

基于频繁子树模式的评价对象抽取

田卫东, 苗惠君

(合肥工业大学 计算机与信息学院, 合肥 230009)

摘 要: 现有的评价对象抽取方法多基于启发式规则或者基于词性、词形等特征的机器学习方法, 未能较好地利用依存分析所揭示出的深层句法关联关系。为此, 基于从依存关系树库所挖掘的频繁子树模式, 提出一种针对中文评论性短文本的评价对象抽取方法。该方法基于依存关系频繁子树模式进行短文本的初始标注, 采用错误驱动框架的方法提炼出能反映评价对象特征的频繁子树模式有序模式规则集, 并利用该规则集进行评价对象的抽取。实验结果表明, 该方法具有较好的稳定性与准确性, 在召回率和 F1 值等评价指标上优于基于支持向量机的方法。

关键词: 依存句法; 短文本; 频繁子树模式; 错误驱动; 支持向量机

中文引用格式: 田卫东, 苗惠君. 基于频繁子树模式的评价对象抽取[J]. 计算机工程, 2017, 43(4): 222-227.

英文引用格式: Tian Weidong, Miao Huijun. Extraction of Opinion Targets Based on Frequent Sub-tree Pattern[J]. Computer Engineering, 2017, 43(4): 222-227.

Extraction of Opinion Targets Based on Frequent Sub-tree Pattern

TIAN Weidong, MIAO Huijun

(School of Computer and Information, Hefei University of Technology, Hefei 230009, China)

[Abstract] Most existing opinion target extraction methods are based on the heuristic rules or machine learning using features such as part of speech, morphology and etc., but the defect of these methods is that deep association relationship mined by dependency syntax analysis is not used. In order to solve this problem, a novel opinion target extraction method for Chinese short critical texts is proposed based on frequent tree patterns mined from dependency relation tree bank. First, this method labels the initial tagging opinion target based on frequent sub-tree patterns, and then it trains out an ordered rule set based on error-driven TBL framework which can be related to the combination of opinion targets. Finally, opinion target is extracted based on the ordered rule set. Experimental results show that this method has good stability and precision, and is better than Support Vector Machine(SVM)-based method on indicators such as recall and F1-score.

[Key words] dependency syntax; short text; frequent sub-tree pattern; error driven; Support Vector Machine(SVM)

DOI: 10.3969/j.issn.1000-3428.2017.04.038

0 概述

文本的情感挖掘, 即对文本进行挖掘分析以辨别出文本所表达的情感信息。互联网上评论性文本分布广泛, 针对评论性文本的情感挖掘被广泛应用于电子商务、舆情监控等领域。评价对象抽取作为一项基本的情感挖掘任务, 引起学者们的广泛关注, 国内的中文倾向性分析评测 COAE 早在 2008 年就已将评价对象抽取设为一项独立的子任务^[1]。

针对评价对象抽取工作, 目前国内外已经有了诸多研究成果, 抽取方法主要包括基于启发式规则

的方法和基于机器学习的方法。

在启发式规则方面, 文献[2]首先自定义评价对象的模板, 然后采用基于人工规则的方式抽取出评论目标; 文献[3]认为出现频率较高的名词很可能是评价对象, 因此, 利用关联规则进行评价对象的识别; 文献[4]认为评论如果包含中心词就是评价对象; 文献[5]通过评价对象与评价词的共现频率等特征进行识别工作; 文献[6]则结合 Wordnet、统计分析、电影理论等建立知识库, 并基于此提出了一种启发式的算法来抽取产品特征。研究表明, 启发式规则在限定领域内抽取效果较好。

基金项目: 国家“863”计划项目(2012AA011005); 国家自然科学基金(61273292); 情感计算与先进智能机器安徽省重点实验室开放课题(ACAIM2015xxx)。

