基于CABOSFV聚类算法的汉语词汇类别知识挖掘研究

王东波[1] 朱丹浩[2]
(南京农业大学信息科学技术学院 南京 210095)[1] (联合国大学-国际软件技术研究所 澳门 3058)[2]

摘 要 在清华大学973汉语树库的基础上,根据汉语词汇的句法功能分布状况,构建了句法功能分布知识库。在构建的句法功能分布知识库基础上,使用CABOSFV(Clustering Algorithm Based On Sparse Feature Vector)聚类算法,从中挖掘了汉语词汇的类别知识,并对这些类别知识逐一进行了分析。

关键词 973汉语树库,句法分布功能,知识库,CABOSFV
中图法分类号 TP301.6,TP309 **文献标识码** A

Research of Mining Word Category Knowledge Based on CABOSFV

WANG Dong-bo[1] ZHU Dan-hao[2]
(School of Information Sciences and Technology, Nanjing Agricultural University, Nanjing 210095, China)[1]
(The United Nations University International Institute for Software Technology, Macao 3058, China)[2]

Abstract According to the Chinese word syntactic function distribution, the paper constructed syntactic function distribution knowledge base based on Tsinghua 973 treebank. The Chinese word category knowledge was mined by using the CABOSFV(Clustering Algorithm Based On Sparse Feature Vector) based on syntactic function distribution knowledge base. The Chinese word categories were analyzed one by one.

Keywords 973 Chinese treebank, Syntactic function distribution, Knowledge base, CABOSFV

1 引言

随着自然语言处理和文本挖掘技术的发展,从非结构化文本中挖掘和抽取相应的专门或通用知识以便更好地服务于基础和应用研究日益成为一种趋势。本文基于清华大学树库,通过CABOSFV(Clustering Algorithm Based On Sparse Feature Vector)聚类算法,在汉语词汇句法功能分布知识库的基础上,挖掘词汇具体类别知识的研究正是在这一趋势下的一种尝试。本文挖掘的类别知识不仅可以应用到中文信息处理的汉语句法结构歧义消解上,而且对于构建大规模精确度更高的汉语树库具有重要的促进作用。

Arats[1]使用手工的方法,在72000个英语惯用法调查的词汇基础上,考察了名词短语的句法功能分布,不仅得出名词短语的分布受到句法功能的影响,并且量化地得出作主语的轻小名词短语占到整个名词短语的85.33%。该研究以最原始的统计方法考察了句法功能与相关语言单位的关系,为后续者的研究提供了新的视角。Haan[2]探讨了句法功能与名词短语种类的关系,考察了名词短语在主语、宾语和介宾等句法上的分布情况,但该研究仅为少量的统计分析,数据证明不充分。Maestre[3]使用统计的方法,在由《时代周刊》上的文章标题组成的语料库中,根据名词短语在句法功能中的分布情况把复杂名词短语分成了带前置修饰词、带后置修饰词和带前后置修饰词3类。该研究虽然使用了语料库,但由于技术条件的限制,该语料库不是经过深加工的树库,因此对复杂名词短语的句法功能的考察和分类来说是欠缺的和不完善的。Maienborn[4]在修饰词的研究上给出了句法和语义的证据,即关于句法的基本位置和解释上,在言语领域有3种不同的位置修饰词,并且三者之间存在一定的差异,从句法和语义界面的存在条件中推断出的位置修饰词的句法分布在研究中也有所说明。Uzuner和Katz[5]从语言信息可以用于改进相似文档的评价这一角度入手,把句法功能分布的特征和知识应用到了书籍识别和其作者归属的分类实验中,在使用句法结构分布的实验中,个人书籍的识别正确率达到了76%。Kalampakas[6]提出直接的欧拉图集合是不被识别的,而带有基本图的直接欧拉图集合容易被识别,同时计算出了这种图语言的句法复杂度功能,并与连接图进行了相关的对比。Stepanov和Tsa[7]证明了不同的词条呈现不同的句法分布,并给出了"how"和"why"的不同的句法分布情况。

