

基于多粒度视角下的 D-S 证据理论融合策略

林国平^{1,2,3} 梁吉业^{1,2} 钱宇华^{1,2}

(计算智能与中文信息处理教育部重点实验室 太原 030006)¹

(山西大学计算机与信息技术学院 太原 030006)² (闽南师范大学数学与统计学院 漳州 363000)³

摘要 D-S 证据理论与多粒度粗糙集是两类不同的信息融合方法。对该理论展开详细的论述,找出两者之间的联系。根据这两者之间的互补性,提出了一种多粒度与证据理论相结合的新的融合策略,称之为基于多粒度视角的 D-S 证据理论的粒度融合方法。最后,通过实例说明了该融合算法的有效性。

关键词 多粒度, D-S 证据理论, 粗糙集, 信息融合

中图法分类号 TP18 **文献标识码** A

Multigranulation View Based Fusing Strategy of D-S Evidence

LIN Guo-ping^{1,2,3} LIANG Ji-ye^{1,2} QIAN Yu-hua^{1,2}

(Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, China)¹

(School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)²

(School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China)³

Abstract D-S evidence theory and multigranulation rough set theory are different information fusion methods. The relationship between these theories was addressed and the completeness property was found. A new fusion strategy called the combination fusion of D-S theory and multigranulation rough set theory was presented. An example was employed to illustrate the effectiveness of the proposed fusion method.

Keywords Multigranulation, D-S evidence theory, Rough set, Information fusion

1 引言

在即将来临的信息化时代,进入系统的信息越来越多,来源越来越广,相关层次越来越复杂。因此,在信息处理中心,绝不允许多源的观测信息仅仅是持续简单的堆积,而需要对其进行有效的融合处理。而信息表现形式的多样性、信息容量以及信息的处理速度等要求都已大大超出了人脑的信息综合能力,因此,信息融合技术便应运而生。

所谓多源信息融合^[1]是通过适当的融合策略或算法,实现具有相关性和互补性的多源信息的有效综合与利用,以期得到一个比单源信息更优的结果。多源信息分别是对被观测系统各种属性或特征以及背景或环境信息给出的定量表示或定性描述。多源信息系统恰好可以用来表示这样的多源信息。因此对多源信息的组合问题可以转化为对多源信息系统的数据分析问题。多源信息系统,顾名思义,是由一族单源信息系统构成的,并且每一个源都是一个单独的信息系统。形式化地,一个多源信息系统是由一族子信息系统构成的。从粒计算的角度来看^[2,3],对每一个子信息系统,根据属性描述所诱导的二元关系决定所对应的粒结构。换句话说,多源信

息系统中来自不同源的信息可以看成不同的粒空间,从而多源信息融合问题也变为多粒空间融合问题。

D-S 证据理论^[4,5]是信息融合技术中使用较广的一种方法^[1]。该理论具有很强的处理不确定信息的能力。它给出了将多个证据合成的公式,为信息融合提供了解决手段。近年来,许多学者针对 D-S 证据理论在解决冲突问题时产生的弊端展开了广泛研究,并提出了许多改进规则来代替原有的 Dempster 组合规则。较为经典的是 Smets、Yager、Dubois 以及 Smarandache 和 Dezert 等著名学者提出的以其名字命名的规则^[6-10],国内许多学者也提出了改进的合成规则^[11-14],比如,杨善林等^[11]通过配置面向具体问题的可变参数来修正原始证据模型。在应用时,该方法的关键是如何确定相关焦元的修正系数。可使用神经网络遗传算法等不同的方法得到不同的修正系数。

1982 年波兰学者 Pawlak 提出粗糙集理论^[15,16],其无需提供所需处理数据集合之外的任何先验信息,用论域的已知划分来刻画目标概念以进行规则获取。近二十年,Pawlak 粗糙集得到不同深度和广度的推广。需要指出的是钱宇华、梁吉业等^[17,18]于 2006 年提出多粒度粗糙集,它不仅是经典粗

到稿日期:2013-05-20 返修日期:2013-07-29 本文受国家自然科学基金资助项目(10671173,10971186),国家自然科学基金重点项目(71031006),国家重大基础研究计划(973)子课题(2013CB329404),教育部新世纪优秀人才支持计划(20120301),高等学校博士学科点专项科研基金(20121401110013)资助。

