

文章编号: 1003-0077(2015)01-0001-12

共指消解研究方法综述

宋 洋, 王厚峰

(北京大学 计算语言学教育部重点实验室, 北京 100871)

摘 要: 共指消解作为自然语言处理中的一个重要问题一直受到学术界的重视。二十多年来, 基于规则的和基于统计的不同方法被提出, 在一定程度上推进了该问题研究的发展, 并取得了大量研究成果。该文首先介绍了共指消解问题的基本概念, 并采用形式化的方法对该问题做了描述; 然后, 针对国内外近年来在共指消解研究中的方法进行了总结; 之后, 对共指消解中重要的特征问题进行了分析与讨论; 最后, 历数了共指消解的各种国际评测, 并对未来可能的研究方向进行了展望。

关键词: 共指消解; 指代消解; 有指导学习; 无指导学习

中图分类号: TP391

文献标识码: A

A Survey of Coreference Resolution Research Methods

SONG Yang, WANG Houfeng

(Key Laboratory of Computational Linguistics (Ministry of Education), Peking University, Beijing 100871, China)

Abstract: Coreference resolution, as a challenging issue, has been noted by NLP researchers for a long time. In recent twenty years, many kinds of advanced NLP techniques have been applied on this problem, and some of them have achieved significant improvements. In this paper, we first introduce some basic concepts and formalized this issue. Then we summarize different research strategies adopted by researchers in recent decades. We highlight the feature engineering, which lies in the core of coreference resolution. Finally we describe the recent evaluations for this task and discuss some key issues and prospects in the future.

Key words: coreference resolution; anaphora resolution; supervised learning; unsupervised learning

1 引言

共指消解^[1-2]是自然语言处理的重要问题, 它的有效解决对于机器翻译、信息抽取、关系抽取、自动文摘以及问答系统等应用都有重要意义。在机器翻译中, 有效的共指消解尤其是代词消解将帮助机器翻译提高准确率。例如, 对于英文中的代词“they”, 在英汉机器翻译系统中根据其实际所指向的不同实体的性别, 可以翻译成为“他们”、“她们”或“它们”。而对于信息抽取任务, 在识别出共指的名词短语之后, 将有利于挖掘出更多有用的实体信息与属性信息, 关系抽取亦是如此。对于自动文摘以及问答系统, 共指消解也能够语言生成时起到重要地作用。

例如, 对于采用相同名词短语作为主语连续若干个句子, 如果将除首句以外的其他句子的主语选择性地用代词来进行替换, 则既能保持文字的通顺和连贯, 又减少了一些不必要的表述, 使文字的适读性得到增强。

1.1 问题定义

共指消解有很多不同定义, 但这些定义的实质都是相同的。这里沿用国际著名评测 ACE (Automatic Content Extraction, 自动内容抽取) 的定义: 所谓共指消解即为文本中的表述 (Mention, 或称为指称语) 确定其在真实世界中所指向的实体 (Entity) 的过程。这里, 实体是一个抽象的概念, 而其在文本中的具体体现则是各种不同或相同的表述。表

收稿日期: 2013-03-09 定稿日期: 2013-05-15

基金项目: 国家自然科学基金(61370117, 61333018); 国家社科重大项目(12&ZD227)

述主要分为 3 类: 普通名词短语、专有名词和代词(本文不讨论事件共指问题, 也即不把动词看作表述)。从语言学的观点看, 表述还可以有很多其他更细致的分类, 例如, 缩略语、别名以及同位语等, 甚至还有中文中经常出现的省略语指代(即零形指代)。考虑如下几个例子。

例 1 [张三]对人很热情, 大家都叫[他][张哥]。[张哥]是一名医生, [他]工作非常认真负责, 同时, [他]也是[一个好父亲]。

例 2 [美利坚合众国总统][巴拉克·奥巴马]将于 11 月 15 日至 18 日对中国进行国事访问。

例 3 两会闭幕后, 今年下半年将召开[中国共产党第十八届全国代表大会]。目前, [十八大]的各项准备工作正有条不紊地进行。

这 3 个例子中包含了前面提到的各种表述实例, 比如例 1 中, [他]表示代词, [张三]表示专有名词, [张哥]可以看作是别名, [一个好父亲]是普通名词短语; 例 2 中, [美利坚合众国总统]与[巴拉克·奥巴马]构成同位语; 例 3 中, [十八大]是[中国共产党第十八届全国代表大会]的缩略语。

共指消解问题的一个相关问题是指代消解。指代消解从定义上来说, 是指为文本中出现的回指成分(Anaphora)确定其先行语(Antecedent)的过程。从广义地角度讲, 指代消解与共指消解这两个问题可以看成是同一个问题。

1.2 形式化表示

假设文本中全部指称语(即候选表述)构成集合 $A = \{m_1, m_2, \dots, m_N\}$, 则共指关系实际上是集合 A 上的一个等价关系。等价关系的定义如下:

设 A 不是空集, R 是 $A \times A$ 的子集, 若 R 是自反的、对称的和传递的, 则称 R 为 A 上的等价关系。

文本中名词短语间的共指关系在没有其他约束的情况下, 可以看作是等价关系。非空集合 A 上的等价关系 R , 决定了 A 的一个划分。同样地, 集合 A 上的一个划分确定了 A 的元素间的一个等价关系。

共指消解问题可以看作是集合 A 的划分问题。假设集合 A 包含有 N 个元素(即 N 个候选表述), 那么集合 A 的划分问题实际上就是 N 个元素的划分问题。对于 N 个元素来说, 如果划分的等价类数目为 m , 则全部的划分数 $S(n, m) = S(n-1, m-1) + m * S(n-1, m)$, 此亦被称为第二类 Stirling 数, 如式(1)所示。

$$S(n, m) = \frac{1}{m!} \sum_{i=0}^m (-1)^{m-i} C_m^i i^n \quad (1)$$

同时, m 的取值范围可以从 1 取到 N , 因此 N 个元素的全部可能的划分数目是随着 N 的增长而呈指数增长的, 如式(2)所示。

$$N = \sum_{m=1}^n \left(\frac{1}{m!} \sum_{i=0}^m (-1)^{m-i} C_m^i i^n \right) \quad (2)$$

现实中, 针对共指消解问题, 很少有研究者从搜索最优的等价类划分的角度来考虑, 因为这样做会使得问题的复杂度太高。因此, 无论是提出分类模型还是聚类模型解决共指问题时, 大多都基于问题本身提炼了一系列的限制条件, 以降低问题的复杂度, 使得共指问题在一个合理的时间范围内可以求解。