作者简介: 田卫东(1970—), 男, 副教授, 主研方向为智能计算、数据挖掘; 苗惠君, 硕士研究生。

收稿日期: 2016-03-21 **修回日期:** 2016-04-27 **E-mail:** 582457334@qq.com

在机器学习方面,文献[7]曾尝试利用词语的部分词法、语义角色标注等特征,并结合最大熵模型进行目标对象的抽取;文献[8]则将词性、位置作为特征引入最大熵模型进行评价对象抽取;文献[9]较早提出使用CRF模型来抽取评价对象,将该工作视为一项序列标记任务;之后文献[10]提出在CRF中引入浅层句法和启发式位置这2个特征;文献[11]通过词性、共现频率等特征获得候选的评价对象集,而后使用频率过滤、PMI等方法得到最终的评价对象集。现有的机器学习方法一般是围绕自然语言的句法语法特征进行,而对语言成分之间的关联关系,尤其是依存句法关系,缺乏关注。

本文提出将基于频繁子树模式^[12]的方法应用于评价对象抽取中,从而将汉语依存句法分析所能揭示的深层次关联关系与词的属性信息相结合,以改善评价对象抽取性能,提出了基于转化的频繁子树学习(Frequent Tree Transformation-based Learning, FTTBL)方法。该方法首先利用频繁子树模式进行初始标注,然后基于错误驱动算法框架训练出频繁子树有序模式规则集,再对初始标注进行修正。

1 相关概念

1.1 情感评价对象

在评论性文本中,即使是同一个评价词,与不同的评价对象搭配也可能产生不同的情感倾向,例如“物价涨得很快”和“博尔特跑得真快”这2个评论句,评价词同样为“快”,但显然可以看出,前者的情感倾向是负向,而后者则为正向。因此,情感评价对象的抽取是情感挖掘中的必要步骤。

本文的研究是基于以下3种形式的评价对象进行的:

- 1) 对象的整体,如“这辆越野车真贵呀”中的“越野车”。
- 2) 对象的某个成分或某一部件,如“这个车的底盘太低了”中的“底盘”。
- 3) 对象的属性或附加特征,如“这车的颜色我不喜欢”中的“颜色”。

1.2 依存关系树

依存句法是一种被广泛使用的句法形式,最早在《结构句法基础》中被提出,认为词语之间的关系是有向的,具有支配和被支配关系,并将这种关系称为依存关系。依存句法则是通过分析语言单位(词)之间的依存关系(表示为依存关系树)来揭示句子结构。

依存关系树是一棵树 $T = (F, B)$, 其中, F 为结点集合; B 为边集合。边是有方向的,若边从结点 v

指向结点 w , 且边标注为依存方式 r , 则可记为 $v <_r w$ 。在依存关系树中,边通常从父结点指向其孩子结点。

依存关系树基于句子内部的依存关系所建立,完整地保留了句子的语义信息,能将中文语句所独有的句法结构特征直观地表达出来,因此,可以将本来对句子的处理转化为对其相应的依存关系树的处理。图1给出了依存关系树的例子。

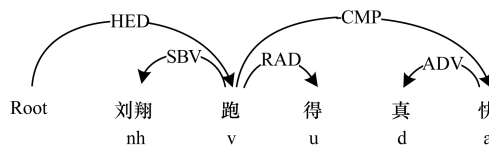


图1 依存关系树示例

2 FTTBL 方法流程

本文提出的FTTBL方法是基于文献[13]提出的基于转化的错误驱动学习(Transformation Based Error-driven Learning, TBL)的算法框架构造。

FTTBL方法主要分为3个阶段:基于频繁子树模式的评论短文本初始标注,基于转换的频繁子树有序模式规则集的训练,基于有序模式规则集对初始标注的修正。

方法的具体流程如下:

- 1) 首先对语料库进行人工标注。
- 2) 采用 PlainFT (Plain Frequent Tree) 方法进行训练语料的初始标注,即使用频繁子树模式标注训练语料。
- 3) 基于初始标注中标注错误的部分,使用预先设好的标注转换模板生成对应频繁子树模式和规则,构造初始模式规则集。
- 4) 对初始模式规则集的所有初始模式规则,分别用其对训练语料进行标注,依据结果评价函数计算得分。
- 5) 选择结果评价函数得分最高的初始模式规则,将其放入有序模式规则。
- 6) 反复执行步骤4)、步骤5),直到某轮迭代中所有规则的结果评价函数得分都不满足条件为止。每一轮迭代都会得到一条模式规则,最后形成有序模式规则集。
- 7) 按照有序模式规则的先后顺序标注测试语料,得到最终标注结果。