林国平(1978—),女,博士生,讲师,主要研究方向为粒计算及信息融合,E-mail:guopin@163.com;梁吉业(1962—),男,教授,博士生导师,主要研究方向为粒计算及数据挖掘、人工智能;钱宇华(1976—),男,教授,博士生导师,主要研究方向为粒计算及数据挖掘、人工智能。

粗糙的有意义的推广而且还作为一类新的定性的融合方法。在多粒度粗糙数据分析过程中,根据“求同存异”和“求同排异”的策略分别发展了乐观多粒度粗糙集和悲观多粒度粗糙集。“求同存异”策略指的是:每个决策者根据自己的粒空间进行决策,而不反对其他决策者所给出的粒空间的决策,是一种乐观的决策策略。“求同排异”的策略指的是:有决策者使用共同满意的方案进行决策,而存在分歧的方案则不能用于决策,是一种悲观或保守的决策策略。本文将根据粗糙集与证据理论之间的关系^[19,20],在多粒度视角下拟发展一类新的融合策略:基于多粒度视角下的证据理论的融合方法。

2 基本概念

2.1 多源信息系统

一个给定的多源复杂数据通常可由一个多源信息系统来表示。多源信息系统,顾名思义,是由一族单源信息系统构成的,并且每一个源都是一个单独的信息系统。形式化地,表示为 $MS = \{IS_i | IS_i = (U, AT_i, \{V_a\}_{a \in AT_i}, f_i)\}$, 其中 $IS_i = (U, AT_i, \{V_a\}_{a \in AT_i}, f_i)$ 是一个四元信息系统, $U = \{u_1, u_2, \dots, u_N\}$ 为论域; $AT_i = \{a_{i1}, a_{i2}, \dots, a_{i|AT_i|}\}$ 为属性集; $\{V_a\}_{a \in AT_i}$, V_a 表示属性 $a \in AT_i$ 的值域; $f_i: U \times AT_i \rightarrow \{V_a\}_{a \in AT_i}$ 是一个信息函数,即对 $\forall u \in U, a \in AT_i$, 有 $f(u, a) \in V_a$ 。

对每一个子信息系统 IS_i , 由 AT_i 的描述所诱导的二元关系(等价关系、相容关系、邻域关系或者优势关系)决定所对应的粒结构(粒空间)。即从粒计算的角度看,根据某个特定二元关系,得到论域上的一个划分或者覆盖,从而构成一个粒空间。因此,对多源信息系统的不确定性数据融合问题可以转化为多粒空间的融合问题。

2.2 D-S 证据理论

D-S 证据理论^[4,5]基于人们对客观世界的认识,根据人们掌握的证据和知识,对不确定事件给出不确定度量。在 D-S 理论中,习惯上把证据的信任度类似于物理对象的质量去考虑,即证据的质量(Mass)支持了一个信任,关于质量这一术语也被称为基本概率赋值(BPA, the Basic probability assignment)或简称为基本赋值(Basic assignment)。

定义 1 对给定的有限论域 U , 我们可定义一个函数 $M: 2^U \rightarrow [0, 1]$, 满足下列条件:

- (1) $M(\Phi) = 0$;
- (2) $\sum_{A \subseteq U} M(A) = 1$

则称 M 为 2^U 上的基本概率分配函数(BPA); $M(A)$ 为 A 的基本概率数。BPA 表示在当前环境下证据源对某个子集的信任程度。若 A 由多个元素构成, 则 $M(A)$ 表示对子集 A 中所有元素的总体信任程度。如果令 $Bel(A) = \sum_{B \subseteq A} M(B)$, $\forall A \subseteq U$, 则称 $Bel: 2^U \rightarrow [0, 1]$ 为信任函数。若 $A \in 2^U$ 且 $M(A) > 0$, 则称 A 为 Bel 的焦元, 所有焦元的并成为它的核。若定义函数 $Pl(A) = 1 - Bel(\sim A)$, $\forall A \subseteq U$, 则称 pl 为似然函数, 它包含了所有与 A 相容的那些集合(命题)的基本可信度。事实上, $[Bel(A), pl(A)]$ 用于表示当前环境下子集 A 的不确定区间。

3 基于多粒度视角的 D-S 证据理论融合策略

D-S 证据推理是处理不确定问题的重要方法, 但实用中存在着许多问题, 如假设集的基本概率指派往往由专家事先确

定, 带有较强的主观性。利用粗糙集理论和证据理论的关系^[19,20], 计算了多源信息的有关证据的基本概率指派。这里, 将从多粒度角度出发, 拟提出一类新的融合方法: 基于多粒度的 D-S 证据理论的粒空间融合方法。

