

基于模式识别的文本知识点深度挖掘方法

温浩¹ 温有奎² 王民¹

(西安建筑科技大学信息与控制工程学院 西安 710055)¹ (西安电子科技大学经济管理学院 西安 710071)²

摘要 针对目前大数据知识获取存在的噪声大的问题,提出了文本知识点深度挖掘方法。首先构建了学术论文创造性特征的“问题,方法,结果”三元组本体模型;其次利用模式识别等技术对学术论文文摘进行统计分析、特征提取、机器学习、模式判定分析;最后对学术论文创造性核心知识的三元组进行深度挖掘。实验结果表明,该方法能大大过滤掉学术文献大数据检索的噪声,便于用户快速定位大型学术文献数据库论文的研究问题,采用的新方法和得到的结果能判断学术论文的阅读价值,并为大数据深度知识挖掘和关联发现研究提供基础。该类方法未见有公开的文献报道,属于一种探索性研究和实验。

关键词 模式识别,文本挖掘,语义三元组,直接知识获取

中图分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.3.052

Approach to Text Knowledge Depth Mining Based on Pattern Recognition

WEN Hao¹ WEN You-kui² WANG Min¹

(School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China)¹

(School of Economics and Management, Xidian University, Xi'an 710071, China)²

Abstract A new method of text mining was presented to make up for the disadvantages of big data knowledge acquisition. Firstly, we constructed triple ontology model about academic inventive features “problem, methods, results”. Secondly, pattern recognition techniques were used for statistical analysis, feature extraction, machine learning and pattern determination analysis. Finally, depth mining of triples of the creative core academic knowledge was realized. The Experimental results show that the new method can effectively reduce the retrieval noise of academic literature, which is convenient for users to quickly locate the research problem. The methods and results can determine the reading value of papers and provide a basis for depth knowledge mining of large data and related discovery. The method has not been reported in the literature, and it is a kind of exploratory research and experimentation.

Keywords Pattern recognition, Text mining, Semantic triples, Direct knowledge acquisition

1 引言

大数据的到来引起了国际科技界对科学研究方法论的重新审视,正在引发科学研究思维与方法的一场革命。李国杰院士对新科研模式做了简洁描述,即面对海量数据,科研人员只需从数据中直接查找或挖掘所需要的信息、知识和智慧,甚至无需直接接触需研究的对象^[1]。文献[2]认为科学家们需要访问文献中的关键图和知识,以验证自己的研究成果,并产生新的假设。而目前人们不能从大型学术文献数据库中直接获取有用知识是由于两方面的问题,一方面,文献知识的存储与检索方式以本、篇为传播单位,这些知识分散在其包含冗余信息内容的全文中,几乎总是不能被人直接获取^[3]。为此,早在1980年,B. C. Brookes就提出认知地图的构想,将知识对象的粒度由文章细化到文章中微观的知识元^[4],图书馆学、情报学、信息资源管理等多个领域的专家学者对知识元的概念、

结构、模型、抽取等方法进行了研究,世界著名期刊 Nature 已开始对学术论文中的创新点进行语义标注^[5]。另一方面,现有大多数信息检索技术都基于词汇(term-based)方法^[6],从关键词检索技术本身来讲,这种简单匹配方法噪声很大,常常会检索出很多非相关文档^[7]。多年的实验也表明,基于词组(phrase-based)的方法可以比基于词汇的方法携带更多“语义”信息的假设,并没有表现得更好,词组中的冗余和噪声并不小^[8]。最近一个新的基于概念模式(concept-based model)^[9]的方法被提出,该方法的优点是能分析句子和文档层面的词汇,其架起了 NLP 和文本挖掘之间的桥梁。这个方法由 3 个部分组成。首先分析句子的语义结构;接着建立描述语义结构的本体图(COG);最后根据前两步抽取主题概念(top concepts),建立特征向量。其优点是能描述一个句子的有效词汇和不重要词汇。在解决信息检索带来的非相关文档的噪声问题方面,文献[10]认为 KeyGraph 考虑了关键词一

