

# 基于机器学习与语义知识的动词隐喻识别

赵红艳<sup>1 2</sup> 曲维光<sup>1 2 3</sup> 张 芬<sup>1 2</sup> 周俊生<sup>1 2</sup>

(1. 南京师范大学 计算机科学与技术学院 江苏 南京 210046)

(2. 江苏省信息安全保密技术工程研究中心 江苏 南京 210097)

(3. 南京师范大学 语言信息科技研究中心 江苏 南京 210097)

**[摘要]** 在自然语言中,动词的用法主要包含 3 种:字面、转喻和隐喻。动词的隐喻用法在隐喻的研究中占有重要的地位。动词的隐喻计算包含两个子任务:隐喻识别和隐喻理解。本文采用条件随机场、最大熵模型来识别动词的隐喻搭配,并在此基础上,加入了两种语义信息《同义词词林》的同义词信息和《知网》的语义信息来识别动词的隐喻表达,实验识别正确率可达 94.47%。

**[关键词]** 动词隐喻识别 机器学习 知网 同义词词林

**[中图分类号]** TP391 **[文献标志码]** A **[文章编号]** 1672-1292(2011)03-0059-06

## Chinese Verb Metaphor Recognition Based on Machine Learning and Semantic Knowledge

Zhao Hongyan<sup>1 2</sup> ,Qu Weiguang<sup>1 2 3</sup> ,Zhang Fen<sup>1 2</sup> ,Zhou Junsheng<sup>1 2</sup>

(1. School of Computer Science and Technology ,Nanjing Normal University ,Nanjing 210046 ,China)

(2. Jiangsu Research Center of Information Security and Privacy Technology ,Nanjing 210097 ,China)

(3. Research Center of Language and Informatics ,Nanjing Normal University ,Nanjing 210097 ,China)

**Abstract:** In a natural language ,the expressions of verbs contain three types: literal ,metonymic and metaphoric. Verb metaphor plays an important role in the field of metaphor study. The computing of verb metaphor can be divided into two subtasks: metaphor recognition and metaphor interpretation. This paper uses two models: Conditional Random Fields model and Maximum Entropy model to recognize verb metaphor. And in these basic models ,this paper adds two kinds of semantic knowledge: synonymy information of Tongyici cilin and semantic information of HowNet to recognize verb metaphor. The accuracy of experiment reaches 94.47% .

**Key words:** verb metaphor recognition , machine learning ,HowNet ,tongyici cilin

认知语言学认为隐喻是一种认知现象,在日常生活中随处可见。概念隐喻<sup>[1]</sup>认为,隐喻是人的概念系统中源域到目标域的映射。在自然语言处理中,隐喻计算是一个不可回避的问题。其中,动词隐喻计算在隐喻研究中占有重要的地位。束定芳<sup>[2]</sup>指出动词性隐喻是指话语中使用的动词与逻辑上的主语或者宾语构成冲突所形成的隐喻。Wilks<sup>[3]</sup>最先以违反语义选择限制作为判定隐喻的条件。Fass<sup>[4]</sup>对违反语义选择限制的现象进行了细分,进一步分为转喻、隐喻和异常。他不仅识别出动词的隐喻用法,还识别出除了隐喻用法外的所有非字面用法。Mason<sup>[5]</sup>从大规模的语料库中获取动词的语义选择知识,来识别动词的隐喻用法。贾玉祥<sup>[6]</sup>主要采用基于语义选择限制的动词隐喻识别。唐旭日<sup>[7]</sup>提出了一种基于语义关系模式的隐喻识别模型。在基于机器学习的方法中,Gedigian<sup>[8]</sup>使用最大熵模型判断动词在一个句子中是隐喻用法还是字面用法。王治敏<sup>[9]</sup>采用基于最大熵模型的机器学习方法对形如“ $n + n$ ”(知识的海洋)等名词隐喻进行识别。Wang Zhimin<sup>[10]</sup>利用语法知识库 GKB(Grammatical Knowledge Base of Contemporary Chinese)和 CCD(Chinese Concept Dictionary)来建立源域到目标域的映射,构建汉语名词隐喻知识库,利用知识库,采用基

收稿日期: 2011-05-30.

