

# 基于有监督关联聚类的中文共指消解

刘未鹏<sup>1</sup> 周俊生<sup>2</sup> 黄书剑<sup>1</sup> 陈家骏<sup>1</sup>

(南京大学计算机软件新技术国家重点实验室 南京 210093)<sup>1</sup>

(南京师范大学计算机科学系 南京 210097)<sup>2</sup>

**摘要** 共指消解是文本信息处理中的一个重要问题。提出了一种有监督的关联聚类算法以实现对中文实体提及的共指消解。首先将共指消解过程看成图的关联聚类问题,从全局的角度实现对共指等价类的划分,而不是孤立地对每一对名词短语分别进行共指决策;然后给出了关联聚类的推导算法;最后设计了一种基于梯度下降的特征参数学习算法,使得训练出的特征参数能够较好拟合关联聚类的目标。在ACE中文语料上的实验结果显示,该算法优于传统的“分类-聚类”共指消解学习算法。

**关键词** 共指消解,关联聚类,损失函数

## Coreference Resolution with Supervised Correlation Clustering

LIU Wei-peng<sup>1</sup> ZHOU Jun-sheng<sup>2</sup> HUANG Shu-jian<sup>1</sup> CHEN Jia-jun<sup>1</sup>

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)<sup>1</sup>

(Department of Computer Science, Nanjing Normal University, Nanjing 210097, China)<sup>2</sup>

**Abstract** Coreference resolution plays an important role in natural language processing. A supervised correlation clustering algorithm for coreference resolution was proposed. Firstly, coreference resolution was treated as a graph correlation clustering problem, which partition the coreference relation from the global view, rather to make pairwise coreference decisions independently of each other. Then, the inference algorithms for correlation clustering were presented. Finally, a learning algorithm based on gradient descent was proposed to make the features parameters be trained from the training corpus, so that the learned parameters can better fit the objective of the correlation clustering. The experimental results on the ACE Chinese corpus demonstrate that the proposed method achieves better performance, compared with the traditional approaches.

**Keywords** Coreference resolution, Correlation clustering, Loss function

## 1 引言

共指是自然语言中一种非常普遍和常见的语言现象,只有通过共指消解,建立起各个短语之间的语义关系,才有助于对句子乃至篇章的理解。随着自然语言处理应用的日益广泛,共指消解的作用愈来愈突出,它几乎是任何一个自然语言处理的应用领域都需要解决的问题,如信息抽取、问答系统、文本摘要和对话解释系统等等。

虽然共指消解对于自然语言处理具有重要作用,但共指消解的实现却是个非常困难的问题。事实上,共指消解已经被人们广泛认为是最具挑战性的人工智能问题之一。近年来,研究者们正在尝试应用各种机器学习方法来解决共指消解的问题。Soon et al. [1], Ng and Cardie [2], Strube et al. [3], Yang et al. [4]等先后都提出了基于不同的机器学习方法的共指消解系统,但这些方法的中心思想都是将共指消解问题转换为一个二值分类任务。具体做法是:首先训练一个二值分

类器,然后应用这个分类器判断位于同一篇文档内的两个实体提及(Entity mention)是否共指;接下来再使用一个独立的聚类机制来协调前面成对的共指分类过程中所产生的冲突和矛盾,从而对一个给定的实体提及集合产生一个划分,其中的每个聚类对应于一个实体。虽然这些“标准”方法产生了一定的效果,但这些方法本身存在着一些固有的缺陷与不足。首先,由于在标准方法中共指分类和聚类是两个截然分开的过程,对共指分类器的训练和优化是完全独立于聚类过程的,因此对共指分类准确性的改进并不一定能保证聚类层在准确性上也能产生相应的改进;其次,标准方法中的共指分类器是成对(pairwise)分类器,即分类器每次均是对一对实体提及进行分类决策,这种成对分类器一般是假定各对之间是满足独立同分布(i. i. d)假设的,但实际上各对实体提及的分类决策之间可能不是相互独立的,如果能依据多个短语之间的相互依赖关系进行集体性的决策,就可以更好地避免多个成对分类器的结果不一致性问题。

