

基于二分网中心节点识别的产品评论特征-观点词对提取研究^①

刘 臣, 吉 莉, 唐 莉

(上海理工大学 管理学院, 上海 200093)

通讯作者: 吉 莉, E-mail: 18761800337@163.com

摘 要: 本文以电商平台上的产品评论文本为研究对象, 针对产品评论中特征词和观点词的识别问题进行了研究. 首先构建特征-观点对二分网络, 再给出特征-观点对二分网络中节点重要性排序算法, 最后将此算法应用到实际的评论文本数据中以检验算法的有效性.

关键词: 特征-观点词对提取; 二分网; 中心节点识别; 产品评论

引用格式: 刘臣, 吉莉, 唐莉. 基于二分网中心节点识别的产品评论特征-观点词对提取研究. 计算机系统应用, 2018, 27(11):9-16. <http://www.c-s-a.org.cn/1003-3254/6630.html>

Research on Product Feature-Opinion Extraction Based on Center Node Recognition in Bipartite Network

LIU Chen, JI Li, TANG Li

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: This study takes the product review texts on the e-commerce platform as the mining object, and focuses on the identification of feature words and opinion words in reviews. First, we build bipartite network with feature-opinion words, and give the sorting algorithm of node importance in this network. At last, the algorithm is applied to the actual review text data to verify the effectiveness of the algorithm.

Key words: feature-opinion extraction; bipartite network; center node recognition; product review

1 引言

近年来, 随着互联网技术的迅猛发展, 催生了电子商务这种购物模式. 消费者购买和使用产品之后会在网上发表对产品的评论, 因此电商平台上产生了大量的商品评论文本数据. 这些评论中的观点词是把握消费者情感倾向的关键, 而观点词所修饰的特征词则反映了消费者对于产品关注的焦点. 这些特征观点词不仅影响着消费者的购买意向, 同时也可以作为商家了解竞争对手的一个窗口, 从而提高产品质量, 更好地为消费者服务. 如何从这些海量评论文本中有效地提取

商品特征词和观点词, 更好为消费者跟商家服务, 是意见挖掘领域中的热点问题. 在这些特征词观点词中又有高频词和低频词之分, 高频词更能准确地反应消费者关注产品的焦点, 所以本文重点挖掘出产品评论中高频特征观点词.

近些年有很多学者针对产品特征词观点词提取进行了研究. Zhao 等人^[1]提出 MaxEnt-LDA 为产品特征词和观点词联合建模, 并利用句法特征使两者分离. 但在实际数据中, 却很难识别出评论文本中出现的高频特征词. Hu 等人^[2]利用关联规则算法, 将名词中的频繁

^① 基金项目: 国家自然科学基金 (71401107)

Foundation item: National Natural Science Foundation of China (71401107)

收稿时间: 2018-03-27; 修改时间: 2018-04-24; 采用时间: 2018-05-08; csa 在线出版时间: 2018-09-30

项集提取出来作为候选特征词,再将产品特征词所在句子中的形容词提取出来作为观点词.这种单纯的将名词作为候选特征词的方法,会产生许多不相关的特征词,降低结果的准确率. Popescu 等人^[3]将 Hu 等人的方法做了改进,首先用 PMI 算法将停用词过滤,再通过句法依存关系和特征词来提取观点词.李实等人^[4]基于对关联规则算法的改进对产品评论的特征信息进行挖掘.马柏樟等人^[5]提出基于潜在狄利特雷分布模型的产品特征提取方法. Qiu 等人^[6]、Hai 等人^[7]基于双向传播算法,利用特征观点词之间的修饰关系或依存关系,观点词提取特征词、特征词提取观点词的双向传播模式.实验结果表明,利用双向传播算法提取特征词和观点词的召回率较高,但随着迭代的深入开始出现较多的无关词,导致准确率较低.

孙晓等人^[8]提出了基于条件随机场模型和支持向量机的层叠模型,提取产品评论中的特征词和观点词.刘臣等人^[9]则是将评论中的名词组块作为产品特征,动词组块作为观点词来提取特征观点词.刘通等人^[10]依据 N-Gram 的边界平均信息熵的指标和子串依赖关系对候选项进行过滤并提取特征. Jin 等人^[11]采用 HMMs 模型识别特征观点词.李志义等人^[12]在条件随机场模型 (CRFs) 的基础上,通过分析特征词和观点词之间存在的依存关系抽取特征观点词. Titov 等人^[13]利用多粒度主题模型,提取出按主题自动聚类的特征词和观点词.彭云等人^[14]提出语义关系约束的主题模型 SRC-LDA (Semantic Relation Constrained LDA), 用来提取细粒度特征和情感词. Kamal 等人^[15]对评论文本进行语言学 and 语义分析,利用相关规则实现评论文本的产品特征观点词对的提取.