基金项目: 国家自然科学基金(60773173 61073119)、国家哲学社科基金(10CYY021)、江苏省自然科学基金(BK2010547)、江苏省教育厅自然科学基金(10KJB520009)、江苏省高校社科基金(06SJB71007)。

通讯联系人: 曲维光,博士,教授,研究方向: 自然语言处理. E-mail: wgqu@njnu.edu.cn

于最大熵的机器学习方法识别隐喻。

在基于语义知识的方法中, Eric<sup>[11]</sup>将语义角色标注与类型独立句法分析联合用于隐喻识别计算中。他认为, 语义角色标注可有效地识别概念隐喻, 语义角色标注与当前研究中独立运用句法分析对隐喻识别计算均有自己的优势。Shutova<sup>[12]</sup>通过名词和动词的无监督聚类, 来进行隐喻的自动识别, 该方法只需要一小部分经过人工标注(隐喻/非隐喻)的种子集。在标注中, Shutova<sup>[13]</sup>对源域到目标域的映射进行了统计, 标注了最经常发生隐喻的源域到目标域的映射原则。对名词和动词的自动聚类采用了在高维空间中效果较好的谱聚类算法。其特征的计算, 依赖于一定的句法分析。这种通过聚类的识别方法, 在大规模语料的隐喻自动识别中取得了很好的效果。Shutova<sup>[14]</sup>认为, 隐喻的源域与目标域不是一种意义上的相似, 而是一种句法结构上的相似。

在中文的隐喻计算研究中, 没有相对比较成熟的句法或者语义角色分析工具, 因此本文旨在通过现有的语义知识资源, 应用于动词隐喻识别中。动词隐喻是一种源域不出现, 直接赋予目标域的动作或者行为。

例

(1) 目睹/v 白发苍苍/i 的/b 老头/n 在/c 花坛/n、/w 墙角/n 吃力/a 地/u 弯腰/v 捡拾/v 他们/r 丢  
弃/v 的/b 烟头/n、/w 纸屑/n。/w

(2) 一个/m 墨/n 守/v 传统/n 的/u 民族/n 是/v 没有/v 活力/n 的/u 民族/n 而/c 一个/m 丢  
弃/v 传统/n 的/u 民族/n 是/v 没有/d 希望/v 的/u 民族/n。/w

在现代汉语词典中, 丢弃的解释为:

[丢弃]diu1qi4 扔掉; 抛弃: 虽是旧衣服, 他也舍不得~。

在这两个例子当中, 实际抽取的短语搭配为“丢弃/v 烟头/n”与“丢弃/v 传统/n”。从搭配的短语来看, 在“丢弃/v 烟头/n”的搭配中, 宾语“烟头”为具体概念, “丢弃/v 传统/n”的搭配中, 宾语“传统”为抽象概念。在动词语义的形成过程中, 最先出现的语义为动词的字面义。根据词典解释, 例(1)“丢弃”的宾语为“烟头”, 为字面用法; 例(2)的宾语为“传统”, 为隐喻用法。

在本文中, 通过不同的模型来进行隐喻的识别, 分析了两种模型, 即: 条件随机场、最大熵模型, 并提出了基于语义知识的隐喻识别方法, 给出了基于条件随机场和基于最大熵模型的实验结果以及加入了语义知识后的实验结果, 并对两个实验结果进行了分析与比较。

