

# 基于上下文翻译的有监督词义消歧研究

杨陟卓

(山西大学计算机科学与技术学院 太原 030006)

**摘要** 针对目前有监督词义消歧方法存在的数据稀疏问题,提出一种基于上下文翻译的词义消歧方法。该方法假设由歧义词上下文的译文所组成的语境与原上下文语境所表述的意义相似。根据此假设,首先,将译文所组成的上下文生成大量的伪训练语料;然后,利用真实训练语料和伪训练语料训练一个贝叶斯消歧模型;最后,利用该消歧模型决策歧义词的词义。实验结果表明,与传统消歧方法相比,所提出的方法消歧准确率提高了4.35%,并且超过了参加SemEval-2007测评的最好的有监督消歧系统。

**关键词** 词义消歧,上下文扩充,机器翻译,伪训练语料,贝叶斯模型

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.04.053

## Supervised WSD Method Based on Context Translation

YANG Zhi-zhuo

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

**Abstract** In order to overcome the data sparseness problem for supervised WSD methods, this paper presented a WSD method based on context translation. The method assumes that the context consisted of the ambiguous words has the similar meaning as the context in the original. Under this assumption, first, a large number of pseudo training data are generated in the context of the target text. Then the Bayesian model is trained by utilizing both authentic and pseudo training data. Finally, the method performs word sense disambiguation by using Bayesian model. Experimental results show that the proposed method can significantly improve traditional WSD accuracy by 4.35%, and outperforms the best participating system in the SemEval-2007 evaluation.

**Keywords** Data sparseness, Context expansion, Machine translation, Pseudo training data, Bayesian model

## 1 引言

词义消歧是确定多义词在给定上下文语境中的意义,它是自然语言处理领域中重要的研究课题之一。相关研究表明,词义消歧在机器翻译、信息检索、文本分析、自动文摘、知识挖掘等多个方面均具有十分重要的作用<sup>[1]</sup>。

目前,基于语料库的词义消歧方法可分为有监督方法和无监督方法<sup>[2]</sup>两类。无监督方法无需训练语料,但该方法的消歧效果却不尽如人意,很难达到实用目的。有监督方法的消歧效果要明显优于无监督方法,但该方法需要大规模高质量的训练语料支持,而获取大规模高质量的训练语料费时费力,严重阻碍了有监督词义消歧方法的大规模应用<sup>[3]</sup>。为了解决此问题,很多学者开始研究自动生成有标注语料的方法<sup>[4]</sup>。该方法提出了一种假设:歧义词词义可以用语义相关的词汇描述。该方法通常先利用字典和大规模无标注的语料库自动生成有标注数据,然后采用有监督方法训练消歧模型,进行消歧。

最早提出该方法的是 Leacock<sup>[5]</sup>,他先利用 WordNet 查

找歧义词中的单义词词汇,并建立单义词词库;随即在大规模语料库中查找包含这些单义词的语句,然后用这些语句中的单义词替换掉歧义词,这些语句即可看作有标记数据训练的有监督模型。

Mihalcea 将歧义词各词义的单义、同义词及注释作为查询片段<sup>[6-7]</sup>,通过搜索引擎查找语句片段。Agirre 也采用相同的方法<sup>[8]</sup>,利用搜索引擎生成有标注实例。随后,Brody 提出了基于字典的方法和基于语料库的方法,用于查找歧义词不同词义的相关词<sup>[9]</sup>。实验比较了两种方法的消歧准确率,结果表明:基于语料库的方法的消歧结果优于基于字典的方法。通常,双语语料库中包含了许多消歧知识。Brown 提出了一种基于双语语料库的词义消歧方法<sup>[10]</sup>,该方法采用目标词标注合适的译词进行词义消歧。首先,将双语语料进行词对齐,根据上下文信息为目标词选择最合适的译文,生成有标注数据;然后,利用有标注数据训练有监督模型。