到稿日期:2008-10-14 返修日期:2009-07-02 本文受国家自然科学基金项目(60673043),国家社科基金(07BY0),江苏省高校自然科学基金(07KJB520057)资助。

刘未鹏 硕士研究生,主要研究方向为自然语言处理;周俊生 博士,副教授,主要研究方向为自然语言处理、机器学习和信息抽取;黄书剑 博士生,主要研究方向为自然语言处理、机器学习;陈家骏 教授,博士生导师,主要研究方向为自然语言处理、机器翻译和软件工程。

相对于英语,对中文中的共指问题的研究明显显得不够深入和活跃。王厚峰等提出了一种基于权值计算方法的中文人称代词消解方法<sup>[5]</sup>。该方法不需要学习的过程,主要还是利用基于规则的思想解决中文人称代词的共指消解问题。庞宁等提出了基于最大熵模型的中文共指消解方法<sup>[6]</sup>。该方法类似于英文中的标准共指消解方法,使用最大熵模型作为成对分类器分别对各对短语进行单独分类,因而存在着成对分类器的固有缺陷。李元龙等提出在成对分类器的基础上使用关联聚类方法实现共指等价类的全局划分,取得了一定的效果<sup>[14]</sup>。

本文中采用图来对实体提及的共指消解问题进行建模,将共指消解看成图的关联聚类过程,从而将共指消解问题转化为图划分问题,这样使得共指消解过程并不是孤立地对每一对名词短语分别进行共指决策,而是充分考虑了多个待消解项之间的相关性,从全局的角度实现对共指等价类的划分。本文给出一种有监督的关联聚类算法以实现对中文实体提及的共指消解之方法。

## 2 关联聚类(Correlation Clustering)

关联聚类问题是指,给定一个包含  $n$  个顶点的图,对于图中的每条边  $(u, v)$ ,按照其两个顶点  $u, v$  之间是否相似或不相似,将该边标注为  $\langle + \rangle$  或  $\langle - \rangle$ 。另外每条边都关联一个非负的权值  $C_e$ ,表示该边的两个顶点之间的相似性(或不相似性)的一种信任测度。权值越大表示信任度也越大。关联聚类问题的目标则是对于这样一个带权图,生成一个最佳划分,使得聚类的一致性权值达到最大或者不一致性权值达到最小<sup>[7]</sup>。具体地说,一致性权值是指各个聚类内部的  $\langle + \rangle$  边权值与各个聚类之间的  $\langle - \rangle$  边权值之和。而不一致性权值则是指各个聚类内部的  $\langle - \rangle$  边权值与各个聚类之间的  $\langle + \rangle$  边权值之和。

对图中每一个顶点对  $(u, v)$  引入一个相应的 0-1 变量  $\theta_{uv}$  (当  $(u, v)$  是图中的一条边  $e$  时,  $\theta_{uv}$  也可表示为  $\theta_e$ ), 对于一个给定的图划分,当顶点  $u$  与  $v$  位于同一个聚类时,  $\theta_{uv} = 1$ , 否则  $\theta_{uv} = 0$ 。这样,若采用最大化一致性权值作为聚类的目标,关联聚类的推导问题可以表示为如下的整数规划形式<sup>[8]</sup>。

$$\begin{aligned} \max \quad & \sum_{w: \theta_{uw} > 0} \theta_{uw} w_w - \sum_{w: \theta_{uw} < 0} (1 - \theta_{uw}) w_w \\ \text{s. t.} \quad & \theta_{uv} \in \{0, 1\} \\ & \theta_{uv} + \theta_{vw} \leq \theta_{uw} + 1, \forall u, v, w; \\ & \theta_{uv} = 1, \forall u \end{aligned}$$