关于依据汉语词汇句法功能对整个汉语词汇进行分类的研究,相关的研究者进行了多方面的探究。在朱德熙提出利用词汇的句法功能进行分类的设想基础上,从词汇句法功能要为汉语完全自动句法分析服务这一目的出发,陈小荷[8]给出了主要依据汉语词汇的句法功能分布对汉语主要实词词汇进行分类的理念。从上述词汇句法功能的设想和理念出发,

到稿日期:2012-09-29 返修日期:2013-02-26 本文受863计划项目(2011AA01A206),自然科学基金面上项目(71273126)资助。
王东波(1981—),男,博士,讲师,主要研究方向为中文信息处理与文本挖掘、信息计量;朱丹浩(1986—),男,助理研究员,主要研究方向为自然语言处理。

基于人民日报大规模语料，徐艳华[9]在计算机的辅助下，通过手工逐一分析了一般名词、动词、形容词以及高频副词的语法功能，并完成了3514个常用汉语词汇的简单语法功能数据库的构建。

上述研究者在基于词汇句法功能的基础上，从理念、人工分类和应用的角度上，都对整个词汇的分类问题进行了探索和研究，本文的研究正是上述研究的一个延续。在具体的研究过程中，本文使用了相应的聚类算法从词汇句法功能知识库中挖掘出相应的词汇知识，关于聚类算法，国内外研究者在单语和双语上都进行了多方面的探究。在单语文本的聚类研究中，国外的相关学者[10-12]给出了各种文本聚类的方法以及相应的改进算法。在国内，崔尚卿[13]从基于密度的聚类算法本身存在的问题出发，提出了不均匀密度的自动聚类算法。王伟[14]针对文本聚类的技术从各个角度进行了综述，关注了文本聚类中的文本聚类定义、文本预处理、文本特征抽取和文本聚类的各算法中的各种方面的问题。王舵[15]针对聚类速度慢、初值对结果的影响大、易陷入局部最优等问题，提出了基于相似度测度和覆盖方法来解决问题。潘章明[16]将基于差分进化的自动聚类算法和半监督方法结合起来，在具体的实验中，启发式地引入部分已经获得了带标签的数据，聚类的精确度也得到了有效的提高。于洪[17]从面向知识的聚类框架出发，使用决策粗糙集理论对聚类算法进行了研究，并且考虑了一种特定的损失函数。

2 句法功能知识库构建

2.1 清华大学树库简介和结构转换规则制定

按照完整的层次结构描述方法，在16种汉语短语成分标记和27种句法结构关系的前提下，清华大学计算机系智能技术与系统国家重点实验室在以文学、学术、新闻、应用4种体裁构成的汉语平衡语料基础上构建了一个100万字的汉语树库。清华汉语树库的词性标记是依据“依句辨品”理论进行的。清华树库对每个句子进行了细致和全面的句法分析，对每一个短语结构给出了句法和功能两种标记。具体例子如下：

[zj-XX [dj-ZW 海南/nS [vp-ZZ [pp-JB 在/p [sp-FW [np-DZ 人/n 的/u 脑海/n] 里/f]] [vp-ZZ 总是/d [vp-PO 留下/v [np-DZ [dj-ZW 对比/v 强烈/a] 的/u 印象/n]]]]] 。/。]

为了确保所获取的汉语词汇句法功能的真实性、充分性和精确性，依据相关的原则对树库中的句法结构进行了有针对性的调整。本研究所提到的句法结构调整主要是指删除某些不含句法意义的结构、拆分涵盖范围过宽的句法结构、合并从实际句法需求考虑无需细分的结构这三大类，在清华树库中就删除了含有“－NONE－”的这类不包含任何词汇的句法结构标记。制定相关的转换规则，在清华大学原树库的基础上完成相应句法结构的转换。具体的清华汉语树库结构转换流程如下：