近年来,在国内的研究中,何径舟提出基于特征选择和最大熵的词义消歧模型<sup>[11]</sup>,实验结果表明,不同的词义消歧组合特征对不同的歧义词具有不同的消歧效果;鹿文鹏先后提

出基于依存适配度<sup>[12]</sup>和领域知识<sup>[13]</sup>的词义消歧模型;杨陟卓先后提出基于词语距离<sup>[14]</sup>和基于异构关系网络图<sup>[15]</sup>的词义消歧模型。由以上研究可知,语料库中隐藏着大量的词义消歧知识,如果能够挖掘并有效利用这些知识,将能进一步提升词义消歧效果。

本文提出一种自动生成有标注语料的方法,思想来源于基于双语语料库的方法,但是与前人方法的不同之处在于:本文方法利用小规模有标记语料库,从另一个角度生成有标注语料,扩充有标记语料句子中歧义词附近的上下文,通过机器翻译系统将扩充的译文生成有标注语料库。相比之前的方法,该方法可以有效克服知识获取瓶颈问题,生成较高质量的有标记语料库。另外,应当指出的是,这种方法与以往的词义消歧方法并不矛盾,可以同时应用在词义消歧任务中。

本文首先阐述贝叶斯模型;然后重点介绍基于上下文翻译的词义消歧方法,包括提出的假设、上下文翻译原理以及词义消歧方法;随后通过一个国际测评语料集测试本文的方法,并且给出实验结果分析;最后对本文的工作进行总结,并提出下一步的工作计划。

## 2 贝叶斯模型

本文采用贝叶斯模型进行消歧训练<sup>[16]</sup>,如式(1)所示:

$$p(s_i | w_{-L} \dots w_b \dots w_L) \propto p(s_i) \prod_{f_j \in F} p(f_j | s_i) \quad (1)$$

其中,  $w_b$  表示歧义词,  $f_j$  表示歧义词的某个上下文特征,  $F$  表示上下文的特征集合。这些特征包括词法层面特征、语法层面特征以及语义层面特征。由于本文的目标并不是考察各个特征对词义消歧的影响,而是侧重考察扩充语料的消歧效果,因此只选用了一些最常用的特征<sup>[11]</sup>,这些特征如表 1 所列。

表 1 词义消歧特征模板

特征类型	Feature template	Description
词语和词性	$W_i (-L \leq i \leq L, i \neq 0)$	Words in the window
	$P_i (-L \leq i \leq L, i \neq 0)$	Pos for each word in window
组合特征	$W_i W_{i+1} (-L \leq i \leq L-1)$	2-gram words in the window
	$P_i P_{i+1} (-L \leq i \leq L-1)$	2-gram pos in the window
主题特征	$W_i P_i (-L \leq i \leq L, i \neq 0)$	Word and its pos in window
	$W_i (-5 \leq i \leq 5, i \neq 0)$	Topic words in the window

$p(f_j | s_i)$ 表示词义与特征的条件概率,利用式(2)进行估计:

$$p(f_j | s_i) = \frac{c(f_j, s_i)}{c(s_i)} \quad (2)$$

其中,  $c(s_i)$ 表示词义  $s_i$  在语料库中出现的次数,  $c(f_j, s_i)$ 表示特征  $f_j$  与词义  $s_i$  在训练语料中的共现次数。

## 3 基于上下文翻译的词义消歧方法

机器翻译,是指利用计算机将一种自然语言转化成另一种自然语言的过程。近年来,随着统计机器翻译技术的快速发展,译文质量也逐渐得到提高。目前,国内外著名的机器翻译系统有谷歌翻译、百度翻译、微软翻译、有道翻译等。本文采用机器翻译系统输出译文,扩展一定大小窗口下歧义词所处上下文,为词义消歧提供更多的知识。在以往词义消歧研究中,最常用的假设是:歧义词在一段上下文中通常表达相似的词义<sup>[3]</sup>。进一步将该假设推广,可以得出,歧义词在相似上下文中表达的词义相似。