关联聚类的目标是求出一个  $\theta_{uv}$  的有效赋值,使得目标函数数值最大。如果  $\theta_{uv} \in \{0, 1\}$ , 且满足三角不等式,则该赋值是有效的。

## 3 基于关联聚类的有监督共指消解

### 3.1 建模

将文本中出现的各个实体提及看成是图中的各个顶点,则实体之间的共指消解过程可看成是基于图的关联聚类过程。使用  $\text{Sim}_e$  表示边  $e$  的相似度:

$$\text{Sim}_e = w^T F(x_e)$$

其中,  $F(x_e) \in R^n$  是边  $e$  的特征向量表示,即表示边  $e$  的两个顶点所构成的实体提及对的特征向量,  $w$  为相应的特征权值向量。这样,对应于共指消解过程的关联聚类目标函数可以表达如下:

$$w^T F(x, y) = \sum_{e: \text{Sim}_e > 0} \theta_e w^T F(x_e) - \sum_{e: \text{Sim}_e < 0} (1 - \theta_e) w^T F(x_e)$$

即使不相似的实体对(即权值为负的实体对)也可以位于同一个聚类中,只要包含它们的聚类的整体相似度值为正。类似地,相似的实体对(即权值为正的实体对)可能不在同一个聚类中,如果包含它们的聚类的整体相似度值为负。若定义:

$$\begin{aligned} s^* &= \max_{y \in Y} \sum \theta_e w^T F(x_e) \\ s^- &= \sum_{e: \text{Sim}_e < 0} w^T F(x_e) \\ s^+ &= \sum_{e: \text{Sim}_e > 0} w^T F(x_e) \end{aligned}$$

则最大化目标函数的值可表示为:  $s^* - s^-$ 。

这样,对于一个实体集合的共指消解过程就是使聚类中的相似度值最大化的求解过程:

$$\begin{aligned} \operatorname{argmax}_y w^T F(x, y) &= \operatorname{argmax}_y \sum_{e: \text{Sim}_e > 0} \theta_e w^T F(x_e) - \sum_{e: \text{Sim}_e < 0} \\ & (1 - \theta_e) w^T F(x_e) = \operatorname{argmax}_y s^* - s^- \end{aligned}$$

### 3.2 损失函数的设计

给定一个实体提及集合  $M$ , 及其正确的共指关系划分  $\text{Key}(M)$ , 若使用关联聚类算法对集合  $M$  进行共指关系划分, 得到一个划分结果  $P(M)$ 。如何比较  $P(M)$  与  $\text{Key}(M)$  这两个聚类之间的差异即损失值并不是一个直接的问题, 但可以将两者之间的损失值定义问题转化为求所有提及对在  $P(M)$  与  $\text{Key}(M)$  之间的不一致性问题, 即将损失函数定义为集合  $M$  中的任何两个实体提及的成对错误的和。通过分析可知, 在  $P(M)$  与  $\text{Key}(M)$  之间的实体提及的成对错误可分为两类: 错误划分与错误归并。其中错误划分是指若集合  $M$  中的两个提及  $a$  和  $b$ , 在  $\text{Key}(M)$  中属于同一个聚类, 而在  $P(M)$  中分属于两个不同的聚类; 反之, 错误归并则是指  $a$  和  $b$  在  $\text{Key}(M)$  中属于不同的聚类, 而在  $P(M)$  中被划在同一个聚类中。根据关联聚类的目标函数可以推出: 在应用关联聚类方法进行共指关系划分时引起的错误归并的原因可以归结为是由于当前的相关特征权值过大而引起; 而错误划分的原因则可以归结为由于当前的相关特征权值过小而引起。因而为了修正这两类错误, 需要对它们的特征权值进行相应的调整。这样可以定义  $P(M)$  与  $\text{Key}(M)$  之间的损失函数如下:

$$L_M(\text{Key}, P) = \frac{1}{|M|^2} \sum_{m_a, m_b \in M} (\text{sim}_{e-m}(m_a, m_b) - \text{sim}_{e-d}(m_a, m_b))$$

