

南京航空航天大学
论文集

(二〇一〇年) 第35册

经济与管理学院

(第1分册)

南京航空航天大学科技部编

二〇一一年五月



NUAA2011039792

Z427
1033 (2010) - (35)

经济与管理学院

教师



2011039792

目 录 (教师)

序号	姓 名	职 称	单位	论 文 题 目	刊物、会议名称	年、卷、期	类别
1	张明宝	副教授	092	领域本体在信息检索中的应用研究	情报学报	2010.29.2	
2	胡明礼	讲师	091	Study on Evolution Mechanism of Fateful Stampede Accident based on Graphical Evaluation and Review Technique	2010 IEEE International Conference on System, Man and Cybernetics	2010	
3	周鹏	教授	091	Total factor carbon emission performance: A Malmquist index analysis	Energy Economics	2010.32.1	
4	周鹏	教授	091	Weighting and Aggregation in Composite Indicator Construction: a Multiplicative Optimization Approach	Social Indicators Research	2010 .96.1	
5	周鹏	教授	091	Data aggregation in constructing composite indicators: A perspective of information loss	Expert Systems with Applications	2010 .37.1	
6	周鹏	教授	091	An extension to data envelopment analysis with preference structure for estimating overall inefficiency	Applied Mathematics and Computation	2010 .216.3	
7	周鹏	教授	091	Measuring economy-wide energy efficiency performance: A parametric frontier approach	2010 International Conference on Applied Energy, Singapore		
8	李珊	讲师	092	个性化服务中用户兴趣建模与更新研究	情报学报	2010. 29. 1	
9	李珊	讲师	092	prediction of cyclosporine blood concentration and dosage in kidney transplantation patients	comparison of multiple linear regression and neural networks	2010	
10	李珊	讲师	092	基于Web的国防科研情报采集系统设计与实现	2010国防科技信息学会情报学术专业委员会研讨会 (会议论文)	2010	
11	徐海燕	教授	091	Using Matrices to Link Conflict Evolution and Resolution in a Graph	European Journal of Operational Research	2010年207卷	SCI
12	徐海燕	教授	091	Matrix Representation and Extension of Coalition Analysis in Group Decision Support	Computers and Mathematics with Applications	2010年60,卷	SCI
13	徐海燕	教授	091	An Integrated Algebraic Approach to Conflict Resolution with Three-level Preference	Applied Mathematics and Computation	2010年216卷	SCI
14	徐海燕	教授	091	Combining Strength and Uncertainty for Preferences in the Graph Model for Conflict Resolution with Multiple Decision Makers	Theory and Decision	2010年69卷4期	SCI
15	王英	副教授	095	劳动密集型出口经济、FDI与中国经济发展	财贸研究	2010, (4)	
16	王英	副教授	095	基于灰色关联理论的FDI和中国区域经济发展差距研究	系统工程理论与实践	2010, 30 (3)	
17	王英	副教授	095	中国货物贸易对于服务贸易的促进作用	世界经济研究	2010, (7)	
18	王英	副教授	095	FDI对于中国服务部门产业内贸易的影响	价格月刊	2010, (3)	
19	王英	副教授	095	影响中国服务贸易国际竞争力的内外因素比较	企业经济	2010, (3)	
20	王英	副教授	095	加入WTO后我国服务贸易竞争力影响因素分析	黑龙江对外经贸	2010, (4)	

21	章玲 周德群	副教授	091	A study on R-OWGA operators	IEEE International Conference on Grey Systems and Intelligent Services	2010	
22	查冬兰 周德群	副教授	091	Driving forces of residential CO ₂ emissions in urban and rural China: An index decomposition analysis	Energy Policy	2010, 38 (7)	
23	查冬兰	讲师	95	基于CGE模型的中国能源效率回弹效应研究	数量经济技术经济研究	2010. 27. 10	
24	周德群 孙立成	教授	091	考虑石油替代品和关税配额政策的石油储备模型研究	中国管理科学	2010, (01)	
25	张浩	副教授	093	消费者行为研究新范式及基于行为主体的计算营销学评介	外国经济与管理	2010. 3	
26	赵湘莲	教授	094	董事会特征对自愿信息披露影响的实证研究	中国会计学会2010年度学术会议	2010	
27	赵湘莲	教授	094	寿险公司投资连接保险收益研究——基于投资者视角	中国财务管理学会2010年度学术会议	2010	
28	李宗植	教授	096	我国技术进步、技术效率的地区差异分析	商丘师范学院学报	2010. 1	
29	吴和成	教授	091	指数型贮备系统可靠性的精确置信下限	系统工程学报	2010. 25. 3	
30	吴和成	教授	091	制造业R&D效率测度及对策研究——基于中国17个制造行业的数据	科研管理	2010. 31. 5	
31	党耀国	教授	091	Study on Modeling Method for Sequences with Law of Grey Exponent	International Conference on Systems, Man and Cybernetics	2010	
32	陈毅然	教授	090	管理工效学的创新思路	人类工效学	2010. 16. 4	
33	耿成轩	教授	094	基于内生结构与外部环境的家族企业融资行为研究	管理世界	2010. 5	
34	刘益平	教授	094	灰色聚类方法在创业板股票投资价值中的应用	经济师	2010. 8	
35	刘益平	教授	094	新会计准则对上市公司综合财务状况的影响	价值工程	2010. 11	
36	刘益平	教授	094	双倍余额递减法下年折旧费用倒置问题探索	财会通讯	2010. 1	
37	马晓平	副教授	095	近阶段食用油价格上涨的原因分析及对策探讨	价格月刊	2010. 3	
38	陈其霆	副教授	095	An Empirical Study of Fiscal Support Policy on farmers	2010年International Conference on E-Business	2010	
39	陈其霆	副教授	095	江苏省城市规模分布的演化分析	生态经济,	2010. 2	
40	罗正军	讲师	092	基于灰色关联的制造型企业ERP实施过程评价研究	价值工程	2010. 29. 5	
41	罗正军	讲师	092	制造型企业ERP系统实施后评价研究	统计与决策	2010. 2	
42	柯宏发	博后	091	基于灰关联度的多目标规划新求解算法【EI: 20101712897061】	系统工程与电子技术	2010. 32.3	
43	柯宏发	博后	091	无人机飞行轨迹的实时显示预测算法【EI: 20105113504386】	第29届中国控制学术会议	2010	
44	柯宏发	博后	091	基于灰色模型的电子装备故障预测精度分析	Proceedings of CCAST(World Laboratory) Workshop, Grey System Theory and its Applications	2010	
45	谢乃明	副教授	091	Novel methods on comparing grey numbers	Applied mathematical Mod	2010. 34(2)	
46	谢乃明	副教授	091	NDGM模型的性质及预测效果分析	系统工程与电子技术	2010. 32 (9)	
47	陈晔	教授	091	Dynamic Multiple Criteria Decision Analysis with Application in Emergency Management Assessment	Proceedings of 2010 IEEE International Conference on Systems, Man and Cybernetics	2010	