循环读入每一个句子 PSen

循环抽取子结点，从中取出结点中的结构名称 PStr

如果(－NONE－ *)∈PStru 且(－NONE－ *)不含任何词汇

删除(－NONE－ *)整个结构

如果(－NONE－ *)∈PStru 且(－NONE－ *)含有一个以上的词汇

删除结构中的(－NONE－ *)成分，并且把其兄弟结点提升为父结点

如果(－)∈PStru

删除结构中(－)后面的表示语义的成分

如果(＝)∈PStru

删除结构中(＝)后面的表示语义指向、数量的成分

如果 PStru＝＝RM

删除 RM 这一结构名称和内部所有不表示句法和语义的成分

如果 PStru＝＝SBARQ

删除 SBARQ 这一结构名称并且不统计其内部“?、?”的句法成分

同时把 SBARQ 中的相应结构提升一级

如果 PStru＝＝S

删除 S 这一结构名称，同时把 S 中的相应结构提升一级

如果 PStru＝＝PRN

删除 PRN 这一结构名称，同时把 PRN 中的相应结构提升一级

如果 PStru＝＝IP

删除 IP 这一结构名称，并把 IP 内部没有任何句法和语义的成分删除

如果 PStru＝＝RS

删除 RS 这一结构名称，并把 RS 内部没有任何句法和语义的成分删除

如果 PStru＝＝EDITED

删除 EDITED 这一结构名称，并把 EDITED 中的相应结构提升一级

如果 PStru＝＝ADV

删除 ADV 这一结构名称，并把 ADV 内部错误的标注成分去掉

如果 PStru＝＝VB

删除 VB 这一结构名称，并把 VB 内部错误的标注成分去掉

如果((``)、(‘’)、(,,)、(:——)、(..)、(:;)、(.?)、(.!))∈Stru

不用统计无句法和语义的标点的句法功能

如果((PU 。)、(PU,)、(PU 、)、(PU ?)、(PU !)、(PU ———)、(PU “)、(PU ”)、(PU ‘)、(PU ’)、(PU《)、(PU 》)、(PU;)、(PU:))∈Stru

不用统计无句法和语义的标点的句法功能

输出经过结构转换后的句子 T-PSen

清华大学汉语树库结构转换完成

2.2 汉语词汇句法功能的统计与词汇句法功能知识库的构建

多叉树被用来存放整个句法树，并使用树遍历算法遍历每一棵句法树，从而可以便捷地进行句法树的短语结构或词汇功能存储、统计、编辑和删节等。

2.2.1 多叉树概念及特点

作为非线性结构中的重要结构，树的一个重要特点是其具有递归性，这也是采用树结构存储每一个经过完全句法分析后的句子的原因之一。与二叉树相比，多叉树中每个节点的子节点数没有限制，更便于存储具有多个子节点或叶子的

短语结构、词汇和词性。图 1 是一棵常见的多叉树。T 是整棵树，T_1、T_2、T_3 分别都是子树。

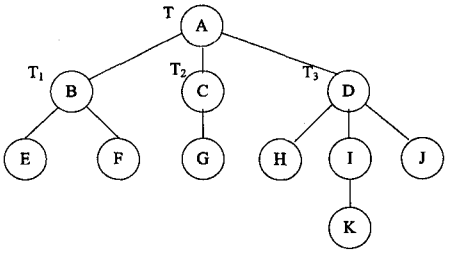

图 1　多叉树图示

2.2.2　清华汉语句法结构多叉树实现

汉语句子树使用“（）”对称的标记符号标记了句法结构的起始和结尾。从多叉树数据结构特点出发，结合语料库标记，构建句法多叉树的流程如下：汉语句子构成一颗多叉树；每个短语结构标记构成多叉树中的非叶节点；每个词汇或者标点符号构成叶节点。汉语句法多叉树的生成流程和算法实现见图 2。

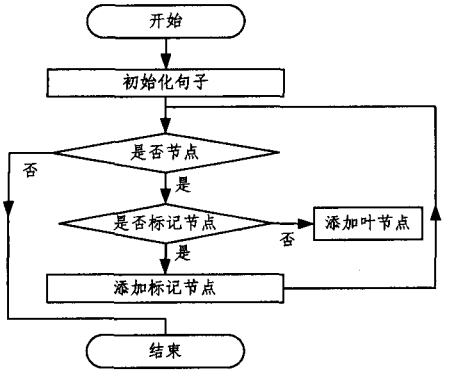

图 2　汉语语句子树生成流程图

汉语句法树构造算法如下：

步骤 1　开始扫描整个汉语语句子。

步骤 2　第一个节点（左边没有其它字符标记）若为汉语句法节点，则视为 root 节点，将 root 节点 ID、值等属性压入栈。层深 layer 为 0。否则退出，不具备构建英汉句法树条件。