2 共指消解研究方法

2.1 基于规则的方法

2.1.1 基于句法结构的方法

Hobbs^[3]提出的 Hobbs 算法是最早的代词消解算法之一。该算法主要基于句法分析树进行相关搜索, 它包含两种算法: 一种方法是完全基于句法知识的, 也称朴素 Hobbs 算法; 另一种既考虑句法知识又考虑语义知识。该算法不仅是一个具体的算法, 同时更是一个理论模型, 具体的算法流程和示例可以参见相关文献^[2-3]。

Haghighi 和 Klein^[4]采用了丰富的句法语义知识作为规则, 在 MUC 和 ACE 数据集上的实验结果显示该方法超过了大部分无指导学习方法, 并达到了有指导学习方法同样的效果。

2.1.2 基于语篇结构的方法

中心理论(Center Theory)^[5]是由 Grosz 等人提出的一种局部篇章连贯性理论。该理论主要关注篇章结构中的焦点转移、表述形式选择以及话语一致性等问题。它的主要目标是跟踪句子中实体的焦点变化。由于共指消解中所研究的代词消解问题往往就是寻找代词所指向的某个焦点实体, 因此, 中心理论一经提出便常用于代词消解研究。

Brennan 等^[6]提出了一种基于中心理论的代词消解算法, 即 BFP 算法(该算法以其 3 位作者 Brennan、Friedman 和 Pollard 姓氏的首字母来命名), 它能够用来寻找给定句子中代词所指向的先行语, 具体流程为: 顺序遍历所有可能的候选先行语, 选择能够同时满足词汇句法(Morphosyntactic)、约束

(Binding)和类型标准(Sortal criteria)的那个表述作为先行语。

Poesio 等^[7]采用一种参数化的方法来实际检验中心理论,其算法表明,在进行代词消解优选性考察时,回指中心的唯一性约束是极其重要的。他们同时认为,由于原始的中心理论对于什么是话语(Utterance)以及如何计算实体排序等问题均没有明确回答,需要根据具体的语言环境而有所设定。

2.1.3 基于突显性计算的方法

Lappin 和 Leass^[8]提出 RAP(Resolution of Anaphora Procedure)算法,该算法采用句法信息来识别第三人称代词和具有反身特征与共指特征的先行语,其先通过槽文法(Slot Grammar)进行句子分析,再使用句法知识计算候选先行语的突显性,最后选择突显性打分最高的作为先行语。

Kennedy 和 Boguraev^[9]对 RAP 算法进行了简化。他们首先通过浅层分析得到必要的文法信息(主要是词性标注结果和句法功能标识);再根据所得到的各种句法信息对各候选先行语计算突显性(采用加权求和的方式,为不同句法特征赋予不同的权值);然后通过突显性得分来确定先行语。

1998 年 Mitkov^[10]提出了一种“有限知识”的指代消解方法。该方法只需要利用语法信息(获得句子的词性标注结果),结合一些指示性特征(如是否为限定性名词短语或名词短语是否重复出现等)来计算候选先行语的突显性,再经过一些诸如性别、单复数等一致性检验后,选取最佳候选作为先行语。

近年来,斯坦福大学自然语言处理研究小组的 Raghunathan 等人^[11]提出了一种基于多遍过滤的共指消解方法,他们基于准确率由高到低构建了多个筛子(Sieve)用于为文本中出现的不同表述选取可能的先行语,同时每层过滤均基于之前得到的共指结果来进行,该方法保证了准确率较高的特征能够在消解过程中被强化,并且聚类结果中不同表述的属性信息可以在不同的筛子中得到共享。尽管方法很简单,但是他们的系统在 CoNLL-2011 的英文共指消解评测中获得了最佳的结果^[12],甚至超出了很多基于统计机器学习方法的系统。

2.2 基于学习的方法

基于机器学习的共指消解方法的核心在于距离准则的学习(Distance Metric Learning),这里所指的“距离”,既可以是表述对之间的距离,也可以是实体与表述之间的距离,不同的定义方式与相应的消

解粒度以及解决问题的框架有关。更详细的又分为有指导的学习模型与无指导的学习模型,其差别就在于学习“距离”准则时是否有训练数据的参与。另外,一些半指导(或者称为弱指导)的学习模型(例如,自训练、互训练)也曾被应用于共指消解研究^[13],但本文并不特别讨论这部分内容。

2.2.1 有指导的学习模型

有指导的共指消解学习模型,可以依照不同的范式进行划分。这里以消解粒度及消解框架来进行划分,包括表述对模型、实体—表述模型、表述排序模型以及实体排序模型,如图 1 所示。当然,不同的消解粒度对无指导的学习模型和基于规则的共指消解方法都是适用的。为了清晰地描述有指导的共指消解问题的解决框架,本文以表述对模型为例进行详细介绍,其余后续模型除了消解粒度不同外,在很多方面与表述对模型都参考了同样的处理策略。

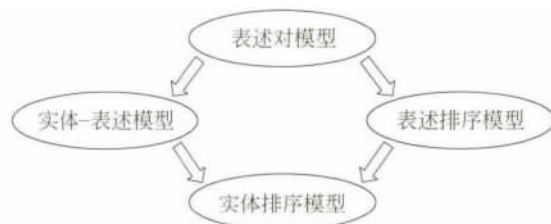


图 1 有指导的共指消解学习模型分类

表述对模型: 最常见的共指消解框架即表述对模型,该模型将共指消解问题看成是表述对之间的二元分类问题。表述即实体在文本中的具体表现形式,而共指消解则是确定文本中的不同表述分别指向哪些实体的过程。最朴素的解决方案就是判断任意两个表述共指或不共指,然后基于所有的表述对二元分类结果生成最终的共指聚类结果。表述对模型的共指消解框架如图 2 所示,具体消解流程如下。

- 表述检测: 共指消解的第一步是表述检测(也被称为指称语检测,英文即 Mention Detection 或 Anaphoricity Determination),即识别出文本中可能产生共指的全部候选表述(候选表述来自于全部的名词短语集合的某个子集)。一般的方法包括基于规则的方法^[12,14]、基于学习的方法^[15-16]以及规则和统计相结合的方法^[17]。至于究竟采取何种办法,与选择的语料库有关,有些共指数据集(如 ACE 数据)标注了仅含有单一表述的实体(Singleton Entity),则采用基于学习的方法进行表述检测可能会取得较好的效果并有助于后续共指消解任务结果的提升;另一些数据集(如 MUC 和 OntoNotes)则未