48	张钦	副教授	091	1990-2007年中国矿产资源型城市全要素生产率的动态实证分析	系统工程	2010. 28. 202	
49	张钦	副教授	091	带有时间窗约束的车辆路径问题的一种改进算法	系统管理学报	2010. 19. 5	
50	张钦	副教授	091	江苏省秸秆发电现状分析与对策	中国软科学	2010. 1	
51	朱建军	教授	091	基于区间偏好的政府节能政策贡献度测算模型	统计与决策	2010. 1	
52	陈洪转	副教授	091	Category of inter-grey non-symmetric evolutionary game chain model of supervision on research funds of colleges and universities	IEEE International Conference	2010. 1	
53	陈洪转	副教授	091	科技工作者视角下的高校科研经费使用问题与对策研究	科技进步与对策	2010. 27. 21	
54	彭 灿	教授	093	知识资本、社会资本与研发团队的有效性：关系模型与理论假设	技术经济	2010年10期	
55	彭 灿	教授	093	团队智力资本、团队社会资本及其关系研究	现代管理科学	2010年10期	
56	菅利荣	教授	091	杂合灰色聚类与扩展优势粗集的概率决策方法	系统工程学报	2010. 4	
57	任君卿	副教授	91	Game Analysis of Cooperative Advertising Based on "One to Many" Supply Chain Model	ICSSSM 10 International Conference Service System and Service Management		
58	庄品	副教授	091	Coordination Mechanisms in a Dominant-Retailer Supply Chain with Asymmetric Information and Disruption	the 2 nd international conference	2010	
59	庄品	副教授	091	考虑制造商质量水平的供应链质量契约研究	物流技术	2010. 29. 20	
60	李帮义	教授	093	基于客户感知质量差异的再制造优化决策	南京航空航天大学学报（自然科学版）	2010年06期	
61	李帮义	教授	093	作为阻止战略的再制造决策研究	控制与决策	2010年11期	
62	楚岩枫	副教授	091	The Analysis on the self-adaptation and the stability of the Microscopic subject in the logistics industry	会议论文	2010	
63	米传民	副教授	092	Empirical study on relationship between wuxi city's S&T investment and economy growth based on grey incidence analyzing	会议论文	2010	
64	李敏	副教授	093	文化产业全球价值网络的运作特性及功能研究	工业技术经济	2010年第8期	
65	李敏	副教授	093	企业国际营销体系的要素构成、体系优势及价值创造	2010年江苏省社会科学联合会	2010年	
66	李敏	副教授	093	企业国际营销体系的要素构成、体系优势及价值创造	2010年江苏省社会科学联合会	2010年	
67	刘思峰	教授	091	一种新型多目标智能加权灰靶决策模型	控制与决策	2010. 25. 8	
68	刘思峰	教授	091	基于核和灰度的区间灰数运算法则	系统工程与电子技术	2010. 32. 2	
69	刘思峰	教授	091	基于相似性和接近性视角的新型灰色关联分析模型	系统工程理论与实践	2010. 30. 5	

doi:10.3772/j.issn.1000-0135.2010.02.004

领域本体在信息检索中的应用研究¹⁾

张明宝 马 静 施秀丽

(南京航空航天大学经济与管理学院信息管理与电子商务系, 南京 210016)

摘要 为了改进基于关键词的信息检索方法的局限性, 论文研究了一种综合利用领域本体改善信息检索性能的方法。该方法强调通过交互式的方式引导用户一步步逼近其真实的、潜在的检索需求, 使用基于编辑距离的词汇匹配方法辅助用户查询本体词汇, 使用基于概念空间的检索词联想方法帮助用户扩充检索词。使用基于领域本体的词义识别算法来确定文档中的词汇词义。使用 XML 技术实现用户查询需求和文档标注的规范化标注。实验表明, 该方法会有效提升查全率并且会改进查准率。

关键词 信息检索 领域本体 概念空间

Research on Field Ontology Applications in Information Retrieval

Zhang Mingbao, Ma Jing and Shi Xiuli

(Department of Information Management and Electronic Business, College of Economy and Management,
Nanjing University of Aeronautics and Astronautics, Nanjing 210016)

Abstract In order to improve the performances of keywords-based information retrieval method, a method of using field ontology in information retrieval is introduced. It focuses on guiding user to find their actual and potential search requirements step by step which depends on two methods, one is using edit distance based method to search word in ontology base and the other is using concept space based method to help user to extend key words. Moreover it utilizes the relations between words built by filed ontology to find word's actual meanings in it's context, and finally one XML scheme is described simply which is used to index key words selected by user and processed document. Experiment results show that this method can improve precision ratio to a certain extent, but recall ratio greatly.

Keywords information retrieval, ontology, concept space

1 前言

传统的基于关键词的信息检索方法难以获得令用户满意的查准率和查全率, 这主要是由于用户查询需求模糊、自然语言中的“一词多义”和“一义多词”现象以及文档标引的不完善等问题所引起。为

了解决这些问题, 基于语义的信息检索研究逐渐兴起。

基于词典的信息检索研究是其中开展得较早, 并且成果较为丰富的一个方向。文献[1]以 WordNet 为基础研究了语义处理技术在英文信息检索过程中的应用问题, 文献[2]提出基于 HowNet 来研究中文信息检索问题。在查询扩展方面, 文献[3]提出通过

收稿日期: 2008年11月26日

作者简介: 张明宝, 男, 1973年生, 博士, 南京航空航天大学副教授, 主要从事企业集成、信息检索、知识挖掘等方面的研究。E-mail: zmb7381@126.com。马静, 女, 博士, 南京航空航天大学副教授, 主要从事信息检索和知识挖掘的研究。施秀丽, 女, 南京航空航天大学经济与管理学院硕士研究生, 研究方向: 知识挖掘。

1) 基金项目: 南京航空航天大学引进人才基金(1009-234039); 某国防技术基础项目。

建立概念空间来自动扩展用户的检索词。在文档准确标引方面,文献[4]提出使用词汇链技术实现词汇语义的聚合,以过滤无关词汇,文献[5]研究了词汇链的不同构建方法。

近年来,随着本体论技术的快速发展,将本体引入信息检索正在成为研究的新热点。文献[6]利用多智能体和本体论设计了一种信息检索服务器,它能利用本体对文档进行领域分类,同时对用户的查询信息进行规范。文献[7]开发了一个基于本体的智能化知识检索原型 Kretreival,它依据本体组织领域知识,实现动态启发式概念扩展算法和相关反馈,对情报学和人工智能领域的文献知识进行检索。文献[8]介绍了一个基于本体的信息检索系统 MELISA,用于在医疗专业领域检索参考文献。文献[9]提出了一个基于本体论的信息检索模型,该模型能够实现用户查询的导引和目标文档的标注,但是该模型还是停留在关键词匹配查询的基础上,没有有效发挥本体的作用。文献[10]提出了一种将本体应用于信息检索的具体方法,但是该方法比较粗糙,同时该方法没有提及本体对查询需求的作用。文献[11]研究了本体对查询需求的作用,并且提出了基于本体的查询策略和查询方法。文献[12]研究了在信息检索过程中通过利用领域本体实现语义消歧来改善信息检索性能的方法,实验数据显示通过词义消歧可以有效提高查准率。

综合来看,目前基于本体论的信息检索研究主要从基于智能代理的体系结构、查询需求处理、文档标引以及基于本体的推理能力应用等方面单独展开。但是,从多方面应用本体技术实现完整的信息检索方法的报道还很少见。本文在借鉴上述工作的基础上,提出一种将领域本体综合应用于信息检索过程的方法,该方法强调使用交互的方式帮助用户选择本体词汇准确描述其检索需求,使用本体词汇之间的逻辑关系实现词义消歧。

2 关于领域本体在信息检索中应用的几个基本观点

领域本体是对领域共享概念的明确的、形式化的和规范化的说明,它的目标在于捕获相关领域的知识,提供对该领域知识的共同理解,确定该领域内共同认可的词汇,并且按照层次关系给出这些词汇和词汇之间相互关系的明确定义,以实现软件系统对这些概念的共享和重用。

2.1 现阶段领域本体在信息检索中应用的特点

通过形式化和规范化地表示领域概念,本体获得了两大优势:①通过规范化的概念表示,实现了软件系统之间、人与软件系统之间对这些概念的共同理解,这对实现语义级别的机-机和人-机信息交换意义重大;②通过形式化的概念表示,实现了机器自动求解的推理过程,这对智能化的问题求解意义重大。