## 1 机器学习模型

### 1.1 最大熵模型(Maximum Entropy Model)

最大熵模型是基于信息论的模型, 在过去的几年里, 在自然语言处理中得到广泛的应用。王治敏<sup>[15]</sup>利用最大熵模型对名词隐喻的识别进行了实验。该模型通过给定的上下文  $c$  计算出  $m$  的条件概率  $p(m|c)$ , 求出符合  $c$  条件下的  $m$  概率分布。随机变量  $c$  表示当前词是否为隐喻表达, 其中  $m$  有两种取值,  $m = \text{metaphor}$  表示隐喻用法,  $m = \text{nometaphor}$  表示字面用法。本文用的最大熵模型是 Zhang Le 编写的工具包([http://homepages.inf.ed.ac.uk/lzhang10/maxent\\_toolkit.html](http://homepages.inf.ed.ac.uk/lzhang10/maxent_toolkit.html))。本文实验中, 最大熵模型特征采用了搭配词与词性以及当前词词性。

### 1.2 条件随机场(Conditional Random Fields)

条件随机场是由 Lafferty<sup>[16]</sup>第一次提出的, 它是一个在给定输入节点(观察值)条件下, 计算输出节点的条件概率的模型, 其条件概率公式为:

$$P(y|x) = \frac{1}{Z(x)} \exp \left[ \sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, x) + \sum_{i,k} \mu_k g_k(y_i, x) \right] \quad (1)$$

式中  $f_k(y_{i-1}, y_i, x)$  是观察序列  $x$  中在位置  $i$  和  $i-1$  的输出节点的特征,  $g_k(y_i, x)$  是位置为  $i$  的输入节点和输出节点的特征,  $\lambda$  和  $\mu$  是特征函数的权重,  $Z$  是归一化因子:

$$Z(x) = \sum_y \exp \left( \sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, x) + \sum_{i,k} \mu_k g_k(y_i, x) \right) \quad (2)$$

在本实验中采用的条件随机场模型, 使用 TakuKudo 编写的工具包“CRF++ 0.50”(<http://crfpp.sourceforge.net>) 来进行训练和测试。在本实验中, 条件随机场模型的特征模板如表 1 所示。特征模板的每

一行表示一个模板,用来确定一个输出数据的序列 $x$ 为输入的观察法.

表1 条件随机场模型模板

Table 1 The template of CRF model

特征模板	含义
$U1: \%x[-1\ \rho]$	左边第一个词
$U2: \%x[0\ \rho]$	当前词
$U3: \%x[1\ \rho]$	右边第一个词
$U4: \%x[-1\ \rho]/\%x[0\ \rho]$	左边第一个词与当前词
$U5: \%x[0\ \rho]/\%x[1\ \rho]$	当前词与右边第一个词
$U6: \%x[-1\ \rho]/\%x[1\ \rho]$	左边第一个词与右边第一个词
$U7: \%x[-1\ \downarrow]$	左边第一个词词性
$U8: \%x[1\ \downarrow]$	右边第一个词词性
$U12: \%x[0\ \downarrow]$	当前词词性
$U9: \%x[-1\ \downarrow]/\%x[0\ \downarrow]$	左边第一个词词性与当前词词性
$U10: \%x[0\ \downarrow]/\%x[1\ \downarrow]$	当前词词性与右边第一个词词性
$U11: \%x[-1\ \downarrow]/\%x[1\ \downarrow]$	左边第一个词词性与右边第一个词词性
$U13: \%x[-1\ \rho]/\%x[-1\ \downarrow]$	左边第一个词与词性
$U14: \%x[0\ \rho]/\%x[0\ \downarrow]$	当前词与当前词词性
$U15: \%x[1\ \rho]/\%x[1\ \downarrow]$	右边第一个词与词性

### 1.3 基于《同义词词林》和《知网》的隐喻识别算法

#### 1.3.1 《同义词词林》

《同义词词林》是梅家驹等人<sup>[17]</sup>于1983年编纂的一部关于词语的同义词词典,该词典包含了3 925个词群,每个词群包含了词语的同义词或同类词.之后,哈尔滨工业大学信息检索实验室在此基础上进行了扩展,形成了《哈工大信息检索研究室同义词词林扩展版》,它收录的词语接近70 000条,是一部比较详尽的同义词词典.本文主要利用该词典获取测试集中搭配词的同义词集合.