其他一些学者基于节点排序算法,将特征词和观点词进行重要性排序.例如郝亚辉^[16]将评论中的特征词和观点词间的句法依存关系模式作为 HUB 节点,再利用 HITS 算法对候选特征词和观点词进行排序,提高了特征词和情感词的准确率. Liu 等人^[17]提出了一种协同排序算法来估计每个候选词的可信度,并提取出具有较高可信度的候选词作为候选目标词. Zhang 等人^[18]对特征候选进行特征重要性排序,由特征相关性和特征频率两个因素决定,利用 HITS 算法查找重要特征并将其排序.但这些研究中,都是以等权重的方式处理候选特征词和观点词节点,没有考虑到节点权重的大小对节点重要性排序的影响.

本文将从二分网络的节点重要性排序角度来识别特征观点词,建立特征-观点对二分网络.针对网络是否加权,分为无权网络和加权网络.首先在无权网络中提出了 B-核分解算法, B-核分解算法是将两类节点的度值作为度量值对节点的重要性进行排序.后针对无权网络的缺陷改进算法,提出了 BW-核分解算法. BW-核分解算法则是将节点的权值作为度量值对节点进行重要性的排序.本文从京东上选取了四种产品的评论数据集作为研究对象,评价指标采用目前广泛接受的准确率 (Precision) 和召回率 (Recall)、F 值 (F-measure) 来衡量算法的有效性.

2 特征-观点对二分网络的构建

2.1 特征-观点对二分网络的表示

本文从二分网络的角度来识别高频特征观点词,因此首先构建特征-观点对二分网络.二分网络是由两种类型的节点构成,不同类型的节点之间才有连边.在复杂网络研究中,对于给定的网络 $G = (N, E)$, 如果节点集 N 可以分为两个互不相交的非空子集 X 和 Y , 使得 E 中的每一条边的两个端点中必定有一个属于 X , 另一个属于 Y . 则将 $G = (N, E)$ 称为二分网络 $B = (X, Y, E)$, 其中 $X = \{n_1, n_2, n_3\}$, $Y = \{n_4, n_5, n_6\}$. 在特征-观点对二分网络中,包括特征词和观点词两类节点. X 中的元素表示特征词节点, Y 中的元素表示观点词节点, E 表示连边即特征词和观点词的修饰关系.典型的特征-观点对二分网络如图 1 所示.

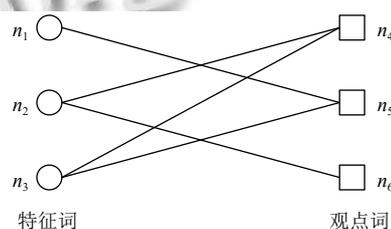


图 1 特征-观点对二分网络图

2.2 特征-观点对二分网络中的度和点权

节点的度是单点网络中常见的基本性质,通常是指与该节点连接的边的数量.在二分网络中,一个节点的度同样也是指与该节点连接的边的数量,且两类节点的度之和相等^[19].在特征-观点对二分网络中,一个特征词的度即为与其相连的观点词的个数,一个观点词的度即为与其相连的特征词的个数.所有特征词节点的度之和等于所有观点词节点的度之和.用公

式表示即为:

$$k_i = \sum_{j=1}^N a_{ij} = \sum_{j=1}^N a_{ji} \quad (1)$$

其中, a_{ij} 表示给定网络的邻接矩阵:

$$A = (a_{ij})_{N \times N} \quad (2)$$

对于给定的特征-观点对二分网络 $B = (X, Y, E)$, $E(B)$ 表示网络中所有连接边的集合. 令 $e_i \in E(B)$, 若任意 $|e_i| = 1$, 则该网络为无权网络, 否则为加权网络. 在加权特征-观点对二分网络中, 点权定义为:

$$S_i = \sum_{j \in N_i} w_{ij} \quad (3)$$

点权 S_i 表示与节点 i 关联的边权之和, 其中 N_i 是节点 i 的近邻集合.