当我们将本体技术引入信息检索领域,一个重要的问题便一直困扰着研究者,即本体对信息检索到底起到什么样的作用。笔者认为,在信息检索中使用领域本体,现阶段由于无法解决准确地、自动化地实现文档标引和知识抽取等问题,所以基于本体的推理对信息检索的作用是发挥不出来的或者说是很受局限的。当前,在信息检索中引入本体最大的好处在于:通过领域本体规范了用户和计算机系统对词汇所表达概念的理解,消除了人-机语义级别的理解分歧。从这一点来说,领域本体的作用和传统的分类系统、语义词典、词表等没有什么本质的区别。从应用的角度来说,由于伴随本体技术研究的深入,出现了各类本体构建工具和领域本体,使得传统分类系统、词典、词表所不能覆盖的范围可以使用本体工具和本体构建方法快速构建相应的领域本体来覆盖,这是现阶段使用本体最主要的优势。

将领域本体引入信息检索过程会极大地提高信息检索的性能,主要体现在如下几个方面:首先,由于本体包含了对领域对象的规范化描述,所以,当用户无法明确其检索需求时,可以使用本体词汇引导用户逐步使其需求清晰化;其次,由于本体词汇描述中包含了该对象的同义词,所以,检索过程中的“一义多词”问题很容易解决;第三,本体词汇间的关系描述了词汇语义之间的联系,可以利用这种关系来确定文档中词汇的真实含义;第四,本体词汇间的语义联系会反映词汇在应用时的共现倾向,可以利用这种特征来过滤掉文档中不相关的词汇,以减少无关词对检索结果的负面影响。

2.2 基于领域本体的信息检索与基于词典的信息检索的关系

首先,二者是有密切联系的。领域本体与自然语言词典要解决的问题是类似的,它们在本质上是一致的。既然,基于语义词典的信息检索研究已经表明使用规范化的词表词汇可以明显地提高信息检

索的查准率和查全率^[1,2],那么,将领域本体应用于信息检索也应具有同样的作用。基于词典的信息检索方法对研究基于本体的信息检索方法具有重要的借鉴价值。

其次,二者又是有显著区别的。自然语言词典收录的词汇局限于公共的自然语言词汇,缺乏领域词汇,所以,它远远不能满足各类专业技术或研究领域的信息检索需要。由于词典和本体的构建方法有很大的不同,所以,基于词典的信息检索方法不能直接移植到基于本体的信息检索方法之中。此外,基于本体的信息检索一般应用于专业领域,这些领域的信息检索过程往往属于“深度”检索,检索者需要全面、精确地表达需求,倾向于追求更高的查准率和查全率,而不太关注检索过程是否简单。这些都需要我们根据领域本体的特征来研究合适的信息检索过程和所使用的方法。

最后,二者必须要综合应用。任何专业领域的文档都是由自然语言词汇和各类领域词汇所构成。所以,将自然语言词典和领域本体综合应用于信息检索是不可避免的。在研究基于领域本体的信息检索算法时要考虑到这一特征,设计适合自然语言词典和各种领域本体配合使用的可行算法。

2.3 领域本体的不完备性对信息检索的影响

当前,领域本体的构建主要采用人工或半自动的方式进行,领域本体的不完备性一直是困扰其广泛应用的主要问题。我们不能等到构建领域本体的技术成熟才开始应用本体技术。将领域本体应用于信息检索必须要考虑其不完备的特征。所设计的算法既要立足于可以获得的领域本体,又要能够充分发挥它的作用实现对信息检索系统性能的改进。

3 基于领域本体的信息检索的关键问题

将领域本体应用于信息检索必须要充分结合信

息检索技术的特点来研究领域本体的使用方法。不管什么样的信息检索过程最终还是要依赖于词形、词位等语法层面的特征匹配和综合处理来实现。所以在信息检索领域应用领域本体,不仅要注重领域本体所表达的内容,还要特别重视领域本体表达的形式。按照上述几个基本观点,我们建立了如下基于领域本体的信息检索的形式化描述。

定义 1 领域本体结构 O 可以表示为一个三元组: $O = \{C, R, A^0\}$ 。其中, C 表示概念的集合; R 表示关系的集合; A^0 表示本体公理的集合。

定义 2 按照领域本体结构 O 来表达某一具体领域本体的词汇空间 L 可以表示为一个四元组: $L = \{L^C, L^R, F, G\}$ 。其中, L^C 是用来表达领域本体概念的词汇的集合; L^R 是用来表达领域本体关系的词汇的集合; F 是建立 L^C 与 C 中实体之间对应关系的映射的集合; G 是建立 L^R 与 R 中实体之间对应关系的映射的集合。

需要说明一点, L^C 与 C 以及 L^R 与 R 之间的映射并不要求严格一一对应。即可以存在若干词汇表达某一概念或关系,也可以存在若干概念或关系都可以以某一词汇表达。前者即“一义多词”现象,后者即“一词多义”现象。

定义 3 领域本体 ψ 可以表示为一个二元组: $\psi = \{O, L\}$ 。其中, O 是领域本体结构; L 是用于表示领域本体的词汇空间。

定义 4 用户的查询需求 Q 是对领域本体的一个映射,可以表示为一个二元组: $Q = \{W^Q, V^Q\}$ 。其中, W^Q 是检索词汇的集合,满足 $W^Q(w) = \{\forall w \in W^Q: w \in L^C \vee w \in L^R\}$; V^Q 是检索词汇与领域本体的关系的集合,满足 $V^Q(v) = \{\forall v \in V^Q: v \in F \vee v \in R\}$ 。

定义 5 目标文档 D 是对领域本体的一个映射,可以表示为一个二元组: $D = \{W^D, V^D\}$ 。其中, W^D 是文档特征词汇的集合,满足 $W^D(w) = \{\forall w \in W^D: w \in L^C \vee w \in L^R\}$; V^D 是文档特征词汇与领域本

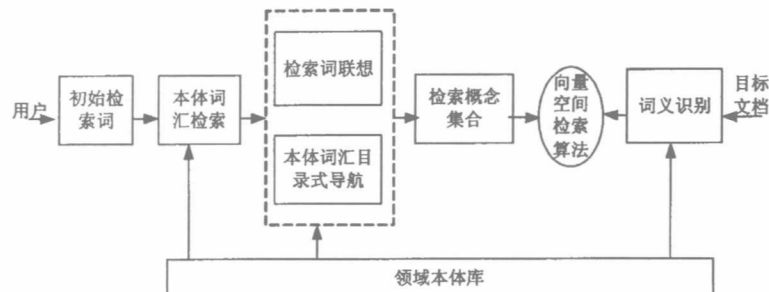


图1 基于本体的信息检索过程

体的关系的集合,满足 $V^D(v) = \{ \forall v \in V^D : v \in F \vee v \in R \}$ 。

定义6 基于领域本体的信息检索方法 *Ontology Based Retrieval* 可以表示为一个三元组: $Ontology Based Retrieval = \{ T^Q, T^D, M^{Q \rightarrow D} \}$ 。其中, T^Q 表示将用户实际查询需求转换为使用领域本体表示的查询需求 Q 的方法的集合; T^D 表示将原始文档转换为使用领域本体表示的目标文档 D 的方法的集合; $M^{Q \rightarrow D}$ 表示实现 Q 和 D 之间匹配的算法的集合。

定义7 匹配算法集合 $M^{Q \rightarrow D}$ 包含任何基于词形匹配的信息检索匹配算法。

上述形式化定义揭示了基于领域本体的信息检索方法的特征,即基于领域本体的信息检索方法是以传统信息检索算法为基础的,其要解决的关键问题在于实现具体的 T^Q 和 T^D 。

具体来说,将领域本体应用于信息检索必须要解决如下两个关键问题:

首先,如何用领域本体中的词汇及其含义来准确地描述用户的检索需求。随着领域本体的规模扩大,用户使用本体词汇将变得更加困难,所以需要建立专门的工具辅助用户快速地选择所需要使用的本体词汇,需要采用规范化的方法准确表达用户所选词汇的确切含义。

其次,如何使用领域本体中的词汇及其含义来准确地描述文档的主题。文档中词汇所表达的语义是与其上下文密切相关的,所以需要研究基于领域本体的词义识别算法。由于文档中会包含大量与文档主题不相关的词汇,这些词汇会降低信息检索的性能,所以需要研究文档词汇过滤算法以滤除这些不相关词汇。在此基础上,还需要解决使用规范化的方法准确表达文档主题的方法。

4 一种基于领域本体的全文信息检索方法

为了解决上述问题,建立了图1所示的信息检索过程框架,该框架提供了一种确定 T^Q 和 T^D 的解决方案。该方法从两个方面强调了领域本体的作用:一方面,使用交互的方式帮助用户选择本体库中的相关词汇(表达确切含义)来准确描述其检索需求;另一方面,使用本体库中词汇(表达确切含义)之间存在的逻辑关系实现文档词汇的词义识别。使用该方法,用户的检索需求和目标文档的标引都使用一系列本体库中具有确切含义的词汇的集合来描

述,所以,仍然可以采用传统的向量空间检索模型来实现检索需求与目标文档的匹配计算。

查询扩展和文档反馈等方法强调自动地扩展用户查询关键词来提高检索系统的查全率,但会明显降低查准率。该方法不太适用于基于领域本体的用户查询扩展^[13]。首先,基于领域本体的信息检索适用于专业领域,这些领域往往不具备大量的文档集合或者检索历史经验用于运算;其次,用户的查询需求更加具有个性化的特征;最后,这些领域的信息检索更加强调查准率。所以,使用基于人机交互的查询扩展方法更加适合于基于本体的信息检索过程。

图1所示的方法中,使用两个步骤来实现对用户输入的检索词的扩展。第一步,根据用户输入的检索入口词,使用词形相似度计算方法来定位与检索入口词最相似的本体词汇,即从本体库中选择出与该入口词最为相似的本体词汇供用户选择;第二步,一旦用户选择了某一词汇,系统使用两种方法来呈现相关联的本体词汇。一是使用目录导航方式以用户确定的本体库节点为出发点,可以选择向上级展开,也可以向下级展开,从而选择所需的相关词汇;二是使用基于概念空间的检索词联想方法,系统自动推荐与入口词汇相关的其他本体词汇,用户通过选择确定所需词汇。重复使用上述两个过程,可以帮助用户快速遍历本体库寻找到所需要的本体词汇集合来准确、全面地描述其检索需求。

传统的文档标引方法使用文档中出现的所有有意义的实词来标引文档,这会带来两个问题:一是无法解决“一词多义”的问题;二是无法解决大量无关词汇的影响问题。图1所示的方法中仅仅引入了词义识别功能,确定文档词汇在上下文环境中的确切含义,解决“一词多义”问题。尽管现在出现了一些主题词提取算法,譬如,词汇链方法和概念空间方法等。但是,由于这些方法的应用效果还不是很理想,从尽可能不降低现有检索系统性能方面考虑,图1没有解决无关词过滤问题。这是我们进一步工作需要解决的问题。

5 基于领域本体的信息检索方法实现的关键技术

5.1 领域本体词汇检索

领域本体词汇检索的作用在于提供一种辅助手段帮助用户选择所需要的初始领域本体词汇。根据用户输入的初始检索词,可以利用字符串匹配技术

来检索与该词字形上相似的本体词汇。根据领域本体词汇检索的需求以及各种字符串匹配技术的特征,选择应用最为广泛的编辑距离方法实现领域本体词汇检索^[14]。

编辑距离是指将一个字符串转换成另一个字符串所需要进行插入、删除、替换等相关编辑操作的次数。可以通过计算两个字符串之间的编辑距离来比较两个字符串之间的字形相似性。设两个字符串间的编辑距离为 d , m 为相比较的两个字符串的字符串长度最大值,则两字符串的字形相似度为 $Sim = \frac{d}{m} \times 100\%$ 。可设定一相似度阈值 γ_0 , 将 $Sim \leq \gamma_0$ 的字符串作为与目标字符串字形相似的字符串。

基于编辑距离的字符串匹配方法关键是计算两个相比较字符串的编辑距离 d 。编辑距离 d 的计算可以使用动态线性规划方法。设 $S = s_1 s_2 \cdots s_n$ 为待匹配的字符串,而 $W = w_1 w_2 \cdots w_m$ 为目标字符串,其中 n, m 分别为 S 和 W 的字符串长度。动态计算 S 和 W 的编辑距离的过程可以通过一个 $n \times m$ 的矩阵来描述。这个矩阵最右下角的值即为所求的 S 和 W 之间的编辑距离 d 。该矩阵元素的计算方法如下:

设 $D(s_i, w_j)$ 为 S 中前 i 个子串和 W 中前 j 个子串间的编辑距离,是矩阵中的一个值。置初值 $D(s_0, w_0)$ 为 0, 则

$$D(s_i, w_j) =$$

$$\min \begin{cases} D(s_{i-1}, w_{j-1}), & \text{当 } s_i \text{ 与 } w_j \text{ 相同时} \\ D(s_{i-1}, w_{j-1}) + 1, & \text{当在 } S \text{ 中用 } w_j \text{ 替换 } s_i \text{ 时} \\ D(s_i, w_{j-1}) + 1, & \text{当在 } S \text{ 中插入 } w_j \text{ 时} \\ D(s_{i-1}, w_j) + 1, & \text{当在 } S \text{ 中删除 } w_j \text{ 时} \end{cases}$$

5.2 基于概念空间的检索词联想

传统的基于树结构的层层展开法帮助用户通过遍历的方式查找所需词汇。这种方法虽然简单,但是不能体现本体词汇在文档库中使用的规律,而且效率较低。为此,图 1 中引入了基于概念空间的检索词联想方法。

概念空间方法就是自动构造概念语义网络,并以此为基础进行概念检索的一种方法。它的目标是建立一个知识丰富的、可理解的概念空间,用来代表所在信息空间中的概念及两两之间的关联。概念空间是指由词语及其权重关联形成的网络。信息空间是指要处理的文档集合。目前国内外利用概念空间

进行信息检索的研究很多,它在扩展检索词方面的应用效果较为理想^[3,13,15]。

基于概念空间的领域本体词汇联想方法的实现分为四步。

第一步,确定领域文档集合,即构建信息空间。由于基于领域本体的信息检索主要应用于专业领域,所以,可以直接将需要检索的文档集合作为处理对象。

第二步,确定概念集合。可以直接使用领域本体中的本体词汇。然后,使用这些词汇进行预处理,取出文档集合的每一篇文档中所有出现的本体词汇及其出现次数,并作标注。

第三步,进行概念共现分析,建立概念间的语义联系网络。这一步的关键在于计算任意两个词汇之间的关联权值。设词汇对 T_j 与 T_k 间的关联权值分别为 $ClusterWeight(T_j, T_k)$ 和 $ClusterWeight(T_k, T_j)$, 其计算方法为

$$\begin{aligned} ClusterWeight(T_j, T_k) &= \frac{\sum_{i=1}^N d_{ijk}}{N} \times WeightingFactor(T_k) \\ ClusterWeight(T_k, T_j) &= \frac{\sum_{i=1}^N d_{ikj}}{N} \times WeightingFactor(T_j) \end{aligned}$$

