

结合规则与语义的中文人称代词指代消解

张文艳¹ 李存华² 仲兆满² 王艺² 李莉²

(1. 中国矿业大学计算机科学与技术学院, 徐州, 221116; 2. 淮海工学院计算机工程学院, 连云港, 222005)

摘要: 指代消解是一种为了确定文章中出现的指代词与前文中出现的内容是否为同一事物的技术, 在海量信息文本智能处理中具有重要的作用, 而人称代词在各种指代词集合中占有相当一部分比例。本文采用规则与语义相结合的方法对中文人称代词进行指代消解, 在基础的语法过滤规则之上新增同位语规则过滤指代词的候选消解项; 提出更精确的同义词距离计算方法, 利用同义词词林和知网对人称代词的关联词与候选先行词的关联词进行语义关系计算, 选择关联度最高的候选先行词作为最终的指代结果。通过不同方法的对比实验和在真实语料数据集上的实验表明, 本文所提方法获得了较好的效果。

关键词: 指代消解; 人称代词; 规则; 候选先行词; 语义特征

中图分类号: TP391.1 **文献标志码:** A

Coreference Resolution of Chinese Personal Pronouns With Combination of Semantics and Rules

Zhang Wenyan¹, Li Cunhua², Zhong Zhaoman², Wang Yi², Li Li²

(1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, 221116, China; 2. School of Computer Engineering, Huaihai Institute of Technology, Lianyungang, 222005, China)

Abstract: Coreference resolution is a widely used technology to judge whether pronouns can match with the entity existing before in the text, which plays a crucial role in intelligent processing for massive text information on internet. A coreference resolution study, especially on the frequently-used Chinese personal pronouns, was carried out with the result of a developed algorithm with the combination of semantics and rules. Based on fundamental filtration rules, an improved mechanism specific to apposition was also adopted. To raise the accuracy of calculating the synonyms distances, the algorithm identified the associative-word of personal pronouns and selected antecedents based method for analyzing semantic relations and selecting high relevancy antecedent, which was realized with the aid of Tongyici Cilin and HowNet. Comparison experiments with different methods and experiments on the real corpus dataset were conducted, and results show that the presented method achieves higher validity and obvious improvement.

Key words: coreference resolution; person pronouns; rules; antecedent; semantic relations

引言

在汉语中,人们经常为了表达的多样性而使用代词指代前文已经提到的事物,达到行文流畅的效果。但是,这无疑给计算机处理信息带来难度:自然语言中最大的问题就是计算机不能像人脑一样通过固有的知识结合上下文的语义环境判断某个代词具体指代什么。指代消解的出现更好地解决了词义消歧、词义标注等问题,为机器翻译、信息自动抽取、多语言信息处理和文本摘要等应用系统的发展提供了支持。

指代消解就是为了明确自然语言中两个成分是否代指同一个事物的过程,主要研究人称代词的消解,为进一步的文本处理提供完整信息。目前,指代消解的主要技术分为基于句法的方法和基于语料库^[1]的方法:前者主要利用句法层面的知识,将启发式规则应用到指代消解当中^[2],比较典型的有 RAP 算法;而后者则主要运用统计学的方法和机器学习^[3,4]的方法,其中包括简单同现方法、决策树、聚类^[5]以及其他机器学习^[6]方法等。

近年来指代消解逐渐成为计算机语言学研究的重点之一。1995 年美国麻省大学的 McCarthy 等^[7]运用决策树的方法进行共指消解;1997 年的 EACL 和 1999 年的 ACL 年会设立了指代消解专题会议;1999 年 Cardie 等^[8]提出通过聚类方法进行名词短语的共指消解;《Computational Linguistics》学报在 2001 年出了指代消解的专辑。与国外相比,国内的指代消解研究相对滞后。2000 年王凌飞等研究基于中心理论的指代消解在汉英机器翻译中的应用;王厚峰等^[9]在 2002 年提出了消解人称代词的基本方法和技术;李国臣等^[10]在 2005 年采用决策树机器学习并结合优先选择的策略进行中文人称代词指代消解。与英语相比,汉语指代消解的研究才刚刚起步,还存在较大的差距。汉语词语间不像英语那样有明显的分界,因此分词的好坏对于指代消解将是一个重要的因素。同时由于语言之间的差异,使得直接运用英语指代消解的方法解决汉语指代消解问题变得困难。

