XXXXXX(XXXXX) 第52卷 第1期 2016年1月

Acta Scientiarum Naturalium Universitatis Pekinensis, Vol. 52, No. 1 (Jan. 2016)

doi: 10.13209/j.0479-8023.2016.009

基于**《**现代汉语语义词典**》**的未登录词  
语义预测研究

XXX1,2 XXX1,2,† XXX3 XX3 XXX1,2 XXX1,2

1. XXXXXXXXXXXXXXXX, XX 210023; 2. XXXXXXXXXXXXXXXXX, XX 210023;   
3. XXXXXXXXX, XX 210097; † 通信作者, E-mail: XX@njnu.edu.cn

摘要基于《现代汉语语义词典》, XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX, 最终得到带有未登录词语义义项标注的语料资源。

关键词汉语未登录词; 语义预测; 语义标注; 集成学习

中图分类号TP391

Research on the Sense Guessing of Chinese Unknown Words   
Based on “Semantic Knowledge-base of Modern Chinese”

XXX1,2, XXX 1,2,†, XXX3, XX3, XXX1,2, XXX1,2

1. XXXXXX, XXX, XX 210023; 2. XXX   
XXX & XX, Nanjing 210023; 3. XXXXXX, Nanjing 210097;   
† Corresponding author, E-mail: [gu@njnu.edu.cn](mailto:gu@njnu.edu.cn)

**Abstract** Based on the research issue of sense guessing of Chinese unknown words, XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX Finally, corpus resources with the sense annotation of unknown words were obtained.

**错误！未定义书签。**

国家自然科学基金(61272221, 61472191)、国家社会科学基金(11CYY030, 10CYY021)、江苏省社会科学基金(12YYA002)和江苏省高校自然科学基金(14KJB520022)资助

收稿日期: 20150619; 修回日期: 20150903; 网络出版日期: 20150930

**Key words** Chinese unknown words; sense guessing; semantic annotation; ensemble learning

语义问题一直是自然语言处理领域的研究热点。XXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX, 如机器翻译、信息检索、语义分析、词典编纂等有重要意义。

汉语未登录词语义预测的研究难度较大,XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXX进行未登录词语义预测, 并对2000年《人民日报》语料中的未登录词进行语义预测和标注。

1 相关研究

在对汉语未登录词的语义预测研究中, XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX, 比如Chen等[3]和Lua[14]的研究。

关于XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX方法。

1) 基于XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXX, 还有基于《知网》的模型[1516]。

2) 基于XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXX类别。

3) 基于XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX相似度。

早期的XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX模型, 取得较好的预测效果。

2 语义资源及词典构建

汉语XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX究。

2.1 语义资源介绍

本文未登录词语义预测研究使用的语义资源是《现代汉语语义词典》, XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX。由于名词库的分类较为详细, 因此本文主要研究名词库的词语。

根据XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX, 如表1所示。

2.2 词典构建

利用XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX。本文分别构建3个SKCC词典。

2.2.1 第一级语义类别SKCC1

将XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX2所示。

2.2.2 第二级语义类别SKCC2

语义词典XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

表3 SKCC2的各个语义类别及类别下词语数目

Table 3 Word number under different semantic levels

of SKCC2

|  |  |
| --- | --- |
| 类别 | 词数 |
| 方位 | 210 |
| 心理特征 | 691 |
| 领域 | 725 |
| 相对时间 | 669 |
| 非生物 | 15241 |
| 构件 | 1474 |
| 动机 | 39 |
| 自然现象 | 172 |
| 生理 | 669 |
| 信息 | 757 |
| 绝对时间 | 109 |
| 属性 | 2944 |
| 法规 | 318 |
| 生物 | 8238 |
| 事件 | 1657 |
| 处所 | 2492 |

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 包含属于二级及以下类别的词语, 并且将这些词语都向上划分到第二级父节点上的词语。

第二级语义类别分为16个。SKCC名词库划到

155w052.tif

矩形表示语义类别, 椭圆形表示词语

图1 XXXXXXXXXXXX

Fig. 1 XXXXXXX

表2 XXXXXXXXXXX

Table 2 XXXXXXXX

levels of SKCC1

|  |  |
| --- | --- |
| 类别名 | 词数 |
| 过程 | 1908 |
| 时间 | 906 |
| 抽象事物 | 8643 |
| 空间 | 3195 |
| 具体事物 | 25149 |

第二级的16个语义类别的词语数目如表3所示。

2.2.3 第三级语义类别SKCC3

语XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX, 并且把这些词语都向上划分到第三级父节点上。

第三XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX别中,

表2 XXXXXXXXXXXXXXX

Table 4 XXXXXXX

levels of SKCC3

|  |  |
| --- | --- |
| 类别 | 词数 |
| 自然物 | 2003 |
| 颜色 | 88 |
| 外形 | 307 |
| 身体构件 | 1040 |
| 人 | 5923 |
| 情感 | 102 |
| 意识 | 574 |
| 植物 | 1155 |
| 模糊属性 | 2352 |
| 量化属性 | 448 |
| 微生物 | 76 |
| 可听现象 | 35 |
| 非生物构件 | 0 |
| 排泄物 | 100 |
| 可视现象 | 108 |
| 动物 | 1063 |
| 人工物 | 12811 |