其中, N 表示文档集合中所有文档的数目; d_{ij} 表示词汇 T_j 在文档 i 中的权重; d_{ik} 表示词汇 T_k 在文档 i 中的权重; d_{ijk} 和 d_{ikj} 表示 T_j 和 T_k 在文档 i 中的联合权重; $WeightingFactor(T_j)$ 和 $WeightingFactor(T_k)$ 为 T_j 与 T_k 的惩罚因子,该词越常用,则其惩罚因子值越小。

d_{ij} 的计算方法为 $d_{ij} = tf_{ij} \times \log\left(\frac{N}{df_j} \times \omega_j\right)$ 。其中, tf_{ij} 为词频,即词汇 T_j 在文档 i 中出现的次数; df_j 表示在文档集合中出现词汇 T_j 的文档数目; ω_j 表示词汇 T_j 的字符串长度。 d_{ik} 的计算方法类似。

d_{ijk} 的计算方法为 $d_{ijk} = tf_{ijk} \times \log\left(\frac{N}{df_{jk}} \times \omega_j\right)$ 。其中, tf_{ijk} 表示词汇 T_j 和词汇 T_k 共同在文档 i 中出现的次数; df_{jk} 表示在文档集合中同时出现词汇 T_j 和词汇 T_k 的文档数目。 d_{ikj} 的计算方法类似。

$WeightingFactor(T_j)$ 的计算方法为

$$\text{WeightingFactor}(T_j) = \frac{\log \frac{N}{df_j}}{\log N}; \text{WeightingFactor}(T_k) \text{ 的}$$

$$\text{计算方法为 } \text{WeightingFactor}(T_k) = \frac{\log \frac{N}{df_k}}{\log N}。$$

第四步,应用所建立的概念空间进行概念检索。在概念空间中进行概念检索最为成功的算法是Hopfield网络算法。该算法依赖于一个激活和迭代的过程,主要计算公式为

$$\mu_j(t+1) = f_j \left[\sum_{i=0}^{n-1} t_{ij} \mu_i(t) \right], 0 \leq j \leq n-1$$

其中, $\mu_j(t+1)$ 是词语 j 在第 $t+1$ 次迭代时的激活值, t_{ij} 是词语 i 到词语 j 的共现权值,即为上述所计算的关联权值 $\text{ClusterWeight}(T_j, T_k)$, f_j 是持续的S形转换函数,它把每个给定值规格化为0到1之间的值。当用户输入检索入口词以后,首先使用Hopfield算法激活它们的强关联词,然后进一步激活这些强关联词的相邻强关联词,以此类推,直到算法收敛为止。最后,输出与初始词汇相关联的所有词汇。

5.3 基于领域本体的文档词汇词义识别

文档中词汇所表达的语义是与其上下文词汇密切相关。引入领域本体以后,可以利用领域本体库中词汇之间的规范化关系来实现文档词汇语义的自动识别,这可以消除一词多义现象对信息检索的不利影响。将领域本体库简化为一个具有层次结构的词汇语义关联网络,则词义识别的目的本质上就是确定文档中词汇所描述的语义应该对应这个语义关联网络中的哪一个节点。

设: p_1 和 p_2 为所构建的词汇语义关联网络中的两个具有明确含义的词;若 p_1 或 p_2 不属于概念词汇,则 P_1 和 P_2 为其所属的概念词汇;若 p_1 或 p_2 属于概念词汇,则 P_1 和 P_2 即为其自身; P_1 和 P_2 的最近父节点为 P_f , P_f 到树根 P_{root} 的距离为 $\text{dis}(P_f, P_{\text{root}})$, 树的深度为 M ; P_1 到 P_2 的最短路径长度为 $\text{dis}(P_1, P_2)$ 。则 p_1 和 p_2 之间的词义关联度为

$$\begin{aligned} \text{RelationValue}(p_1, p_2) \\ = \frac{\text{dis}(P_f, P_{\text{root}})}{M} \times \frac{\alpha}{\text{dis}(P_1, P_2) + \alpha} \end{aligned}$$

其中, α 是可调参数,取决于本体结构的大小。

设文档中某一词汇 W_0 在词义关联网络中对应

n 个节点,即该词具有 $S_1, S_2, \dots, S_i, \dots, S_n (1 \leq i \leq n)$ 个可能的词义。取该词上下文中在词义关联网络中出现的 m 个词汇 $W_1, W_2, \dots, W_j, \dots, W_m (1 \leq j \leq m)$, 若存在语义不确定的词,可将其滤除,即保证这 m 个词汇都具有唯一确定的词义。设词义 S_i 与上下文词义的聚合度为 $\text{SenseClusterValue}_i$, 则

$$\text{Sense Cluster Value}_i = \sum_{l=1}^m \theta_l$$

$$\theta_l = \{1 \text{ if } \text{RelationValue}(S_i, W_l) \geq \phi_0, 0,$$

其中, ϕ_0 为词义关联度阈值,可通过实验确定。

在词汇 W_0 的 n 个词义中,若某词义 S_q 的 $\text{Sense Cluster Value}_q$ 满足 $\text{Sense Cluster Value}_q = \max_{1 \leq \beta \leq n} (\text{Sense Cluster Value}_\beta)$ 并且 $\max_{1 \leq \beta \leq n} (\text{Sense Cluster Value}_\beta) \geq \bar{\omega}_0$, 则该词汇在该上下文中取词义 S_q ; 否则,该词汇的词义为不确定。其中, $\bar{\omega}_0$ 为词义识别阈值,可通过实验确定。引入 $\bar{\omega}_0$ 是为了考虑领域本体不完备性的影响,可能会出现某些词汇的词义领域本体没有定义的情况。所以,对于文档中出现的所有领域本体词汇,不管其是否存在多义现象,都要使用上述算法确定其语义。

5.4 基于XML的词汇标记方法

在确定了查询词汇和文档词汇的确定词义以后,还需要使用一种规范化的方法来描述词汇与其所表达的确切含义(在领域本体库中的确定位置)之间的这种对应关系,以便于实现计算机的自动处理。为此,设计了一种基于XML的词汇标记方法,该方法依赖于如下所定义的XML Scheme。

在上述模式中,最重要的Word类型定义了词汇标记的主要属性,通过这些属性值来建立词汇与其在领域本体中所表达含义的对应关系;OntologyLexiconType类型可以是枚举类型,包含所有领域本体描述所用的词汇;RelationNameType类型可以是包含“kind of”、“instance of”、“synonymy”等关系词汇。ConceptIDType类型定义了领域本体中概念唯一性编号的表示方法。

6 实验及性能分析

防务快讯挖掘系统是南航经管院承担的一项国防技术基础项目的关键内容,该系统的主要目的在于从大量的防务快讯中挖掘有价值的信息为决策人员提供参考。系统的核心是航空领域本体构建以


```

<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="unqualified"
attributeFormDefault="unqualified">
  <xs:element name="SearchWords" type="WordsType"/>
  <xs:element name="SearchedDocument" type="WordsType"/>
  <xs:complexType name="WordsType">
    <xs:choice minOccurs="0" maxOccurs="unbounded">
      <xs:element name="Word">
        <xs:complexType>
          <xs:simpleContent>
            <xs:extension base="OntologyLexiconType">
              <xs:attribute name="IsConcept" type="AnswerType"/>
              <xs:attribute name="ConceptID" type="ConceptIDType"/>
              <xs:attribute name="IsRelation" type="AnswerType"/>
              <xs:attribute name="RelationName" type="RelationNameType"/>
              <xs:attribute name="BelongToConcept" type="ConceptIDType"/>
            </xs:extension>
          </xs:simpleContent>
        </xs:complexType>
      </xs:element>
    </xs:choice>
  </xs:complexType>
  <xs:simpleType name="OntologyLexiconType">
    .
    .
    .
  </xs:simpleType>
  <xs:simpleType name="AnswerType">
    .
    .
    .
  </xs:simpleType>
  <xs:simpleType name="RelationNameType">
    .
    .
    .
  </xs:simpleType>
  <xs:simpleType name="ConceptIDType">
    .
    .
    .
  </xs:simpleType>
</xs:schema>