其中,  $\text{sim}_{e-d}(m_a, m_b)$  表示错误划分的实体提及对的相似度值,  $\text{sim}_{e-m}(m_a, m_b)$  表示错误归并的实体提及对的相似度值。  $\text{sim}_{e-d}(m_a, m_b)$  与  $\text{sim}_{e-m}(m_a, m_b)$  均可以表示为特征权值向量  $w$  的线性积。

### 3.3 特征参数的训练算法

在给出一组训练数据的条件下, 训练过程就是逐步减小损失函数值的过程, 训练的目标就是学习一组最佳的特征权值向量, 使得基于该特征向量值进行关联聚类, 划分产生的损失函数值最小。

$$w^* = \operatorname{argmin}_w L(\text{Key}, P)$$

设计了一种基于梯度下降的特征参数训练算法, 训练过程的收敛性则依赖于损失函数的凸性。由于所定义的  $P(M)$  与  $\text{Key}(M)$  之间的损失函数  $L_M(\text{Key}, P)$  实际上是一个特征权值向量的线性函数, 因而训练过程在理论上可以保证收敛到一个全局极小值。

算法的输入是标注文档的实体提及集合  $M_i$  和其相应的

共指关系划分  $Key(M_i)$  ( $i=1, \dots, N$ ), 其中的集合  $M_i$  表示第  $i$  篇文档中的所有实体提及的集合。算法的输出为学习的特征权值。详细算法描述如下。

1. 初始化权值, 将所有特征参数的初值设置为 0;

2. 在  $t$  上进行迭代 ( $t=1, 2, \dots, T$ ):

2.1 对于每一个实体提及集合  $M_i$  和其相应的共指关系划分  $P(M_i)$ , 分别进行下列处理:

① 基于当前的特征参数值使用关联聚类算法对集合  $M_i$  进行划分, 得到一划分结果  $P(M_i)$ ;

② 依据损失函数, 计算  $P(M_i)$  与  $Key(M_i)$  之间的损失值;

③ 按照梯度值修改当前的特征参数向量值:

$$w^t = w^{t-1} - \eta \cdot \frac{\partial L_{M_i}(key, P^{t-1})}{\partial w^{t-1}}$$

④ 对特征参数进行归一化处理,

$$w^t = \frac{w^t}{Z}, Z = \|w^t\|$$

3. 当迭代次数  $t > T$  时, 算法结束。

算法中的迭代参数  $T$  和学习率  $\eta > 0$  为可调节参数。算法中的学习率  $\eta$  不可设置过小, 否则会使学习的过程变得非常缓慢; 但也不可设置过大, 否则可能会引起训练过程中出现震荡现象。受文献[9]的启发, 引入了参数平均化的技术来避免这种问题。如果用  $w_i^t$  表示特征  $f_i$  在经过  $t$  次迭代之后, 再输入第  $i$  个训练样本之后的值, 则采用参数平均化的方式计算特征  $f_i$  的权值  $w_i$  为  $w_i = \sum_{t=1, \dots, T, i=1, \dots, n} w_i^t / nT$ 。这种平均方式可以使用最后一次输出作为最终的模型参数结果。

#### 3.4 关联聚类的近似推导

关联聚类的推导问题实际上可归结为一个整数规划问题, 但这个整数规划问题是 NP-hard 问题。文献[8]中给出了一种 LP 近似推导算法, 但对于我们的共指消解问题所产生的 LP 规划的约束集合过于庞大, 会导致内存需求问题难以解决, 从而实际计算不可行。文献[10]中则引入一种 LP 组块技术来解决在求解这种大规模 LP 规划时所面临的内存需求问题。这种 LP 近似推导算法是一种整体求优的思想, 虽然进一步通过引入 LP 组块技术使得算法可以实际运行, 但算法的计算开销仍然巨大, 因而本文引入一种贪心的近似推导算法。虽然贪心算法不一定能产生整体最优解, 但其最终结果却是最优解的很好近似。