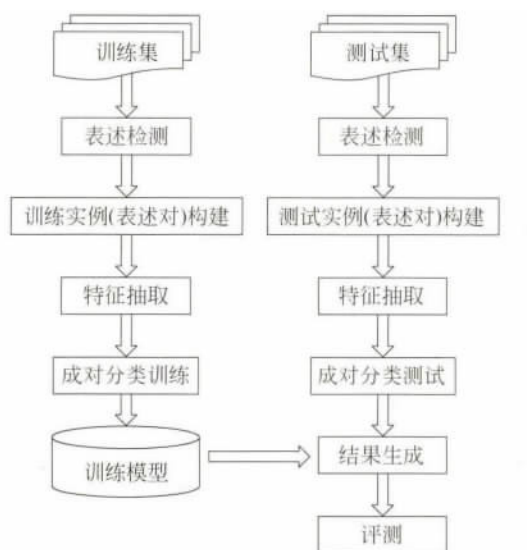


图2 表述对共指消解模型框架

标注单一表述实体,因此采用基于规则的方法在尽量提高表述识别召回率的情况下,方能够对最终的共指消解任务的实验效果起到积极的作用。

• 训练(或测试)实例(表述对)构建:即从训练语料(或测试语料)中构建用于分类器的输入实例。在训练语料上构建二元分类的训练实例时,需要考虑如何构建正例和负例。最朴素的表述对实例构建方式来自于 McCarthy 和 Lehnert^[18],该方法将文本中任何两个不在同一实体(或被称为共指链)中的表述构成负例,任何两个位于同一实体中的表述构成正例。由于这种方法产生的训练实例数量巨大,而且负例数量远远大于正例数量,会严重影响机器学习算法的效率同时产生严重的不平衡问题,因此后来很少被采用。一些研究人员采用一些特殊的用于实例构建的规则来降低正负例的不平衡,使模型训练更趋合理。Soon 等^[19]采用的方法是将文本中的每一个表述 m_i 与其前面最近的一个共指的表述 $m_j (j < i)$ 构成正例,其间的每个表述都和 m_i 构成负例。Bengtson 和 Roth^[20]采用 Soon 同样的方式构建正例,而构建负例时则是将表述 m_i 与其前面一直到文本开头为止都不共指的全部表述均作为负例。Ng 和 Cardie^[21]采用 Soon 同样的方式构建负例,而构建正例时的特殊之处在于,当表述 m_i 是代词时,需要寻找其前不是代词且共指的表述 m_j 与其构成正例。Gasperi^[22]则是采用主动学习的方法对表述对进行选择,来解决正负例分布不平衡的情况。测试阶段的实例构建,一般是针对文本中的每个表述 m_i 与其前面一直到文本开头的表述 m_j 都构建表述对。一般地,为了在学习中有效降低问题的

复杂性,也可以考虑设定表述对的最大句子距离间距,或对表述对个数、正负例的比例进行限制。

• 特征抽取:详细的共指消解特征的介绍参见第3节。

• 成对分类:即表述对的二元分类模型。当前用于表述对二元分类的机器学习算法主要有朴素贝叶斯 (Naïve Bayes)^[23]、决策树 (Decision Tree)^[18-19,21]、最大熵 (Maximum Entropy)^[24-25] 以及支持向量机 (Support Vector Machine)^[26-27] 等。

• 结果生成(也被称为表述聚类):在获得了所有表述对的二元分类结果后(通常来说该二元分类结果不一定必须是 0-1 取值,也可以是取 1 时的共指概率结果),需要对其进行进一步处理以获取实体聚类结果。常见的聚类结果生成方法包括:最近最先 (Closest-First),即在所有满足共指条件(如共指概率大于某给定阈值)的候选先行语中选择与当前表述最近的那个;最优最先 (Best-First),即在所有候选先行语当中选择与当前表述相互共指的可能性最大的那个(如共指概率大于某给定阈值里最大的那个);以及传递性约束(3 个表述中的任意两对共指,则第 3 对也共指)。在表述对模型的共指消解框架下强化传递性约束的具体方法包括关联聚类^[28-30]、图划分^[31-32] 以及谱聚类^[33] 等。

• 评测:共指消解的各种评价方法可参考相关论文^[34-36]。

实体—表述模型:与表述对模型不同,实体—表述模型^[26,37] 能够整合已经形成的实体(或共指链,即当前已有的共指聚类结果)。该模型训练一个分类器,用以判断当前表述 m_i 是否与其前面已经形成的某个实体 c_k 共指。在训练实例构建上,该模型与表述对模型有本质的不同,其实例构建方式为将当前表述 m_i 与其前面已形成的实体中相互共指的 c_k 构成正例,将 c_k 中与 m_i 最接近的一个表述之间出现的所有表述所在的实体构成负例。在特征抽取方式上,与表述对模型只刻画两个表述之间的特征不同,实体—表述模型需要考虑表述与已形成实体间的特征,而实体是由表述聚类而成的,因此只需要将所有表述对层次的特征刻画到实体一层就可以了,例如,之前的表述对层面的二元特征 {不出现 (False), 出现 (True)} 刻画到实体一层则体现为 {均不出现 (None), 大部分不出现 (Most-False), 大部分出现 (Most-True), 全部出现 (All)}。表述对模型下常用的机器学习方法也可以直接迁移到实体—表述模型中来使用。值得注意的是,与表述对模型完

全不同的是,实体—表述模型中实体聚类结果会通过测试过程自然而然的产生。这是源于特殊的测试实例构建方式。首先,表述是按其在文本中出现的顺序从左到右依次处理的;其次,对于每一个表述 m_i ,测试实例基于该表述与前面按序处理过程中已经形成的实体 c_k 进行构建。在结果生成方面,也可以选择最近最先或最优最先,分别是选择所有与当前表述 m_i 共指的实体中最近的一个或最优的一个。

表述排序模型:表述对模型中针对某一表述 m_i ,在考虑其他可能候选先行语 m_j 的时候是独立进行的,也就是独立判断不同表述对的共指结果,而不能针对表述 m_i 同时考虑所有其他候选表述。表述排序模型将排序学习应用于候选先行语的选择中,针对每一个表述 m_i ,构建训练实例时,令与其真实共指的表述的排序值(例如,设定为 2)高于其他与其不发生共指的表述的排序值(例如,设定为 1)。基于这样的思想,Yang 等^[38]首先提出了基于竞争的双候选模型(Twin-candidate Model),该模型针对给定表述 m_i 同时考察两个候选先行语以确定哪一个与给定表述 m_i 共指的可能性更大,经过对所有可能的双候选对进行考察后确定 m_i 的候选先行语,详细的算法描述可参见相关文献^[38-39]。Rahman 和 Ng^[26]扩展了 Yang 等人的思路,将双候选模型拓展为基于全部候选先行语的排序选择,并采用排序支持向量机(RankSVM)^[40]完成候选先行语的排序学习。有关该模型的表述检测、实例构建以及结果生成部分可参考表述对模型中的相应介绍。