```

及该领域本体在信息挖掘中的综合应用,我们按本文提出的方法设计了其中的信息检索子模块。通过实验发现:

首先,该方法会极大地提高查全率。对查全率指标的改进主要依赖于所提供的对用户查询需求的辅助处理手段。由于我们使用交互式的方式引导用户一步步逼近其真实的、潜在的检索需求,所以,用户所使用的准确的检索词会比较丰富,它可以在不降低查准率的情况下提高查全率。

其次,该方法会改进查准率,但其改进效果不如对查全率的改进效果显著。主要由于:一、专业领域的“一词多义”现象不如“一义多词”现象普遍,这就使得文档词义识别功能的发挥受到限制;二、查准率的改善主要取决于词义识别的准确率,而词义识别的准确率受 5.3 节方法中阈值 α , ψ_0 , $\bar{\omega}_0$ 大小以及领域本体完备性的影响。原型系统中词义识别的准确率在 65% 左右。实验发现领域本体规模越大,则阈值 α , ψ_0 , $\bar{\omega}_0$ 越具规律性,词义识别的准确率就越高。

最后,该方法会带来检索效率的降低。检索效率的降低体现在两个方面:①用户检索所用时间增加,由于用户需要学习领域本体词汇并对其进行评价;②系统对文档处理过程的开销增大,主要由于概念空间计算和词义识别计算需要较高的资源消耗,特别是文档库规模大时这一问题较为明显。这一特征使得该方法更加适合于构建信息推送服务系统,在防务快讯挖掘系统中就使用了这种模式。

7 结束语

论文通过对基于领域本体的信息检索技术的形式化定义得出了两点结论:①现阶段基于领域本体的信息检索主要利用本体论中规范化的词汇及其含义来解决“一词多义”和“一义多词”问题,它要以传统的基于词形匹配的信息检索方法为基础;②为了有效应用领域本体,必须要解决用户查询需求和目标文档向领域本体映射问题以及基于领域本体的用户查询需求和目标文档的规范化标注问题。按照这两条结论,论文建立了一种综合应用领域本体改善信息检索性能的方法。该方法使用基于编辑距离的词形匹配方法辅助用户查询本体词汇,然后,使用基于概念空间的检索词联想方法帮助用户扩充检索词。通过交互式的方式引导用户一步步逼近其真实的、潜在的检索需求。在文档标记方面,本文提出基于领域本体的词义识别算法,来确定文档中的词汇词义。实验表明,本文建立的方法会有效提升查全率指标,会改进查准率指标。

从人工智能领域发展起来的本体论技术之最大的作用在于实现基于机器推理的智能化处理。如何利用这种推理能力实现智能化的信息检索是我们目前还没有解决的问题。通过本文的工作,我们认识到基于推理的智能化检索必须要以文档的自动、准确、规范的标引为基础,这需要首先建立一种文档标引本体。利用这种文档标引本体实现文档标引表述的规范化,借助于各种领域本体实现文档含义表述

的规范化,并在此基础上研究基于推理的信息检索方法,是我们下一步工作的内容。

参 考 文 献

- [1] 丁晟春,岑咏华,顾德访. 基于 Ontology 的语义检索研究[J]. 情报学报, 2005, 24(6): 702-707.
- [2] 余正涛,宋丽哲,樊孝忠. 基于本体的个性化领域信息服务[J]. 计算机工程, 2005, 31(5): 22-25.
- [3] 朱晓华. 基于概念空间方法的信息检索技术研究[J]. 大学图书馆学报, 2003(2): 47-53.
- [4] Stairmand M A. Textual context analysis for information retrieval[C]//Proceedings of the 20th Annual ACM SIGIR Conference on Research and Development in IR. 1997: 140-147.
- [5] Silber H G, McCoy K F. Efficiently Computed Lexical Chains as an Intermediate Representation for Automatic Text Summarization[J]. Computational Linguistics, 2002, 28(4): 487-496.
- [6] 武成岗,焦文品,田启家,等. 基于本体论和多主体的信息检索服务器[J]. 计算机研究与发展, 2001, 38(6): 641-647.
- [7] 周宁,张玉峰,张李义. 信息可视化与知识检索[M]. 北京: 科学出版社, 2005.
- [8] Abasolo J M. MELISA: An ontology_ based agent for information retrieval in medicine[DB/OL]. Proceedings of the First International Workshop on the Semantic Web (SemWeb2000), Lisbon, Portugal, 2000.
- [9] 郭祥文,刘惟一,钱民,等. 基于本体论的信息检索[J]. 云南大学学报(自然科学版), 2003, 25(4): 324-327.
- [10] 宋峻峰,张维明,肖卫东,等. 基于本体的信息检索模型研究[J]. 南京大学学报(自然科学), 2005, 41(2): 189-197.
- [11] Stojanovic N, Studer R, Stojanovic L. An Approach for Step-By-Step Query Refinement in the Ontology-based Information Retrieval[C]// Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence. 2004: 36-43.
- [12] 胡艳丽,白亮,张维明,等. 知识网格中基于领域本体的智能检索[J]. 计算机科学, 2007, 34(8): 202-207.
- [13] 朱连花. 基于概念空间的信息检索方法[J]. 情报科学, 2004, 22(4): 470-474.
- [14] 孙海霞,成颖. 信息集成中的字符串匹配技术研究[J]. 现代图书情报技术, 2007(7): 22-26.
- [15] 郑毅,吴斌,史忠植. 基于概念空间的文本检索系统[J]. 计算机工程于应用, 2002(12): 67-71.

(责任编辑 芮国章)

Study on Evolution Mechanism of Fateful Stampede Accident based on Graphical Evaluation and Review Technique

Mingli Hu

College of Economics and Management
Nanjing University of Aeronautics and Astronautics
Nanjing, China
huml@nuaa.edu.cn

Zhigeng Fang

College of Economics and Management
Nanjing University of Aeronautics and Astronautics
Nanjing, China
zhigengfang@nuaa.edu.cn

Sifeng Liu

College of Economics and Management
Nanjing University of Aeronautics and Astronautics
Nanjing, China
sfliu@nuaa.edu.cn

Wei Chen

Department of Information Management
Nanjing Audit University
Nanjing, China
chenweich@nau.edu.cn

Baohua Yang

College of Economics and Management
Nanjing University of Aeronautics and Astronautics
Nanjing, China
mathhua@126.com

Abstract—Fateful stampede accident has become one of hot spots in public security field because of its sudden and serious consequence. Actual cases show that crowding is not a sufficient condition for fateful stampede accident, but a necessary condition. Some emergencies play a role of trigger in stampede accident, such as fire, explosion, individual injury, earthquake, etc. Evolution mechanism of fateful stampede accident is discussed based on graphical evaluation and review technique (GERT) in the paper. Escape panic caused by emergencies greatly increases the risk of stampede accident. In the presented model, risk of fateful stampede accident is measured in terms of the difference between “desired velocity” and “actual velocity” of crowd. Finally, a GERT model is set up to describe the evolution mechanism of a fateful stampede accident at Miyun county.

Keywords—fateful stampede accident, evolution mechanism, GERT

I. INTRODUCTION

In recent years, more and more fateful stampede accidents were reported. According to incompletely figures, more than 2800 people were killed in stampede accidents all over the world during the first decade of 21st century. Table 1 shows part of fateful stampede accidents occurred in 21st century. Fateful stampede accident has become one of hot spots in public security field because of its sudden and serious consequence.