3 特征-观点对提取

本文在无权特征-观点对二分网络中, 首先提出了将两类节点的度值作为度量值来评价节点重要性的算法, 我们将之称为 **B-核分解算法**. 一般来说, 如果仅用节点的度值作为度量值来评价节点重要性是不够精确的. 这是由于现实生活中, 许多网络都是加权网络, 权重及其分布会对网络的属性和功能产生重要影响. 权重的大小, 代表了两个节点之间联系的紧密程度. 即当两个节点同时出现的次数越多时, 两者之间存在某种关联的可能性越大. 例如当候选特征观点词对总是共同出现时, 说明两者是固定搭配的可能性越大, 就越有可能是真正的特征观点词. 而度只能用来表示两类节点共同出现过, 但共同出现的词对不一定就是真正的特征观点词对. 因为在候选特征观点词集中, 错误的特征观点词对也会共同出现. 相对于度值来说, 将权重作为度量值可以更有效地诠释节点的重要性. 所以本文对 **B-核分解算法**进行了调整, 提出将权值大小作为评价节点重要性排序的度量值, 我们称为 **BW-核分解算法**. 上述两种算法的目的是对特征-观点对二分网络中的节点进行重要性排序, 从而识别出中心节点, 找出特征观点词.

3.1 B-核分解算法

首先计算网络中每个节点的度值, 确定网络中所有节点的最小度值 b_{\min} . 通过递归地移除网络中所有度值小于或等于 b_{\min} 的节点, 从而将网络分成若干层. 被去除的节点的集合, 称为网络的 **B-shell(B-壳)**, 简称 B_s .

B-shell 同时作为节点重要性排序指标, B_s 值越大, 节点重要性越大. 剩下的节点的集合称为网络的 **B-核**. 以下是 **B-核算法**.

算法 1. B-核算法

CFO: 候选特征观点词集.

B : 无权特征-观点对二分网络. i 表示网络中的节点.

Ranking set: 新特征观点词排序集.

Step 1: Input: CFO

Step 2: 构建网络 B

Step 3: For i in B :

E is empty set

$b_{\min} = \min_degree(B)$

If i is feature:

If $i.degree < b_{\min}$:

i is inserted into E

If i is opinion:

If $i.degree \leq b_{\min}$:

i is inserted into E

E is inserted into Ranking set

E is deleted

Update B

Every node are recalculated

Step 4: Output: Ranking set

通过 **B-核分解算法**能够确定所有节点在网络中所处的层级, 并给出节点的重要性排序, 识别出二分网络中的中心节点. 下面我们用实例对 **B-核算法**进行更加直观地解释. 首先构建一个无权特征-观点对二分网络, 如图 2 所示. 该网络包含特征词和观点词两类节点, 连边表示它们之间的修饰关系. 例如节点 A 表示的特征词是“质量”, 那么与它有连边的节点 H、I、J 可以分别表示为观点词“好”、“差”和“不错”. 每个节点连边的个数表示此节点的度, 例如节点 A 的度值为 $b = 3$, 节点 L 的度值为 $b = 2$.

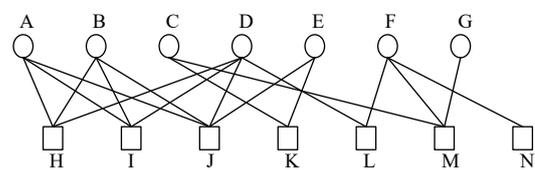


图 2 无权特征-观点对二分网络图

再根据 **B-核分解算法**对节点重要性进行排序, 排序结果如图 3 所示. 该网络被划分成 3 个不同的层, 每一层节点的 B_s 值相等. 通过 **B-核分解算法**确定网络中的核心节点, 即 B_s 值最大的节点是最具有影响力的节点. 在此实例中, 网络中的核心节点分别是特征词节点

A、B、D 和观点词节点 H、I、J. 这六个节点是该网络中的中心节点, 同时也最有可能是我们要找的特征观点词. 如图 2 所示, 特征词节点 A 和 B 分别与观点词节点 H、I、J 一同出现过, 特征词节点 D 分别与观点词节点 H、I、L 一同出现过. 当某个候选观点词同时跟几个候选特征词共同出现时, 说明候选观点词 H、I、J 有可能是真正的观点词. 同理, 当某个候选特征词同时跟几个候选观点词同时出现时, 候选特征词 A、B、D 也可能是真正的特征词. 例如, 节点 A 为候选特征词“质量”, 那么节点 H、I、J 就有可能是候选观点词“好”、“不错”、“差”. 通过人工分析我们知道“质量”、“好”、“不错”、“差”都是真正的特征观点词. 同理得出特征词节点 B、D 和观点词节点 H、I、J 也有可能是真正的特征词和观点词.