实体排序模型:为了组合实体—表述模型与表述排序模型的优点,提出了实体排序模型。该模型类似于实体—表述模型,只不过为当前表述确定其所共指的某个实体的时候,采用的不是分类学习算法而是排序学习算法,即为当前表述 m_i 在之前按序处理形成的多个实体(或共指链)中寻找排序值得分最高的那个。Rahman 和 Ng^[26]采用排序支持向量机实现了该模型。模型中的其他部分可参照之前介绍的 3 个模型中的相应内容来理解。Rahman 和 Ng 在其论文^[41]中指出,在采用基本一致的共指特征条件下,不同模型的效果为:实体排序模型>表述排序模型>实体—表述模型>表述对模型。

2.2.2 无指导的学习模型

无指导的学习模型用于共指消解问题有其本质的优势,因为它克服了有指导的学习模型中需要大量人工标注训练语料的问题。作为篇章一级的自然语言处理任务,共指消解的标注工作的复杂性要远

远大于句子一级的词性标注(POS Tagging)、命名实体识别(NER)以及句法分析(Parsing)等任务。因此无指导的学习模型用于共指消解问题有着重要的研究意义,以下介绍一些主要的工作。

Cardie 和 Wagstaff^[42]用特征向量来刻画每个候选表述,然后采用凝聚式的层次聚类 HAC(Hierarchical Agglomerative Clustering)来对这些表述进行迭代式地合并,不过其中的距离准则函数和特征权重均通过启发式的方法指定。

Wagstaff 和 Cardie^[43]提出了一种约束聚类算法应用于共指消解任务。该算法规定了两种约束“必须链接(Must-Link)”和“不能链接(Cannot-Link)”,分别限制了哪些候选表述必须共指和必须不共指。在他们的实验中,大部分约束都是“不能链接”类型的,主要实现了一些语言学约束,比如性别、单复数以及语义类别一致性等。

Haghighi 和 Klein^[44]采用非参数贝叶斯模型将共指消解问题刻画成实体生成表述的过程,其基本模型描述了实体如何生成表述的中心词,但为了解决较为特殊的代词消解问题以及候选先行语的优选问题,该算法将代词模型以及突显度模型结合到表述的整个生成过程中,同时,由于采用了非参数贝叶斯方法,该模型不需要预先指定实体(聚类)类别数,全部实体的个数是通过推断过程自动产生的。

周俊生等^[32]采用基于图的方法对共指消解问题进行建模,将共指消解过程转化为图划分过程,使得每一对名词短语并不是孤立地进行共指判断,而是充分考虑多个待消解项之间的相关性。同时,引入有效的模块函数作为图聚类的目标函数,从而自动确定合适的聚类数目,实现名词短语等价类的自动划分。实验结果表明,该算法是一种有效可行的无监督共指消解方法。

Ng^[45]提出了期望最大化的聚类算法用于共指消解,并与改进的非参数贝叶斯模型进行了比较,取得了更好的效果。

Poon 和 Domingos^[46]采用联合无指导的马尔科夫逻辑网络框架解决共指消解问题,该方法将共指的传递性约束结合到实体—表述模型中,并采用基于规则方法获得的 3 个重要特征(中心词匹配、同位语和谓语主格)来指示共指。实验结果显示,该模型的实验效果能够与部分有指导的模型相媲美。

Haghighi 和 Klein^[47]采用生成模型分别刻画了实体类型、实体以及表述的生成过程,并基于 3 者的联合概率分布进行近似的参数估计和求解。实验结

果显示他们的模型甚至超出了 Rahman 和 Ng^[26] 的有指导的实体排序模型。

3 共指消解特征分析

对于共指消解问题来说,无论是基于规则的方法还是基于学习的方法,如何选择特征往往对于问题的有效解决起着至关重要的作用。共指消解的特征大体可以分为两类:一类是优先型特征(Preferences);另一类是约束型特征(Constraints)。前者

包括字符串匹配优先、近距离优先以及句法平行优先等等,后者包括性别一致性约束、单复数一致性约束以及语义类别一致性约束等。这种区分方式的主要依据是共指特征的指示性强弱。本文从语言学的角度出发,将共指特征分为词法特征、语法特征、距离和位置特征以及语义特征,如表 1 所示。由于实体与表述之间的特征可以由实体中的每个表述与当前表述之间的特征来刻画,因此表 1 只介绍表述(m_i 与 m_j 且 $j < i$)间的特征。

表 1 共指消解常用特征

词法特征
EXACT_MATCH: 表述 m_i 与 m_j 精确字符串匹配成功返回 T, 否则返回 F
PRO_MATCH: 表述 m_i 与 m_j 均为代词且字符串匹配成功返回 T, 否则返回 F
PN_MATCH: 表述 m_i 与 m_j 均为专有名词且字符串匹配成功返回 T, 否则返回 F
NONPRO_MATCH: 表述 m_i 与 m_j 均为非代词且字符串匹配成功返回 T, 否则返回 F
HEAD_MATCH: 表述 m_i 与 m_j 中心词字符串匹配成功返回 T, 否则返回 F
HEAD_WORD: 表述 m_i 与 m_j 中心词对, 特征取值为词对字符串
SUB_MATCH: 表述 m_i 与 m_j 子字符串匹配成功返回 T, 否则返回 F
语法特征
GENDER_MATCH: 表述 m_i 与 m_j 性别一致返回 T, 否则返回 F
NUMBER_MATCH: 表述 m_i 与 m_j 单复数一致返回 T, 否则返回 F
GN_MATCH: 表述 m_i 与 m_j 性别与单复数均一致返回 T, 否则返回 F
CLOSEST_MATCH: 表述 m_i 与 m_j 性别与单复数均一致且之间无其他表述返回 T, 否则返回 F
ANIMACY: 表述 m_i 与 m_j 动物性一致返回 T, 否则返回 F
NESTED: 表述 m_i 与 m_j 相互嵌套返回 T, 否则返回 F
C_COMMAND: 表述 m_i 与 m_j 相互 C-Command 返回 T, 否则返回 F
MAXIMALNP: 表述 m_i 与 m_j 由相同的名词短语管辖返回 T, 否则返回 F
IN_QUOTE: 表述 m_i 与 m_j 均在引用句中则返回 T, 否则返回 F
MEN_TYPE: 表述 m_i 或 m_j 的表述类型, 特征取值为表述类型字符串
REFLEXIVE: 表述 m_i 或 m_j 是否自反, 是返回 T, 否则返回 F
POSSESSIVE: 表述 m_i 或 m_j 是否是所有格, 是返回 T, 否则返回 F
DEF_NP: 表述 m_i 或 m_j 是有限定词的名词短语, 是返回 T, 否则返回 F
INDEF_NP: 表述 m_i 或 m_j 是没有限定词的名词短语, 是返回 T, 否则返回 F
DEM_NP: 表述 m_i 或 m_j 是指示性的名词短语, 是返回 T, 否则返回 F
距离和位置特征
SENT_DIS: 表述 m_i 与 m_j 之间的句子距离, 特征取值为句子距离数值
MENT_DIS: 表述 m_i 与 m_j 之间的表述个数, 特征取值为表述个数值
FIRST: 表述 m_i 或 m_j 是否为该句子的首个表述
语义特征
AILAS: 表述 m_i 与 m_j 之间是别名关系, 是返回 T, 否则返回 F
SAME_SPEAKER: 表述 m_i 与 m_j 来自相同的陈述者, 是返回 T, 否则返回 F
SAME_SRL: 表述 m_i 与 m_j 有相同的语义角色, 是返回 T, 否则返回 F
ENT_MATCH: 表述 m_i 与 m_j 的实体类型一致, 是返回 T, 否则返回 F
ENT_TYPE: 表述 m_i 或 m_j 的实体类型, 特征取值为语义类型字符串