本文将对基于频繁子树模式的评价对象初始标注和基于转换的频繁子树有序模式规则集的训练这2个阶段进行详细阐述,有序模式规则集对初始标注的修正阶段因为较为简单,不再赘述。

3 基于频繁子树模式的评价对象初始标注

3.1 频繁子树模式

分析依存树库可以发现,有些依存关系树的子树结构是经常出现的,这些经常出现的子树结构,实际反映了人们在自然语言使用上的习惯和常用表达方式。因此,本文引入了频繁依存子树模式^[14]。

定义 1(子树) 给定树 $T = (F, B)$ 及 $T_1 = (F_1, B_1)$, 若有 $F_1 \subset F$, 对于 $v, w \in F_1, v <_r w \in B_1$, 有 $v <_r w \in B$ 成立, 则称树 T_1 是树 T 的子树。

定义 2(子树的支持度) 给定依存树库 D 以及子树 T , 则树 T 的支持度 $\text{supp}(T) = |p(T)|/|D|$, 其中, $p(T)$ 是指 D 中包含子树 T 的依存树集合, $|p(T)|$ 也称为子树 T 的支持数。

定义 3(频繁子树) 给定支持度阈值 minsupp , 当子树 T 的支持度 $\text{supp}(T) \geq \text{minsupp}$ 时, 称子树 T 为依存树库中的频繁子树。

通常来说句子中的每个词语均有多个维度的属性, 如词性、依存关系、位置等, 在评价对象抽取时需要考虑多个属性的共同影响才能有效的提高抽取准确率, 因此, 本文使用多属性的组合作为子树的结点, 即每个结点都是一个属性集合。本文使用的候选属性集合如表 1 所示。

表 1 属性列表

属性	含义
序号	词的位置编号
词	当前词本身
依存词	当前词所依存的词
词性	当前词的词性
依存词词性	依存词的词性
依存关系	词与依存词之间的关系
标注	是否是评价对象

3.2 评价对象初始标注方法

利用频繁子树进行评价对象的初始标注主要分为 2 个阶段: 频繁子树模式的挖掘和频繁子树规则的生成。

1) 频繁子树模式挖掘阶段

针对所使用的评论性短文本库, 通过子树筛选算法生成候选子树集, 而后计算候选子树的支持数, 挖掘出符合指定支持度阈值的所有频繁子树, 随着子树的结点维度的增加, 实验的时间成本急剧增加, 并且经过实验可以发现, 表 1 中的各属性对标注效果的影响有高低之差, 因此本节只选取“词性”和“依存关系”这 2 个对结果影响最为明显的的属性进行组合。

对图 1 中的句子进行解析, 得到的频繁树结构如图 2 所示。

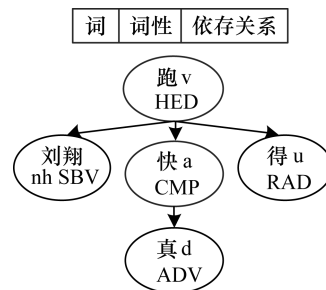


图 2 子树结构示例

2) 频繁子树规则生成阶段

将挖掘出的每一个频繁子树模式都与训练集进行逐句对比。假设 $T_s = (F_s, B_s)$ 表示某频繁子树, $T = (F, B)$ 表示某训练语料生成的依存树, 如果 $T_s \in T$, 且 F_s 中含有评价对象, 则将子树 T_s 作为规则前件, 生成对应的候选规则, 加入候选规则集。全部规则生成完毕后, 将满足置信度要求的候选规则加入最终的频繁子树规则集, 将不满足要求的规则舍弃。频繁子树规则如图 3 所示。

0	vHED	-1
1	ATT	2
2	SBV	3
3	vVV	0

2	f	0.67

图 3 频繁子树规则示例

利用生成的频繁子树规则对测试集进行标注, 以图 3 规则为例, 若句子能匹配到该规则的规则前件, 即子树, 则将子树中序号为 2 的结点标注为 f, f 即是评价对象的标识符。