借鉴梁、钱等人^[21]提出的知识之间的距离, 我们给出了证据之间的距离。

定义 2 (粒度距离) 设 $MS = \{IS_i | IS_i = (U, AT_i, \{V_a\}_{a \in AT_i}, f_i)\}$ 是一个多源信息系统, 由每一个子系统 IS_i 的属性子集 AT_i 诱导得到的一个等价类集合相应构成一个粒度空间 G_i , 记 $G_i = \{G_i(x_1), G_i(x_2), \dots, G_i(x_n)\}$ 。则粒空间距离定义为:

$$d(G_i, G_j) = \frac{1}{|U|^2} \sum_{x=1}^{|U|} |G_i(x_i) \oplus G_j(x_i)|$$

式中, $|G_i(x_i) \oplus G_j(x_i)| = |G_i(x_i) \cup G_j(x_i)| - |G_i(x_i) \cap G_j(x_i)|$, $|\cdot|$ 表示集合的基数。

这里, 若把粒空间视为证据空间, 那么粒度距离就是证据空间距离。

选定一个很小的正数 δ (一般情况取 $0.5^{[22]}$), 当 $0 \leq d(G_i, G_j) \leq \delta$ 时, 表示证据 G_i 和 G_j 是依赖的; 当 $\delta < d(G_1, G_2) \leq 1$, 表示证据 G_1 和 G_2 是冲突证据。

利用 δ 和 K-Means^[23] 聚类算法, 将 k 个粒度空间进行聚类。它们满足类内的粒空间是依赖的, 相应地应用经典的证据理论融合规则进行融合。而类间的粒空间是冲突的, 则相应地应用改进的证据理论融合规则进行融合。

定义 3 (粒度距离矩阵) 设 $MS = \{IS_i | IS_i = (U, AT_i, \{V_a\}_{a \in AT_i}, f_i)\}$ 是一个多源信息系统。为了便于讨论, 假设子系统的个数为 k 个, 那么粒度距离矩阵可以定义为 $G_{k \times k} = (g_{ij})_{k \times k}$, 其中 $g_{ij} = d(G_i, G_j)$ 。

利用粒度距离判断证据间是否冲突, 并提出相应的融合规则。基于对证据理论冲突性证据合成的讨论以及对多粒度粗糙集两个极端情况的分析, 并且根据粗糙集与证据理论之间的关系, 提出一种基于证据合成的多粒度融合函数。

定义 4 (融合函数) 设 G_1, G_2, \dots, G_k 是一个多源信息系统 $MS = \{IS_i | IS_i = (U, AT_i, \{V_a\}_{a \in AT_i}, f_i)\}$ 的 k 个粒度空间, 融合函数定义如下:

$$f(G_1, G_2, \dots, G_k) = M_1^{\oplus} \oplus M_2^{\oplus} \oplus \dots \oplus M_k^{\oplus}$$

M_i^{\oplus} : 表示不冲突(互相依赖)时, 用经典的证据理论融合规则进行乐观融合。

\oplus : 表示冲突(互相不依赖)时, 用改进的证据理论融合规则进行悲观融合。

3.1 基于多粒度视角的 D-S 证据理论融合算法

对于给定的一个多源信息系统 $MS = \{IS_i | IS_i = (U, AT_i, \{V_a\}_{a \in AT_i}, f_i)\}$, 假设该系统有 k 个子系统, 并且用 $|U|$ 表示多源信息系统的对象个数, 那么基于多粒度视角的 D-S 证据理论融合算法流程如下。

算法 1 基于多粒度视角的 D-S 证据理论融合算法

输入: 多源信息系统 $MS = \{IS_i | IS_i = (U, AT_i, \{V_a\}_{a \in AT_i}, f_i)\}$ 和一个目标集合 X ;

输出: X 的必然测度和与似然测度。

Step 1 将每一个子系统通过属性集合诱导得到一个等价类集合, 可以将其看成是一个单粒空间, 根据粗糙集与证据理论之间的关系, 确定每个粒结构中每个粒的信任度指派(公式);

Step 2 通过计算粒空间之间的距离 $d(G_i, G_j) = \frac{1}{|U|^2} \sum_{i=1}^{|U|} |G_i(x_i) \oplus$

$G_j(x_i)|$, 建立粒度距离矩阵;

Step 3 按照一个很小的正数 δ , 将 k 个粒度空间分成 l 类, 类内粒空间是不冲突证据 (当 $0 \leq d \leq \delta$ 时); 类间的粒空间是冲突证据 (当 $\delta \leq d \leq 1$ 时);