现阶段的指代消解技术仍然面临着指代范围不全面、指代对象不准确等问题。本文提出了规则和语义相结合的汉语人称代词指代消解方法,涵盖了 12 种人称代词,消解效果有了一定的提高。

1 规则与语义结合的消解方法

1.1 总体结构

规则与语义相结合的消解方法主要体系结构如图 1 所示。指代消解流程包括以下部分:(1)语料处理:对文本进行预处理,包括文本格式的统一、中文分词、词性标注以及对语料中各种类型人称代词的统计;(2)规则过滤:首先利用代词过滤集去掉不需要消解的代词,然后根据候选先行词过滤规则,确定满足条件的候选先行词组成待消解对;(3)语义相似度计算:利用同义词词林和 HowNet 计算代词的关联词与候选先行词的关联词在文本中的语义关系度量值,确定人称代词指代结果。

1.2 人物及关联词提取

采用中科院分词工具 ICTCLAS 对文本进行分词和词性标注。参照在线新华词典中对汉语人称代词的定义提取人称代词及人名,包括:第一人称(我、我们),第二人称(你、你们),第三人称代词(他、他们、她、她们)和特殊指代词(这人、这个人、那

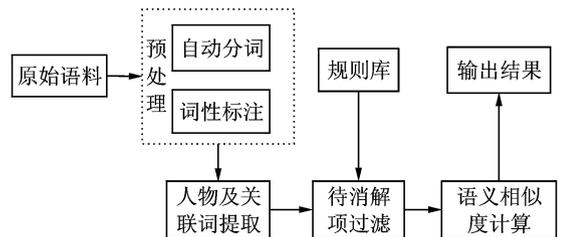


图 1 指代消解体系结构

Fig. 1 Architecture of coreference resolution

人、那个人),其中代词“自己”作为特殊指代也加入到第一人称代词中。

受语言表达形式多样性及分词系统影响,要排除那些被分词系统识别为人称代词但是并不需要消解的成语和俗语,如:“顾而言他、敌众我寡、你唱我和”等,这些成语中含有人称代词,但是作为成语被引用时并不需要进行消解。本文利用《汉语成语词典》提取包含人称代词的成语和俗语,主要为第一人称代词和第三人称代词,被识别为人称代词的成语加入到人称代词过滤集,作为过滤规则引用。利用舆情监测系统(研究室自主研发)抽取当前热门的3个人物、5个事件共300篇文本作为实验语料进行分析。统计语料中各类人称代词的比例如表1所示。

表1 人称代词比例

Tab.1 Proportion of personal pronouns

代词类别	人物/个		事件/个		比例/%	
	单数	复数	单数	复数	单数	复数
第一人称代词(自己)	150(127)	88	486(148)	111	24.42(10.56)	7.64
第二人称代词	33	0	174	15	7.95	0.58
第三人称代词	410	51	708	96	42.93	5.64
特殊指代词	1	0	6	0	0.27	0
总数	721	139	1522	222	86.14	13.86

从表1看出,在中文人称代词的分布中,第一人称和第三人称占有较大的比例,单数人称代词出现的次数要远远超过复数人称代词,因此对第一人称和第三人称代词消解的正确与否将直接影响最后效果。通过对第一人称代词的分析,发现第一人称和第二人称在语料中主要以两种形式出现:第一类以直接或间接引用的方式描述说话内容,是需要消解的;第二类是在祈使句中出现,是不需要消解的。

选择人名和代词周围特定词性的词语作为人物关联词,根据关联词判断人名和代词之间的指代关系,词性包括:nz,ng,an,vd,vn,vg,i,j,l,v,n,具体字母的含义参见表2。标点符号作为划分关联词归属的标志,无标点符号则扩大关联词提取范围至邻近的人物。

表2 词性标注集

Tab.2 Set of speech tagging

符号	nz	ng	an	vd	vn	vg	i	j	l	v	n	r
名称	其他专名	名语素	名形词	副动词	名动词	动语素	成语	简称略语	习用语	动词	名词	代词

1.3 候选先行词过滤

代词所代指的词即为候选先行词,能够作为代词候选先行词的词包括出现在人称代词前部的人名或者是已经被消解的人称代词。根据语料统计结果,结合汉语中的语法、语义特征,提出了同位语特征的消解规则和人物对话的语法规则,配合性别、单复数、距离等属性提取代词的候选先行词。