步骤 3　顺序扫描下一个节点，判断其是否为汉语句法标记，是则跳至步骤 4，否则跳至步骤 5。

步骤 4　将汉语句法节点标记节点 ID、节点值、父节点等属性压入栈，根据上一个节点改变 layer。

步骤 5　根据汉语的“(（”标记取得叶子节点，将节点属性压入栈。当上一个节点为汉语句法标记，则 layer＋1，其它情况，layer 不变。

若 layer 不为 0，跳至步骤 3。若为 0，汉语句法树构建成功，退出程序。

基于汉语的句法生成树，使用根节点深度优先的方法，通过树检索遍历了整个树结构，当遍历到叶子节点的时候，抽取词汇，并且记录其父节点所在的句法结构。如基于图 3，通过上述统计流程，本文获取了更为细致的 26 种汉语短语结构中 55508[1)] 个汉语词汇所充当的句法功能，并构建了汉语词汇句法功能知识库。具体选取的 10 个汉语词汇的句法功能例子见表 1。

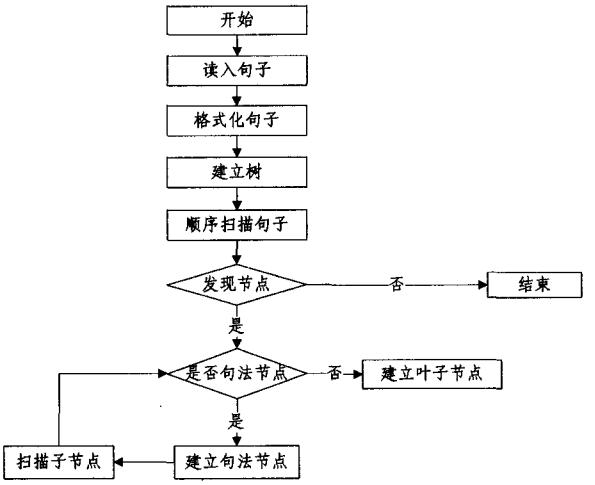

图 3　汉语词汇句法功能统计流程图

表 1　10 个汉语部分结构的词汇句法功能样例

词汇＼功能	的	哲学	和	发展	地区	认为	思想	客观	静	这样
ZW	0	40	0	22	0	36	0	0	4	0
DZ〈de〉	2075	83	2	54	1	0	19	2	0	10
DZ	0	313	0	28	2	0	41	21	1	4
SL	0	0	0	0	0	0	0	0	0	2
ZZ〈de〉	0	0	0	1	0	0	0	0	0	0
ZZ	0	0	0	13	0	2	0	1	3	4
SB	0	0	0	0	0	0	0	0	0	0
SB〈de〉	0	0	0	0	0	0	0	0	0	0
SB〈bu〉	0	0	0	0	0	0	0	0	0	0
PO	2	4	0	6	0	123	2	0	5	1
FW	0	8	0	1	0	0	0	0	0	0
JB	0	7	31	0	0	0	4	0	2	0
LW	0	0	0	1	0	0	0	0	0	0
JY	0	0	0	0	0	3	0	0	0	0
AD	145	0	0	6	0	0	0	1	1	0
KS	104	0	0	0	0	0	0	0	0	0
SX	0	0	0	0	0	0	0	0	0	0
LH	2	9	321	16	0	0	0	1	9	2
FH	0	0	2	0	0	0	0	0	1	1
FJ	0	0	0	0	0	0	0	0	0	0

3　CABOSFV 聚类算法及实现流程

CABOSFV 是基于密度的子空间划分方法，核心在于提出了一个新的概念 Sparse Feature Vector 用以说明 CABOSFV 用一个稀疏特征向量描述聚类集合。

在集合的基础上，CABOSFV 完成了对聚类中的相似度的计算。Sparse Feature Vector 保留了聚类集合中有关稀疏的信息以及集合中的数据对象，其定义如下：

$$SFV(X)=(|X|,S(X),NS(X),SFD(X))$$

在设定稀疏度为 μ 的前提下，利用 CABOSFV 算法对 n 个对象完成聚类，取出一个对象，并将其定义为第一个类。后续取的对象被加入到已有类的集合中，若 n 个类的值大于稀疏度的值 μ，则表明此对象不能和现有任意一类聚合在一起，那么就将其放入其他新的类中，反之，把该类归入稀疏差异度最小的一类中。算法流程如图 4 所示。