基于以上假设,提出本文的假设:由上下文的译文所组成的语言片段与原上下文语言片段表达的意义相似;同时,出现在这两种相似上下文语言片段中的歧义词表达的意义相同。例如,语句“稳妥地推进中医医疗体制改革”,其中“中医”是歧义词,作为名词它在知网中的词义有两个,分别是“医学”和“医生”。从歧义词所处的上下文推测,歧义词“中医”的词义是“医学”。将该语句进行机器翻译,其译文扩展上下文词语的过程如图 1 所示。

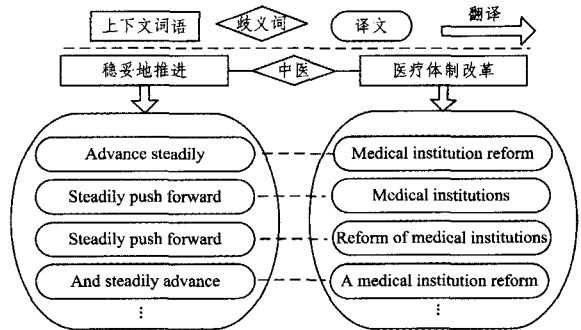


图 1 基于上下文翻译的词义消歧方法示例

已有相关研究表明<sup>[11,14-15]</sup>,歧义词附近的词语对歧义词的词义影响是最大的,因此在图中只列出了附近上下文词语“稳妥地”、“推进”、“医疗体制”和“改革”,并且只对这几个上下文词语进行机器翻译。为了综合不同机器翻译系统的优点,采用 4 种机器翻译系统(谷歌翻译、百度翻译、微软翻译和有道翻译)分别对歧义句进行翻译,利用这些译文又可以组合成多个包含歧义词的语句。例如“Steadily push forward Medical institution reform”,“Advance steadily Medical institution reform”,“And steadily push forward Reform of medical institutions”等。在所有由译文组合而成的语言片段中,歧义词“中医”的词义仍然表示“医生”。不难看出,译文的引入可以为词义消歧提供额外的消歧知识,如果能利用这些译文训练词义消歧模型,将会进一步提升词义消歧效果。

主题词语特征是指歧义词附近一定大小窗口内的词语。本文提出的第一种方法,将歧义词上下文的译文当作主题词语特征(Topic feature),并加入贝叶斯模型进行训练。如果训练语料中的词语在歧义词附近出现一次,就认为相应词语的译文在训练语料中出现一次。训练好的词义消歧模型在对测试语料进行词义决策时,首先对新的实例进行翻译,如果新的实例包含上下文的译文,则根据训练语料中的词义决策新实例歧义词词义。例如,在测试语料中歧义词“中医”附近的上下文包含“Medical institution reform”或“Steadily push forward”时,词义消歧模型就很可能将该实例中的歧义词“中医”的词义推断为“医生”。这是最直观的一种将译文加入到训练模型的方法。但是,应注意,这种方法存在以下缺点:真实的训练实例与由译文组建的训练实例,对歧义词决策的影响力度应当是不同的。真实的训练实例是经过人工标注的,具有较高的质量,在消歧时应当发挥较大的作用;而由译文构建的训练实例是机器翻译系统自动生成的,并没有经过人工标注,它与人工语料相比,还有一定的噪声,因此应当发挥相对小的作用。

本文将由上下文词语扩充的译文所组成的语言片段当作伪训练实例,综合利用真实训练实例和伪训练实例进行词义消歧。伪训练实例可以弥补真实训练实例知识不足的缺点,在采用贝叶斯式(1)进行词义决策时,通过两种类型的语料共同决策词义,在估计歧义词与特征的条件概率时,采用式(3)进行计算:

$$p(f_j | s_i) = \frac{c_t(f_j, s_i)}{c_t(s_i)} + \lambda \frac{c_p(f_j, s_i)}{c_p(s_i)} \quad (3)$$