(1) 同位语规则

通过对语料的分析发现,在文本中出现“\S+/nr+/r+\S”和“\S+/r1+/r2+\S”的表达形式,例如:“林森浩:我也一直跟公安说我这个人矛盾。”,这种以同位语形式出现的指代词根据同位语规则直接消解:表达一中 $\langle r \rangle = \langle nr \rangle$,表达二中 $\langle r1 \rangle = \langle r2 \rangle$ 。这一规则可以消解掉大部分的特殊指代词和第一人称中“自己”的指代。

(2) 语法规则

根据第一人称和第二人称在文本中出现的两种表达形式,结合人物对话特征,设定语法规则:出现

如下形式“\S+/\nr1+/\v+/\w+/\w+\S+/\r1+\S+/\w+/\w+\S+/\nr2+/\v+/\w+/\w+\S+/\r2+\S+/\w+/\w”，表达中若 r_1, r_2 为第一人称则 $\langle r_1 \rangle = \langle nr_1 \rangle, \langle r_2 \rangle = \langle nr_2 \rangle$ ，若 r_1, r_2 为第二人称则 $\langle r_1 \rangle = \langle nr_2 \rangle, \langle r_2 \rangle = \langle nr_1 \rangle$ 。

(3) 性别一致性

收集具有明显性别特征的词分别组成男女性别词库。如女士、小姐为女性特征词，先生、兄弟为男性特征词，其余无明确特征的词则标注为无性别类。候选先行词与人称代词性别一致则为 1，有一项无性别为 0.5，不一致为 0。

(4) 单复数一致性

词语的单复数分为单数、复数和无单复数三类。人称代词的单复数可直接判断，候选先行词则根据“和、每个、大部分、许多”等特征词判断，无明显特征的标注为无单复数类。

(5) 距离属性

经过对待消解项的分析统计，发现汉语中指代词与先行词的距离大多在三句话之内，设定候选先行词与指代词的距离上线为 3，过滤掉大于 3 的候选先行词。

1.4 语义特征判断

语义特征的判断主要是对待消解项中人称代词的关联词与候选先行词的关联词语义关系的判断，包括关联词一致性、关联词相似性和关联词相关性。其中，关联词一致性用代词关联词与候选先行词关联词的字符串匹配判断，利用同义词词林和 HowNet 确定关联词相似性^[11]和相关性。

首先根据同义词词林计算待消解项中关联词的相似度，在小于设定阈值的情况下再进一步计算关联词在 HowNet 中的相关度。对知网中未收录的词语从同义词词林中找出近义词，利用近义词进行相关度计算。

1.4.1 语义相似度计算

《同义词词林》是梅家驹等人于 1983 年编纂而成，后来由哈尔滨工业大学信息检索实验室进一步扩展成为了一部具有汉语大词表的《哈工大信息检索研究室同义词词林扩展版》。词林中收录词语近 7 万条，是一部同义类词典，词典采用层级体系，具备 5 层结构，存储在同义词词林中的词语都采用 8 位标记法，具体编码格式如表 3 所示。

表 3 词语编码表
Tab. 3 Word coding table

编码位	1	2	3	4	5	6	7	8
符号举例	D	a	1	5	B	0	2	= / # / @
符号性质	大类	中类	小类	词群	原子词群			
级别	第一级	第二级	第三级	第四级	第五级			

词林级别越高词义越详细，第 5 级词群中很多只有一个词语并且不可再分。若两个词语有相同的前三级编码，其相似程度已非常地接近，但是本文通过研究发现并非所有处于相同三级的词语都具有很高的相似性，这与两词语所在的第四级的个数和编码距离有关。统计发现第四类的编码最大范围为 $A \sim J$ ，个数为 10，当两词语的编码距离在 4 以内时，词语的相似度较高，若超过则需要进一步的相似计算。相似度与距离成反比，用距离评测词语间的相似关系^[12]。

给定两词 A, B ，其语义距离计算方法为

$$\text{Dist}(A, B) = \min_{a \in P, b \in Q} \text{dist}(a, b) \quad (1)$$

式中: P, Q 分别为 A, B 两词在同义词词林中的所有语义的集合。语义 a, b 之间的距离为

$$\text{dist}(a, b) = \begin{cases} (5 - n(a, b)) * 2^{(i-10)} & n \geq 4 \\ 5 - n(a, b) & n < 4 \end{cases} \quad (2)$$