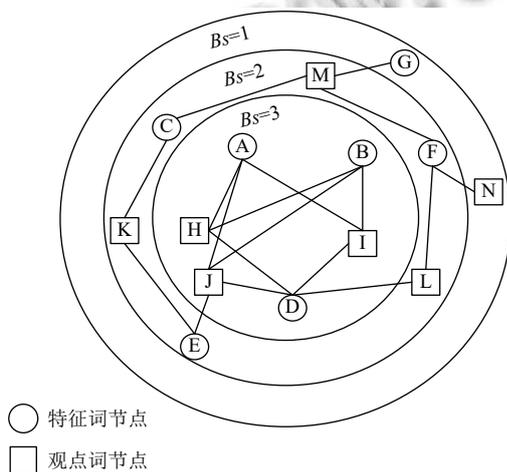


图 3 节点重要性排序图

3.2 BW-核分解算法

首先计算网络中每个节点的权值, 确定网络中最小的权值 bw_{min} . 通过递归地去除网络中所有权值小于或等于 bw_{min} 的节点, 从而将网络分成若干层. 被删除的节点集合称为 Bw-shell (Bw-壳), 简称 Bws . Bw-shell 同时作为节点重要性排序指标, Bws 值越大, 节点的重要性也就越大. 剩余的节点集合称为 BW-核.

在本文的加权网络中, 我们将权值的大小设置为整数, 即 bw_{min} 的起始值为整数 1. 然而在实际生活中, 权值的大小并不全是整数, 更多的是随机数. 即一个加权网络中权值有可能是整数, 也有可能是小数. 所以在本文算法中, 我们将参数值设为 $a \geq bw_{min}$. 即当参数值 a 大于或等于网络中最小权值时, BW-核算法才会以权

值为整数进行分解. 以下是 BW-核算法.

算法 2. BW-核算法

CFO: 候选特征观点词集.
 B: 加权特征-观点对二分网络. i 表示网络中的节点.
 Ranking set: 新特征观点词排序集.

```

Step 1: Input: CFO
Step 2: 构建网络 B
Step 3: For i in B:
E is empty set
 $bw_{min} = \min\_weight(B)$ 
 $a \geq bw_{min}$ 
If i is feature:
    If  $i.weight \leq bw_{min}$ :
        i is inserted into E
If i is opinion:
    If  $i.weight \leq bw_{min}$ :
        i is inserted into E
E is inserted into Ranking set
E is deleted
Update B
Every node weights are recalculated
Step 4: Output: Ranking set
    
```

通过 BW-核分解算法能够确定所有节点在网络中所处的层级, 并给出节点的重要性排序, 识别出此网络的中心节点. 下面我们同样用实例来阐述 BW-核分解算法. 首先构建一个加权特征-观点对二分网络, 如图 4 所示. 在该网络中, 节点的权值是指与该节点相连边的权重之和. 其中边的权重定义为特征-观点对在数据集集中出现的次数, 简称边权. 例如节点 A 的权值 bw 等于 A-I 和 A-J 的边权之和. 假设 A-I 的边权 $bw = 8$, A-J 的边权 $bw = 4$, 那么节点 A 的权值 $bw = 12$.

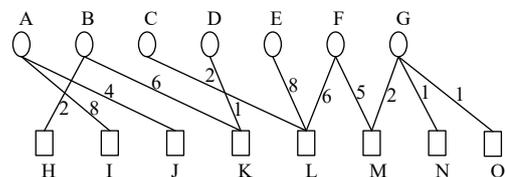


图 4 加权特征-观点对二分网络图

再根据 BW-核算法对加权特征-观点对二分网络中的节点进行重要性排序, 排序结果如图 5 所示. 该网络被分成 6 层, 其中处于第 6 层的节点属于该网络的核心节点, 也就是影响力最大的节点. 通过该分解图我们还可以发现权值大的节点, 并不一定就越接近核心层. 例如特征词节点 G, 它的权值 $bw = 4$, 但却和