##### 3.4.1 贪心近似推导方法

关联聚类的推导也可以采用贪心方法进行近似求解。初始时, 每一个实体提及分别属于一个独立的聚类, 然后重复地查找两个能使目标函数值增长最大的聚类  $y_i, y_j \in y$ , 将它们进行合并, 合并过程不断重复, 直到不能发现新的两个聚类的合并能够增长目标函数值为止。通过这种基于贪心法的聚类合并过程最终可以给出一个小于实际目标函数极值的近似聚类结果。

## 4 实验

### 4.1 实验数据与评测指标

使用 2005 年 ACE 评测的训练数据作为实验数据。ACE 评测是指由美国国家标准技术研究院(NIST) 组织的自动内容抽取(Automatic Content Extraction) 评测, 其研究的主要内容是自动抽取新闻语料中出现的实体、关系、事件等内容。2005 年 ACE 评测提供的语料包括英文、中文和阿拉伯文。语料数据的来源为广播新闻和新华社新闻。ACE 评测的训

练数据不但标注了实体、实体的各种属性、实体关系以及关系的属性, 还标注了对应于同一个实体的各个名词短语之间的共指关系, ACE 语料是目前唯一可公开利用的中文共指标注语料。

由于 ACE 语料提供的是未经切分的中文语料, 在共指消解实验之前, 需要进行一些预处理。首先使用一个基于条件随机场的中文分词工具进行分词, 接着进行了词性标注和基本名词短语识别。由于 ACE 语料标注的待消解项中出现了大量的非基本名词短语, 如待消解项“一位控诉他从非法赌博业者收取贿赂的过去朋友”根本就无法被基本名词短语识别器正确识别, 因此, 为了验证共指消解算法本身的有效性, 根据 ACE 训练语料的标注对名词短语进行了调整, 从而使得每个名词短语都是真正的待消解项。

为了评测实验结果, 本文采用 MUC-6 中所定义的召回率(Recall)、准确率(Precision) 和  $F$  值(其中计算  $F$  值中的  $\beta$  参数, 取为 1) 作为评测指标<sup>[11]</sup>, 具体计算方法如下:

$$\text{准确率 } P = \frac{\text{正确识别出的共指对数}}{\text{识别出的共指对数}} \times 100\%$$

$$\text{回率 } R = \frac{\text{正确识别出的共指对数}}{\text{实际共指对数}} \times 100\%$$

$$F = \frac{R \times P \times 2}{R + P}$$

从 ACE 中文语料中随机选择 229 篇文档组成了训练语料, 69 篇文档组成了测试语料。

### 4.2 特征选择

为了实现共指消解, 首先必须要选择和设计一个合适的特征集, 使其能有效地描述两个实体提及之间的共指信息。而且, 这些特征必须要有足够的通用性, 可以用于不同的领域之间, 还要能够根据不同类型的实体提及给出不同的共指决策。许多重要的信息源有助于对共指关系的刻画, 如两个实体提及之间的距离特征、性别信息、单复数信息和语义类别信息等。

参考了文献[1]中定义的特征集合, 定义和选择了 10 个适合于中文的特征构成特征集合, 这 10 个特征不仅满足了上述的一些要求, 而且相对容易从中文中获取。每个特征都是针对一对实体提及  $i$  和  $j$ , 其中提及  $i$  是位于提及  $j$  之前的先行语。具体的特征集合描述如下:

(1) 距离特征: 它可能的取值有 0, 1, 2, ..., 距离特征就是获取  $i$  和  $j$  之间的距离, 如果  $i$  和  $j$  在同一个句子中, 那么特征值取 0; 如果它们所处的位置之间相差一个句子, 那么特征值取 1, 以此类推。