该类别没有词语和子类别。

3 模型构建

根据XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX模型。

3.1 基于重叠字的模型

根据XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX类别。

对于XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXX*X*XXXX

XXXXXXXXXXXXXXXXXX*X*XXXXXXX*X*XXXXX*XX*XXXXXX*X*XX*X*XXXX*XX*XXX*X*XX*XX*XX*X*XXXXXXXXXXXXXXXXXXXXXX*XX*X*XX*XXXXXXXXXX*X*XXXXXXXXX*XX*X*XX*XX*XX*XXXXXXXXXX*XX*X*XX*XXXXX*X*XXXXXXXXX的总数; *Npi*表示在类别Cat中, 位于位置*pi*的字的总数; *Nw*表示在类别Cat中词的 总数。

变式1: XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX得到。

变式1a: ; (1)

变式1b: 。 (2)

变式2: XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX到。

变式2a: ; (3)

变式2b: 。 (4)

变式3: 变式3a中, XXXXXXXXXXXXXXXXXX*XX*XXXXXXXX*XX*XXXXXXXXXXXXXXXXXXXXXXXXXXXXX权得到。

变式3a: ; (5)

变式3b: 。 (6)

变式1XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX。每一个变式, 得分最高的类别被推荐为未登录词的类别。

3.2 基于字****类别关联的模型

字类别XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX*X*2, 如式(7)~(9)所示:

, (7)

, (8)

。 (9)

其中, XXXXXXXXXXXXXX*X*XXXXXXXXXXXXXXXXX*X*XXXXX*X*X*X*XX*X*X*X*XXXXX*X*XXXXX率。

计算XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX示:

, (10)

其中, XXXX*X*XXX*X*XX*X*XXXXXX*X*XXXX*X*XXXXX*XX*XXXXXX*X*XXXX*X*XXXXXX重, *λi*的和为1。

3.3 基于规则的模型

基于XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX”, 如表5所示。

表5 XXXXXXXXXXX

Table 5 XXXXXXX”

|  |  |  |
| --- | --- | --- |
| 词语 | 规则 | 语义类别 |
| 文学家 | A+BC: “文”+“学家” | 具体事物 |
| 神学家 | A+BC: “神”+“学家” | 具体事物 |
| 农学家 | A+BC: “农”+“学家” | 具体事物 |
| 史学家 | A+BC: “史”+“学家” | 具体事物 |
| 医学家 | A+BC: “医”+“学家” | 具体事物 |

4 模型实验

4.1 实验语料与预处理

实验XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX的词语。

4.2 实验与分析

从SKCC1中随机抽取3000个测试词,XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX, 计算语义分类的正确率。

实验**1** 基于XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX到的77.0%。

实验 **2** 基于XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX*X*XXXXX*X*XXXXXXXXXXXXX, 为74.3%。

实验**3** 基于规则模型的未登录词语义预测正确数和正确率。所抽取的未登录词总数为861, 即

表6 基于SKCC1词典的XXXXXXXXX

Table 6 XXXXXXXX

based on SKCC1 dictionary

|  |  |  |
| --- | --- | --- |
| 变式名 | 正确数 | 正确率/% |
| 1a | 2179 | 72.6 |
| 1b | 2134 | 71.1 |
| 2a | 2309 | 77.0 |
| 2b | 1694 | 56.5 |
| 3a | 2293 | 76.4 |
| 3b | 1963 | 65.4 |

表7 基于XXXXXXXXXXX模型预果

Table 7 Prediction results based on SKCC1 dictionary

under word-type models

|  |  |  |
| --- | --- | --- |
| 统计量 | 正确数 | 正确率/% |
| MI | 1702 | 56.7 |
| *χ*2 | 2230 | 74.3 |

表8 基于XXXXXXXXXXXXXXX

语义预测结果

Table 8 Results of unknown words based on SKCC1

dictionary

|  |  |  |  |
| --- | --- | --- | --- |
| 有语义返回词数 | 正确数 | 正确率/% | 召回率/% |
| 861 | 770 | 89.4 | 28.7 |

在XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX低。

实验**4** 多模型的集成。XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX。集成模型的预测语义由以下两条确定。

1) 如果XXXXXXXXXXXXXXXXXX将这个语义作为混合模型预测语义。

2) 如XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX语义。

集成XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX性能。

表9 XXSKCC1XXXXXXXXXXXXX

Table 9 Results of intergration models based on SKCC1

|  |  |  |
| --- | --- | --- |
| 有语义返回词数 | 正确数 | 正确率/% |
| 3000 | 2337 | 77.9 |

5 汉语未登录词语义预测应用

在基于XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX低。XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX, 所得到的语料示例如表10所示。