$bw = 2$ 的特征词节点 C 和 H 在同一层级. 这是由于该候选特征词很可能是大多数用户在评论时的习惯用语, 虽然出现的次数较多, 但并不是真正的特征词. 例如“方面”这个词语, 大多数用户在评价某产品的特征词时会习惯地带上“方面”. 比如当某个用户想表达“质量不错”这个特征观点时, 往往在评论时会写成“质量方面不错”. 这时, 利用 SBV 关系不仅能识别出“质量-不错”这一对正确的特征观点词, 也会识别出“方面-不错”这一对错误的特征观点词. 所以利用 BW-核算法可以将此类节点排在影响力较小的外层.

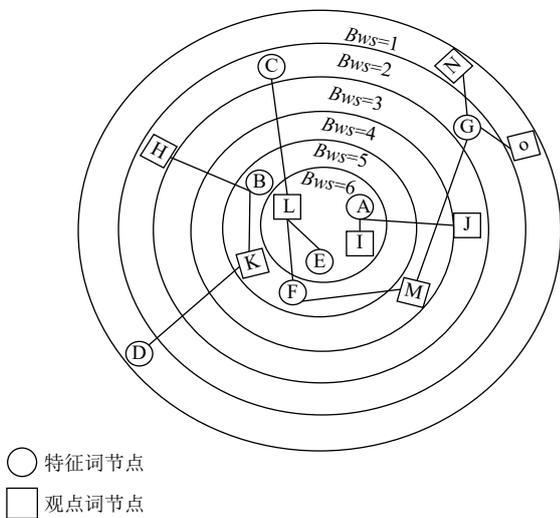


图5 节点重要性排序图

4 实验

本文根据二分网络中节点重要性排序算法即 B-核跟 BW-核分解算法, 对候选特征观点词进行排序. 为了验证此算法在识别特征词和观点词方面的有效性, 本文将来自京东商城的四种商品的评论文

本作为实验数据集进行对比分析. 分别是乐视手机、洗面奶、华为手机、羽毛球拍.

4.1 实验数据集

本文首先对评价文本进行依存句法分析. 基于产品评论特征词, 利用依存关系提取出与产品特征相关的观点词, 构成候选特征观点词对集. 图6是以乐视手机举例说明, 利用哈尔滨工业大学语言云的句法解析结果. 图中 n 代表名词, a 代表形容词, d 代表副词. 利用 SBV 关系识别出[屏幕-不错]、[质量-好]、[内存-大]这三组候选特征观点词对.

根据基于 SBV 关系识别出的候选特征观点词对构建二分网络, 特征-观点对二分网络数据集如表1所示. 表1中给出了网络的一些详细的统计性质. I-/I+分别表示为无权无向网络跟加权无向网络, N_N 中f表示特征节点总数, o表示观点词节点总数, N_E 表示边数.

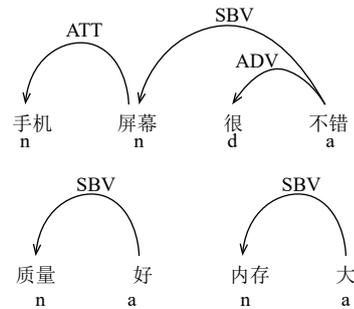


图6 句法分析结果

表1 特征-观点对二分网络数据集

网络	网络类型	N_N	N_E
Le mobile	I-/I+	f299	808
		o228	
Face cleaner	I-/I+	f115	291
		o134	
Racket	I-/I+	f222	497
		o151	
Huawei mobile	I-/I+	f182	458
		o128	

在特征-观点对二分网络中, 在确定了网络中各个节点的度值之后, 我们可以把网络中节点的度数按照从小到大排序, 从而得到满足度为k的节点总数. 我们将这种排序方法称为节点的度分布. 特征词的度分布即与每个特征词相连接的观点词数量的分布, 结果如图7所示; 观点词的度分布即与每个观点词相连接的特征词数量的分布, 结果如图8所示. 从图中我们可以看出在特征-观点对二分网络中, 随着度数的增大, 两类节点数均不断减小, 这类具有较高的度值且数量不多的节点就是我们要找的高频特征观点词. 比如特征词中的“外观”“质量”“价格”等, 他们都是具有高连接的节点, 即具有较高的度值. 比如观点词中的“好”、“不错”、“可以”等也都是具有高连接的节点. 这些具有高度值的节点大多都是高频特征观点词, 能准确地代表消费者对产品的关注焦点.