到稿日期:2015-05-29 返修日期:2015-08-10 本文受国家自然科学基金项目(70373946)资助。

温浩(1979—),男,博士后,讲师,主要研究方向为模式识别与智能系统,E-mail:smczg@126.com;温有奎(1951—),男,博士,教授,主要研究方向为智能推理,E-mail:wykui123@126.com;王民(1959—),男,教授,主要研究方向为智能信息处理,E-mail:wangmin1329@163.com。

同出现的问题,能成功地过滤无关的文件,并能识别大噪声社交媒体集合中的事件。但人们使用术语关系分析、共现分析和 WordNet 并没有得到预期的效果,这表明不适宜的相关术语包含的噪声信息难以避免^[11]。文献[12]认为文本挖掘或文本知识发现是一个解决信息检索噪声问题的有效的潜在工具,并将挖掘分为两类:在文件浏览中采用粗粒度的挖掘技术,而在更深层次的内容挖掘中采用文本句法和语义分析技术。文献[13]提出基于模式挖掘的解决方法。文献[14]认为在自然语言的使用中,模式确实是存在的,但不是存在于词汇级别,而是存在于语义级别上,必须去挖掘文章所蕴涵的语义信息以及这个语义的表达形式,并且用这个表达形式进行语义模式的检测与识别。文献[15]认为许多实验并不支持基于模式(或词组)的方法比基于词汇的方法好的假设,提出了一种包括模式部署和模式演变(pattern deploying and pattern evolving)过程的模式发现技术。然而,过多的特征属性使得被识别出来的模式只能适用于小部分的数据,这将使得这个模式变得毫无意义。为了更加有效地进行句子的语义模式的检测与识别,文献[16,17]对句子内部成分的依赖关系进行分解以产生语法子集,即采用主语、谓语及宾语这3个主要组成部分作为句子的表示来降低句子识别的复杂性。采用主、谓、宾句代表句子主要结构的方法为文本的语义模式的检测与识别提供了可行性,但还不能有效解决对学术性文摘的创造性知识点进行挖掘的模式问题。为此,本文建立了学术文摘的创造性知识点的本体模型,构建了以“问题,方法,结果”为三元组的创造性知识点模式,以学术期刊文摘为挖掘对象,实现了基于模式识别的文本知识点深度挖掘方法。

2 构建学术文摘创造性知识点本体模型

2.1 学术文摘本体构建

(1)本体构建步骤

本体被广泛应用于信息科学和计算机领域,并已成为当今信息科学研究的一个热点。在知识发现和管理、知识库设

计和集成、信息检索和抽取等领域,本体扮演着越来越重要的角色。领域本体是用于描述特定领域实体概念、领域属性概念、领域属性值及相互关系,以及该领域所具有的特性和规律的一种形式化描述。构建领域本体需要经过下面5个步骤:①需求分析;②罗列领域中重要的概念、术语;③构建领域本体框架;④对领域本体编码、形式化;⑤领域本体的确认和评价。

(2)学术期刊文摘结构统计分析

首先,选择国内11种中文学报的论文文摘,如表1所列。其中包括电子、计算机、软件、药学、社会科学、法学、管理、财经、农业、交通运输、教育,文摘总量为41316篇。然后对11种核心学术期刊文摘进行句子结构统计分析,结果表明,平均每篇学术期刊文摘有4.4个由句号组成的句子,平均每个句子有2.1个逗号组成的子句。

表1 国内11种中文核心科技期刊

学报名称	文献数量	统计年份
11种学报	41316	1999-2014
电子学报	10302	2004-2014
计算机学报	6000	1978-2014
软件学报	5062	1990-2014
中国药科大学学报	4119	1987-2014
中国社会科学	3970	1980-2014
中国法学	3254	1984-2014
管理学报	2345	2004-2014
东北财经大学学报	2373	1999-2014
南方农业学报	1468	2011-2014
交通运输工程学报	1462	2001-2014
教育学报	961	2005-2014

(3)学术期刊文摘动词语义关系统计

为了调查期刊文摘创造性知识点表达的语义关系,我们对文摘的动词进行统计分析。首先,建立停用词表;其次,采用中科院 ICTCLAS50 分词软件进行分词处理;再次,统计动词词频和概率。概率=出现的动词频率/总文摘数。表2列出了11种学报文摘动词词频数量分布较高的前30个动词词频及概率。

