

以“的”字结构为核心的最长名词短语识别研究

钱小飞

QIAN Xiao-fei

中国传媒大学 文学院 北京 100024

School of Chinese Language and Literature, Communication University of China, Beijing 100024, China

E-mail: qierflying@163.com

QIAN Xiao-fei. Recognition of MNP with “De-Phrase” core. Computer Engineering and Applications 2010 46(18): 138-141.

Abstract: The MNP with “De-Phrase” core is a special subclass of MNP. The identification of the phrase in this paper gives a new subsumption to the task of Chinese MNP recognition. The paper first analyzes the distribution and the structure feature of the phrase, then it advances a strategy of “Identify the right boundary first, then identify the left one”. Furthermore, it adopts the method “Boundary Distribution Probability” to recognize the phrase. A corpus (about 0.85 million Chinese Characters) of news is used for the automatic identification training and another (about 0.42 million Chinese Characters) is used for test and the experiment achieves 80.63% in precision and 75.68% in recall.

Key words: Maximal Noun Phrase (MNP); De-Phrase; identification; shallow parsing

摘要: 以“的”字结构为核心的最长名词短语是汉语最长名词短语的一个特殊子类。以该短语的自动识别为基础重新分化了汉语MNP的识别任务。在考察其结构和分布特征的基础上,提出“先识别右边界,识别成果参与左边界识别”的策略,并使用边界分布概率模型分治了左右边界。实验基于85万字的新闻语料上进行训练,并在42万字的同质语料上进行了开放测试,取得了80.63%的正确率和75.68%的召回率。

关键词: 最长名词短语; “的”字结构; 识别; 浅层句法分析

DOI: 10.3778/j.issn.1002-8331.2010.18.044 文章编号: 1002-8331(2010)18-0138-04 文献标识码: A 中图分类号: TP391

1 引言

最长名词短语(MNP)是指句子中不被其他名词短语直接包含的名词短语。它的识别能够为自动句法分析、机器翻译系统提供有力的支持。如对于句法分析而言,最长名词短语内部结构比较复杂,几乎涉及了汉语所有的词类和句法结构,如果能够识别出句子中所有的最长名词短语,那么计算机就能较好地构建出句子的句法树或森林。以往的研究表明,最长名词短语识别对于结构长度较为敏感,长度大于等于5的复杂最长名词短语的识别精度一般要比简单最长名词短语低16个百分点,并且复杂定语规定的左边界识别难度远远大于右边界^[1]。

复杂最长名词短语识别精度较低的原因值得进一步考虑,从语法上说,复杂最长名词短语之所以复杂是因为:(1)汉语缺乏形态标记,词类和句法成分不是一一对应的,如动词既能作述语,又能作定语,可造成动词宾语位上的最长名词短语左边界模糊;(2)语法递归性,如定语可以由复杂句法结构充当。两者的合力使问题更为突出,特别是“的”字短语参与构成的最长名词短语,几乎容纳了所有的词类和句法结构,而这部分最长名词短语的长度也是较大的。

“的”字结构的参与方式有两种,一种是参与最长名词短语修饰语的构成,一种是参与最长名词短语句法核心的构成。该文将研究第二类最长名词短语的识别问题。

2 以“的”字结构为核心的最长名词短语的界定

以“的”字结构为核心的最长名词短语(dePMNP)是“的”字结构参与最长名词短语句法核心的构成的名词最长名词短语。“的”字结构(deP)的定义是界定该短语的关键。

文献[2]将“的”划分为副词性语法单位的后附成分“的₁”,形容词性语法单位的后附成分“的₂”,名词性语法单位的后附成分“的₃”。然而,“的₂”和“的₃”的区分有时并不十分明晰。如“更+A+的”中“的”应该是“的₂”,但有的形式却能出现在主语或宾语的位置上指称事物,如“我要更好的。”文献[3]区分了表述功能的词汇层面和句法层面,在此基础上把“的”看成是词汇层面的“饰词标记”,上述“更好的”则是该短语在句法层面的指称化。