Actual cases show that crowding is a necessary condition for fateful stampede accident. In earlier research on crowd behavior, a curve describing the relationship between average marching velocity and human density of crowd was drawn in [1]. BFIREs is known as the first model for computer simulation of human behavior in building fires[2]. Force-Information-Space-Time(FIST) model is the first model for describing stampede accident in the world. [3] presented a continuous pedestrian flow model to describe crowding and stampede respectively. Social force model presented in [4] is capable of describing the self-organization of several observed collective effects of pedestrian behavior. In [5], the concept of “reject force” in Helbing’s social force model was studied in depth. [6] used a model of pedestrian behavior to investigate the mechanisms of panic and jamming by uncoordinated motion in crowds. Some crowd phenomena under uncoordinated situation in emergencies were studied in [7].

There are also some related achievements published in China. [8] founded dynamical equations of evacuation speed for personnel. Causes as well as prevent methods of stampede accident are discussed in [9]. Thesis[10] studied mechanism of stampede accident in outdoor large-scale social activities. Four phases model for estimating stampede accident risk in large-scale public places was set up in [11]. GERT model presented in [12] revealed the mechanism of disaster evolution in terms of the natural evolution of disaster and the action of emergency rescue.

TABLE I. FATEFUL STAMPEDE ACCIDENTS OCCURRED IN 21ST CENTURY

ID	Date	Description of stampede accident
1	April 11, 2001	43 people were crushed in the Ellis Park Stadium disaster in Johannesburg, South Africa.
2	February 17, 2003	21 people were killed in the stairway exit to E2, a nightclub in Chicago, after a pepper spray use on an upper-story dance floor.
3	February 4, 2004	At least 37 people were died with 15 were injured, when a crowd stampede, during Lantern Festival in Mihong Park, Miyun County, Beijing, China.
4	August 31, 2005	965 people were killed in a Baghdad bridge stampede, Iraq.
5	January 12, 2006	345 people were killed at Jamarat Bridge in Mecca during the stoning of the devil.
6	November 11, 2007	3 people were killed and more than 30 injured at the Supermarket Carrefour in Chongqing, China when the shop was offering 20% discounts on cooking oil.
7	September 30, 2008	147 people were killed during the Chamunda Devi stampede at the Chamunda Devi temple in Jodhpur, India.
8	March 29, 2009	The Houphouët-Boigny Arena stampede. 19 people killed and 130 injured in a stampede at a football stadium in Ivory Coast as fans try to squeeze into the stadium for a World Cup qualifier.

Some basic concepts of GERT model are provided in Section II. Evolution mechanism of fateful stampede accident is discussed in Section III. Then, a GERT model of the evolution of a real case is given in section IV. Finally, Section V is the summary of the paper.

II. GERT MODEL: GRAPHICAL EVALUATION REVIEW TECHNIQUE SIMULATION MODEL

GERT is a technique that has been developed for the analysis of stochastic networks. As shown in Figure 1, there are two types of GERT network nodes: the deterministic node with round tail and the stochastic node with triangular tail. When a deterministic node is attained, all arrows starting from the node will begin. On the contrary, in the case of a stochastic node, only one of the arrows will be triggered. Every node in the GERT network have three parameters. The right-hand number in each node represents the arrow number. The left-hand and upper side number is the initial number of jobs required to release this node, while the lower side number is the subsequent number. When the left-hand and upper side number is 1, the node is an OR-node. If the number is more than 2, the node is an AND-node.

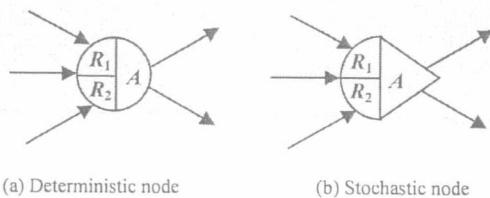


Figure 1. Deterministic node and stochastic node in GERT network

III. EVOLUTION MECHANISM OF FATEFUL STAMPEDE ACCIDENT

Actual cases show that some emergencies play a role of trigger in fateful stampede accident, such as fire, explosion, individual injury, earthquake, etc. Inductive effect is defined to describe this type of evolution mechanism in stampede accident as following.

Definition 1 (Inductive Effect): As shown in Figure 2, Inductive effect means that event A result in some emergencies (denoted by $N_i (i=1,2,...,n)$) with probability p_i respectively. R_1 is the initial number of jobs required to release this node, while R_2 is the subsequent number. For example, emergency N_1 is the consequence of event A with probability p_1 .

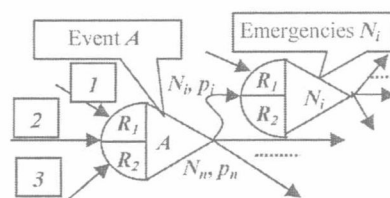


Figure 2. Inductive effect in GERT network

According to the research on human behavior in crowding place, we assume that the relationship between human density and marching speed is linear as following.

Definition 2 (Actual Velocity): If human density of crowd is denoted by ρ (unit: people per m^2), then actual velocity of crowd, denoted by V_a is the linear function of ρ [10]. It is formulated in (1), in which V_0 is the human velocity in the free walk (unit: $m s^{-1}$), and B is a constant.

$$V_a = V_0 - B \cdot \rho \quad (1)$$

In emergency situation, people are eager to escape from crowding place. People move or try to move considerably faster than normal. Individuals start pushing, and interactions among people become physical in nature. The physical interactions in the jammed crowd add up and cause dangerous pressures up to $4450 N m^{-1}$ which can bend steel barriers or push down brick walls[6].

Definition 3 (Desired Velocity of Crowd): Desired velocity of crowd (denoted by V_d) means the velocity people try to move in crowd. Magnitude of desired velocity lies on the impatient extent of people in crowd. As shown in (2), impatient extent (denoted by $f(A)$) is the function of circumstance around people in crowd (here only consider the influence of emergency event A). K is a constant.

$$V_d = K \cdot f(A) \quad (2)$$

It is suggested that escape panic caused by emergencies greatly increases the probability of fateful stampede accident. So, it is reasonable that probability of fateful stampede accident

(FSA) is measured in terms of the difference between "desired velocity" and "actual velocity" of crowd.

Definition 4(Risk Assessment Model of FSA): if the risk of FSA is denoted by P ($0 \leq P \leq 1$), risk assessment model of FSA is given in (3). According to (1) and (2), (3) is equal to another form shown in (3').

$$P = \left| \frac{V_d - V_a}{V_d} \right| \quad (3)$$

$$P = \left| 1 - \frac{V_0 - B \cdot \rho}{K \cdot f(A)} \right| \quad (3')$$

(1)-(4) are static description of human behavior in crowd. V_d as well as V_a , ρ , P vary with time t . Then the continuous risk assessment model of FSA is given by a equation as following.

$$P_t = \left| 1 - \frac{V_0 - B \cdot \rho_t}{K \cdot f_t(A)} \right| \quad (4)$$

Human density ρ_t is observable. V_0 and B are constant in a certain crowd. If the impatient extent $f_t(A)$ is measured accurately, then parameter p_t in GERT network can be estimated by (4). It is significant for early warning and emergency decision making in fateful stampede accident.

IV. A CASE STUDY

At least 37 people died and 15 others were injured at a Lantern Festival gathering at a Miyun County Park 65 kilometres northeast of downtown Beijing on February 4, 2004. Thousands of people had come to watch the specially prepared lanterns featured at the annual festival. The accident occurred at 7:45 pm at Mihong Park when a visitor tumbled on an overcrowded bridge and in the confusion people were crushed in an oncoming throng. Many of the victims were crushed and suffocated to death by crowd[13].

Obviously, the event that a visitor tumbled on the bridge is an inductive factor of the stampede accident. Here, a GERT model on the evolution of this accident is described by shown in Figure 3.