表2 11种学报文摘动词词频和概率

动词名称	频率	概率	动词名称	频率	概率	动词名称	频率	概率
提出	24488	0.593	利用	7788	0.188	计算	5114	0.124
进行	23007	0.557	提高	7609	0.184	能够	5004	0.121
研究	21279	0.515	实验	7347	0.178	得到	4744	0.115
分析	16434	0.398	设计	7343	0.178	相关	4577	0.111
具有	12135	0.294	给出	6665	0.161	结合	4575	0.111
应用	10190	0.247	存在	6413	0.155	证明	4528	0.110
发展	9687	0.234	建立	5967	0.144	使用	4482	0.108
影响	9683	0.234	处理	5834	0.141	表明	4384	0.106
实现	9620	0.233	提供	5608	0.136	解决	4198	0.102
采用	8573	0.207	优化	5577	0.135	描述	4069	0.985

统计结果:(1)11种学报总文摘数为41316,总字数为7821154,总句子数为176832,总动词数为1057177。平均每篇文摘189.30个汉字,平均每篇文摘4.28个句子,平均每篇文摘25.59个动词,平均每个句子有5.98个动词。(2)所统计的大多数学报文摘的高频动词分布规律与表1的顺序大致一致,有些学报出现的频率顺序虽有不同,但高频动词都在表1的范围内。(3)“提出”是11种学报文摘动词

词频数量最高的动词,平均每篇文摘中“提出”这个动词出现的频率达0.593,在《电子学报》平均每篇文摘中出现频率达1.024,在《软件学报》达0.897,在《计算机学报》文摘中达0.884。

对11种核心期刊文摘的句子结构、动词词频的分布统计和对每一种期刊文摘的单个动词词频的统计,为我们深度挖掘学术文摘的创新性知识点提供了可靠的依据。

2.2 学术文摘创造性知识点的本体模型

(1) 学术期刊文摘创造性知识点表达方式

统计表明学术期刊文摘平均每篇有 4.28 个句子,摘要表达创造性知识点的模式一般由 3 个主要部分组成:①说明当前存在的问题;②提出解决问题的创新方法;③对研究结果的说明。

(2) 学术期刊文摘创造性知识点的本体语义关系模型

根据学术期刊文摘创造性知识点表达方式,构建了学术期刊文摘创造性知识点本体语义关系模型,如图 1 所示。

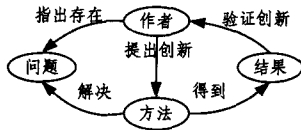


图 1 学术期刊文摘创新点本体语义关系模型

设 O 为学术期刊文摘创造性知识点本体, $O = \{P, M, R\}$, P, M, R 分别表示问题类、方法类、结果类。

根据学术期刊文摘创造性知识点本体语义关系模型,我们把 P, M, R 每个类的语义关系表示成 5 个主谓宾结构。主谓宾结构的语义关系如表 3 所列。

表 3 主谓宾结构的语义关系

主—谓—宾	语义关系图
①作者针对存在的问题	问题 → 针对 → 作者
②作者提出了新方法	作者 → 提出 → 方法
③新方法解决了问题	方法 → 解决 → 问题
④问题得到了结果	问题 → 得到 → 结果
⑤结果验证了方法	结果 → 验证 → 方法

3 文摘创新性知识点挖掘

3.1 建立文摘创新性知识点特征引导词表

(1) 特征引导词定义

根据本文建立的文摘创新知识点本体 $O(P, M, R)$, 建立问题(P)、方法(M)、结果(R)的特征引导词。

定义问题特征引导词 P_V : $P_V = \{ \text{针对, 存在, 不足, 效率不高, } \dots \}$ 。

定义方法特征引导词 M_V : $M_V = \{ \text{提出了, 研究了, 分析了, 设计了, 给出了, 构建了, 发现了, 引入了, 采用了, 建立了, } \dots \}$ 。

定义结果特征引导词 R_V : $R_V = \{ \text{表明, 证明, 提高, 提升, 改善, 解决了, 得到了, 验证, 实现, } \dots \}$ 。