式中: $n(a, b)$ 为两词语的编码开始分支的级数; i 为第四级中两词语的距离差。

不同方法计算的词语相似度如表4所示。计算词语在同义词词林和知网中的相似度量,HowNet相似度参照李峰等^[13]的语义相似度计算,在 $i=4$ 时,相似距离相同但知网中的相似度相差较大, $i < 4$ 时相似度普遍较高。因此设定 $i=3$ 为阈值,在距离大于 $1 * 2^{-7}$ 时,利用知网计算相似度。

表4 词语相似度

Tab. 4 Word similarity

类别	词语	编码	距离	Hownet 相似度
	美女	Ac03A01=	0	1
“美女”与其他词语相似度	灰姑娘	Ac03E02#	$1 * 2^{-6}$	0.735 1
	丑八怪	Ac03C01=	$1 * 2^{-8}$	0.681 5
	国人	Ad02A01=	0	1
“国人”与其他词语相似度	殖民	Ad02E01@	$1 * 2^{-6}$	0.054 75
	唐人	Ad02D01=	$1 * 2^{-7}$	0.579 2

1.4.2 语义相关度计算

知网(Hownet)是一个以汉语和英语词语所代表的概念为描述对象,揭示概念与概念之间以及概念所具有的属性之间的关系为基本内容的常识知识库。影响语义相关度计算的因素有:词语上下位关系所体现的相似度、横向关系所体现的相关度。采用李生琦^[14]的语义相关度计算方法:相关度=相似度*(权值1)+关联度*(权值2)+实力影响因素*(权值3),其中第三项的值为0。

(1) 词语相似度计算

《知网》中的所有义原根据上下位关系构成了一个树状义原层次结构,利用距离计算相似度^[13]。假设两个义原在层次体系中的路径长度为 d ,定义两个义原的语义距离为

$$\text{Sim}(p_1, p_2) = \frac{\alpha}{\alpha + d} \quad (3)$$

式中: p_1 和 p_2 代表两个义原(primitive); d 为 p_1 和 p_2 在义原层次体系中的路径长度; α 为一个可调节的参数。

两个概念之间的相似度为

$$\text{Sim}(S_1, S_2) = \sum_{i=1}^4 \beta_i \prod_{j=1}^i \text{Sim}_j(S_1, S_2) \quad (4)$$

式中: $\text{Sim}_j(S_1, S_2)$ 为概念的4个不同义原相似度; $\beta_i(1 \leq i \leq 4)$ 为可调节的参数,且有 $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1, \beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$ 。

两个词语 W_1 和 W_2 的相似度取各个概念相似度的最大值,即

$$\text{Sim}(W_1, W_2) = \max_{\substack{i=1, \dots, n \\ j=1, \dots, m}} \text{Sim}(S_{1i}, S_{2j}) \quad (5)$$

式中 S_{1i}, S_{2j} 为词语的概念。

(2) 词语关联度计算

词语关联度^[15]为义原与解释义原之间的横向关系度量。解释义原由若干义原组成,义原之间有12种关系:RelaSet={*, @, ?, !, ~, #, MYM, %, ^, &, +, null},对应权值:{0.7, 0.6, 0.7, 0.4, 0.75,

$0.9, 0.7, 0.5, -1, 0.8, 0.9, 0.5\}$ 。义原关联度计算公式为

$$\text{Rele}(P_1, P_2) = \sum_{p_i \in \text{exp}(P_1)} W_i \times \text{Sim}(p_i, P_2) / n + \sum_{p_j \in \text{exp}(P_2)} W_j \times \text{Sim}(P_1, p_j) / m \quad (6)$$

式中: $\text{exp}(P_1)$ 和 $\text{exp}(P_2)$ 分别为义原 P_1 和 P_2 解释义原集合; W_i, W_j 为义原关系对应权值; $\text{Sim}(p_i, P_2)$ 和 $\text{Sim}(P_1, p_j)$ 代表义原语义距离; n 和 m 为义原 P_1 和 P_2 解释义原个数。

概念 S_1 和 S_2 的关联度近似为两概念的义原关联度最大值, 即

$$\text{Rele}(S_1, S_2) = \max_{\substack{i=1, \dots, n \\ j=1, \dots, m}} \text{rele}(p_i, p_j) \quad (7)$$

词语关联度取概念关联度最大值, 计算公式为

$$\text{Rele}(W_1, W_2) = \max_{\substack{i=1, \dots, n \\ j=1, \dots, m}} \text{Rele}(S_i, S_j) \quad (8)$$