4 基于转换的有序模式规则集训练

图 4 是基于错误驱动算法框架的有序模式规则集训练过程, 主要分为 2 个阶段: 1) 初始模式规则集挖掘阶段; 2) 有序模式规则集训练阶段。初始模式规则集挖掘阶段利用频繁树模式从初始标注中挖掘初始模式规则, 有序模式规则集训练阶段在初始模式规则集上通过不断迭代的方式进行自学习, 以训练有序模式规则集。

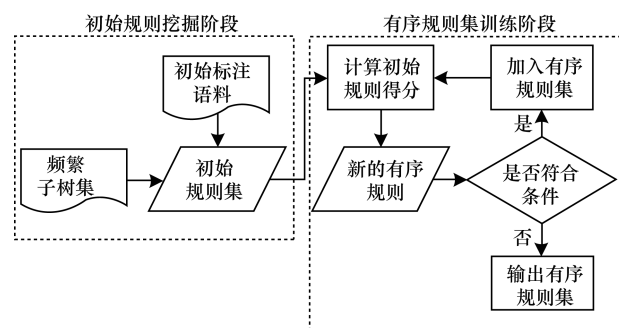


图 4 有序模式规则集训练流程

4.1 初始模式规则集生成阶段

Brill 的错误驱动算法,其初始模式规则的格式是以模板的形式由人工制定的。文献[15]在中心词的标注问题中,根据这种思想制定了规则模板,对每一个错误的初始标注都对应生成 24 条初始模式规则。这种启发式方法以经验为主导,主观性强,可能存在模板覆盖面完备性问题。

本文使用频繁子树模式代替人工模板作为初始模式规则的触发条件,通过频繁模式与其所属类别的强关联性来指导初始模式规则的生成,并在此基础上进行对应转换动作的执行。本节使用的频繁子树模式来源于由训练语料中挖掘出的频繁子树库。

Brill 的错误驱动算法对挖掘出的各初始模式规则并不加以筛选,虽然这样可以保证初始模式规则集的覆盖率,但也使得其规模较为庞大,导致有序模式规则训练阶段的工作效率明显下降。因此,本文引入置信度阈值进行初始模式规则的筛选,只有规则的置信度超过指定阈值时,才能加入初始模式规则集。

本文使用的初始模式规则格式为: *condition index from_tag, to_tag*, 即一条规则的内容包括触发条件 *condition* 和之后的转换动作这 2 个方面。其中,触发条件是一棵频繁子树,即针对初始标注产生错误的语料,分析其所含有的全部频繁子树,得出的每一棵频繁子树都可以作为一条初始模式规则的触发条件。在转换动作方面,本文的转换动作均是对下标为 *index* 的词语进行转换标注,若其原标注可以匹配 *from_tag*,则将其修改为 *to_tag*。转换动作具体分为 2 种:动作 1 为将原本的标注为评价对象的词转换为普通词;动作 2 则是将原本标注为普通的词转换为评价对象。

该过程的伪代码如下所示。

输入 1) *Tree*: 频繁子树集; 2) *Train*: 训练集; 3) *mincf*: 最小置信度阈值

输出 *InitRules*: 初始模式规则集

```
function TblInitRule ( Tree, Train, mincf );
    for I ← 0 to Tree. size() - 1 do
        supp[i] ← 0;
        targetword[i] ← 0;
    for j ← 0 to Tree[i]. size() - 1 do
        targetword[i][j] ← 0;
    for k ← 0 to Train. size() - 1 do
        if TreeEqual( Tree[i][j], Train[k] ) then
            supp[i] + +;
        for m ← 0 to Train[k]. size() - 1 do
            if Train[k][m][7] ≠ Train[k][m][8] then
                targetword[i][j] + +;
    for j ← 0 to targetword[i]. size() - 1 do
        cf ← targetword[i][j] / supp[i];
```

```
if cf < mincf then
    continue;
FillRule( rule, Train, Tree[i], targetword[i][j] );
Initrules ← Initrules ∪ { rule };
return rules;
```

4.2 有序模式规则集训练阶段

针对第一步生成的初始模式规则集,将其每一条规则应用于当前语料库的标注。当某一句语料满足规则的触发条件时,规则成立,就执行对应的转换动作对标注的结果进行修改。修改标注结果时,主要有 2 种情况:

1) 原错误标注被改正为正确标注,称为“错变对”。

2) 原正确标注被修改为错误标注,称为“对变错”。

本文设置 2 个得分指标 $g(r)$ 和 $b(r)$, 前者表示某一初始模式规则 r 的正确得分,即该规则导致“错变对”的句子的个数;而后者则表示 r 所导致的“对变错”的句子个数。

当计算规则 r 的得分时,将其逐句作用于训练集中,计算 $g(r)$ 和 $b(r)$ 的值各是多少。 r 的最后得分就是 $g(r) - b(r)$ 的结果。

依据此方法计算初始模式规则集中每一条规则的分值,从而选出分数最高且满足指定阈值条件的规则作为有序模式规则集的第一条规则,用该规则重新标注语料库,并将该规则从初始模式规则集中删除。然后用新的初始模式规则集重复以上步骤,每重复一次都可以选出一条新的有序模式规则加入有序模式规则集中,直到找不到满足条件的规则为止。

该过程伪代码如下所示。

输入 1) *Initrules*: 初始模式规则集; 2) *Train*: 训练集

输出 *OrderedRules*: 有序模式规则集

```
function OrderedRule ( Initrules, Train );
    while Initrules is not null do
        foreach rule r in Initrules do
            score[i] ← 0;
            for k ← 0 to Train. size() - 1 do
                if RuleSuitable ( r, Train[k] ) then
                    ChangeScore( score[i] );
            for I ← 0 to Initrules. size() - 1 do
                if score[i] is max then
                    maxindex ← i;
            if maxindex >= 0.0001 × Train. size() then
                TaggedCorpus( Initrules[ maxindex ], Train );
                OrderedRules ← OrderedRules ∪ Initrules[ maxindex ];
                Initrules. erase( maxindex );
            else
                break;
    return OrderedRules;
```

5 实验结果与分析

5.1 实验数据集和评价方法

目前文献所涉及的依存句法标注体系并未统一,本文实验使用的是哈尔滨工业大学语言技术云平台(LTP-cloud)的标注体系。同时,本文所使用的文本预处理工具也来自该平台。

实验所用语料分别来源于 COAE2014 任务 4 微博语料以及电商网站京东商城中关于电脑产品的部分评论语料。

本文用准确率(Precision),召回率(Recall)及 F1 值(F1-score)3 个评价指标来评价实验的结果,其中,准确率 $P = \text{正确标注的评价对象数} / \text{全部的评价对象数}$,召回率 $R = \text{正确标注的评价对象数} / \text{标注出的评价对象数}$,F1 值 $F = (2 \times P \times R) / (P + R)$ 。

5.2 结果及分析

为了保证实验结果的客观准确,实验采用 4:1 方式划分训练集与测试集,并随机分割 10 次,得到 10 组内容各不相同的数据集。

为了验证频繁子树的支持数与规则置信度对实验的影响,本文设置了一组对比实验,其中支持数 α' 设置了 1,2,3,5 不同的阈值,同时规则置信度则从 0.01 ~ 0.4 共取 6 个阈值进行对比,如图 5 所示。

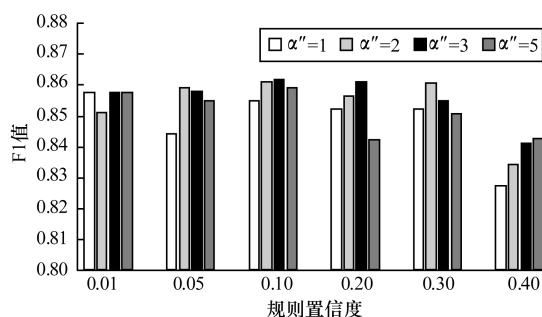


图5 支持数及置信度的影响

可以发现,当支持数阈值设为 3 且频繁子树规则的置信度设为 0.1 时,所得到的初始标注的 F1 值是最高的。虽然这种参数取值不一定是最优的,但为统一起见,本文相关的实验均采用此阈值。

在生成有序模式规则集的过程中,初始模式规则集的置信度设为 0.1,有序模式规则得分阈值设为 2,即如果没有规则的最终得分大于或等于 2,则不再进行有序模式规则的抽取。