因此“的”字结构理论上可定义为“X+的₃”和“X+的₂”的指称化形式两种形式,其中X是指词或短语。前者如“正在游泳的”,后者如“那红红的”,出现在主语位置上的“更重要的”。技术上参考周强主持的清华大学973树库的标注方法,将“X+的₂”的模式也统一地标注为“的”字短语。

3 分布特征考察

3.1 句法分布

句法分布以短语所在的句法结构作为考察其句法功能的依据,提供了从外部观察短语特征的窗口。文献[4]为实词分类

作者简介: 钱小飞(1981-)男,博士研究生,主要研究领域为计算语言学、汉语语法学。

收稿日期: 2008-12-23 修回日期: 2009-03-13

拟定了一个含 8 种句法结构的考察框架, deMNP 作为一种名词性短语, 句法功能与名词(实词)类似, 因此该框架也被认为基本适用于 deMNP 的句法功能考察; 所不同的是, 由于介宾结构与动宾结构的句法功能并不等同, 如对于是否可以出现的主语的位置上, 动宾结构可以, 而介宾结构一般不可, 并且 dePMNP 更倾向于出现在动词宾语而非介词宾语的位置上, 将述宾结构重新调整为动宾结构和介宾结构, 见表 1。

表 1 句法分布考察框架

结构名称	直接成分 1	直接成分 2	结构名称	直接成分 1	直接成分 2
主谓结构	主语	谓语	定中结构	定语 1	中心语 1
动宾结构	述语 1	宾语 1	定“的”中结构	定语 2	中心语 1
介宾结构	述语 3	宾语 2	状中结构	状语 1	中心语 2
述补结构	述语 2	补语 1	状“地”中结构	状语 2	中心语 2
述“得”补结构	述语 2	补语 2	-	-	-

依据这个标准考察发现, dePMNP 主要出现在主语和宾语 1 的位置上, 而在述补结构、状“地”中结构等结构中的出现概率几乎为零, 见表 2。

表 2 dePMNP 句法分布考察

统计指标	句法位置				
	主语	宾语 1	宾语 2	定语 1	谓语
频次/例	41	55	2	6	1
比例(%)	39.05	52.38	1.90	5.71	0.95

可见, dePMNP 句法上有着较为集中的分布, 因此识别 dePMNP 需要特别关注两个句法位置: (1) 宾语 1, 从词类特征看, 即动词后面的位置; (2) 主语, 这个位置上的结构与宾语 1 中的 dePMNP 有着相联系的特征。

3.2 线性分布

dePMNP 的线性特征是指在一定上下文中, dePMNP 与其他词语或句法成分共现关系, 包括线性邻接关系。在下文的讨论中, 对于言语成分序列“ $c_1 c_2 \dots c_j w_i c_k \dots c_n$ ”, 把 $c_j c_k$ 分别称作 w_i 的一元邻接成分, 把 c_{k+1} 称作 w_i 的二项共现成分, 相应地, 把 $\langle c_{j-1} c_j \rangle$ 、 $\langle c_k c_{k+1} \rangle$ 分别称作 w_i 的二元邻接成分。

从标注了 dePMNP 的训练语料中提取出 dePMNP 的邻接成分及共现成分, 形成各自的成分集合, 进行统计分析, 发掘 dePMNP 的上下文特征, 从而制定合理的识别方案。

3.2.1 上文特征

上文特征主要研究某一语言成分的上文(左部)线性关系, 以期寻找到该成分的边界区分条件。从词语和语义两个角度观察边界词内部的分布比例。在 L_1Set 中共出现 80 个词型和 734 个词例, 词型的频次分布相当集中, 出现频次排在 L_1Set 集合前两位的是“是”、“#”, 其中“是”出现 344 次, 句首标记词“#”出现 248 次, 两者占据了 80% 以上的比例。

从语义上看, L_1Set 的语义频次分布更为集中。充当 dePMNP 边界词的主要有“是非类”、“像如类”以及连接词类、句首标记词等:

“是非”类: 是, 正是, 总是, 说是, 而是, 乃是, 但是, 并非, 非, 凡, 所谓

“像如”类: 像, 如, 正如, 有如, 不如

由此, 在识别 dePMNP 的左边界过程中可以有两种选择: (1) 利用词型参与统计; (2) 利用词型的语义类参与统计。该文选择前者, 因为同一个语义类中的不同词型作为边界时也有分布概率上的差别, 利用词型参与统计有利于分辨这一点。

从边界词的概率分布来看, L_1 位置上的边界词概率分布极不均匀, “是非”类、“像如”类等语义类都有较高的分布水平。但

是, 大多数其他类别的边界词都处于较低分布水平, 这为左边界的识别带来了一定的难度。

以上的分析和讨论建立在能够判别当前位置上存在一个 dePMNP 或不存在 dePMNP 的基础上。然而在语料中, dePMNP 与含“的”字偏正结构的最长名词短语(deSMNP)的形式区分是模糊的, 都以 de 为标志。如何区分 dePMNP 与 deSMNP, 希望在右边界处找到答案, 因为 dePMNP 与 deSMNP 右边界必然是不一致的, 识别出 dePMNP 的右边界 de, 也就等于判定了当前位置存在一个 dePMNP。

3.2.2 下文特征

dePMNP 的下文特征的分布规律与上文有所不同, 不仅表现在邻接关系中, 在与局部下文的共现关系上也有所体现。前者如“是”等动词经常出现在 dePMNP 下文一元邻接词位置 R_1 上; 后者如“是”之前有时会加上一些副词, 表达程度的不同或委婉的语气等, 使得“是”后退到下文二项共现词位置 RC_2 上, 如“最令人喜爱的还是”。造成这种情况的有两种因素: (1) dePMNP 常位于主语位置上, R_1 位置上也常常出现动词; (2) 汉语中存在一条语法规则: 动词可以受副词等语言成分修饰。下面就从下文一元邻接词集合 R_1Set 、二项共现词集合 RC_2Set 及位于前的修饰词三个方面讨论 dePMNP 的下文特征。

R_1 作为下文特征的典型分布位置, 可以在熵值中得到验证, R_1 上的信息熵为 3.25, 远小于 RC_2 上的 6.30。 R_1 上的这种确定性同样只能解释为受到某种句法因素的影响, 其中包含了更多的特征项。出现频次排在 R_1Set 集合前三位的是“。”、“;”, “是”。其中“。”出现 235 次, “;”出现 201 次, “是”出现 118 次, 三者占据了 75% 以上的比例。

从语义上看, 这些特征可以归纳为“是非类”、“有无类”、连接词类以及标点符号等类别。如“是非”类有: 是, 正是, 只是, 还是; “有无”类包括: 有, 还有, 具有, 等。

从边界词的概率分布来看, R_1 位置上的边界词概率整体有着较高的分布水平, 大部分下文边界词的边界概率都在 80% 以上, 相对于 L_1 位置特征更为明显。

相对于 R_1 特征项在 RC_2 上的分布要稀疏得多, 主要表现为一些谓词性的语言成分(词语)。把这些能够同时出现在 R_1 和 RC_2 位置上的边界特征称之为强特征 f , f 主要由动词, 以及形容词充当, 主要有: 是, 有。

以 R_1Set 和 RC_2Set 相减可以发现, 某些副词只能出现 R_1Set 中而绝少出现在 RC_2Set 中, 这些副词包括: 太, 还, 一定, 等等。通过语料验证表明, 这些副词基本都可以进入“dePMNP+副词 D +被修饰词 $W|W \in f$ ”的描写格式。在 W 为非动词, 如形容词的时候, 某些副词如“还, 更”甚至比 W 更能区分 dePMNP 和 deSMNP, 比较下面两个例子:

(1) 我的不错, 他的更好。

(2) 原谅他的不好。

W 同样为“好”, 从语法上说, 虽然“更好”和“不好”是谓词性的, 但由于副词性质的差异, 使得“不好”能够在 deS 中心语的位置上实现指称化, 而“更好”则不能。

dePMNP 的上下文特征分析表明, L_1 和 R_1 位置都有比较明显的特征分布, 而 R_1 位置的确定性更高。集中的特征词词型分布和较高边界词概率水平使人们有理由预测基于词型边界分布概率的方法也能取得较好的效果。