续表

语义特征
SAME_VERB: 表述 m_i 与 m_j 受相同的动词支配, 是返回 T, 否则返回 F
APPOSTIVE: 表述 m_i 与 m_j 是同位语关系, 是返回 T, 否则返回 F
PREDICATE_NOM: 表述 m_i 与 m_j 由系动词连接, 是返回 T, 否则返回 F
WORD_NET: 表述 m_i 与 m_j 是在 WordNet 中的距离, 特征取值为具体距离值

表 1 所示的特征均被视为平面特征(Flat Features)。近年来,随着支持向量机中的核方法研究的不断深入,各种基于表述间句法结构的结构化特征(Structured Features)被以核(Kernel)的形式提出来,Yang 等^[48]和孔芳等^[49-50]分别将这样的思想应用在了中英文代词消解和中文的零形指代问题上。

盲目地扩大特征规模并不一定能有效地提高共指消解系统的性能。事实上,当系统的语料规模受限时,并不是选用的特征越多得到的效果越好。对于机器学习方法这种情况更加突出。因为语料受限决定了可以构建的训练实例受限,这时如果特征越多将导致特征空间中的各种相关参数训练就越不充分,从而出现数据稀疏的情况,使得训练得到的机器学习模型的泛化能力较差,影响实验效果。因此,有效地基于机器学习模型进行特征选择就显得尤为重要了。Hoste 和 Daelemans^[51]采用反向消除和双向爬山的方法进行共指消解的特征选择,取得了不错的效果。Bengtson 和 Roth^[20]研究了各种特征对于共指消解任务的贡献,提出了即使在很弱的分类器下,如果采用合理的特征组合方式依然能够达到不错的实验结果。Saha 等^[52]将基于遗传算法的多目标优化技术应用于共指消解问题,实现了基于不同评测准则目标优化的特征选择,实验结果表明效果提升显著。

4 共指消解评测

最早开始的共指消解评测是消息理解系列会议 MUC(Message Understanding Conference),该会议由美国国防高级研究计划委员会(DARPA)资助。MUC 主要包括信息抽取相关的评测任务。从 1987 年到 1998 年,该会议一共举办了 7 届。共指消解相关的任务出现在 1995 年举行的 MUC6 和 1998 年举行的 MUC7 中,当时提供的评测语料库全部为英文。

从 2000 年开始,由美国国家标准与技术研究院

(NIST)组织的自动内容抽取(Automatic Content Extraction,ACE)评测会议取代了之前的 MUC 系列会议。ACE 评测的主要任务之一是实体检测与跟踪。该任务将篇章中出现的各种表述指向其对应的实体,本文所采用的共指消解定义正是来源于 ACE 评测中该任务的定义。从 2003 年开始,ACE 评测提供了中文语料的共指消解任务,开启了中文共指消解国际评测的先河。ACE 评测一直持续到 2008 年,从 2009 年开始被 TAC(Text Analysis Conference)评测所取代,共指消解任务也过渡到 KBP(Knowledge Base Population,即基于维基百科的实体链接)任务。

2010 年的 SemEval 国际评测增加了多语言共指消解任务,其采用 OntoNotes 2.0 数据集。该数据集不同于 ACE 评测所使用的 ACE 数据集,主要区别在于 OntoNotes 数据集不包含单表述实体(即单一表述的实体聚类,Singleton Entity),而 ACE 是包含的。这也就导致了 ACE 标注了大量指称表述,而 OntoNotes 仅标注那些发生共指关系的表述。

2011 年的自然语言处理国际著名评测 CoNLL 举办了英文的共指消解评测,采用了更新版本的 OntoNotes 4.0 数据。除了不包含单表述实体以外,该数据集中对于同位语以及谓语主格(Predicate Nominative)也均不看作为共指关系。同时,事件名词与动词的共指关系也做了标注。在这次评测中,排名前三的系统分别来自斯坦福大学 Lee 等的系统^[12]、加泰罗尼亚理工大学 Sapena 等的系统^[53]以及伊利诺伊香槟分校 Chang 等的系统^[54]。Lee 等的系统^[12]采用基于规则并层层筛选的方法取得了第一名的好成绩,他们以准确率由高到低构建了一系列的筛子迭代地为不同的表述选取先行语,分别基于 MUC、B-cubed 和 CEAF-E 3 项评测指标获得了 59.57%、68.31%和 45.48%的 F 值,3 项指标的平均 F 值为 57.79%,同时针对表述检测子任务获得了准确率 66.81%、召回率 75.07%以及 F 值 70.70%。Sapena 等的系统^[53]首先基于决策树

C4.5 学习共指特征权重,然后采用松弛标记(Relaxation Labeling)的方法迭代地为每个表述(Mention)标记其所属实体(Entity),分别基于 3 项评测指标获得了 59.55%、67.09%和 41.32%的 F 值,平均 F 值为 55.99%,同时针对表述检测子任务获得了准确率 28.19%、召回率 92.39%以及 F 值 43.20%。Chang 等的系统^[54]基于一个单独训练的成对共指消解模型,采用整数线性规划(Integer Linear Programming)技术执行最优最先聚类(Best-First Clustering),其在 3 项评测指标中分别获得了 57.15%、68.79%和 41.94%的 F 值,平均 F 值为 55.96%,同时在表述检测子任务上获得了准确率 61.96%、召回率 68.08%以及 F 值 64.88%。

2012 年的 CoNLL 依旧举办了共指消解评测,这次的任务采用了 OntoNotes 5.0 的数据集,并同时提供英文、中文以及阿拉伯文的语料进行多语言的共指消解评测,以研究不同语言的共指消解研究的差别与共性。