其中,  $c_t(f_j, s_i)$  表示词义  $s_i$  与特征  $f_j$  在真实训练语料中的共现次数;  $c_t(s_i)$  表示词义  $s_i$  在训练语料中出现的次数;  $c_p(f_j, s_i)$  表示特征与歧义词在伪训练语料中共现的次数;  $c_p(s_i)$  表示词义  $s_i$  在伪训练语料中出现的次数。  $\lambda$  的取值范围为  $0 < \lambda < 1$ , 用于调整伪训练语料对歧义词词义影响的作用, 如果重视伪训练语料的决策作用, 则  $\lambda$  取较大的值; 反之, 取较小的值。可以看出, 在新消歧模型中, 伪训练实例对歧义词词义的决策作用总是小于真实的训练实例。

本文所提方法的总体流程图如图 2 所示。该算法分为 5 步: ①采用机器翻译系统将训练实例的上下文进行翻译, 生成伪训练实例; ②同时采用训练实例和伪训练实例构建贝叶斯消歧模型; ③采用机器翻译系统输出测试实例的上下文译文; ④将测试实例和上下文译文输入有监督模型; ⑤根据式(3), 综合利用训练实例和伪训练实例来决策歧义词的词义。

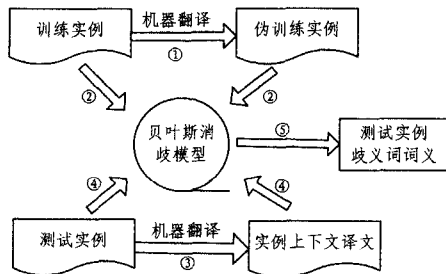


图 2 基于上下文翻译的词义消歧算法流程图

在对上下文词语进行译文扩展时, 遇到一个问题: 并不是上下文词语的所有译文都适合用来扩充而用于构建新的训练语料。例如图 1 中的译文中包含一些噪声“*A*”和“*And*”等, 这些噪声并不适合加入训练语料。

为了解决这个问题, 本文采用投票的方法去除机器翻译系统的噪声。具体的做法是, 利用多个翻译系统同时对歧义词的上下文进行翻译, 统计翻译出的译文的出现次数  $threshold\_cooc$ , 只利用具有一定共现次数的译文训练消歧模型。通过译文次数的限制, 可以很大程度地解决机器翻译系统的噪声问题。共现次数阈值  $threshold\_cooc$  的选取将在实验中说明。

## 4 实验

### 4.1 测试语料评价标准与方法介绍

利用 ACL2007 的一个组成部分即 SemEval-2007<sup>[17]</sup> 国际语义评测的中英文词汇任务 (task # 5 multilingual Chinese English lexical sample task) 对本文方法进行评估。该任务共含 40 个歧义词, 语料由训练语料和测试语料两个部分组成, 如表 2 所列。同时, 采用 SemEval-2007 提供的标准评测工具

及相应评价指标 (macro average accuracy) 来进行评价, 计算方法为:

$$p_i = \frac{m_i}{n_i}, p_{avr} = \frac{\sum_{i=1}^N p_i}{N} \quad (4)$$

其中,  $N$  为所有的目标词数,  $m_i$  是对每一个特定的词所标注的正确的例句数,  $n_i$  是对该特定词所有的测试例句数。

表 2 训练语料与测试语料

	歧义词数目	训练实例数目	测试实例数目
名词 (19)	2.45	1019	364
动词 (21)	3.57	1667	571

实验比较了 2 个 baseline 方法和本文的方法。2 个 baseline 方法如下:

- 1) Original, 即原始的贝叶斯词义消歧模型, 该模型利用式(1)和式(2)进行词义决策, 并利用表 1 的特征进行训练;
- 2) SRCP\_WSD<sup>[18]</sup>, 该系统是参加 SemEval-2007 评测的有监督系统, 获得了 SemEval-2007 Task # 5 评测第一名 ( $p_{avr} = 74.9\%$ )。

本文所提出的方法如下:

- 1) Method\_1, 即本文提出的方法 1, 该方法在原始的贝叶斯词义消歧模型基础上加入了译文的特征, 并将译文当作主题特征加入模型进行训练;
- 2) Method\_2, 该本文提出的方法 2, 该方法同样利用贝叶斯模型进行词义消歧, 但是采用式(3)计算词义条件下的特征概率, 综合利用真实训练语料和伪训练语料训练消歧模型。

## 4.2 实验结果

### 4.2.1 各种方法实验结果的比较

从表 3 数据可以看出, 消歧效果最好的是本文提出的 Method\_2 方法, 其次是 SRCP\_WSD 和 Method\_1, 最差的是传统的贝叶斯方法。本文提出的方法 2 不仅超过了传统的贝叶斯方法, 而且超过了参加 SemEval-2007 测评的最好的有监督方法, 该结果说明译文的引入确实可以有效提高词义消歧的准确率。

表 3 各方法实验结果的比较

	Original	Method_1	SRCP_WSD	Method_2
平均准确率	0.7280	0.7447	0.7490	0.7597
提高百分比	4.35	2.01	1.42	0

为了更加客观地考察各种方法对各个歧义词的消歧准确率, 将 Original, Method\_1 和 Method\_2 这 3 种方法对 40 个词的消歧精度结果整理在表 4 和表 5 中。其中, 表 4 为 19 个名词的结果, 表 5 为 21 个动词的结果, 两表的最后一行是 Method\_1 和 Method\_2 方法对 Original 方法名词和动词性能分别提升的百分比, 表中粗体表示该词消歧性能的最好结果。从表中可以看出, 虽然在一少部分词语的消歧准确率上, 传统的贝叶斯模型好于本文所提出的方法, 但是在大多数词语上本文所提方法的消歧效果要优于传统的贝叶斯模型, 尤其是方法 2 在名词和动词上的消歧准确率分别提高了 1.8% 和 5.07%。

表 4 各种方法名词消歧实验结果( $p_{mar}$ )

名词	词义数	Original	Method_1	Method_2
本	3	0.840	0.840	0.840
表面	2	0.778	<b>0.833</b>	<b>0.833</b>
菜	2	0.752	<b>0.842</b>	<b>0.842</b>
长城	3	<b>0.904</b>	0.857	<b>0.904</b>
单位	2	0.823	0.823	0.823
道	2	0.722	0.722	0.732
队伍	3	<b>0.500</b>	0.454	0.484
儿女	2	<b>0.850</b>	0.800	<b>0.850</b>
机组	2	1.000	1.000	1.000
镜头	2	0.666	<b>0.866</b>	<b>0.866</b>
面	3	0.782	<b>0.826</b>	<b>0.826</b>
牌子	2	0.764	<b>0.823</b>	0.705
旗帜	3	0.666	0.666	0.666
气息	2	<b>0.857</b>	0.714	0.785
气象	2	0.875	0.875	0.875
日子	3	0.468	0.468	<b>0.531</b>
天地	3	<b>0.900</b>	0.840	0.880
眼光	2	0.571	<b>0.642</b>	<b>0.642</b>
中医	2	<b>1.000</b>	0.812	0.937
平均准确率( $p_{mar}$ )		0.774	0.774	<b>0.788</b>
提高百分比/%		1.800	1.800	0.000

表 5 各种方法动词消歧实验结果( $p_{mar}$ )