4 识别方法研究

采用统计的方法, 将上下文特征融合到统计模型中去, 通

过“训练-测试”的方式进行识别 dePMNP。相对于规则方法,它能更好地适应不同类型的文本,只要给出训练语料,它就能学习到相应的识别数据,同时也避免了人工规则的主观性和规则维护的困难。

如上文所析,由于 dePMNP 的右边界能够较为有效地分辨 dePMNP 与 deSMNP,提出“先识别右边界,识别成果参与左边界识别”的策略。

4.1 边界分布概率

边界分布是指任意某种边界在语言成分序列中的某一位置上出现或不出现,它服从二项分布。边界分布概率是指 n 个 ($n \in N$) 语言成分之间或前后出现某种边界的概率。它的计算公式如下:

$$P(\text{bound}_i|c_b, \epsilon_1, \epsilon_2, \dots, \epsilon_i, \epsilon_{i+1}, \dots, \epsilon_n, \epsilon_e) = \frac{f(c_b, \epsilon_1, \epsilon_2, \dots, \epsilon_i, \text{bound}, \epsilon_{i+1}, \dots, \epsilon_n, \epsilon_e)}{f(c_b, \epsilon_1, \epsilon_2, \dots, \epsilon_i, \epsilon_{i+1}, \dots, \epsilon_n, \epsilon_e)} \quad (1)$$

其中 $b=0, \epsilon=n+1, i \in [1, n], j \in [1, n], c_b, \epsilon_1, \epsilon_2, \dots, \epsilon_i, \epsilon_{i+1}, \dots, \epsilon_n, \epsilon_e$ 表示特征语言成分序列, ϵ_b, ϵ_e 是虚设的起始符和终结符, f 表示频次。分子表示该序列在第 i 个和第 $i+1$ 个成分之间出现边界的频率,分母表示该序列在语料库中出现的频率。该公式反映了在特征语言成分序列的第 i 个和 $i+1$ 个成分之间出现边界的概率。

4.2 统计模型

以往的研究常常为研究对象的不同方面选择统一的特征建立统计模型,提高了程序的执行效率,但是,从完全句法分析到浅层句法分析再到组块识别,体现的是一种分析的思维方式,当把一个综合的任务分解为子任务时,通常也凸现出子任务的特征。因此由于 dePMNP 上下文特征分布的差异,在特征的选取上必然应该有所差别。同时,特征的选择不仅仅局限于上文分析的主要特征中,其他特征也有一定的贡献,比如词性等。这里选择 <word tag> 序偶作为基本特征单位构造特征,然后向主要的上下文特征作线性回退。

4.2.1 基本模型

dePMNP 识别遵循“先识别右边界,在识别左边界”的策略,将识别的右边界作为 dePMNP 存在于当前位置的证明,亦即左边界识别的触发特征。正如上文所分析的,dePMNP 的主要上下文特征同时表现于一元邻接词和二项邻接词集合中,关涉到其后的两个词。因为 dePMNP 右边界必然以 de 终止,识别只需要对当前 de 后是否出现右边界作一个二值判定。对于三元组 $\langle de_1, c_2, \epsilon_3 \rangle$, 设 $L(de_1, \epsilon_2, \epsilon_3)$ 表示 de_1, ϵ_2 之间的位置, $L_c(de_1, \epsilon_2, \epsilon_3)$ 表示位于 de_1, ϵ_2 之间的右边界位置,设定阈值 δ , 右边界识别模型用于确认 de_1 和 c_1c_2 之间出现边界概率是大于 δ 的右边界位置:

$$L_c(de_1, \epsilon_2, \epsilon_3) = \arg \max_{L(de_1, \epsilon_2, \epsilon_3)} \{P(\text{bound}_i|de_1, \epsilon_2, \epsilon_3) > \delta\} \quad (2)$$