综合上述计算, 词语相关度的计算公式为

$$R(W_1, W_2) = \gamma_1 \times \text{Sim}(W_1, W_2) + \gamma_2 \times \text{Rele}(W_1, W_2) \quad (9)$$

式中: $\gamma_1 + \gamma_2 = 1$ 。参数设置如下: $\alpha = 1.6, \beta_1 = 0.5, \beta_2 = 0.2, \beta_3 = 0.17, \beta_4 = 0.13, \gamma_1 = 0.23, \gamma_2 = 0.77$ 。

2 实验结果及分析

选取舆情系统收集到的 300 篇文章作为实验语料, 手工标注这 300 篇语料的指代消解结果。标注格式为“她 158 叶云芬”, 其中, “她”为文章中待消解的人称代词, 158 为该代词在文中的位置, “叶云芬”为指代消解的标准答案。测试主要分为两个方面: 一是比较改进的方法与原来方法的效率, 检测改进方法的可靠性; 另一方面是测试本方法对不同类型人称代词的指代效果。本文采用 MUC-6^[16]中所定义的评测方法。指代消解性能评测计算方法为

$$F - \text{Score} = \frac{2PR}{P + R} \quad (10)$$

式中: $P = M/N$ 为准确率, $R = M/C$ 为召回率, M 为正确消解的代词个数, N 为实际消解代词个数, C 为应该消解的代词个数。

表 5 为不同方法的测试结果, 包括单纯的知网语义计算方法(SC)、加入改进的同义词词林相似距离计算的方法(ISC)以及加入同位语等语义规则(AR)的方法。

表 5 各种方法的实验结果

Tab. 5 Results of various methods

方法	准确率 P	召回率 R	F 值	时间/ms
SC	0.743	0.736	0.739	93 423
ISC	0.774	0.768	0.771	96 232
SC+AR	0.762	0.739	0.750	94 192
ISC+AR	0.813	0.790	0.801	100 046

通过不同的方法实验发现, 在语义计算的基础上加入细化的同义词相似距离计算, 比单独的语义计算效果好, 而在语义计算之前增加规则可以进一步提高消解效果。综合评测, 将规则和改进的语义计算方法相结合取得了较好的指代结果。在时间的消耗上, 由于算法的执行时间是算法复杂度的重要指标, 因此, 加入细化的同义词相似距离的算法要比单独的语义计算消耗时间多, 而规则的加入也同时增加了算法执行时间。利用本文所提方法对 300 篇语料中的不同人称代词测试效果如表 6 所示。通过实验, 可以看到第一人称和第二人称的准确率高, 这是因为语料中的第一和第二人称在引用中出现的次数较

多,并且“自己”作为第一人称代词多以同位语的形式出现在语料中。第三人称的指代结果有了一定的提高,相比第一第二人称情况复杂,因此准确率稍低。特殊指代在语料中出现较少并且都是以同位语形式出现。综合分析实验结果,影响实验结果的因素有分词系统中人名识别错误、对于出现多个人名的复数指代准确度不高以及对于前面代词指代结果的错误继承等。

表 6 指代消解实验结果

Tab. 6 Results of coreference resolution

称谓	应该消解	实际消解	正确消解	准确率(P)	召回率(R)	F 值
第一人称	847	816	695	0.852	0.821	0.836
第二人称	183	169	147	0.870	0.803	0.835
第三人称	1 250	1 214	920	0.758	0.736	0.747
特殊	7	6	5	0.833	0.714	0.769
总计	2 287	2 205	1 767	0.801	0.773	0.787

3 结束语

指代消解是自然语言处理中一个重要的部分,对于提高文本的可读行、分析的准确性都有不可忽视的作用。本文主要采用了规则和语义结合的指代消解方法,通过添加代词的同位语规则、精确同义词相似距离计算,提高了指代消解系统的性能。实验结果表明改进的方法是有效的,但是第三人称代词的消解结果还有待进一步提高。在之后的研究中将会尝试采用不同的语义相似度和语义相关度计算方法,研究第三人称代词出现的语法特征等,以期提高第三人称代词的消解效率。

参考文献:

- [1] Dagan I, Itai A. Automatic processing of large corpora for the resolution of anaphora references[C]//Proceedings of 13th Conference on Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 1990:330-332.
- [2] 周炫余,刘娟,卢笑. 篇章中指代消解研究综述[J]. 武汉大学学报:理学版,2014,60(1):24-36.
Zhou Xuanyu, Liu Juan, Lu Xiao. Intra-document anaphora resolution: A survey[J]. Journal of Wuhan University: Natural Science Edition, 2014,60(1):24-36.
- [3] Ngai G, Wang C S. A knowledge-based approach for unsupervised Chinese coreference resolution[J]. Computational Linguistics and Chinese Language Processing, 2007,12(4):459-484.
- [4] 陈恩红,邱思语,许畅,等. 单词嵌入——自然语言的连续空间表示[J]. 数据采集与处理,2014,29(1):19-29.
Chen Enhong, Qiu Siyu, Xu Chang, et al. Word embedding: Continuous space representation for natural language[J]. Journal of Data Acquisition and Processing, 2014,29(1):19-29.
- [5] Lee H, Chang A, Peirsman Y, et al. Deterministic coreference resolution based on entity-centric, precision-ranked rules[J]. Computational Linguistics, 2013,39(4):885-916.
- [6] 邸鹏,段利国. 一种新型朴素贝叶斯文本分类算法[J]. 数据采集与处理,2014,29(1):71-75.
Di Peng, Duan Ligu. New native Bayes text classification algorithm[J]. Journal of Data Acquisition and Processing, 2014, 29(1):71-75.
- [7] McCarthy J F, Lehnert W G. Using decision trees for coreference resolution[C]//Proceedings of the 14th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufman Publishers Inc, 1995:1050-1055.
- [8] Cardie C, Wagstaff K. Noun phrase coreference as clustering[C]//Joint Conference on Empirical Methods in Nlp & Very Large Corpora. Maryland: University of Maryland, 2002:277-308.
- [9] 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.
Wang Houfeng. Survey: Computational models and technologies in anaphora resolution[J]. Journal of Chinese Information

Processing, 2002,16(6):9-17.

- [10] 李国臣, 罗云飞. 采用优先选择策略的中文人称代词的指代消解[J]. 中文信息学报, 2005,19(4):24-30.
Li Guochen, Luo Yunfei. Chinese pronominal anaphora resolution via a preference selection approach[J]. Journal of Chinese Information Processing, 2005,19(4):24-30.
- [11] 葛斌, 李芳芳, 郭丝路, 等. 基于知网的词汇语义相似度计算方法研究[J]. 计算机应用研究, 2010,27(9):3329-3333.
Ge Bin, Li Fangfang, Guo Silu, et al. Word's semantic similarity computation method based on HowNet[J]. Application Research of Computers, 2010,27(9):3329-3333.
- [12] 车万翔, 刘挺, 秦兵, 等. 基于改进编辑距离的中文相似句子检索[J]. 高技术通讯, 2004,14(7):15-19.
Che Wanxiang, Liu Ting, Qin Bing, et al. Similar Chinese sentence retrieval based on improved edit-distance[J]. High Technology Letters, 2004,14(7):15-19.
- [13] 李峰, 李芳. 中文词语语义相似度计算——基于《知网》2000[J]. 中文信息学报, 2007,21(3):99-105.
Li Feng, Li Fang. An new approach measuring semantic similarity in HowNet 2000[J]. Journal of Chinese Information Processing, 2007,21(3):99-105.
- [14] 李生琦, 田巧燕, 汤承. 基于《知网》词汇语义相关度计算的消歧方法[J]. 情报学报, 2009,8(5):706-711.
Li Shengqi, Tian Qiaoyan, Tang Cheng. Disambiguating method for computing relevancy based on HowNet semantic knowledge[J]. Journal of the China Society for Scientific and Technical Information, 2009,28(5):706-711.
- [15] 王小林, 王义. 改进的基于知网的词语相似度算法[J]. 计算机应用, 2011,31(11):3075-3077.
Wang Xiaolin, Wang Yi. Improved word similarity algorithm based on HowNet[J]. Journal of Computer Applications, 2011,31(11):3075-3077.
- [16] Vilain M, Burger J, Aberdeen J, et al. A model-theoretic coreference scoring scheme[C]//Proceedings of the 6th Conference on Message Understanding. San Francisco: Morgan Kaufmann, 1995:45-52.

作者简介:



张文艳(1989-),女,硕士研究生,研究方向:人工智能,
E-mail: wenwen112@126.com.



李存华(1963-),男,教授,研究方向:数据挖掘、人工智能、图像处理。



仲兆满(1977-),男,副教授,研究方向:信息检索、文本信息挖掘、事件本体。



王艺(1992-),女,本科,研究方向:人工智能。



李莉(1992-),女,本科,研究方向:人工智能。