在英文任务中,巴西里约热内卢天主教大学 Fernandes 的系统^[55]获得了第一名的成绩。他们采用潜共指树(Latent Coreference Trees)结构以及基于熵的特征选择(Entropy Guided Feature Induction)方法,在最大间隔结构化感知器(Large Margin Structure Perceptron)的框架下进行共指消解学习与推断,分别基于 MUC、B-cubed 和 CEAF-E 3 项评测指标获得了 70.51%、71.24%和 48.37%的 F 值,三项指标的平均 F 值为 63.37%,同时针对指标表述检测子任务获得了准确率 83.45%、召回率 72.75%以及 F 值 77.73%。来自德国海德堡理论研究所的 Martschat 等的系统^[56]采用基于图的方法,将文本中的表述看作图中的节点,将表述间的不同共指特征所表示的关系看作边,并尝试采用谱聚类和贪心聚类进行共指消解,在 3 项评测指标上分别获得了 66.97%、70.36%和 46.60%的 F 值,3 项指标的平均 F 值为 61.31%,同时针对表述检测子任务上获得了准确率 76.10%、召回率 74.23%以及 F 值 75.15%。来自斯图加特大学 Bjorkelund 等的系统^[57]采用 Resolver Stacking 的方法将 BF(Best-First,最优最先)、PCF(Pronoun Closest First,代词最近最先)和 AMP(AverageMaxProb,最大化平均概率)3 种共指推断方法进行层叠组合,同时在识别候选表述的时候基于一个分类器针对代词(比如 it、you 和 we)进行特殊的指代性(Referential)判断,实验结果显示他们的系统在三项评测指标上分别获得

了 67.58%、70.26%和 45.87%的 F 值,平均 F 值为 61.24%,同时在表述检测子任务获得了准确率 77.09%、召回率 73.75%以及 F 值 75.38%。

在中文任务中,获得前三名的系统分别来自于德州大学达拉斯分校 Chen 等的系统^[58]、哈尔滨工业大学 Yuan 等的系统^[59]以及斯图加特大学 Bjorkelund 等的系统^[57]。Chen 等的系统^[58]借鉴了 2011 年斯坦福大学 Lee 等的方法,将多层筛选的规则化模型应用在了中文共指消解任务上,并取得了第一名的好成绩,分别基于 MUC、B-cubed 和 CEAF-E 三项评测指标获得了 62.21%、73.55%和 50.97%的 F 值,3 项指标的平均 F 值为 62.24%,同时针对表述检测子任务获得了准确率 72.16%、召回率 71.12%以及 F 值 71.64%。Yuan 等的系统^[59]依然是主要借鉴了 2011 年斯坦福大学 Lee 等的方法,他们将规则和统计相结合的方法应用在表述检测和共指消解的多个相关任务中,在 3 项评测指标上分别取得了 60.33%、72.90%和 48.83%的 F 值,平均 F 值为 60.69%,同时针对表述检测子任务获得了准确率 64.09%、召回率 72.75%以及 F 值 68.15%。Bjorkelund 等的系统^[57]采用了他们在英文任务中相同的框架和思路,同时结合中文语言中的一些特殊性作为规则特征,他们在 3 项评测指标上分别取得了 58.61%、73.10%和 48.19%的 F 值,平均 F 值为 59.97%,同时针对表述检测子任务获得了准确率 63.54%、召回率 69.45%以及 F 值 66.37%。

5 当前研究趋势

近年来,在学术界的共同努力下,共指消解研究取得了不错的成绩,越来越多的科研人员开始关注这一领域的研究。以下从模型构建以及共指特征等角度分别介绍共指消解任务的当前研究趋势。

模型构建:共指消解问题是由实体表述检测和共指消解两个子任务共同组成的,特别适合使用多任务的联合学习与联合推断框架。Denis 和 Baldridge^[60]采用整数线性规划联合推断名词短语的指代性检测(即表述检测)和共指消解。在 ACE 数据上,该方法使得 MUC 评测准则的 F 值提高了 3.7%~5.3%。针对类似的问题,Rahman 和 Ng^[26]采用排序支持向量机(RankSVM)模型联合学习篇章新表述检测(Discourse-New Detection)和共指消解,获得了比独立学习两个子任务更好的效果。

基于表述对模型的共指消解问题,在表述检测已经执行的情况下,通常分为成对分类和表述聚类两个子任务(如 2.2.1 节所述)。因此基于这两个子任务的联合模型也大量应用在共指消解问题上。Finkel 和 Manning^[28]与 Chang 等^[54]利用整数线性规划技术,基于预测得到的表述对共指概率值,分别采取传递性约束(Transitivity Constraints)和最优最先约束(Best-First Constraints),针对成对分类和表述聚类进行联合推断,取得了很好的效果。但这种联合推断框架的劣势在于,预测表述对间共指概率值的特征权重是通过单独学习得到的。为了使学得特征权重更加准确,Finley 和 Joachims^[61]、McCallum 和 Wellner^[62]以及 Song 等人^[63]分别采用有指导的聚类框架、条件模型以及马尔科夫逻辑网络实现了将成对分类和基于传递性约束的表述聚类进行联合学习的过程。

共指特征:近年来,规则方法用于共指消解的实验效果越来越好,而大量机器学习方法对共指消解问题的效果提升越发有限,因此更多的学者开始关注如何基于深层次的语言知识以及背景知识来进行共指消解研究。

获取深层次的语言知识及背景知识可以通过以下几种途径。首先,基于常规知识库。这种方法需要借助一些已有的知识词典,例如,英文的 WordNet,中文的《现代汉语语法信息词典》和 HowNet 等;其次,从大规模语料库中挖掘模式信息。这种方法主要是根据一些已有模板,然后在大规模语料库上统计各种信息。例如,Bergsma^[64]在一个经依存分析的语料库上获取了大量的共指信息,实现了名词短语的性别和单复数信息的自动提取。Yang 和 Su^[65]利用从语料库中挖掘的模板信息来增强共指消解。Ponzetto 和 Strube^[25]同时采用 WordNet、Wikipedia 以及语义角色信息构建多种语义特征用于共指消解问题。Rahman 和 Ng^[66]采用 YAGO、FrameNet、名词对特征、动词对特征以及更为广泛的背景知识来强化共指消解;最后一种方法是将互联网看成是一个巨大的语料库,利用搜索引擎返回的各个查询得到的 Snippets 或结果数来计算相关信息,例如,利用返回结果数来计算两个名词短语间的互信息来考察其关联程度等^[67]。

相关任务:随着共指消解研究的深入开展,与之相似或相关的其他任务的研究也随之被带动。例如,命名实体消歧(或实体链接)任务。继 ACE 评测之后的 TAC 评测有一项 KBP(Knowledge Base

