了一个实验来验证这种匹配关系与研讨绩效之间的关系。实验结果表明,对主意产生型任务,采用线性结构的研讨平台比采用树状结构的研讨平台具有更高的研讨绩效;对于决策型任务,采用树状结构的研讨平台比采用线性结构的研讨平台具有更高的研讨绩效。本章所展示的匹配分析的理论和方法具有一定的扩展性,当有新类型任务到来或新技术被发明时,我们就可以用这种方法去分析新任务和新技术的匹配性,进而预测群体的绩效。

第6章对第5章的研究结果进行了拓展,介绍了一种支持模糊型任务的信息组织结构设计研究。模糊型任务虽然在组织中广泛存在,但是在线研讨环境下支持模糊型任务的信息组织结构研究还很少。本章从任务—技术匹配的角度分析,发现模糊型任务不能被现有的线性或树状信息组织结构有效支持,因此研究提出了一种新型线性+树状信息组织结构。这种新型结构能够在沟通、协调组织和信息处理三个维度上提供高水平的支持,从而满足模糊型任务的需求。本研究设计了一个实验来检验线性+树状结构的有效性。实验结果表明,对于模糊型任务,采用线性+树状结构的研讨平台比采用线性结构或树状结构的研讨平台具有更高的决策效率、更好的决策效果和更高的决策过程满意度。

第7章对言语行为自动分类研究进行了系统性的综述。言语行为自动分类对于对话系统、机器翻译和自动问答系统中问题理解和意图判断具有重要的意义,近

年来已经成为信息检索和自然语言处理领域的一个研究热点。本章较为系统地介绍了言语行为的基本理论、经典分类体系、分类特征、分类算法和评价方法。最后讨论了言语行为自动分类研究未来的研究方向。

第8章介绍了一种面向在线群体研讨的言语行为分类体系设计框架。言语行为分类对于大规模自动化处理研讨文本,实现面向群体研讨平台的自动化辅助具有至关重要的意义。缺乏言语行为分类体系及其构造方法论,已经成为国内言语行为自动化分析研究的一个主要障碍。本章以信息系统设计理论为指导,研究提出言语行为分类体系的设计需求,并以流程为人造物的观点,研究设计出一个包含五个关键步骤的迭代流程来满足设计需求。进一步地,以本章所提方法论为指导,以 E-learning 环境语料为案例,本章研究提出了一个适合在线问答研讨的言语行为分类体系,并验证了该体系的辨别能力和泛化能力。E-learning 语料环境的应用结果表明,本章所提的言语行为分类体系构造方法论具有较好的效果。

第9章介绍了网络舆情冲突场景下的言语行为分析方法。言语行为分析是网络舆情分析中一个重要的环节,对于机器大规模自动分析、识别和刻画网络舆情具有重要意义。国外言语行为分类体系和自动识别研究是基于英语使用规律的,以汉语为主要对象的研究尚未见报道,而中英文语言之间又存在着较大的差异,因此直接套用外语的言语行为分类研究将很难解决中国和中文环境下的问题。矛盾冲突则是网络舆情监管需要重点关注的一类任务,因此本章以网络舆情的冲突类言语行为分析为例,首先初步尝试建立了一个适合中文网络论坛环境的冲突类言语行为分类体系。紧接着研究提出了相应的自动化分类算法,并比较了不同特征集(提示词、n-grams、句法特征和结构特征)解决分类问题时的效果。实验结果表明,本章所提方法可以获得一个较为满意的分类效果,同时也证实新引入的句法特征和结构特征在解决言语行为分类问题上具有积极效果。

第10章介绍了一种基于多阶段转换学习的群体研讨文本言语行为分类方法。在群体支持系统(Group Support Systems)的环境下,群体能够在很短时间内产生大量研讨文本,远远超过了人们对信息处理的能力。因此,迫切需要一种能够自动分析和处理群体研讨文本的方法,言语行为分类就是这类方法中有可能实现并且具有应用价值的一个。本章在分析 Zeno 研讨模型的基础上,提出了适合群体研讨语料的言语行为分类体系。接着采用基于转换学习的办法,通过引入多阶段转换学习的概念,初步解决了群体研讨文本言语行为分类的问题,并且在议题类别和一些表达主张的类别(如支持和反对)上取得了较好的识别效果。研究群体研讨文本的言语行为分类对于拓展群体支持系统,进而研究和开发自动主持人系统具有重

要意义。同时,本研究也为在中文环境下解决其他类型研讨(如网络聊天室、即时聊天工具等)文本的言语行为分类问题提供了参考依据。

第11章对第10章的研究结果进行了拓展,介绍了一种基于多阶段和分层方法的言语行为分类方法。在面向在线研讨的言语行为分类研究中,前指发言类型(即与当前发言形成回复关系的前一条发言的言语行为类别)是一个非常重要的特征。但是由于在测试集上前指发言类型是未知的,因此如何合理利用前指发言的类型信息就成为一个非常关键的问题。本章以 E-learning 语料为例,证实了发言类型的高低层次和分类器的运行顺序对分类效果都有较为显著的影响,并提出了一个基于多阶段分层的分类方法,可以给出一个合理的前指发言层次和分类器运行顺序。在盲测集上的运行结果证实了这组优选的参数能够稳定一致地提高 E-learning 语料上各言语行为类别的分类效果。

第12章介绍了群体研讨支持系统中研讨主题自动聚类及其可视化方法。群体研讨支持系统(Group Argument Support Systems, GASS)的匿名、并行输入及自动化记录群体发言的特征,在辅助群体产生大量有价值观点的同时,也常常导致“信息过载”和“知识断层”。本章介绍了一个自动化聚类工具来增强群体的认知能力并提高电子会议的效率。首先识别了 GASS 环境下自动化主题聚类的一些挑战并回顾了相关研究,结合 GASS 的研讨模式、研讨文本特征及中文文本分析的要求,给出了中文分词、停词表处理以及有效词语识别的文本分析技术;提出基于主题分析的特征向量选择方法,并基于自组织映射的神经网络(Self-Organization Map, SOM)思想,用 Java 语言设计并开发了一个自动聚类工具。实验表明,该工具可以达到 0.28 的聚类准确率和 0.35 的聚类全面率,产生 0.83 的聚类错误率。

参考文献

- [1] 王刊良,席酉民,汪应洛. 群体支持系统理论基础[J]. 管理科学学报,1994(4):1-6.
- [2] DeSANCTIS G, Sambamurthy V, Watson R T. COMPUTER-SUPPORTED MEETINGS-BUILDING A RESEARCH ENVIRONMENT [J]. Large Scale Systems in Information and Decision Technologies, 1987,13(1):43-59.
- [3] Banker R D, Kauffman R J. 50th anniversary article: the evolution of research on information systems: a fiftieth-year survey of the literature in management science [J]. Management Science, 2004,50(3):281-298.
- [4] Bajwa D S, Lewis L F, Pervan G, Lai V S, Munkvold B E, Schwabe G. Factors in the global assimilation of collaborative information technologies: an exploratory investigation in five regions [J]. Journal of Management Information Systems, 2008,25(1):131-166.