其中 ϵ_i 表示序偶 <word_i tag>, f 表示共现频率, $P(\text{bound}_i|de_1, \epsilon_2, \epsilon_3)$ 表示在前面出现 de 的情况下, ϵ_2, ϵ_3 之前出现 dePMNP 右边界的概率。

dePMNP 的主要上文特征表现于一元邻接词集合中,关涉到其前的一个词。由于 dePMNP 可能的左边界位置是不确定的,识别表现为多个可能边界位置的竞争。对于非邻接二元组 $\langle c_1, de_r \rangle$, 设 $L(c_1, de_r)$ 表示 c_1 的右邻接位置, $L_c(c_1, de_r)$ 为 dePMNP 左边界。左边界识别模型用于寻找 c_1 和 de_r 之间出现左边界概率最高的位置:

$$L(c_1, de_r) = \arg \max_{L(c_1, de_r)} P(\text{bound}_i|c_1, de_r) \quad (3)$$

其中 de_r 表示已经确认的 dePMNP 右边界, $P(\text{bound}_i|c_1, de_r)$ 表示在后文出现 dePMNP 右边界的条件下, ϵ_1 之后出现 dePMNP

左边界的概率。

4.2.2 线性回退

真实文本中存在着大量的低频事件,有的反映了语言的规律,有的则是由于语料规模的限制而不出现。因此,采用最大似然估计(MLE)难以准确地估计标注词串的概率。线性回退,是一种普通有效的组合模型的方法,常常用来解决统计数据稀疏的问题,它通过历史等价类的线性组合来预测目标特征。其中,历史等价类可以是 n -gram, 语义类等。

在 dePMNP 右边界基本识别模型中,三元组 $\langle de_1, \epsilon_2, \epsilon_3 \rangle$ 被认为是相关特征,其中 de_1 是固定的触发条件,因此假设与二元组 $\langle c_2, \epsilon_3 \rangle$ 相关的条件概率有以下 9 个,其中 P_2 至 P_9 是 P_1 回退后的结果:

$$\begin{matrix} P_1 P(\text{bound}_i|de_1, \epsilon_2, \epsilon_3) & P_4 P(\text{bound}_i|de_1, t_2, w_3) & P_7 P(\text{bound}_i|de_1, t_3) \\ P_2 P(\text{bound}_i|de_1, w_2, w_3) & P_5 P(\text{bound}_i|de_1, t_2, t_3) & P_8 P(\text{bound}_i|de_1, w_2) \\ P_3 P(\text{bound}_i|de_1, w_2, t_3) & P_6 P(\text{bound}_i|de_1, w_3) & P_9 P(\text{bound}_i|de_1, t_2) \end{matrix}$$

令 $k=9$, 公式(2)回退为:

$$P(\text{bound}_i|de_1, \epsilon_2, \epsilon_3) = \sum_{i=1}^k \lambda_i P_i \quad (4)$$

在 dePMNP 左边界基本识别模型中,序偶 $\langle c_1, de_r \rangle$ 被认为是相关特征,其中 de_r 是固定的触发条件,因此假设与 c_1 相关的条件概率有以下 3 个,其中 P_2 至 P_3 是 P_1 回退后的结果:

$$P_{11} P(\text{bound}_i|c_1, de_r), P_{12} P(\text{bound}_i|\text{word}_1, de_r), P_{13} P(\text{bound}_i|\text{tag}_1, de_r)$$

其中 w 表示词语, t 表示词类。令 $k=3$, 公式(3)回退为:

$$L(c_1, de_r) = \arg \max_{L(c_1, de_r)} \sum_{i=1}^k \lambda_i P_i \quad (5)$$

4.3 算法描述

dePMNP 右边界识别算法如下:

```

从测试语料中读入一个句子 S;
建立 S 的词典索引 wordIndex, 计算句子长度 sLen;
基于 wordIndex 建立 de 索引 deIndex, 计算索引长度 deCount
for(k=0 k< deCount k++)
    for(i=0 i<y i++)
        查表 TRi, 记录当前 Pn 至结构变量 A
        由 A 计算 P(boundi|de1, ε2, ε3)
        if(P(boundi|de1, ε2, ε3)>δ)
            在 wordIndex[k]后标注右边界
左边界识别在识别右边界之后进行,并依赖于右边界识别提供的数据,算法如下:
从测试语料中读入一个句子 S
建立 S 的词典索引 wordIndex, 计算句子长度 sLen
建立右边界索引 rightBoundIndex, 计算索引长度 rbCount
for(k=0 k< rbLen k++)
    定位右边界在 wordIndex 中的位置 rbLoc
    由 rbLoc 向左寻找最近的未标注 de 在 wordIndex 中的位置 deLoc
    for(i=deLoc i≥0 i--)
        for(j=0 j<z j++)
            查表 TLj, 记录当前 Pj 至数组 A
            由 A 计算 P(boundi|c1, der), 记录其值和相应 c1 的位置至数组 B
            计算 B 中最大的 P(boundi|c1, der), 记录相应 c1 的位置
            在 c1 后标注左边界
    
```

5 实验结果及分析

5.1 实验数据

在 85 万字的新闻语料上进行训练和封闭测试,并在 42 万

字的同质语料上进行了开放测试。当右边界评价系数 $\lambda = \{0.02, 0.02, 0.3, 0.03, 0.1, 0.005, 0.02, 0.5, 0.005\}$, 右边界评价系数 $\lambda = \{0.049, 0.95, 0.001\}$, 阈值 $\delta = 0.4$ 时, 实验数据如表 3。

表 3 dePMNP 实验数据

测试类型	左边界		右边界		结构					
	P(%)	R(%)	P(%)	R(%)	T	C	I	P(%)	R(%)	F(%)
封闭测试	90.77	91.27	98.92	99.45	733	669	737	90.77	91.27	91.02
开放测试	80.63	75.68	99.21	93.12	407	308	382	80.63	75.68	78.08

表 3 中, T 表示结构总数, C 表示识别正确数, I 表示识别数, P 表示精确率, R 表示召回率, F 表示调和平均值。

5.2 结果分析

考察了标准标注文本和机器标注文本的自动比对结果, 发现识别错误主要在于以下几个方面:

(1) 语料加工方式。在语料预处理的过程中, 采用了小句切分的方法, 即除顿号、连接符等少数几种标点符号外, 其他的标点都作为切分标志, 标注程序基于小句进行处理。但是, 一部分跨越小句的 dePMNP 使得左边界的实际位置处于程序的标注的有效范围之外:

可以/v 说/v 一定/b 程度/n 上/f 是/v 产品/n 结构/n 障碍/n /w

【生产/vn 和/c 市场/n 衔接/vn 障碍/n 等/u 造成/v 的/De】。/w

(2) 数据稀疏。dePMNP 左右边界的识别主要依赖于左邻接或右邻接词的贡献, 词型信息相对于词性信息更加稀疏, 一定数量的未登录词造成了部分识别错误:

【例如/v 把/p 资本主义/n 使用/v 过/u 的/De】/w

(3) 数据竞争。语法跟概率并不是一回事。概率模型只是将选择了出现概率最高的事件, 而非考虑到语法因素的存在。因此合理的数据竞争并不一定造成正确标注结果:

(上接 129 页)

结果的差异。表中的差距是指其他各类工件与第 4 类工件的综合距离之差。

表 3 距离测度法与三角模糊数法融合结果比较

工件	1	2	3	4	5
该文	0.666 0	0.751 9	0.705 8	0.327 5	0.774 1
差距	0.338 5	0.424 4	0.378 3		0.446 6
文[10]	0.115 3	0.182 7	0.111 9	0.090 4	0.170 7
差距	0.024 9	0.092 3	0.021 5		0.080 3

从表 3 可知, 距离测度法对工件类别 1、2、3、5 的差距均分别大于三角模糊数法的相应差距。很明显, 差距越大, 说明对目标识别的区分程度越高。因此上述分析在一定程度上表明, 提出的距离测度融合法优于三角模糊数法。