如:在搭配“丢弃/v 垃圾/n”中,通过《哈工大信息检索研究室同义词词林扩展版》,获得“垃圾”的同义词集合为:

Ba08B01 = 废物 废料 废品 破烂 破铜烂铁 杂质 下脚 垃圾 渣滓 废弃物 滓 渣 污物 排泄物 污染源 垃圾堆

#### 1.3.2 《知网》

《知网》是由董振东、董强开发的一部语义知识词典(<http://www.keenage.com>),它使用1 500多个义原来刻画概念.例如,律师这个概念的义原描述为:律师={human|人,#occupation|职位,police|警,#law|律法}.它由4个义原“human|人”、“occupation|职位”、“police|警”、“law|律法”组成.其中“human|人”是第一义原,规定了律师所属的语义类,是律师的直接上位概念.“#”是HowNet描述语言中的符号,表示“与其相关”.

本文采用计算测试集中的搭配词的同义词集合(由扩展版的《同义词词林》获得)与训练集中的搭配词的相似度的方法来判断动词的用法.具体算法参看1.3.3节.在算法中,相似度的计算采用刘群<sup>[18]</sup>的基于《知网》的相似度计算方法.

《知网》中词语的相似度计算,如果两个词语 $W1$ 和 $W2$ ,若 $W1$ 有 $n$ 个义项(概念): $S_{11}, S_{12}, \dots, S_{1n}$ , $W2$ 有 $m$ 个义项(概念): $S_{21}, S_{22}, \dots, S_{2m}$ ,则规定 $W1$ 和 $W2$ 是各个概念相似度的最大值,即

$$\text{Sim}(W1, W2) = \max_{i=1..n, j=1..m} \text{Sim}(S_{1i}, S_{2j}). \quad (3)$$

由于所有的概念都最终归结于用义原来表示,所以义原的相似度计算是概念相似度计算的基础.在《知网》中,所有的义原根据上下位关系构成了一个树状的义原层次体系.本文采用简单的语义距离计算相似度的办法,假设两个义原在这个层次体系中的路径距离为 $d$ ,那么义原之间的语义距离为:

$$\text{Sim}(p_1, p_2) = \frac{\partial}{d + \partial}. \quad (4)$$

式中 $P_1$ 和 $P_2$ 为义原, $d$ 是 $P_1$ 和 $P_2$ 在义原层次体系中的路径长度,是一个正整数. $\partial$ 是一个可调节参数,本文实验中 $\partial = 1.6$ .

概念的相似度记作式(5). 其中  $\text{Sim}_1 \sim \text{Sim}_4$  分别为第一基本义原相似度、其他基本义原相似度、关系义原描述相似度、关系符号描述相似度<sup>[18]</sup>.

$$\text{Sim}(S_1, S_2) = \sum_{i=1}^4 \beta_i \text{Sim}_i(S_1, S_2), \quad (5)$$

式中  $\beta_i (1 \leq i \leq 4)$  是可调节的参数, 且有:  $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$   $\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$ . 本文实验中  $\beta_1 = 0.5$   $\beta_2 = 0.2$   $\beta_3 = 0.17$   $\beta_4 = 0.13$ .

### 1.3.3 基于语义知识的动词隐喻识别算法

在基于《同义词词林扩展版》的同义词信息和《知网》语义知识信息下, 本文提出了一种基于语义知识的动词隐喻识别算法. 算法表述如下:

在测试集中依次读入搭配词, 执行以下操作:

**Step 1** 从测试集中读入搭配词  $i$ , 若读入为空, 结束循环; 否则, 若该搭配词能从训练集搭配词集合  $A$  (或  $B$ ) 中找到 (其中  $A$  表示动词字面搭配的集合,  $B$  表示动词隐喻搭配词的集合), 则词  $i$  为字面用法 (或隐喻用法). 否则, 转 2;

**Step 2** 利用《同义词词林扩展版》获取搭配词  $i$  的同义词集合  $C$ , 若集合  $C$  中的词语能在训练集  $A$  (或  $B$ ) 中找到, 则词  $i$  为字面用法 (或隐喻用法); 否则, 转 3;