1)　根据句法功能统计的需要，对一些词汇进行了拆分和细化，从更详细的角度调整了相应的实验。

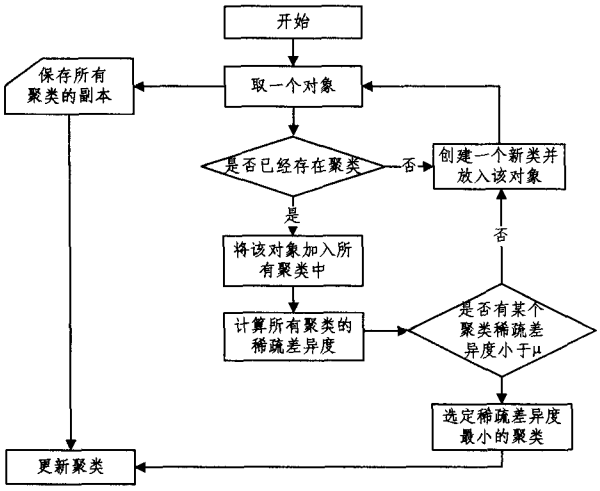

图 4　CABOSFV 算法流程图

4　聚类挖掘的类别结果及分析

4.1　CABOSFV 的类别知识挖掘结果

本文在 CABOSFV 算法的基础上，基于调整过的清华汉语词汇句法功能知识库，设定 SFD 阈值为 0.3，自动聚类后获得 20 类。具体如表 2 所列。

表 2　基于清华汉语树库的聚类结果

编号	词汇数目	SFD 值	编号	词汇数目	SFD 值
1	12757	0.003332	11	527	0.034622
2	9304	0.003929	12	626	0.049683
3	8167	0.005842	13	303	0.039126
4	5798	0.006253	14	197	0.061237
5	3690	0.010334	15	56	0
6	3879	0.009094	16	45	0.018282
7	2904	0.010753	17	85	0.022232
8	2802	0.012807	18	22	0
9	2768	0.013975	19	67	0.027027
10	1468	0.014831	20	43	0.017027

SFD 表示聚类的稀疏差异度，一个聚类的 SFD 值越接近于 0，说明该聚类的内聚性越好，聚类内的数据密集程度越高。如果 SFD 值等于 0，说明该聚类所有的对象分布完全一致，也就是该词条的所有句法功能分布完全一致。由于本文的重心是利用 CABOSFV 聚类算法完成基于词汇句法功能知识库的词汇类别知识获取，对其中的数据稀疏问题就没有展开具体的研究。根据表 1 的分布情况可以得出如下结论：

(1)在统计的所有具有句法功能的词汇中，绝大多数词汇的句法功能倾向于某几类的聚类，在这几个类别中汉语词语在语法功能分布上有着很高的一致性。前 3 类词汇数目为词条 26228 个，占了总词数的 54.46%，远超于已有研究的比重。

(2)而在中间的一些聚类出来的词汇，其处于 1000 以下的类别比较多，这类词汇在一定程度上充分体现出了在句法上所具有的歧义，是自然语言处理应该重点关注的对象。

(3)有两类 SFD 值都是 0，说明这两类中的词汇句法功能是完全一致的，但数量比较少，这在一定程度上说明了，在词汇句法功能上完全一致的词汇数量确实比较少，这也符合自然语言在句子中的分布序列事实。

4.2　具体类别中的词汇简要分析

表 3 给出了基于清华汉语词汇句法功能知识库，在 CABOSFV 聚类算法的基础上挖掘出来的词汇类别知识的前 10 类。

在对基于汉语词汇句法功能知识库挖掘出的词汇的 20 个类别知识的基础上进行分析可以看出，词汇数量分布极为不均匀，这也在一定程度上反映了汉语词汇类别的事实，即汉语词汇主要是集中在名词、动词、形容词和副词等几大类别上，而通过聚类知识获取的类别知识验证了这一事实。下面对词汇数量比较多的前 10 类结果进行一定分析。