Population)任务,是基于跨文档的实体消歧任务,近年来有关该问题的研究工作成果显著^[68-69]。另一项相关任务是事件共指消解。当前的语料库大多都来自于新闻语料,其中的句子包含有大量关于新闻事件的描述,因此如何实现基于事件的共指消解,就显得尤为重要。事件共指消解任务研究已经取得了一定的进展^[70-71]。此外,共指消解也是很多其他自然语言处理任务的基础,例如,文本摘要、信息抽取、关系抽取以及观点挖掘等任务。

6 结论与展望

当前大多数自然语言处理任务脱胎于规则模型,并转进到统计机器学习为主的解决框架。共指消解问题也不例外。但例外的是,在经过了十余年基于机器学习框架下的共指消解研究之后,人们发现,基于规则的系统依然能够取得一致的甚至是更好的效果。导致这一问题的根源在于共指问题的形式化表示与其真正的语言学本质相去甚远。当然,任何一种方法都有其不足之处:基于规则的系统不能从大量已标注数据中获得良好的支持,使得系统的泛化能力一般;基于统计机器学习的系统并不善于精确地解决问题。因此,未来的共指消解研究必然会沿着规则与统计相结合的思路一直走下去。一方面需要寻找更新更有效的特征来强化共指判断,另一方面在现有特征的基础上,需要考虑如何将篇章一级的计算模型融合到规则与统计相结合的系统当中去,以解决当前的形式化表示过于简单且脱离实际的问题。

参考文献

- [1] 郎君,秦冰,刘挺,等. 篇章共指消解研究综述[J]. 汉语语言与计算学报, 2007, 17(4):227-253.
- [2] 王厚峰. 指代消解的基本方法和实现技术[J]. 中文信息学报, 2002, 16(6):9-17.
- [3] J. R. Hobbs. Resolving pronoun references[J]. Journal of Lingua, 1978, 44:311-338.
- [4] A. Haghighi, D. Klein. Simple coreference resolution with rich syntactic and semantic features[C]//Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2009: 1152-1161.
- [5] B. Grosz, A. Joshi, S. Weinstein. Centering: A framework for modelling the local coherence of discourse[J]. Journal of Computational Linguistics,

- 1995, 21(2):203-225.
- [6] Susan E. Brennan, Marilyn W. Friedman, Carl Pollard. A centering approach to pronouns[C]//Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics (ACL), 1987:155-162.
- [7] M. Poesio, R. Stevenson, Barbara Di Eugenio, et al. Centering: A parametric theory and its instantiations [J]. Journal of Computational Linguistics, 2004, 30 (3):309-363.
- [8] S. Lappin, H.J. Leass. An algorithm for Pronominal Anaphora Resolution [J]. Journal of Computational Linguistics, 1994, 20(4):535-561.
- [9] C. Kennedy, B. Boguraev. Anaphora for everyone: Pronominal anaphora resolution without a parser[C]//Proceedings of the 16th International Conference on Computational Linguistics(COLING), 1996:113-118.
- [10] R. Mitkov. Robust pronoun resolution with limited knowledge [C]//Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL), 1998:869-875.
- [11] K. Raghunathan, H. Lee, S. Rangarajan, et al. A multi-pass sieve for coreference resolution[C]//Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2010.
- [12] H. Lee, Y. Peirsman, A. Chang, et al. Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task[C]//Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task, 2011:28-34.
- [13] V. Ng, C. Cardie. Bootstrapping coreference classifiers with multiple machine learning algorithms[C]//Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003:113-120.
- [14] O. Uryupina, S. Saha, A. Ekbal, et al. Multi-metric optimization for coreference: The unitn / iitp / essex submission to the 2011 conll shared task[C]//Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task, 2011:61-65.
- [15] V. Ng. Graph-cut-based anaphoricity determination for coreference resolution [C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies(HLT-NAACL), 2009:575-583.
- [16] Guodong Zhou, Fang Kong. Global learning of noun phrase anaphoricity in coreference resolution via label propagation[C]//Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2009:978-986.
- [17] 孔芳,朱巧明,周国栋. 中英文指代消解中待消解项识别的研究[J]. 计算机研究与发展, 2012, 49(5): 1072-1085.
- [18] J. McCarthy, W. Lehnert. Using decision trees for coreference resolution[C]//Proceedings of the 14th International Joint Conference on Artificial Intelligence, 1995.
- [19] Wee Meng Soon, Hwee Tou Ng, Chung Yong Lim. A machine learning approach to coreference resolution of noun phrases [J]. Computational Linguistics, 2001, 27(4):521-544.
- [20] E. Bengtson, D. Roth. Understanding the value of features for coreference resolution[C]//Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2008.
- [21] V. Ng, C. Cardie. Improving machine learning approaches to coreference resolution[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2002:104-111.
- [22] C. Gasperi. Active learning for anaphora resolution [C]//Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing, 2009.
- [23] Niyu Ge, J. Hale, E. Charniak. A statistical approach to anaphora resolution[C]//Proceedings of the ACL 1998 Workshop on Very Large Corpora, 1998.
- [24] Xiaoqiang Luo, A. Ittycheriah, Hongyan Jing, et al. A mention-synchronous coreference resolution algorithm based on the bell tree[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2004:135-142.
- [25] S. P. Ponzetto, Michael Strube. Exploiting semantic role labeling, wordnet and wikipedia for coreference resolution[C]//Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL), 2006: 192-199.
- [26] A. Rahman, V. Ng. Supervised models for coreference resolution[C]//Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2009:968-977.
- [27] Y. Versley, A. Moschitti, M. Poesio, et al. Coreference systems based on kernels methods[C]//Proceedings of the 22nd International Conference on Computational Linguistics (COLING), 2008: 961-968.
- [28] J. R. Finkel, C. D. Manning. Enforcing transitivity in coreference resolution[C]//Proceedings of the Annual Meeting of the Association for Computational Lin-