Step 4 分别按照证据冲突时的合成公式

$$\begin{cases} m_S(\phi) = k_{12} = \sum_{\substack{X_i, Y_i \in \mathcal{Z}^{\oplus} \\ X_i \cap Y_i = \phi}} m_1(X) m_2(Y) \\ m_S(A) = \sum_{\substack{X_i, Y_i \in \mathcal{Z}^{\oplus} \\ X_i \cap Y_i = A}} m_1(X) m_2(Y) \end{cases}$$

和证据不冲突时的合成公式 $m(A) =$

$$\frac{A_1 \cap A_2 \cap \dots \cap A_n = A \quad m_1(A_1) \dots m_n(A_n)}{1 - K}, \text{ 其中, } K = \sum_{A_1 \cap \dots \cap A_n = \emptyset} m_1(A_1) \dots$$

$m_n(A_n)$, 建立一个多粒度组合融合函数 $f(G_1, G_2, \dots, G_K) = \frac{1}{k} M_{\oplus}^{\cap} + \frac{k-1}{k} M_{\oplus}^{\cup}$.

Step 5 输出 X 的信任测度和与似然测度。

通过对以上算法的分析, 得到该算法的时间复杂度为 $O(k \times |U|^2)$ 。

4 实例分析

下面用一个实际的多源信息系统来分析本文所提算法的有效性。

例 给定一个多源信息系统 $MS = \{IS_i | IS_i = (U, AT_i, \{V_a\}_{a \in AT_i}, f_i)\}$, 它由 5 个病人 3 次临床检查的记录构成, 其中 $U = \{e_1, e_2, e_3, e_4, e_5\}$, 如表 1 所列。

表 1 多源信息系统

病人	D ₁		D ₂		D ₃				
	头痛	肌肉痛	体温	流鼻涕	肌肉痛	体温	打喷嚏	头痛	体温
e ₁	是	是	正常	是	是	正常	是	是	正常
e ₂	是	是	高	否	是	高	是	是	高
e ₃	是	是	高	是	是	很高	是	是	正常
e ₄	否	是	正常	否	是	高	否	是	正常
e ₅	否	否	高	是	否	很高	否	否	高

(1) 根据 3 次临床检查属性分别得到 3 个粒空间, 它们是 $G_1 = \{\{e_1\}, \{e_2, e_3\}, \{e_4\}, \{e_5\}\}; G_2 = \{\{e_1\}, \{e_2, e_5\}, \{e_3\}, \{e_4\}\}; G_3 = \{\{e_1, e_3\}, \{e_2\}, \{e_4\}, \{e_5\}\}$ 。

(2) 根据粗糙集与 D-S 证据理论的关系, 分别得到 3 个粒空间对应的概率分配函数:

$$m_1(\{e_1\}) = \frac{1}{5}; m_1(\{e_2, e_3\}) = \frac{2}{5};$$

$$m_1(\{e_4\}) = \frac{1}{5}; m_1(\{e_5\}) = \frac{1}{5};$$

$$m_2(\{e_1\}) = \frac{1}{5}; m_2(\{e_2, e_5\}) = \frac{2}{5};$$

$$m_2(\{e_3\}) = \frac{1}{5}; m_2(\{e_4\}) = \frac{1}{5};$$

$$m_3(\{e_1, e_3\}) = \frac{2}{5}; m_3(\{e_2\}) = \frac{1}{5};$$

$$m_3(\{e_4\}) = \frac{1}{5}; m_3(\{e_5\}) = \frac{1}{5}。$$

对于任意的 $X = \{e_1, e_2\} \subseteq U$, 通过每一个单一系统 D_1 , D_2 和 D_3 , 分别计算得到

$$Bel_1(X) = \sum_{A \subseteq X} m_1(A) = \frac{1}{5},$$

$$Pl_1(X) = \sum_{A \cap X \neq \phi} m_1(A) = \frac{3}{5};$$

$$Bel_2(X) = \sum_{A \subseteq X} m_2(A) = \frac{1}{5},$$

$$Pl_2(X) = \sum_{A \cap X \neq \phi} m_2(A) = \frac{3}{5};$$

$$Bel_3(X) = \sum_{A \subseteq X} m_3(A) = \frac{1}{5},$$

$$Pl_3(X) = \sum_{A \cap X \neq \phi} m_3(A) = \frac{3}{5}。$$

(3) 计算粒空间的距离

$$d(G_1, G_2) = \frac{1}{|5|^2} \sum_{i=1}^{|5|} |G_1(x_i) \oplus G_2(x_i)| = \frac{4}{25};$$

$$d(G_2, G_3) = \frac{1}{|5|^2} \sum_{i=1}^{|5|} |G_2(x_i) \oplus G_3(x_i)| = \frac{6}{25};$$

$$d(G_1, G_3) = \frac{1}{|5|^2} \sum_{i=1}^{|5|} |G_1(x_i) \oplus G_3(x_i)| = \frac{6}{25}。$$