表10 基于SKCC3标注未登录词语义语料示例

Table 10 Examples of unknown words based on SKCC3

|  |  |
| --- | --- |
| 原始语料 | 标注语料 |
| 20000130-03-009-007/m 正面/b 主景/n 毛/nrf 泽东/nrg 头像/n , /wd 采用/v 手工/d 雕刻/v 凹版/n 印刷/vn 工艺/n , /wd 形象/n 逼真/a 、/wu 传神/a 、/wu 凹凸感/n 强/a ; /wf | 20000130-03-009-007/m 正面/b 主景/n/人工物 毛/nrf 泽东/nrg 头像/n, /wd 采用/v 手工/d 雕刻/v 凹版/n/外形 印刷/vn 工艺/n , /wd 形象/n 逼真/a 、/wu 传神/a 、/wu 凹凸感/n/情感 强/a ; /wf |

表XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX正确。

在XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX进行分析。

假设XXXX*X*XXXXXXXXXXXXXXXXXXXXXXXX*X*XXXXXXXXXXXXXXXXX*X*XXXXXXXXXXXXXXXXXXXXXXXXXXXXX*X*XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX*X*XXXXXXXXXXXXXXXXXXXX*X*XXXXXXXXXXXXXX*X*XXXXXXXXXXXX*X*XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX72.2%。

6 总结与展望

本文XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX中。

参考文献

1. Chen H, Lin C. Sense-tagging Chinese corpus // Proceedings of ACL-2000 Workshop on Chinese Language. Hong Kong, 2000: 7‒14
2. Chen C. Character-sense association and compoun-ding template similarity: automatic semantic classifica-tion of Chinese compounds // Proceedings of the 3rd SIGHAN Workshop on Chinese Language Processing. Barcelona, 2004: 33‒40
3. Chen K, Chen C. Automatic semantic classification for Chinese unknown compound nouns // Proceedings of the 18th International Conference on Computa-tional Linguistics (COLING). Saarbrücken, 2000: 173‒179
4. Lu Xiaofei. Hybrid model for Chinese unknown word resolution [D]. Ohio: The Ohio State University, 2006
5. Lu Xiaofei. Hybrid model for semantic classification of Chinese unknown words // Proceedings of North American Chapter of the Association for Computa-tional Linguistics: Human Language Technologies. Rochester, 2007: 188‒195
6. Tseng H. Semantic classification of Chinese unknown words // Proceedings of the Student Research Work-shop at the 41st Annual Meeting of the Association for Computational Linguistics (ACL). Sapporo, 2003: 72‒79
7. Tseng H, Chen K J. Design of Chinese morphological analyzer // Proceedings of the First SIGHAN Work-shop on Chinese Language Processing. Stroudsburg, 2002: 1‒7
8. Qiu Likun, Wu Yunfang, Shao Yanqiu. Combining contextual and structural information for supersense tagging of Chinese unknown words // Proceedings of CICLing, PartⅠ, LNCS 6608. Tokyo, 2011: 15‒28
9. Qiu Likun, Zhao Kai, Hu Changjian. A hybrid model for sense guessing of Chinese unknown words // Proceedings of 23rd Pacific Asia Conference on Language, Information and Computation (PACLIC). Hong Kong, 2009: 464‒473
10. Cucerzan S. Large-scale named entity disambiguation based on wikipedia data // Procedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Lan-guage Learning. Prague, 2007: 708‒716
11. 周俊生, 戴新宇, 尹存燕, 等. 基于层叠条件随机场模型的中文机构名自动识别. 电子学报, 2006, 34(5): 804‒809
12. 陈钰枫, 宗成庆, 苏克毅. 汉英双语命名实体识别与对齐的交互式方法. 计算机学报, 2011, 34(9): 1688‒1696
13. 冯元勇, 孙乐, 张大鲲, 等. 基于小规模尾字特征的中文命名实体识别研究. 电子学报, 2008, 36(9): 1833‒1837
14. Lua K T. Prediction of meaning of bi-syllabic Chinese compound words using back propagation neural net-work. Computational Processing of Oriental Langua- ges, 1997, 11(2): 133‒144
15. 张瑞霞, 肖汉. 基于《知网》的词图构造. 华北水利水电学院学报, 2008, 29(3): 53‒56
16. 张瑞霞, 杨国增, 闫新庆. 基于《知网》的汉语普通未登录词语义分析模型. 计算机应用与软件, 2012, 29(8): 126‒130
17. 王惠, 詹卫东, 俞士汶. 现代汉语语义词典规格说明书. 汉语语言与计算学报, 2003, 13(2): 159‒176
18. Bai M H, Hsieh Y M, Chen K J, et al. Translating Chinese unknown words by automatically acquired templates // Proceedings of the Sixth International Joint Conference on Natural Language Processing (IJCNLP). Nagoya, 2013: 839‒843