- guistics (ACL), 2008;45-48.
- [29] Shujian Huang, Yabing Zhang, Junsheng Zhou, et al. Coreference resolution using markov logic networks[C]//Proceedings of the 10th International Conference Computational Linguistics and Intelligent Text Processing(CICLing), 2009.
- [30] 刘未鹏,周俊生,黄书剑,等. 基于有监督关联聚类的中文共指消解[J]. 计算机科学, 2009, 36(9):182-185.
- [31] C. Nicolae, G. Nicolae. Bestcut: A graph algorithm for coreference resolution[C]//Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2006;275-283.
- [32] 周俊生,黄书剑,陈家骏,等. 一种基于图划分的无监督汉语指代消解算法[J]. 中文信息学报, 2007, 21(2):77-82.
- [33] 谢永康,周雅倩,黄萱菁. 一种基于谱聚类的共指消解方法[J]. 中文信息学报, 2009, 23(3):10-16.
- [34] Marc B. Vilain, John D. Burger, John S. Aberdeen, et al. A model-theoretic coreference scoring scheme[C]//Proceedings of the Sixth Message Understanding Conference(MUC), 1995;45-52.
- [35] A. Bagga, B. Baldwin. Algorithms for scoring coreference chains[C]//Proceedings of the First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference, 1998;563-566.
- [36] Xiaoqiang Luo. On coreference resolution performance metrics[C]//Proceedings of the joint conference on human language technology and empirical methods in natural language processing(HLT-EMNLP), 2005; 25-32.
- [37] Xiaofeng Yang, Jian Su, Jun Lang, et al. An entity-mention model for coreference resolution with inductive logic programming[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2008;843-851.
- [38] Xiaofeng Yang, Guodong Zhou, Jian Su, et al. Coreference resolution using competition learning approach[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2003;176-183.
- [39] Xiaofeng Yang, Jian Su, Chew Lim Tan. A twin-candidate model for learning-based anaphora resolution[J]. Computational Linguistics, 2008, 34(3): 327-356.
- [40] T. Joachims. Optimizing search engines using click-through data[C]//Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), 2002.
- [41] A. Rahman, V. Ng. Narrowing the modeling gap: A cluster-ranking approach to coreference resolution [J]. Journal of Artificial Intelligence Research (JAIR), 2011;469-521.
- [42] C. Cardie, K. Wagstaff. Noun phrase coreference as clustering[C]//Proceedings of the 1999 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1999.
- [43] K. Wagstaff, C. Cardie. Clustering with instance-level constraints[C]//Proceedings of the Seventeenth International Conference on Machine Learning (ICML), 2000;1103-1110.
- [44] A. Haghighi, D. Klein. Unsupervised coreference resolution in a nonparametric bayesian model[C]// Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2007, 45: 848.
- [45] Vincent Ng. Unsupervised models for coreference resolution[C]//Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2008;640-649.
- [46] H. Poon, P. Domingos. Joint unsupervised coreference resolution with markov logic[C]//Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2008;650-659.
- [47] A. Haghighi, D. Klein. Coreference resolution in a modular, entity-centered model[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies (HLT-NAACL), 2010;385-393.
- [48] Xiaofeng Yang, Jian Su, Chew Lim Tan. Kernel-based pronoun resolution with structured syntactic knowledge[C]//Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics(ACL), 2006;41-48.
- [49] Fang Kong, Guodong Zhou. A tree kernel-based unified framework for chinese zero anaphora resolution [C]//Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2010;882-891.
- [50] 孔芳,周国栋. 基于树核函数的中英文代词消解[J]. 软件学报, 2012, 23(5):1085-1099.
- [51] Véronique H. Optimization Issues in Machine Learning of Coreference Resolution[D]. PhD thesis, University of Antwerp, 2005.
- [52] S. Saha, A. Ekbal, O. Uryupina, et al. Single and multi-objective optimization for feature selection in anaphora resolution[C]//Proceedings of 5th International Joint Conference on Natural Language Processing(IJCNLP), 2011;93-101.

- [53] E. Sapena, Lluís Padró, J. Turmo. Relaxcor participation in conll shared task on coreference resolution [C]//Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task, 2011:35-39.
- [54] K. Chang, R. Samdani, A. Rozovskaya, et al. Inference protocols for coreference resolution[C]//Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task, 2011: 40-44.
- [55] E. Fernandes, Cícero dos Santos, Ruy Milidiú. Latent structure perceptron with feature induction for unrestricted coreference resolution[C]//Proceedings of the Joint Conference on EMNLP and CoNLL Shared Task, 2012:41-48.
- [56] S. Martschat, Jie Cai, S. Broscheit, et al. A multi-graph model for coreference resolution[C]//Proceedings of the Joint Conference on EMNLP and CoNLL Shared Task, 2012:100-106.
- [57] Anders Björkelund, Richárd Farkas. Data-driven multilingual coreference resolution using resolver stacking[C]//Proceedings of the Joint Conference on EMNLP and CoNLL -Shared Task, 2012: 49-55.
- [58] Chen Chen, Vincent Ng. Combining the best of two worlds: A hybrid approach to multilingual coreference resolution[C]//Proceedings of the Joint Conference on EMNLP and CoNLL -Shared Task, 2012: 56-63.
- [59] Bo Yuan, Qingcai Chen, Yang Xiang, et al. A mixed deterministic model for coreference resolution[C]//Proceedings of the Joint Conference on EMNLP and CoNLL Shared Task, 2012: 76-82.
- [60] Pascal Denis, Jason Baldridge. Joint determination of anaphoricity and coreference resolution using integer programming[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL), 2007:236-243.
- [61] T. Finley, T. Joachims. Supervised clustering with support vector machines[C]//Proceedings of the International Conference on Machine Learning (ICML), 2005:217-224,.
- [62] A. McCallum, B. Wellner. Conditional models of identity uncertainty with application to noun coreference[C]//Proceedings of Neural Information Processing Systems (NIPS), 2004:905-912.
- [63] Yang Song, Jing Jiang, Wayne Xin Zhao, et al. Joint learning for coreference resolution with markov logic [C]//Proceedings of the conference on Empirical Methods in Natural Language Processing and Natural Language Learning (EMNLP-CoNLL), 2012: 1245-1254.
- [64] S. Bergsma. Automatic acquisition of gender information for anaphora resolution[C]//Proceedings of the Canadian Conference on Artificial Intelligence, 2005:342-353.
- [65] Xiaofeng Yang, Jian Su. Coreference resolution using semantic relatedness information from automatically discovered patterns[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2007.
- [66] A. Rahman, V. Ng. Coreference resolution with world knowledge [C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2011:814-824.
- [67] M. Poesio, R. Mehta, A. Maroudas, et al. Learning to resolve bridging references[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics(ACL), 2004:143-150.
- [68] Heng Ji, Ralph Grishman. Knowledge base population: Successful approaches and challenges[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics(ACL), 2011:1148-1158.
- [69] S. Singh, A. Subramanya, F. Pereira, et al. Large-scale cross-document coreference using distributed inference and hierarchical models[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2011.
- [70] C. A. Bejan, M. Titsworth, A. Hickl, et al. Non-parametric bayesian models for unsupervised event coreference resolution[C]//Proceedings of Neural Information Processing Systems (NIPS), 2009:73-81.
- [71] Zheng Chen, Heng Ji. Graph-based event coreference resolution[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2009:54-57.



宋洋(1986—),博士,中级工程师,主要研究领域为自然语言处理。
E-mail: ysong@pku.edu.cn



王厚峰(1965—),博士,教授,主要研究领域为语篇分析,情感挖掘,问答系统,语言知识库。
E-mail: wanghf@pku.edu.cn