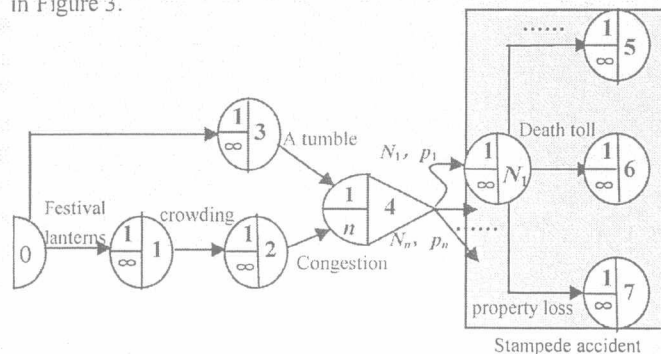


Figure 3. GERT model on the evolution of a fateful Stampede accident at MiYun county

V. SUMMARY

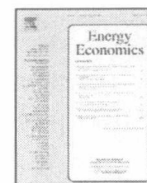
Fateful stampede accidents always result from crowding and congestion. Human behavior in crowd is an ideal breakthrough point to understand the evolution mechanism of stampede accident. Latest achievements on crowd dynamics are embedded into GERT model in the paper. Computability of GERT model provide broad prospects for the application of presented model in early warning and emergency decision making of fateful stampede accidents. How to acquire available parameters of model in its application is an important issue in further research.

ACKNOWLEDGMENT

This work was partially supported by National Natural Science Foundation of China (No.90924022, 70971064 and 70701017). At the same time the authors would like to acknowledge Talents Research Fund of NUAA (No. 1009-909343).

REFERENCES

- [1] Fruin J J. Pedestrian Planning and Design[M]. Metropolitan Association of Urban Designers and Environmental Planners, PO Box 722, Church Street Station, NY 10008, 1971.
- [2] Stahl F. I. A computer simulation of human behavior in building fires: Interim report[R]. U.S. Department of Commerce, Report NBSIR, Washington, DC, USA, 1978, 78-1514.
- [3] Lee R C, Hughes R L. Exploring tramp ling and crushing in a crowd, Journal of transportation engineering[J]. Journal of transportation engineering, 2005, 131 (8): 575-582.
- [4] Helbing, D. & Mulnar, P. Social force model for pedestrian dynamics[J]. Phys. Rev. 1995, E 51, 4282-4286
- [5] Teknomo K. Microscopic Pedestrian Flow Characteristics: Development of an Image Processing Data Collection and Simulation Model[D]. Sendai: Tohoku University, 2002.
- [6] Helbing D, Farkas I, Vicsek. T. Simulating dynamical features of escape panic [J]. Nature, 2000, 407, 487-490.
- [7] Heigeas L, Luciani A, Thollot J. A physically - based particle model of emergent crowd behaviors [R]. Moscow: International Conference on Computer Graphics & Vision, 2003.
- [8] Lu Jun-an, Fang Zheng, Lo Siu-ming, Zhao Chun-mei. Mathematical model of evacuateion speed for personnel in buildings[J]. Engineering Journal of Wuhan University, 2002, 35(2): 66-70.
- [9] Kou Li-ping. Study on analysis causes and prevent method for stampede accident. Journal of Chinese People's Public Security University[J], 2005, (4): 16-22.
- [10] Rui Bai. Study on mechanicsm of stampede accident in outdoor large-scale social activities[D]. Shen Yang Institute of Aeronautical Engineering, 2009, 15-33.
- [11] Zhang Qing-song, Liu Jin-lan, Zhao Guo-min. Discussion on cause mechanism of crowd crushing and trampling accident in large-scale public places[J]. Journal of Natural Disasters, 2009, 18(6): 81-86.
- [12] Fang Zhi-geng, Yang Bao-hua, etc. The GERT network model study of disaster evolution based on bayes inference[J]. Chinese Journal of Management Science, 2009, 17(2): 102-107.
- [13] http://www.chinadaily.com.cn/english/doc/2004-02/06/content_303605.htm.



Total factor carbon emission performance: A Malmquist index analysis

P. Zhou ^{a,*}, B.W. Ang ^b, J.Y. Han ^b^a College of Economics and Management, Nanjing University of Aeronautics and Astronautics, 29 Yudao Street, Nanjing 210016, China^b Department of Industrial and Systems Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore

ARTICLE INFO

Article history:

Received 11 May 2009

Received in revised form 21 September 2009

Accepted 3 October 2009

Available online 13 October 2009

JEL classification:

C61

D24

Q43

Q54

Keywords:

Carbon dioxide emissions

Malmquist index

Data envelopment analysis

Total factor productivity

Bootstrap

ABSTRACT

This paper introduces a Malmquist CO₂ emission performance index (MCPI) for measuring changes in total factor carbon emission performance over time. The MCPI is derived by solving several data envelopment analysis models. Bootstrapping MCPI is proposed to perform statistical inferences on the MCPI results. Using the index the emission performance of the world's 18 top CO₂ emitters from 1997 to 2004 is studied. The results obtained show that the total factor carbon emission performance of the countries as a whole improved by 24% over the period and this was mainly driven by technological progress. The results of a cross-country regression analysis to investigate the determinants of the resulting MCPI are presented.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Global awareness on climate change has created much interest in analyzing the trends of world energy use and carbon dioxide (CO₂) emissions. Policy makers have realized the importance of considering CO₂ emissions in formulating national economic and energy policies. Internationally, it is apparent that the post-Kyoto climate policy will put more emphasis on the reduction of global CO₂ emissions in order to mitigate climate change. These facts bring the need for understanding the patterns of CO₂ emissions and monitoring the emission performance in different countries.

Various indicators have been developed and applied to monitor national CO₂ emission performance trends. For instance, Mielnik and Goldemberg (1999) propose the use of a "carbonization index" (the level of CO₂ emissions per unit of energy consumption) to assess the evolution patterns of developing countries with regard to climate change. Ang (1999) shows that energy intensity (energy consumption per unit of GDP) is as useful as the carbonization index in the study of climate change. Sun (2005) highlights the usefulness of CO₂ emission intensity in measuring decarbonization and assessing energy policies at the national level. Tol et al. (2009) show that both CO₂ emission intensity and CO₂ emissions per person can be considered as a function of per capita income.

The indicators mentioned above may be interpreted as partial indicators since they can only reflect partial aspects of CO₂ emission performance. Ramanathan (2002) points out that a more holistic view is to use the data envelopment analysis (DEA) technique to combine all the relevant indicators such as energy consumption, economic activity and CO₂ emissions into an overall index for performance comparisons. DEA, a nonparametric frontier approach to efficiency evaluation, has been widely applied to assess the relative performance of various entities.¹ Recently, the use of DEA has also been extended to model and analyze CO₂ emission performance. For example, Zaim and Taskin (2000) develop a hyperbolic efficiency measure to calculate the CO₂ emission efficiency of OECD countries. Zofio and Prieto (2001) study the CO₂ emission performance of OECD manufacturing industries by considering several different regulatory scenarios. Zhou et al. (2006) propose a slacks-based efficiency measure for modeling CO₂ emission performance that accounts for economic inefficiency.

A limitation of these previous studies is that they usually study CO₂ emission performance within a cross-sectional data framework and not over time. It is therefore worthwhile to develop a tool for carrying out a formal time-series analysis of CO₂ emission performance. It is for

* Corresponding author. Tel.: +86 25 84896261; fax: +86 25 84892751.
E-mail address: cmzp@nuaa.edu.cn (P. Zhou).

¹ In energy economics, DEA has also been applied to areas such as energy efficiency measurement and productivity analysis of energy sectors. Examples of such studies include Abbott (2006), Barros (2008), Barros and Peypoch (2008), Mukherjee (2008), and Ramos-Real et al. (2009). A recent literature review of DEA in energy and environmental analysis can be found in Zhou et al. (2008a).