4 结束语

针对具有多个特征属性的多目标识别问题, 提出了一种新的距离测度, 并运用于传感器信息融合中。通过求解数学优化模型确定指标的权重向量, 客观准确地衡量了各特征指标在信息融合中的重要性程度, 提高了目标识别结果的客观性和区分程度。算法简单, 便于计算机实时实现, 非常适用于具有多个特征的多目标识别领域。

参考文献:

[1] Chen Tian-lu, Que Pei-wen. Object recognition based on modified

但是/c 【《/w 水浒传/nz 》/w 这些/r 人/n 会/v 武功/n 的/De 不/d 多/a】/w

(4) 语料词切分和词性标注错误和不一致的影响:

【派出所/n 所长/n 陈/nr 书生/nr 一/m 语/Ng 破/v 的/De】

6 结语

从语言学的角度, 重新分化了 MNP 的识别任务, 并以 dePMNP 作为研究课题, 提出“先识别右边界, 识别成果参与左边界识别”的研究策略, 使用边界分布概率模型分治了 dePMNP 左右边界, 取得了较好的效果。在今后的研究中, 将从以下两个方面进行改进:

(1) 深入研究左边界的句法语义特征, 以进一步提高左边界的识别效果。

(2) 探讨 dePMNP 内部结构分析及外部句法依存关系的判别方法。

参考文献:

- [1] 周强, 孙茂松, 黄昌宁. 汉语最长名词短语的自动识别[J]. 软件学报, 2000, 11(2).
- [2] 朱德熙. 说“的”[J]. 中国语文, 1961(12).
- [3] 郭锐. 表述功能的转化和“的”字的作用[J]. 当代语言学, 2000, 2(1).
- [4] 陈小荷. 从自动句法分析角度看汉语词类问题[J]. 语言教学与研究, 1999(3).
- [5] 冯冲, 陈肇雄. 基于条件随机域的复杂最长名词短语识别[J]. 小型微型计算机系统, 2006, 27(6).
- [6] 王立霞, 孙宏林. 现代汉语介词短语边界识别研究[J]. 中文信息学报, 2005, 19(3).
- [7] 朱德熙. “的”字短语和判断句(上)[J]. 中国语文, 1978(1).
- [8] 朱德熙. “的”字短语和判断句(下)[J]. 中国语文, 1978(2).
- [9] combination rule[J]. Journal of Systems Engineering and Electronics, 2006, 17(2): 279-283.
- [10] Begler P. I. Shafer-dempster reasoning with application to multi-sensor target identification system[J]. IEEE Trans on SMC, 1987, 1(6): 156-178.
- [11] 邓勇, 朱振福, 钟山. 基于证据理论的模糊信息融合及其在目标识别中的应用[J]. 航空学报, 2005, 26(6): 754-758.
- [12] Odeberg H. Fusion sensor information using fuzzy measures[J]. Robotica, 1989, 31: 217-242.
- [13] 陈守煜, 胡吉敏. 可变模糊方法及其在工件识别中的应用[J]. 系统工程与电子技术, 2006, 28(9): 1325-1328.
- [14] Pilar A. Francisco A. Applying distance histograms for robust object recognition[J]. Kybernetes, 2007, 36(1): 42-51.
- [15] 万树平. 机器人非视觉多传感器信息融合的 TOPSIS 法[J]. 计算机工程与应用, 2009, 45(7): 209-211.
- [16] 邵远, 何发昌, 彭健. 一种机器人非视觉多传感器信息融合方法[J]. 电子学报, 1996, 24(8): 94-97.
- [17] 车录锋, 周晓军, 徐志农, 等. 可拓方法在多传感器信息融合工件识别中的应用[J]. 系统工程理论与实践, 2000, 20(8): 91-94.
- [18] Dong Jiu-ying. Triangular fuzzy number method for fuzzy multi-sensor information fusion[C]. 2009 International Conference on Computational Intelligence and Natural Computing(CINC'09), Wuhan, China, June 6-7, 2009: 162-165.
- [19] Luo R. C., Lin M., Scherp P. S. Dynamic multi-sensor data fusion system for intelligent robots[J]. IEEE Journal of Robotics and Automation, 1988, 4(4): 386-396.