(2) 建立特征引导词表

基于 2.1 节(3)中的学术期刊文摘动词词频统计分析结果,将特征引导词与 WordNet 进行映射,建立特征引导词的同义词、上位词、下位词、相关词的扩展词。在此基础上,我们请专家对特征引导词给予辨识、选择、补充、确定,建立学术期刊文摘创新性知识点挖掘的特征词表库。

3.2 挖掘流程图

挖掘程序的流程如图 2 所示。流程图分 3 部分,左侧是

领域特征动词学习模块,中间是系统特征动词词表库模块,右侧是挖掘模块。

(1) 特征引导动词预学习

由于涉及的挖掘领域不同,我们借助机器学习的方法,在每次挖掘时先对部分文摘进行动词词频预学习,并与特征词表库进行比较和增补新的特征引导词。

(2) 部分挖掘算法描述

步骤 1 读取用户检索出一批文摘,对前 100 条文摘进行动词词频统计,并经过过滤词表过滤和系统特征动词词表比较,判断是否增补新词。

步骤 2 问题、方法、结果子句的识别与挖掘。

步骤 3 如果文摘中存在方法特征引导词集合 M_V ,且 M_V 的子句不是第一句或最后一句,取出该子句作为方法句 M ;否则,放弃这条文摘的挖掘。

步骤 4 如果方法句之前的子句中存在问题特征引导词集合 P_V ,取出该子句作为问题句 P ;否则取出方法句 M 最邻近的前一个子句作为问题句 P 。

步骤 5 如果方法句 M 之后的子句中存在结果特征引导词集合 R_V ,取出该子句作为结果句;否则取最后一个子句作为结果句 R 。

步骤 6 M, P, R 同时满足,创新知识点抽取成功,抽取结果分别放入数据库相应字段。

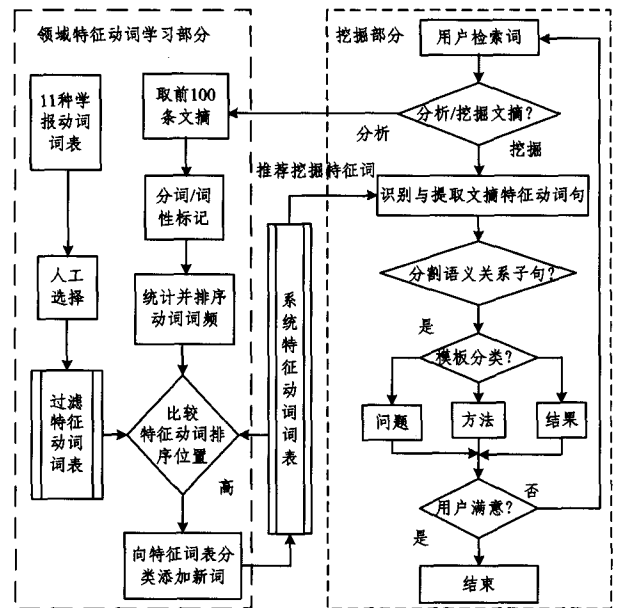


图 2 挖掘流程图

4 挖掘结果分析

4.1 识别率分析

识别率: $\text{识别率} = \frac{\text{挖出正确三元组数}}{\text{总文摘数}}$

对系统而言,正确识别率和正确拒识率都是越高越好,但识别率和拒识率这两组参数是互相矛盾的,一种指标上升必然导致另一种指标下降。本文只对正确识别率进行分析,表 4 和图 3 给出的数据是在特征动词选择在 10~120 个时的正确识别率。为直观分析挖掘结果,根据表 4 的数据绘制出图 3。