在 10 组随机分割成数据集中,使用有序模式规则集对频繁子树方法标注效果的修正效果如图 6 所示。之所以选择 F1 值进行对比,是因为 F1 值兼顾考虑了准确率和召回率这 2 个指标,可以更为平衡地评价各方法标注效果。

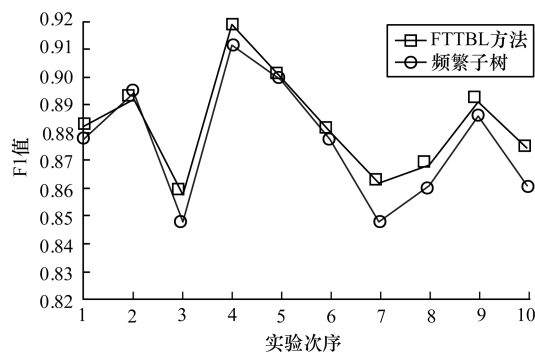


图6 FTTBL 的提升效果

可以发现,FTTBL 方法的效果普遍高于仅使用频繁子树方法的抽取效果,证明了本文基于错误驱动框架训练出的频繁树有序模式规则集在评价对象抽取工作中是有效的。

本文使用基于 SVM 的评价对象抽取方法作为对比实验,使用的工具为林智仁教授团队开发的 LibSVM。但是,为了更为公平地进行实验效果的对比,本文在 SVM 方法中使用与 FTTBL 方法相同的特征集,即“词性”及“依存关系”这 2 个特征的组合。

Plain-FT 方法、FTTBL 方法与 SVM 方法进行对比的结果如表 2 所示。

表2 FTTBL 方法与 SVM 对比 %

方法	准确率	召回率	F1 值
Plain-FT	86.24	86.24	86.24
FTTBL	89.09	87.47	88.27
SVM	92.81	48.56	63.76

从总体上看,本文提出的 FTTBL 方法的实验效果在准确率方面比 SVM 方法有小幅降低,但在召回率和 F1 值方面均明显高于 SVM 方法。

通过各实验结果的对比,可以得出以下结论:虽然 SVM 方法所抽取出的评价对象的准确率较高,但是由于 SVM 方法设置的各特征数据之间是独立无关联的,并不能充分利用中文句子的上下文信息,因此,其所成功标注出的评价对象数量不足,导致 SVM 方法的召回率及 F1 值明显偏低;而本文的 FTTBL 方法则充分利用了句子上下文间的依存关系,能自动抓取与评价对象有关的语言特征及特征之间的组合关系以形成相应的频繁树模式规则,并采用错误驱动的方式来解决规则模式的融合问题,在评价对象抽取问题中,兼顾了准确率和召回率,得到了较为理想的 F1 值。此结果证明了本文提出的方法在评价对象抽取中是有意义的。

6 结束语

评价对象抽取是情感倾向分析的一个关键问题,本文在频繁子树模式的基础上,提出了基于错误驱动框架的评价对象抽取方法。实验结果表明,该方法在中文评论性语料的评价对象抽取工作中可以取得良好的效果。FTTBL方法的实验结果虽然达到预期,但是由于该方法在迭代过程中存在重复计算问题,导致其时间成本较高,因此如何在不牺牲准确率的前提下优化算法的时间效率是下一步研究的重点方向。