(2)  $i$ -代词特征: 它的取值类型为布尔型。如果  $i$  是一个代词, 就返回 true; 否则就返回 false。代词包括反身代词(自己)和人称代词(我、你、他)。

(3)  $j$ -代词特征: 它的取值类型也是布尔型。如果  $j$  是一个代词(如上所述), 那么返回 true; 否则返回 false。

(4) 字符串匹配特征: 它的取值类型是布尔型。由于中文中的实体提及包含大量的修饰语, 因此不能直接对  $i$  和  $j$  的字符串内容进行直接比对, 不过每个实体提及都有中心词, 仅对实体提及  $i$  和  $j$  中心词的字符串进行匹配, 如果两者匹配, 就返回 true; 否则返回 false。

(5) 指示性的实体提及特征: 它的取值类型是布尔型。以指示词“这”、“那”、“这些”或“那些”引导的实体提及就是指

示性的实体提及。如果实体提及  $j$  是一个指示性的实体提及,那么返回 true,否则返回 false。

(6) 单复数一致性特征:它的取值类型是布尔型。如果实体提及  $i$  和  $j$  的单复数信息一致就返回 true;否则返回 false。

(7) 语义类别一致性特征:它可能的取值有 true, false 或 unknown 3 种。如果实体提及  $i$  和  $j$  的语义类别一致,那么特征值就取 true;如果不一致,特征值就取 false;如果其中任何一个的语义类别为“unknown”,那么就比较这两个实体提及的中心词,如果中心词相等,特征值取 true;否则特征值取 unknown。

(8) 性别一致性特征:它可能的取值有 true, false 或 unknown 3 种。如果实体提及  $i$  和  $j$  的性别信息一致,那么特征值取 true;若两者不一致,则特征值取 false。如果实体提及  $i$  和  $j$  其中有一个的性别信息为 unknown,那么特征值也就取 unknown;

(9) 专有名词特征:它的取值类型是布尔型。一般来说,人名、地名和机构名这样的命名实体类型是专有名词。如果  $i$  和  $j$  都是专有名词,那么返回 true;否则返回 false。

(10) 同位语特征:它的取值是布尔型。如果实体提及  $i$  和  $j$  之间具有同位语的关系,则返回 true;否则返回 false。

#### 4.3 Baseline 系统的设计

用类似文献[1,2]中提出的共指消解算法所实现的系统作为本文研究的 Baseline 系统,所采用的分类算法是 C4.5 决策树算法<sup>[12]</sup>和最大熵分类器<sup>[13]</sup>。文献[1]中的系统用来协调分类矛盾的聚类机制是最近优先策略(link-first),最近优先策略选择离指示语最邻近的置信度大于 0.5 的候选先行语作为最终先行语。文献[2]中又提出了对上述系统的改进,它用的聚类机制是最佳优先策略(link-best),最佳优先策略则是从置信度大于 0.5 的候选先行语中选择置信度最大的作为最终先行语。表 1 给出了使用 C4.5 决策树作为共指关系分类器在两种聚类机制下的消解结果,表 2 给出了使用最大熵分类器并分别应用两种聚类机制的消解结果。

表 1 基于 C4.5 决策树分类器的共指消解结果

Approaches	Precision (%)	Recall (%)	F (%)
link-first	75.66	77.55	76.60
link-best	75.99	77.88	76.92

表 2 基于最大熵分类器的共指消解结果

Approaches	Precision (%)	Recall (%)	F (%)
link-first	76.81	75.14	75.97
link-best	76.97	75.31	76.13