表4 特征动词选择在10~120个时的正确识别率

特征词数	11种学报	电子学报	东北财经大学学报	管理学报	计算机学报	交通运输工程学报	教育学报	南方农业学报	软件学报	中国法学	中国社会科学	中国药科大学学报
10	0.297	0.632	0.528	0.319	0.384	0.503	0.085	0.056	0.419	0.116	0.152	0.049
20	0.545	0.646	0.583	0.667	0.549	0.753	0.427	0.631	0.612	0.399	0.483	0.279
30	0.624	0.681	0.614	0.681	0.591	0.806	0.547	0.734	0.650	0.544	0.616	0.392
40	0.635	0.683	0.654	0.690	0.60	0.807	0.60	0.755	0.662	0.609	0.653	0.404
50	0.655	0.684	0.670	0.697	0.602	0.811	0.682	0.755	0.663	0.720	0.679	0.409
60	0.658	0.684	0.684	0.703	0.602	0.812	0.682	0.758	0.663	0.720	0.706	0.417
70	0.661	0.684	0.689	0.709	0.602	0.812	0.682	0.758	0.663	0.720	0.716	0.420
80	0.665	0.684	0.699	0.712	0.602	0.812	0.682	0.758	0.663	0.720	0.725	0.448
90	0.668	0.685	0.708	0.714	0.602	0.812	0.690	0.758	0.663	0.723	0.731	0.450
100	0.699	0.685	0.711	0.714	0.603	0.812	0.691	0.758	0.663	0.729	0.733	0.456
110	0.671	0.685	0.717	0.715	0.604	0.812	0.701	0.758	0.663	0.741	0.738	0.458
120	0.672	0.685	0.720	0.716	0.604	0.812	0.702	0.758	0.663	0.742	0.740	0.464

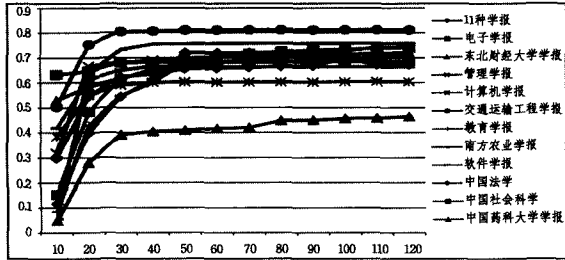


图3 特征动词选择在10~120个时的正确识别率

由图3可以看出挖掘效率有3个明显的阶段:起始上升阶段、阶跃增长阶段、稳定增长阶段。

①起始上升阶段:当特征动词取前10个时,《电子学报》、《东北财经大学学报》、《交通运输工程学报》这3种学报的有效识别率达到了0.5以上;《软件学报》、《计算机学报》、《管理学报》、《中国法学》、《中国社会科学》、11种学报的总和这6种学报都在0.1以上;《中国药科大学学报》和《教育学报》的识别率都在0.1以下。

②阶跃增长阶段:当特征动词取前40个时,10种学报的识别率都达到了0.6以上,《中国药科大学学报》识别率为0.4。

③稳定增长阶段:当选取的特征动词多于40个后,全部进入稳定区。

④最高效率分析:挖掘效率最高的是《交通运输工程学报》,当特征动词取前10个时,其挖掘效率达到0.503,当特征动词取前30个时,识别率达到0.806,特征动词增加到120时,挖掘效率缓慢达到0.81。分析发现,《交通运输工程学报》前30个高频特征动词的每个动词频率都很高,前17个特征动词的词频之和与文摘数之比为4.66。

⑤最低效率分析:效率最低的是《中国药科大学学报》,当特征动词取前10个时,挖掘效率达到0.049,当特征动词取前30个时,识别率达到0.392,特征动词增加到120时,挖掘

效率缓慢达到0.464。分析发现,《中国药科大学学报》前30个特征动词的词频之和与文摘数之比为2.91。此外与其他10个学报相比,《中国药科大学学报》文摘的写作格式与其他学报不同。

⑥对学习功能的挖掘效率分析如表5和图4所示。

表5 学习前 $M_V=40$ 、加学习后, $M_V=120$ 的挖掘效率

学报名称	学习前	学习后	120个特征词
中国药科大学学报	0.404	0.457	0.464
电子学报	0.683	0.686	0.686
管理学报	0.689	0.705	0.716

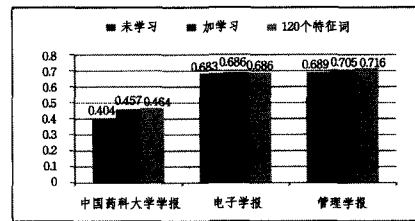


图4 学习前 $M_V=40$ 、学习后, $M_V=120$ 的挖掘效率

由图4可以看出,学习功能对挖掘效率有很大提升,加学习后的结果都超过了取前40个特征动词的效果,接近120个词的效率,也即使用学习功能可以大大减少选择特征动词带来的复杂性。虽然《中国药科大学学报》的挖掘效率不高,但比学习前有很大提升,接近了120个词的效率。