**Step 3** 利用《知网》计算集合  $C$  中每个词语与集合  $A$  中词语的相似度, 其中  $C$  与  $A$  集合中词语的相似度的最大值为  $\text{similarity}_1$ ; 计算集合  $C$  中每个词语与集合  $B$  中词语的相似度, 其中  $C$  与  $B$  集合中词语的相似度的最大值为  $\text{similarity}_2$ ; 若  $\text{similarity}_1 > \text{similarity}_2$ , 则搭配词  $i$  为字面用法, 否则为隐喻用法; 完成后, 转 Step1.

## 2 实验

实验使用的是《人民日报》1997 ~ 1999 年已标注的语料. 标注工具是中科院计算语言所张华平、刘群开发的软件 ICTCLAS (<http://ictclas.org>), 语料没有经过人工校对. 从已标注语料中抽取含有词语“培植、抵御、编织、咀嚼、呵护、品尝、萌芽、提炼、丢弃”9 个词语的搭配, 其中 1997 ~ 1998 年的词语作为训练语料, 1999 年的语料作为测试语料. 各词语的字面与隐喻用法搭配数如表 2 所示 (Baseline 值表示测试语料中动词的隐喻用法所占的比例).

表 2 训练语料与测试语料搭配数

词语	隐喻搭配 (测试)	搭配总数 (测试)	隐喻搭配 (训练)	搭配总数 (训练)	Baseline /%
咀嚼	21	28	43	53	75.00
呵护	22	42	30	63	52.38
品尝	13	35	15	58	37.10
萌芽	18	21	24	26	85.71
提炼	21	41	31	61	51.20
丢弃	17	32	19	55	53.13
培植	17	25	32	44	68.00
抵御	23	28	44	49	82.14
编织	17	26	22	45	65.38

实验的结果采用正确率 (accuracy) 来衡量, 其计算公式为:

$$\text{正确率} = \frac{\text{识别正确的结果数}}{\text{测试语料搭配总数}} \quad (6)$$

### 2.1 未加语义知识的动词隐喻识别

实验 1 采用最大熵、条件随机场模型进行训练, 结果如表 3 所示. 从表 3 可以看出, 最大熵模型分类结果最高可达 85.71%, 最低为 52.38%, 都高于 Baseline 值. 条件随机场模型分类结果最高可达 85.71%, 最低为 40.48%. 表中, 词语“呵护”的条件随机场模型识别结果仅为 40.48%, 原因在于训练语料中的隐喻搭配集合中, 搭配词在左的搭配数远远多于搭配词在右的搭配数 (左搭配: 右搭配 = 1:4), 搭配种类不均, 导致分类结果不理想. 其他词语的分类结果均高于其 Baseline 值. 最大熵模型的分类结果略好于条件随机场模型, 平均正确率为 70.37%, 仅比 Baseline 值高出 7%. 由于动词搭配中包含的信息太少, 导致两种模型的分类结果并不理想.

由此可见, 在前后语境信息较少的情况下, 传统模型的分类结果难以令人满意, 因此实验 2 考虑通过加入语义知识来进行改进.

## 2.2 加入语义知识的动词隐喻识别

实验2加入两种语义知识资源《同义词词林扩展版》和《知网》,利用相似度的计算来进行动词用法的判断,算法详见1.3.3节.采用1.3.3所述算法仅用语义知识训练语料.加入语义知识后,再次利用最大熵、条件随机场模型训练语料,其结果如表4所示(H\_T表示使用词典《知网》和《同义词词林扩展版》进行判断,H\_T+MEM表示加入语义信息的最大熵模型,H\_T+CRF表示加入语义信息的条件随机场模型).