#### 4.4 实验结果比较

采用与 Baseline 系统同样的训练语料与测试语料,对本文所提出的有监督关联聚类算法进行了共指消解实验。表 3 给出了对测试语料的消解结果。表中的第一行(noavg)表示训练过程中未引入参数平均化技术的实验结果,第二行(avg)则表示引入了参数平均化技术的实验结果。从表 3 中的实验数据可以看出,采用本文所提出的算法进行的共指消解结果明显优于表 1 和表 2 所给出的所有 Baseline 系统的结果。其中,采用参数平均化技术进行特征权值训练的结果略优于直接训练的结果,比 Baseline 系统中的最好结果也高出 1.09%。出于计算时间的原因,在实验中采用了贪心算法进行迭代过程中的关联聚类推导,下一步将采用基于整体最优思想的

LP 近似推导算法求解关联聚类问题。

表 3 基于有监督关联聚类的共指消解结果( $T=170, \eta=0.01$ )

Approaches	Precision (%)	Recall (%)	F (%)
noavg	75.97	79.63	77.76
avg	76.06	80.05	78.01

**结束语** 共指消解用于建立一篇文档内的各个短语之间的语义关系,它是文本信息处理中不可缺少的内容。针对传统的用于共指消解的“分类-聚类”学习机制的不足,本文提出一种有监督的关联聚类算法以实现对中国实体提及的共指消解,将共指消解过程看成图的关联聚类,给出了关联聚类的推导算法;然后设计了一种有效的基于梯度下降的特征参数学习算法,实现从训练语料中自动学习各个特征的权值。通过 ACE 中文语料上的实验结果显示,该算法优于传统的“分类-聚类”共指消解学习算法。

在实验过程中,发现各种属性特征识别的准确性会对共指消解的结果产生较大的影响。在目前的系统中,性别、单复数、语义类别等类型特征识别的准确性还有待提高。今后的工作中,需要选择更有效的属性特征和设计更有效的方法实现对各种属性特征的更准确的识别,以进一步提高共指消解的效果。

#### 参考文献

- [1] Soon W M, Ng H T, Lim D. A machine learning approach to coreference resolution of noun phrases[J]. Computational Linguistics, 2001, 27(4): 521-544
- [2] Ng V, Cardie C. Improving machine learning approaches to coreference resolution[C]//Proc. of the ACL. 2002: 104-111
- [3] Strube M, Rapp S, Muller C. The influence of minimum edit distance on reference resolution[C]//Proc. of EMNLP. 2002: 312-319
- [4] Yang X, Zhou G D, Su J, et al. Coreference resolution using competitive learning approach[C]//Proc. of the ACL. 2003: 176-183
- [5] 王厚峰,梅铮. 鲁棒性的中文人称代词消解[J]. 软件学报, 2005, 16(5)
- [6] 庞宁,杨尔弘. 基于最大熵模型的共指消解研究[J]. 中文信息学报, 2008, 22(2)
- [7] Bansal N, Blum A, Chawla S. Correlation Clustering[J]. Machine Learning, 2003, 56(1): 89-113
- [8] Demaine E, Immerlica N. Correlation clustering with partial information[C]//Proc. of 6th APPROX. 2003: 1-13
- [9] Michael C. Discriminative training methods for hidden Markov model; Theory and experiments with the perceptron algorithm [C]//EMNLP2002. Philadelphia, PA, USA, 2002
- [10] Bradley P S, Mangasarian O L. Massive data discrimination via linear support vector machines[J]. Optimization Methods and Software, 2000, 13(1): 1-10
- [11] Vilain M, Aberdeen J, et al. A model theoretic coreference scoring scheme[C]//Proc. of the 6th Message Understanding Conf (MUC6). San Francisco: Morgan Kaufmann Publishers, 1995: 45-52
- [12] Quinlan J R. C4.5: Programs for machine learning[M]. Morgan Kaufmann, 1993
- [13] Berger A L, Pietra S A D, Pietra V J D. A maximum entropy approach to natural language processing[J]. Computational Linguistics, 1996, 22(1): 39-72
- [14] 李元龙,周俊生,陈家骏. 一种基于关联聚类的汉语共指消解方法[J]. 计算机科学, 2007, 34(12)