4.2 挖掘结果存储

通过本文的挖掘方法将学术期刊文摘的内容以“问题-方法-结果”三元组方式挖掘出作者的创造性知识点。挖掘结果以时间、学科、作者、问题、方法、结果、原文链接标记等字段存放在创新性知识点库中。表6列出了关于“人脸识别中的光照问题”挖掘结果的部分存储举例。

表6 创新性知识点库

时间	作者	问题	方法	结果
2012	韩鹏	针对不同光照条件下的人脸识别率较低	提出了一种基于垂直积分投影和高光区域处理的光照方向估计方法	实验结果表明该方法可有效地提高在不同光照条件下的人脸识别率
2012	孙劲光	为提高光照变化下的人脸识别率	提出一种基于局部对比增强(LCE)和局部相位量化(LPQ)的人脸识别方法	证实了所提方法的有效性
2012	赵振华	为了减少光照变化对人脸识别算法的影响	提出了一种基于非下采样Contourlet变换的光照不变特征提取方法	实验结果表明-本文方法能有效减少光照影响
2011	曹雪	针对正面光照人脸识别的难点	提出了一种应用小波变换和去噪模型的光照不变人脸识别算法	实验结果表明-所提算法可以显著提高光照人脸图像的识别率
2011	盛剑会	为了消除光照变化对人脸识别的影响	提出了一种新的基于小波的光照归一化算法	本文方法可有效减弱光照变化对人脸识别的影响
2010	程勇	为了消除光照变化对人脸识别的影响	提出一种结合无下采样轮廓波变换和自商图像的光照不变量提取算法	结果表明-所提出的算法能够有效消除光照变化对人脸识别的影响

- [14] Wu S T, Li Y, Xu Y. Deploying Approaches for Pattern Refinement in Text Mining[C]//Proc. IEEE Sixth Int'l Conf. Data Mining (ICDM '06). 2006;1157-1161
- [15] Liao Jia-bin, Lin Shi-ping. Cluster-based Shallow Semantic Pattern Recognition In Chinese [C]//Fifteenth Annual Academic of Information Theory and First National Annual Academic on Network Coding. Qingdao, China, 2008; 1214-1219(in Chinese)
廖加彬, 林世平. 基于聚类的中文浅层语义模式识别[C]//中国电子学会第十五届信息论学术年会暨第一届全国网络编码学术年会. 青岛, 2008; 1214-1219
- [16] Zhong Ning, Li Yue-feng, Wu Sheng-tang. Effective Pattern Discovery for Text Mining[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(1): 30-44
- [17] Wen Hao, Wen You-kui. Research on Discovery Method in Literature Implicit Knowledge Using Semantic Complementary Reasoning[J]. Computer Science, 2014, 41(6): 171-175(in Chinese)
温浩, 温有奎. 基于语义互补推理的文献隐含知识的发现方法研究[J]. 计算机科学, 2014, 46(6): 171-175
- [18] Huang Cheng-hui, Yin Jian, Hou Fang. Improved Text Retrieve Algorithm Based on Subject-verb-object Structure [J]. Computer Science, 2010, 37(9): 173-176(in Chinese)
黄承慧, 印鉴, 侯防. 一种基于主谓宾结构的文本检索算法[J]. 计算机科学, 2010, 37(9): 173-176

(上接第 245 页)