动词	词义数	Original	Method_1	Method_2
补	3	0.700	0.700	0.700
成立	3	0.814	0.814	0.814
吃	4	0.652	0.608	<b>0.739</b>
出	9	0.584	0.597	<b>0.610</b>
带	8	0.507	0.492	<b>0.537</b>
动	4	0.400	0.450	<b>0.500</b>
动摇	2	0.562	0.562	<b>0.612</b>
发	5	0.722	0.722	0.722
赶	3	0.777	0.777	0.777
叫	4	0.589	0.589	<b>0.615</b>
进	5	<b>0.750</b>	0.727	0.734
开通	2	0.600	0.600	0.600
看	4	0.647	<b>0.676</b>	0.647
平息	2	0.875	<b>1.000</b>	<b>1.000</b>
使	2	0.812	<b>0.875</b>	0.812
说明	2	0.666	<b>0.833</b>	<b>0.833</b>
挑	2	0.643	0.642	<b>0.714</b>
推翻	2	0.800	<b>0.900</b>	<b>0.900</b>
望	2	<b>0.769</b>	<b>0.769</b>	0.615
想	4	0.730	<b>0.810</b>	0.783
震惊	2	0.928	0.928	<b>1.000</b>
平均准确率( $p_{mar}$ )		0.690	0.718	<b>0.725</b>
提高百分比/%		5.070	0.970	0.000

4.2.2 窗口大小对消歧效果的影响

窗口大小对词义消歧效果的影响如图 3 所示。

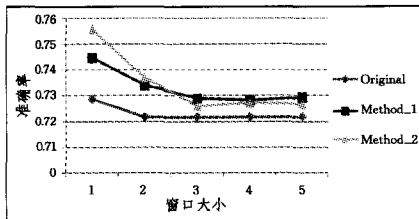


图 3 窗口大小对消歧效果的影响

从图 3 中可以看出,在 SemEval-2007 国际评测语料上,所有的方法均在窗口大小为 1 时取得了最佳的消歧效果,增加窗口大小只会引入较多的噪声。因此,本文实验中,对于所有的歧义词,窗口大小都统一设置为 1。

4.3 词语共现次数阈值  $threshold\_cooc$  对消歧效果的影响

词语共现次数阈值  $threshold\_cooc$  对消歧效果的影响结果如图 4 所示。可以看出,随着共现次数逐渐增大,本文的两种方法消歧准确率不断上升,并且在词语共现次数为 2 时词义消歧效果最佳。由此可见,加入模型中训练的译文都是不同翻译系统输出的相同词语。继续增大词语共现次数阈值,准确率逐渐下降,并最终趋于稳定。

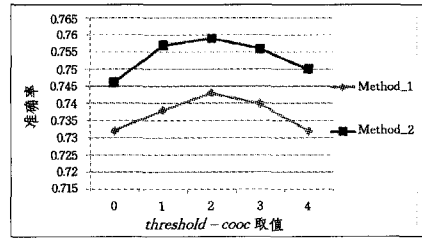


图 4  $threshold\_cooc$  取值不同时方法 1 和方法 2 的消歧性能

实验结果表明,当词语的共现次数阈值设置得比较小时,会将一些噪音译文加入消歧模型中,影响消歧效果;而当共现次数阈值过大时,由于缺乏消歧知识,不足以训练出一个最佳的消歧模型。

4.3.1 方法 2 中  $\lambda$  的取值对消歧效果的影响

方法 2 中  $\lambda$  的取值对消歧效果的影响如图 5 所示。从实验数据可知, $\lambda$  取 0 或 1 时(分别代表伪训练语料不发挥作用或发挥较大的作用)都不能达到最好的消歧效果。一方面,说明伪训练语料可以提供一定的消歧知识,帮助词义消歧系统获得更好的消歧性能;另一方面,伪训练语料也存在一定的噪声,当它与真实训练语料发挥相同的作用时,不利于消歧系统获得最佳的性能。