第 1 类中主要是名词和动词这两类，又主要是名词，选取给出的 10 个样例就是一个证明。该类聚集的基本上是词汇句法功能主要出现在宾语这个位置上的词汇，有些动词也会出现在这个位置，所以从词汇句法功能的角度进行词汇归类，势必把一些动词与名词归在一起。

第 2 类充分体现了汉语名词与动词的兼类词问题。从聚类的结果中可以看出，这一类主要是动名兼类的问题比较突出，从具体句法功能分布的角度考虑，如 vN 这类词与动词的相似度更高一些，所以在第 2 类中出现了一定量的动词。

第 3 类的词汇类别知识聚集充分说明虽然人为地对一些词汇进行了细致的划分，但从充当的句法功能上看，这样的划分不一定便于解决问题，尤其是对于自然语言处理中的歧义问题解决来说。

第 10 类的词汇分类充分说明了汉语中的介词和动词在句法功能上相似度的值是非常接近的，因此在处理动词的句法歧义问题的时候，应该在一定程度上借鉴介词的相应知识。

表 3　基于汉语词汇句法功能知识库挖掘出来的词汇类别

编号	词汇数目	词汇例证
1	12757	财政/n，国用/n，国家/n，思想/n，时期/n，赋税/n，负担/n，财富/n，原则/n，主张/v
2	9304	增长/vN，就业/vN，实现/v，是/vC，形成/vN，预期/vN，预期/v，改革/vN，能/vM，控制/vN
3	8167	中国/nS，苏联/nS，美国/nS，意大利/nS，法国/nS，C·卓别林/nP，祝福/nR，林家/nP，水平/n，方面/n
4	5798	有/v，结合/v，起来/vB，首当其冲/iV，中断/v，创业/v，停顿/vN，生产/vN，发展/vN，改革/vN
5	3690	成立/v，环境/n，垄断/vN，武装/vN，破坏/vN，借鉴/vN，操作系统/n，电影/n，联系/vN，沟通/vN
6	3879	绘画/n，小说/n，戏剧/n，审美/vN，容纳/vN，消费/vN，继承/vN，优秀/a，有益/a，本土/n
7	2904	新/a，蓬勃/a，一般/a，广泛/a，显著/a，重要/a，强有力/a，蓬勃/aD，迅速/aD，直接/aD
8	2802	别是/d，首次/d，很/dD，最先/d，相继/d，先后/d，更加/dD，特别是/d，大大/d，自动/d
9	2768	应用/vN，看到/v，研制/v，发明/v，问世/v，高/a，完整/a，低/a，点/n，水下/n
10	1468	用/p，将/p，有关/p，按/p，从/p，扩展/v，达/v，对于/p，由于/p，所/u

结束语　本文基于清华汉语树库，通过多叉树存储结构，统计了汉语词汇的句法功能，并在统计的词汇句法功能基础上构建了汉语词汇句法功能知识库。基于 CABOSFV 聚类算法，在汉语词汇句法功能知识库的基础上，挖掘出了清华汉语所有词汇的类别知识，并对获取的词汇类别特征通过所选取的前 10 个词汇进行了相应的分析。在下一步的研究中，一

方面将扩大树库的规模，比如把宾州汉语树库通过结构转换增加到清华汉语树库中，同时，把本文获取类别知识的方法应用于以印欧语言为代表的英语当中；另一个方面把通过聚类方法挖掘到的类别知识应用到具体的句法结构消歧中，具体通过该方法获取的词汇类别知识的优越性加以验证。

参 考 文 献

[1] Aarts F. On the distribution of noun-phrase types in English clause-structue [J]. Lingua, 1971(26): 281-293

[2] Haan P D. Postmodifying clauses in the English Noun Phrase: a corpus-based study [M]. Amsterdam: Rodopi, 1989

[3] Maestre M D L. Noun Phrasecom Plexityasasty lemarker: anexe-reiseinstylisti analysis [J]. Atlantis, 1998(2): 91-105

[4] Maienborn C. On the Position and Interpretation of Locative Modifiers [J]. Natural Language Semantics, 2001, 9(2): 191-240

[5] Uzuner, Katz B. A Comparative Study of Language Models for Book and Author Recognition [J]. Lecture Notes in Computer Science, 2005(3651): 969-980