- [3] Ma X, Zhang J F, Yang H F. Heuristic algorithm of attribute reduction based on discernibility matrix[J]. Journal of Computer Applications, 2010, 30(8): 1999-2002(in Chinese)
马翔, 张继福, 杨海峰. 基于区分矩阵的启发式属性约简算法[J]. 计算机应用, 2010, 30(8): 1999-2002
- [4] Yao Y Y, Zhao Y. Discernibility matrix simplification for constructing attribute reducts[J]. Information Sciences, 2009, 179(7): 867-882
- [5] Liu S H, Sheng Q J, Wu B, et al. Research on efficient algorithms for rough set methods[J]. Chinese Journal of Computers, 2003, 26(5): 524-529(in Chinese)
刘少辉, 盛秋戩, 吴斌, 等. Rough 集高效算法的研究[J]. 计算机学报, 2003, 26(5): 524-529
- [6] Ge H, Li L S, Yang C J. Improvement to quick attribution reduction algorithm[J]. Journal of Chinese Computer Systems, 2009, 30(2): 308-312(in Chinese)
葛浩, 李龙澍, 杨传健. 改进的快速属性约简算法[J]. 小型微型计算机系统, 2009, 30(2): 308-312
- [7] Wang G Y, Yu H, Yang D C. Decision table reduction based on conditional information entropy[J]. Chinese Journal of Computers, 2002, 25(7): 759-766(in Chinese)
王国胤, 杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7): 759-766
- [8] Kryszykiewicz M. Properties of incomplete information systems in the framework of rough sets[C]//Polkowski L, Skowron A, eds. Rough Sets in Data Mining and Knowledge Discovery Physica. Heidelberg, 1998, 1: 422-450
- [9] Stefanowski J, Tsoukias A. On the extension of rough sets under incomplete information[C]//Proceedings of the Seventh International Workshop on New Directions in Rough Sets, Data Mining and Granular-Soft Computing. Physica Verlag, Yamaguchi, 1999: 73-81
- [10] Wang G Y. Extension of rough set under incomplete information systems[J]. Computer Research and Development, 2002, 39(10): 1238-1243(in Chinese)
王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238-1243
- [11] Wang W, Xu Z Y, Li X Y. Attribute Reduction Algorithm Based on Object Matrix in Incomplete Decision Table[J]. Computer Science, 2012, 39(4): 201-204(in Chinese)
王炜, 徐章艳, 李晓瑜. 不完备决策表中基于对象矩阵属性约简算法[J]. 计算机科学, 2012, 39(4): 201-204
- [12] Sun L, Xu J C, Tian Y. Feature selection using rough entropy-based uncertainty measures in incomplete decision systems[J]. Knowledge-Based Systems, 2012, 36: 206-216
- [13] Skowron A, Swinarski R, Synak P. Approximation Spaces and Information Granulation[C]//Proc of the International Conference on Rough Sets and Current Trends in Computing. Uppsala, Sweden, 2004; 116-126
- [14] Guan J W, Bell D A, Guan Z. Matrix Computation for Information System[J]. Information Sciences, 2001, 131(1-4): 129-156
- [15] Wang L, Li T R. Matrix-Based Computational Method for Upper and Lower Approximations of Rough Sets[J]. Pattern Recognition and Artificial Intelligence, 2011, 24(6): 757-762(in Chinese)
王磊, 李天瑞. 基于矩阵的粗糙集上下近似的计算方法[J]. 模式识别与人工智能, 2011, 24(6): 757-762
- [16] Jing Y G, Li T R. Reduction Algorithm of Positive Domain for Decision Table Based on Relationship Matrix[J]. Computer Science, 2013, 40(11): 261-264(in Chinese)
景运革, 李天瑞. 一种基于关系矩阵的决策表正域约简算法[J]. 计算机科学, 2013, 40(11): 261-264
- [17] Inuiguchi M, Tsurumi M. Measures based on upper approximations of rough sets for analysis of attribute importance and interaction[J]. International Journal of Innovative Computing, Information and Control, 2006, 2(1): 1-12
- [18] Parthalain N M, Shen Q. Exploring the boundary region of tolerance rough sets for feature selection[J]. Pattern Recognition, 2009, 42: 655-667
- [19] Lu Z C, Zheng Q, Zhang Y Q, et al. A fast feature selection approach based on rough set boundary regions[J]. Pattern Recognition Letters, 2014, 36: 81-88
- [20] Chen Y Y, Chen Y M. Attribute Reduction Algorithm Based on Information Entropy and Ant Colony Optimization[J]. Journal of Chinese Computer Systems, 2015, 36(3): 586-590(in Chinese)
陈颖悦, 陈玉明. 基于信息熵与蚁群优化的属性约简算法[J]. 小型微型计算机系统, 2015, 36(3): 586-590