取 $\delta = 0.5$ (见文献[18]), 因为 $d(G_i, G_j) < 0.5$, 我们认为以上 3 个证据之间不存在冲突, 于是可用经典的 D-S 证据理论合成公式融合。

(4) 融合结果

由 $k_{12} = \sum_{G_{1i}(x_i) \cap G_{2i}(x_i) \neq \phi} m_1(x_i) m_2(x_i) = \frac{10}{25}$, 得 $m_{12}(e_1) = \frac{1}{5}, m_{12}(e_2) = \frac{2}{5}, m_{12}(e_3) = \frac{1}{5}, m_{12}(e_4) = \frac{1}{10}, m_{12}(e_5) = \frac{1}{10}$ 。

再由 $m_{12}(\cdot)$ 与 $m_3(\cdot)$ 融合, 得 $k_{123} = \sum_{G_{12i}(x_i) \cap G_{3i}(x_i) \neq \phi} m_{12}(x_i) m_3(x_i) = \frac{70}{25}$;

所以

$$m_{123}(e_1) = \frac{2}{7}, m_{123}(e_2) = \frac{1}{7}, m_{123}(e_3) = \frac{2}{7},$$

$$m_{123}(e_4) = \frac{1}{7}, m_{123}(e_5) = \frac{1}{7},$$

所以 $Bel_{123}(X) = \frac{3}{7}; Pl_{123}(X) = \frac{3}{7}$ 。由此可见, 基于多粒度的证据融合算法实现了不精确概念的信任度的融合, 得到了比单一系统进行不确定度量更好的效果。

结束语 D-S 证据推理是处理不确定性问题的重要方法, 但其假设空间的基本概率指派往往由专家凭经验、知识和对问题域的了解事先给定, 带有很大的主观性。本文从粗糙集理论和 D-S 证据理论的基本关系入手, 分析了多粒空间的证据的基本概率指派获取等问题。然后根据粗糙集与证据理论之间的互补性, 提出了一种基于多粒度的 D-S 证据合成的融合函数, 它是一种介于多粒度粗糙集和证据理论之间的融合方法, 更适用于多源信息系统的信息融合, 因此, 该方法将会在一定程度上增强处理多源信息系统不确定问题求解的能力。

参考文献

- [1] Waltz E, Llinas J. Multisensor Data Fusion [M]. Boston: Artech House, 1990
- [2] Yao Y Y. Information granulation and rough set approximation [J]. International Journal of Intelligent Systems, 2001(16): 87-104
- [3] Zadeh L A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic [J]. Fuzzy