[6] Kalampakas A. The Syntactic Complexity of Eulerian Graphs [J]. Lecture Notes in Computer Science, 2007(4728): 208-217

[7] Stepanov A, Tsa W D. Cartography and licensing of wh-adjuncts: a cross-linguistic perspective [J]. Natural Language & Linguistic Theory, 2008, 26(3): 589-638

[8] 陈小荷. 从自动句法分析角度看汉语词类问题[J]. 语言教学与研究, 1999(03): 72

[9] 徐艳华. 现代汉语实词语法功能考察及词类体系重构[D]. 南京: 南京师范大学, 2006

[10] Boley D, et al. Partitioning-Based clustering for web document categorization[J]. Decision Support System Journal, 1999, 27(3): 329-341

[11] Mao J, Jain A K. A self-organizing network for hyperellipsoidal clustering [J]. IEEE Trans. Neural Networks, 1996, 7(2): 16-29

[12] Cai W L, Chen S C, Zhang D Q. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation[J]. Pattern Recognition, 2007, 40(3): 825-833

[13] 崔尚卿. 基于不均匀密度的自动聚类算法[J]. 计算机工程, 2008(23): 86-88

[14] 王伟. 文本自动聚类技术研究[J]. 情报杂志, 2009(02): 94-96

[15] 王舵. 一种快速词自动聚类算法[J]. 计算机应用与软件, 2010(08): 277-278

[16] 潘章明. 半监督的自动聚类[J]. 计算机应用, 2010(03): 2614-2616

[17] 于洪. 一种基于决策粗糙集的自动聚类方法[J]. 计算机科学, 2011(1): 221-224

(上接第 177 页)

参 考 文 献

[1] Zdzislaw P. Rough Sets[J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356

[2] Zdzislaw P. Why Rough Sets? [A]// The Fifth IEEE International Conference on Fuzzy Systems[C]. Louisiana, New Orleans: IEEE Press, 1996: 738-743

[3] Richard J, Shen Qiang. Fuzzy-Rough Sets Assisted Attribute Selection[J]. IEEE Transactions on Fuzzy Systems, 2007, 15(1): 73-89

[4] Didier D, Henri P. Rough Fuzzy Sets and Fuzzy Rough Sets[J]. International Journal of General Systems, 1990, 17(2/3): 191-209

[5] Nehad M, Yakout M. Axiomatics for Fuzzy Rough Set[J]. Fuzzy Sets System, 1998, 100(1-3): 327-342

[6] So Y D, Chen De-gang, Eric T C C, et al. On the Generalization of Fuzzy Rough Sets[J]. IEEE Transactions on Fuzzy System, 2005, 13: 343-361

[7] Hu Qing-hua, Zhang Lei, Chen De-gang, et al. Gaussian Kernel based Fuzzy rough Sets: Model, Uncertainty Measures and Applications[J]. International Journal of Approximate Reasoning, 2010, 51: 453-471

[8] Hu Qing-hua, Yu Da-ren, Pedrycz W, et al. Kernelized Fuzzy Rough Sets and Their Applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(11): 1649-1667

[9] Hong T-P, Wang T-T, Wang S-L, et al. Learning a Coverage Set of Maximally General Fuzzy Rules by Rough Sets[J]. Expert Systems with Applications, 2000, 19(2): 97-103

[10] Tsai Y-C, Cheng C-H, Chang Jing-rong. Entropy-Based Fuzzy Rough Classification Approach for Extracting Classification Rules[J]. Expert Systems with Applications, 2006, 31(2): 436-443

[11] Wang Xi-zhao, Tsang E C C, Zhao Su-yun, et al. Learning Fuzzy Rules from Fuzzy Samples Based on Rough Set Technique[J]. Information Sciences, 2007, 177(20): 4493-4514

[12] Li Tian-rui, Ruan Da, Greet W, et al. A rough Sets based Characteristic Relation Approach for Dynamic Attribute Generalization in Data Mining[J]. Knowledge-Based Systems, 2007, 20(5): 485 -494

[13] 陈水利. 模糊集理论及其应用[M]. 北京: 科学出版社, 2006: 10-123

[14] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001: 168-178

[15] Chen Sheng, Cowan C F N, Grant P M. Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks [J]. IEEE Transactions on Neural Networks, 1991, 2: 302-309