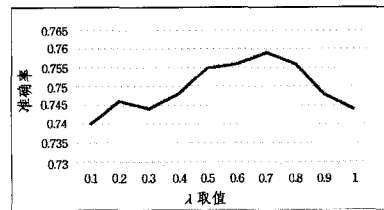


图 5  $\lambda$  取值不同时方法的消歧性能

**结束语** 针对目前有监督词义消歧方法所面临的数据稀疏问题,提出一种基于上下文翻译的有监督词义消歧方法。该方法假设:由上下文词语的译文所组成的上下文语境与原上下文所表达的词义相似,并且出现在这两种语境中的歧义词词义相似。根据该假设,本课题利用词语的译文扩充歧义词所处的上下文。提出两种基于上下文翻译的有监督词义消歧方法,第一种方法将译文当作主题特征加入到模型中进行训练;第二种方法将由译文所组成的上下文当作伪训练语料,并利用真实训练语料和伪训练语料共同决策歧义词的词义。实验结果表明,与传统的消歧模型相比,方法 2 消歧准确率提高了 4.35%,并且超过了参加 SemEval-2007 测评的最好的有监督消歧系统。

本文提出的基于上下文翻译的有监督词义消歧模型是在贝叶斯模型基础上进行实验的,以后将在其他有监督词义消

- [12] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications [J]. *Neurocomputing*, 2006, 70(1-3): 489-501.
- [13] HUANG G B, WANG D H, LAN Y. Extreme learning machines: a survey [J]. *International Journal of Machine Learning & Cybernetics*, 2011, 2(2): 107-122.
- [14] XU Y G, DENG W K, CHEN L D. Online fault diagnosis in wastewater treatment process by kernel-based weighted extreme learning machine [J]. *CIESC Journal*, 2016, 67(9): 3817-3825. (in Chinese)  
许玉格, 邓文凯, 陈立定. 基于核函数的加权极限学习机污水处理在线故障诊断 [J]. *化工学报*, 2016, 67(9): 3817-3825.
- [15] YU B J, ZHU Y L. Transformer fault diagnosis using weighted extreme learning machine [J]. *Computer Engineering & Design*, 2013, 34(12): 4340-4344. (in Chinese)  
遇炳杰, 朱永利. 加权极限学习机在变压器故障诊断中的应用 [J]. *计算机工程与设计*, 2013, 34(12): 4340-4344.
- [16] GAO X, CHEN Z, TANG S, et al. Adaptive weighted imbalance learning with application to abnormal activity recognition [J]. *Neurocomputing*, 2015, 173: 1927-1935.
- [17] YANG Z P. Neural Network Based Classification Methods for Imbalanced Datasets [D]. Shanghai: South China University of Technology, 2015. (in Chinese)  
杨泽平. 基于神经网络的不平衡数据分类方法研究 [D]. 上海: 华东理工大学, 2015.
- [18] DASKALAKI S, KOPANAS I, AVOURIS N. Evaluation of classifiers for an uneven class distribution problem [J]. *Applied Artificial Intelligence*, 2006, 20(5): 381-417.
- [19] KUBAT M, MATWIN S. Addressing the curse of imbalanced training sets: one-sided selection [C] // *Proceedings of the Fourteenth International Conference on Machine Learning*. 2000: 179-186

(上接第 255 页)