表3 两种模型的实验结果

词语	MEM / %	CRF / %	Baseline / %
咀嚼	75.00	75.00	75.00
呵护	52.38	40.48	52.38
品尝	62.87	62.87	37.10
萌芽	85.71	85.71	85.71
提炼	63.40	65.85	51.20
丢弃	59.37	53.13	53.13
培植	72.00	72.00	68.00
抵御	85.71	85.71	82.14
编织	76.92	76.92	65.38
平均正确率	70.37	68.63	63.34

表4 加入语义信息的实验结果

词语	H_T / %	H_T + MEM / %	H_T + CRF / %
咀嚼	89.25	96.43	96.43
呵护	79.06	88.37	88.37
品尝	97.14	97.14	97.14
萌芽	85.71	85.71	100
提炼	95.12	95.12	95.12
丢弃	100	100	100
培植	88.00	96.00	96.00
抵御	92.85	96.42	96.42
编织	69.23	80.76	80.76
平均正确率	88.48	92.88	94.47

从表4可以看出,在加入语义知识后,仅采用1.3.3所述算法,实验结果就得到显著提高,平均正确率可达88.48%.经过分析可见,动词的搭配中,若搭配词的字面类与隐喻类的区分度越高,则算法识别的正确率越高.如词语“提炼”的搭配词集合中,搭配词的字面类是“石油、可卡因、镭、黄金”等具体概念词,而隐喻类是“本质、精神、思想、精华”等抽象概念词,这两类词语类外相似度比较低,概念区分度比较大.因此,在利用《知网》的相似度计算时,同一类中的相似度比较高,不同类中的相似度比较低,再通过《同义词词林扩展版》,尽可能扩大同类搭配词的数目和种类,因此,实验结果得到显著提高.最后,利用加入的语义信息,提高最大熵、条件随机场模型的分类结果,使得分类结果平均正确率最高达94.47%,而条件随机场模型好于最大熵模型.

## 2.3 结果分析与比较

从表3可以看出,未加语义知识前,条件随机场、最大熵模型的实验结果正确率最高只有70.37%.这是因为本文识别的是动词短语中的隐喻搭配,这些短语搭配包含的上下文信息太少,仅含动词与搭配词以及它们的词性,因此实验结果不理想.同时,实验中训练语料的数量和特征的选择也影响到实验的效果.为此,实验2考虑引入一定的语义知识来识别动词短语的隐喻搭配,加入了现有的语义知识资源《同义词词林扩展版》和《知网》,提出了基于语义知识的隐喻识别算法,该算法识别的正确率可达88.48%,比未添加语义信息的以上两个分类模型的最高值70.37%提高了18.11%.通过加入丰富的同义词信息和《知网》的语义信息,增强了分类器的分类能力,两种模型的分类结果最高可达94.47%,比未加语义知识的识别结果至少提高了22%.加入语义知识后,动词搭配的分类结果发生显著提高,实验结果令人满意.综上所述,该方法可以应用在上下文较少的隐喻处理中.

## 3 总结与展望

本文采用了最大熵、条件随机场模型来识别动词搭配的隐喻用法,并提出了基于语义知识资源《哈工大信息检索研究室同义词词林扩展版》和《知网》的隐喻识别算法,最后采用了机器学习与语义知识相结合的方法来识别动词的隐喻用法,实验结果最高可达94.47%,分类结果至少提高了22%.识别结果比较令人满意.

本文实验中,动词短语搭配的语料仅使用了3年的《人民日报》语料,下一步可以考虑将这种识别方法用于大规模的语料中,进一步形成动词短语隐喻知识库.本文实验中识别的是动宾搭配的短语,下一步可以考虑用于其他类型的隐喻识别中.