- [4] Dempster A. Upper and lower probabilities induced by a multi-valued mapping [J]. *The Annals of Mathematical Statistics*, 1967(38):325-339
- [5] Shafer G. *A Mathematical Theory of Evidence* [M]. Princeton University Press, 1976
- [6] Smets P. The combination of evidence in the transferable belief models [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1990, 12(5):447-485
- [7] Smets P, Kennes R. The transferable belief model [J]. *Artificial Intelligence*, 1994, 66(2):191-234
- [8] Yager R R. On the Dempster-Shafer framework and new combination rules [J]. *Information Sciences*, 1987, 41(2):93-138
- [9] Dubois D, Prade H. On the unicity of Dempster rule of combination [J]. *International Journal of Intelligent System*, 1986(1):133-142
- [10] Smsrandache F, Dezert J. *Advances and application of DS_mT for information fusion (Vol. 1)* [M]. Rehoboth; American Research Press, 2004
- [11] Liu W R. Analyzing the degree of conflict among belief functions [J]. *Artificial Intelligence*, 2006, 5:909-924
- [12] 孙全, 叶秀清, 顾伟康. 一种新的基于证据理论的合成公式[J]. *电子学报*, 2000, 28(8):117-119
- [13] 邓勇, 王栋, 李齐, 等. 一种新的证据冲突分析法[J]. *控制理论与应用*, 2011, 28(6):839-844
- [14] 杨善林, 朱卫东, 任明仑. 基于可变参数优化的相关证据合成方法研究[J]. *管理科学学报*, 2003, 6(5):12-16
- [15] Pawlak Z. Rough sets [J]. *International Journal of computer and Information Sciences*, 1982(11):341-365
- [16] Pawlak Z. *Rough Sets; Theoretical Aspects of Reasoning about Data, System Theory, Knowledge Engineering and Problem Solving*[D]. Kluwer, Dordrecht, 1991, 9
- [17] Qian Y H, Liang J Y, Yao Y Y, et al. MGRS: A multi-granulation rough set [J]. *Information Sciences*, 2010(180):949-970
- [18] Qian Y H, Liang J Y. Rough set method based on multi-granulations[C]//*Proceedings of 5th IEEE Conference on Cognitive Informatics*. 2006(D):297-304
- [19] Yao Y Y, Lingras P J. Interpretations of belief functions in the theory of rough sets [J]. *Information Sciences*, 1998, 104(172):81-106
- [20] Wu W Z, Leung Y, Zhang W X. Connections between rough set theory and Dempster-Shafer theory of evidence [J]. *International Journal of General Systems*, 2002, 31(4):405-430
- [21] Liang J Y, Li R, Qian Y H. Distance: A more comprehensible perspective for measures in rough set theory [J]. *Knowledge-Based Systems*, 2012(27):126-136
- [22] 杨风暴, 王肖霞. *D-S证据理论的冲突证据合成方法*[M]. 北京: 国防工业出版社, 2011
- [23] MacQueen J B. Some methods for classification and analysis of multivariate observations[C]//*Proceeding of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. 1967:281-297

(上接第 32 页)

- [45] Jurgens D. *Word Sense Induction by Community Detection*[C]//*Proceedings of the TextGraphs-6 Workshop*. 2011:24-28
- [46] Klapaftis I P, Manandhar S. *Word Sense Induction & Disambiguation Using Hierarchical Random Graphs*[C]//*Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. 2010:745-755
- [47] Clauset A, Moore C, Newman M E J. Hierarchical Structure and the Prediction of Missing Links in Networks[J]. *Nature*, 2008, 453(7191):98-101
- [48] Agirre E, Soroa A. UBC-AS: A Graph Based Unsupervised System for Induction and Classification[C]//*Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*. 2007:346-349
- [49] He Zheng-yan, Song Yang, Wang Hou-feng. Applying Spectral Clustering for Chinese Word Sense Induction [M]. *The first CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP2010)*. 2010
- [50] Slonim, Friedman, Tishby. Unsupervised Document Classification Using Sequential Information Maximization[C]//*Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2002
- [51] Pantel, Patrick, Lin De-kang. Discovering word senses from text [C]//*Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2002:613-619
- [52] Kleinberg J M. Authoritative sources in a hyperlinked environment[J]. *Journal of the ACM*, 1999, 46(5):604-632
- [53] Korkontzelos I, Klapaftis I, Manandhar S. *Graph Connectivity Measures for Unsupervised Parameter Tuning of Graph-Based Sense Induction Systems*[C]//*Proceedings of the NAACL HLT Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics*. 2009:36-44
- [54] Jia Yu-xiang, Yu Shi-wen, Chen Zheng-yan. Chinese Word Sense Induction with Basic Clustering Algorithms [C] // *The first CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP2010)*. 2010
- [55] Zhang Zhen-zhong, Sun Le , Li Wen-bo. ISCAS: A System for Chinese Word Sense Induction Based on K-means Algorithm[C]//*The first CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP2010)*. 2010
- [56] Xu Hua, Liu Bing, Qian Long-hua, et al. Soochow University: Description and Analysis of the Chinese Word Sense Induction System for CLP2010[C]//*The first CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP2010)*. 2010
- [57] Wang Li-sha, Dou Yan-zhao, Sun Xiao-ling, et al. K-means and Graph-based Approaches for Chinese Word Sense Induction Task[C]//*The first CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP2010)*. 2010
- [58] Jin Peng, Sun Rui, Zhang Yi-hao. A Knowledge based Method for Chinese Word Sense Induction[C]//*Genetic and Evolutionary Computing (ICGEC)*. 2010:248-251
- [59] Jin Peng, Zhang Yi-hao, Sun Rui. LSTC System for Chinese Word Sense Induction[C]//*The first CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP2010)*. 2010