歧模型上来验证本文所提方法的有效性, 并且测试其他模型的词义消歧性能, 持续改善有监督词义消歧模型的性能, 力争达到实用的目的。

### 参 考 文 献

- [1] CHAN Y S, NG H T. Scaling up word sense disambiguation via parallel texts [C] // *AAAI*. 2005: 1037-1042.
- [2] WANG R Q, KONG F S. Research on unsupervised word sense disambiguation [J]. *Journal of Software*, 2009, 20(8): 2138-2152. (in Chinese)  
王瑞琴, 孔繁胜. 无监督词义消歧研究 [J]. *软件学报*, 2009, 20(8): 2138-2152.
- [3] NAVIGLI R. Word sense disambiguation: A survey [J]. *ACM Computing Surveys (CSUR)*, 2009, 41(2): 1-69.
- [4] YAROWSKY D. Unsupervised word sense disambiguation rivaling supervised methods [C] // *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 1995: 189-196.
- [5] LEACOCK C, MILLER G A, CHODOROW M. Using corpus statistics and WordNet relations for sense identification [J]. *Computational Linguistics*, 1998, 24(1): 147-165.
- [6] MIHALCEA R, MOLDOVAN D I. A method for word sense disambiguation of unrestricted text [C] // *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. Association for Computational Linguistics, 1999: 152-158.
- [7] MIHALCEA R, MOLDOVAN D I. An automatic method for generating sense tagged corpora [C] // *AAAI/IAAI*. 1999: 461-466.
- [8] AGIRRE E, MARTINEZ D. Unsupervised WSD based on Automatically Retrieved Examples: The Importance of Bias [C] // *EMNLP*. 2004: 25-32.
- [9] BRODY S, LAPATA M. Good neighbors make good senses: Exploiting distributional similarity for unsupervised WSD [C] // *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2008: 65-72.
- [10] BROWN P F, PIETRA S A D, Pietra V J D, et al. Word-sense disambiguation using statistical methods [C] // *Proceedings of the 29th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 1991: 264-270.
- [11] HE Q Z, WANG H F. Chinese Word Sense Disambiguation Based on Maximum Entropy Model with Feature Selection [J]. *Journal of Software*, 2010, 21(6): 1287-1295. (in Chinese)  
何径舟, 王厚峰. 基于特征选择和最大熵模型的汉语词义消歧 [J]. *软件学报*, 2010, 21(6): 1287-1295.
- [12] LU W P, HUANG H Y. Word sense disambiguation based on dependency fitness with automatic knowledge acquisition [J]. *Journal of Software*, 2013, 24(10): 2300-2311. (in Chinese)  
鹿文鹏, 黄河燕. 基于依存适配度的知识自动获取词义消歧方法 [J]. *软件学报*, 2013, 24(10): 2300-2311.
- [13] LU W P, HUANG H Y, WU H. Word Sense Disambiguation with Graph Model Based on Domain Knowledge [J]. *Acta Automatica Sinica*, 2014, 40(12): 2836-2850. (in Chinese)  
鹿文鹏, 黄河燕, 吴昊. 基于领域知识的图模型词义消歧方法 [J]. *自动化学报*, 2014, 40(12): 2836-2850.
- [14] YANG Z Z, HUANG H Y. Graph Based Word Sense Disambiguation Method Using Distance Between Words [J]. *Journal of Software*, 2012, 23(4): 776-785. (in Chinese)  
杨绍卓, 黄河燕. 基于词语距离的网络图词义消歧 [J]. *软件学报*, 2012, 23(4): 776-785.
- [15] YANG Z Z, HUANG H Y. WSD Method Based on Heterogeneous Relation Graph Network [J]. *Journal of Computer Research and Development*, 2013, 50(2): 437-444. (in Chinese)  
杨绍卓, 黄河燕. 基于异构关系网络图的词义消歧研究 [J]. *计算机研究与发展*, 2013, 50(2): 437-444.
- [16] ESCUDERO G, MÁRQUEZ L, RIGAU G. Naive Bayes and exemplar-based approaches to word sense disambiguation revisited [OL]. <http://www.cs.opc.edu/~escudero/wsd/00-ecai.pdf>.
- [17] JIN P, WU Y, YU S. Semeval-2007 task 05: Multilingual chinese-english lexical sample [C] // *Proceedings of the 4th International Workshop on Semantic Evaluations*. Association for Computational Linguistics, 2007: 19-23.
- [18] XING Y. SRCB-WSD: Supervised Chinese word sense disambiguation with key features [C] // *Proceedings of the 4th International Workshop on Semantic Evaluations*. Association for Computational Linguistics, 2007: 300-303.