参考文献

- [1] 赵军,许洪波,黄萱菁,等.中文倾向性分析评测技术报告[C]//第一届中文倾向性分析评测论文集.北京:中国中文信息学会信息检索专委会,2008:1-20.
- [2] Popescu A M, Etzioni O. Extracting Product Features and Opinions from Reviews[M]//Kao A, Poteet S R. Natural Language Processing and Text Mining. Berlin, Germany: Springer, 2007: 9-28.
- [3] Hu Mingqing, Liu Bing. Mining and Summarizing Customer Reviews[C]//Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2004: 168-177.
- [4] Ma Tengfei, Wan Xiaojun. Opinion Target Extraction in Chinese News Comments [C]//Proceedings of International Conference on Computational Linguistics. Beijing, China: [s. n.], 2010: 782-790.
- [5] 徐叶强,朱艳辉,王文华,等.中文产品评论中评价对象的识别研究[J].计算机工程,2012,38(20):140-143.
- [6] Li Zhuang, Feng Jing, Zhu Xiaoyan. Movie Review Mining and Summarization [C]//Proceedings of the 15th ACM International Conference on Information and Knowledge Management. New York, USA: ACM Press, 2006: 43-50.
- [7] Kim S M, Hovy E. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text[C]//Proceedings of ACL/COLING Workshop on Sentiment and Subjectivity in Text. Sydney, Australia: [s. n.], 2006: 1-8.
- [8] 樊娜,蔡皖东,赵煜.基于最大熵模型的观点句主观关系提取[J].计算机工程,2010,36(2):4-6.
- [9] Jakob N, Gurevych I. Extracting Opinion Targets in a Single-and Cross-domain Setting with Conditional Random Fields [C]//Proceedings of Conference on EMNLP. Washington D. C., USA: [s. n.], 2010: 1035-1045.
- [10] 徐冰,赵铁军,王山雨,等.基于浅层句法特征的评价对象抽取研究[J].自动化学报,2011,37(10):1241-1247.
- [11] 刘鸿宇,赵妍妍,秦兵,等.评价对象抽取及其倾向性分析[J].中文信息学报,2010,24(1):84-88.
- [12] 田卫东,黄勇.频繁子树模式在中心词识别中的应用研究[J].微电子学与计算机,2015(11):27-32.
- [13] Brill E. A Simple Rule-based Part of Speech Tagger[C]//Proceedings of the 3rd Conference on Applied Natural Language Processing. [S. l.]: Association for Computational Linguistics, 2002: 152-155.
- [14] Zaki M J. Efficiently Mining Frequent Trees in a Forest: Algorithms and Applications[J]. IEEE Transactions on Knowledge & Data Engineering, 2005, 17(8): 1021-1035.
- [15] 田卫东,李亚娟.基于CRF和错误驱动的中心词识别[J].计算机应用研究,2013,30(8):2345-2348.

编辑 刘冰

(上接第221页)

- [6] 朱莎莎,刘宗田,付剑锋,等.基于条件随机场的中文时间短语识别[J].计算机工程,2011,37(15):164-167.
- [7] Xu Sun, Takuya M, Daisuke O, et al. Latent Variable Perceptron Algorithm for Structured Classification[C]//Proceedings of the 21st International Joint Conference on Artificial Intelligence. San Francisco, USA: Morgan Kaufmann Publishers Inc., 2009: 1236-1242.
- [8] Xu Sun, Takuya M, Li Wenjie. Latent Structured Perceptrons for Large-scale Learning with Hidden Information[J]. IEEE Transactions on Knowledge Data Engineering, 2013, 25(9): 2063-2075.
- [9] Lafferty J D, McCallum A, Pernaudo F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data [C]//Proceedings of the 18th International Conference on Machine Learning. San Francisco, USA: Morgan Kaufmann Publishers Inc., 2001: 282-289.
- [10] Fei Sha, Fernando P. Shallow Parsing with Conditional Random Fields[C]//Proceedings of 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology. Stroudsburg, USA: Association for Computational Linguistics, 2003: 134-141.
- [11] McCallum A. Efficiently Inducing Features of Conditional Random Fields[C]//Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence. San Francisco, USA: Morgan Kaufmann Publishers Inc., 2003: 403-410.
- [12] Sutton C, Rohanimanesh K, McCallum A. Dynamic Conditional Random Fields: Factorized Probabilistic Models for Labeling and Segmenting Sequence Data [C]//Proceedings of the 21st International Conference on Machine Learning. San Francisco, USA: Morgan Kauffmann Publishers Inc., 2004.
- [13] Sutton C, McCallum A, Rohanimanesh K. Dynamic Conditional Random Fields: Factorized Probabilistic Models for Labeling and Segmentation Sequence[J]. Journal of Machine Learning Research, 2004, 8(3): 693-723.
- [14] 莫怡文,姬东鸿,黄江平.基于条件随机场的顿号边界识别[J].计算机应用,2015,35(10):2838-2842.
- [15] Jacob C. A Coefficient of Agreement for Nominal Scales[J]. Educational and Psychological Measurement, 1960, 20(1): 37-46.

编辑 陆燕菲