[参考文献](References)

- [1] Ortony A. Metaphor and Thought[M]. Cambridge: Cambridge University Press ,1993: 202-251.
- [2] 束定芳. 隐喻学研究[M]. 上海: 上海外语教育出版社 2000.  
Shu Dingfang. Studies in Metaphor[M]. Shanghai: Shanghai Foreign Language Education Press 2000. ( in Chinese)
- [3] Wilks Y. A preferential pattern-seeking semantics for natural language inference[J]. Artificial Intelligence ,1975( 6) :53-74.
- [4] Fass D. Met\* : A method for discriminating metonymy and metaphor by computer[J]. Computational Linguistics ,1991 ,17 ( 1) : 49-90.
- [5] Mason Z. CorMet: A computational corpus-based conventional metaphor extraction system[J]. Computational Linguistics 2004 , 30( 1) : 23-44.
- [6] 贾玉祥,俞士汶. 语义选择限制的自动获取及其在隐喻处理中的应用[C]//第四届全国学生计算语言学研讨会( SWCL-2008) . 太原, 2008: 90-96.  
Jia Yuxiang, Yu Shiwen. Automatic acquisition of selectional preference and its application to metaphor processing[C]// The 4nd Student Workshop on Computational Linguistics( SWCL-2008) . Taiyuan ,2008: 90-96. ( in Chinese)
- [7] 唐旭日. 谓词语义计算: 搭配与概念隐喻[D]. 南京: 南京师范大学计算机科学与技术学院 2010.  
Tang Xuri. The computation of predicate sense: collocation and conceptual metaphor[D]. Nanjing: School of Computer Science and Technology ,Nanjing Normal University 2010. ( in Chinese)
- [8] Gedigian M , Bryant J , Narayannan S , et al. Catching metaphors[C]// Proceedings of the 3rd Workshop on Scalable Natural Language Understanding. Stroudsburg ,PA ,2006: 41-48.
- [9] 王治敏. 名词隐喻相似度及推理识别研究[J]. 中文信息学报 2008 22( 3) :37-43.  
Wang Zhimin. A study on metaphorical similarity and metaphorical inference identification[J]. Journal of Chinese Information Processing ,2008 22( 3) :37-43. ( in Chinese)
- [10] Wang Zhimin , Yu Shiwen , Sui Zhifang. The chinese noun metaphors knowledge base and its use in the recognition of metaphors[C]// 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Toronto , 2010: 186-189.
- [11] Eric P S Baumer , James P White , Bill T. Comparing semantic role labeling with typed dependency parsing in computational metaphor identification [C]// Proceedings of the NAACL HLT 2010 Second Workshop on Computational Approaches to Linguistic Creativity. Los Angeles ,2010: 14-22.
- [12] Shutova E , Lin S , Anna K. Metaphor identification using verb and noun clustering[C]// Proceedings of the 23rd International Conference on Computational Linguistics. Beijing ,2010: 1 002-1 010.
- [13] Shutova E ,Teufel S. Metaphor corpus annotated for source-target domain mappings[C]// Proceedings of LREC 2010. Malta , 2010: 3255-3261.
- [14] Shutova E. Models of metaphor in NLP[C]// Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala ,2010: 668-697.
- [15] 王治敏. 名词隐喻的计算研究及识别实验[J]. 语言教学与研究 2008( 2) :68-74.  
Wang Zhimin. The computational research of noun metaphor and recognition experiment[J]. Language Teaching and Linguistic Studies 2008( 2) :68-74. ( in Chinese)
- [16] John Lafferty , Andrew McCallum , Fernando Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]// Proceedings of the 18th ICML. San Francisco: Morgan Kaufmann ,2001: 282-289.
- [17] 梅家驹,竺一鸣,高蕴琦,等. 同义词词林[M]. 上海: 上海辞书出版社, 1993.  
Mei Jiaju ,Zhu Yiming ,Gao Yunqi ,et al. Tongyici Cilin[M]. Shanghai: The Shanghai Lexicographical Press ,1993. ( in Chinese)
- [18] 刘群,李建素. 基于《知网》的词汇相似度计算[C]// 第三届汉语词汇语义学研讨会. 台北 2002: 1-18.  
Liu Qun ,Li Jiansu. Word similarity computing based on HowNet[C]. //The Tenth Chinese Lexical Semantics Workshop. Taipei 2002: 1-18. ( in Chinese)

[责任编辑: 严海琳]