4.3 实验结果

本文采用目前科学研究中广泛使用的准确率 P、召回率 R 以及 F 值来衡量算法的性能, 各指标越高, 说

明算法的性能越好. 它们的计算公式如下所示:

$$P = x/(x+y) \tag{4}$$

$$R = x/(x+z) \tag{5}$$

$$F = (2 \times R \times P)/(R+P) \tag{6}$$

其中, x 、 y 、 z 的含义分别为识别出的真正高频特征观点词数、识别出的非真正高频特征观点词数以及未识别出的真正高频观点词数. $x+z$ 在本文中表示人工手动标记的数据.

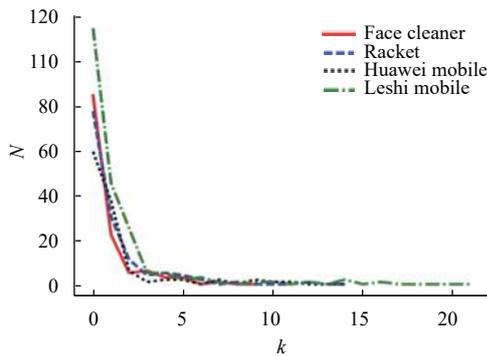


图7 特征节点度分布

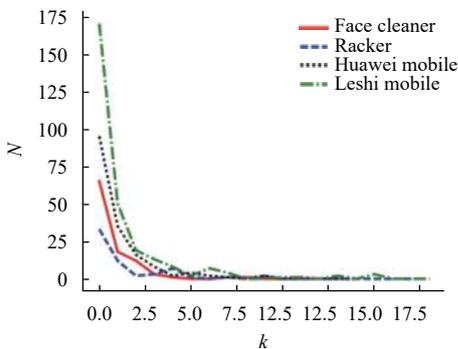


图8 观点词节点度分布

首先根据 B-核分解算法对无权特征-观点对二分网络中的节点进行排序, 即对特征词和观点词进行排序, 识别出高频特征观点词. 同样, 在加权网络中, 根据 BW-核分解算法对加权网络中的节点进行排序, 识别出高频特征观点词. 通过对四类产品的数据集处理之后我们发现, 无论是在无权网络还是加权网络中. 随着层级的增大, 特征词观点词的 P 值是呈上升的趋势, 而 R 值呈下降趋势. 接下来我们将以乐视手机评论的特征词为例, 分析出现这种结果的原因. 在无权网络中的 P 、 R 、 F 值与 B_s 值的关系如图 9 所示. 在加权网络中的 P 、 R 、 F 值与 B_{ws} 值的关系图 10 所示.

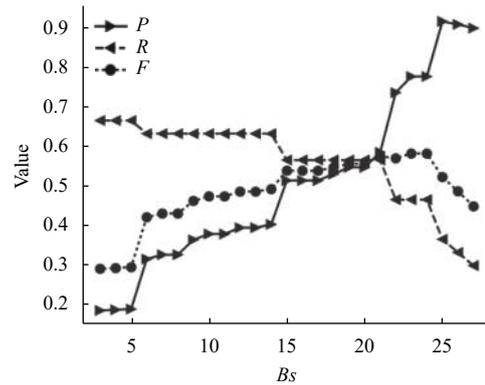


图9 无权二分网络 P 、 R 、 F 值分布

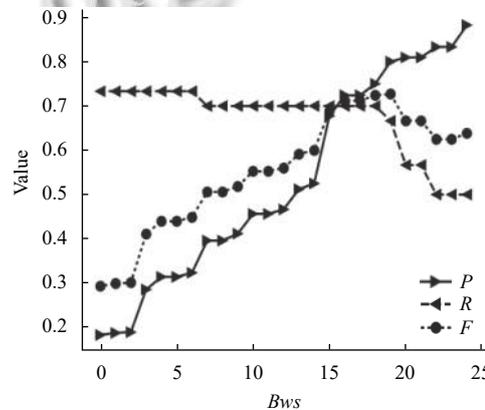


图10 加权二分网络 P 、 R 、 F 值分布

从图 9、图 10 我们可以看出随着 B_s 值、 B_{ws} 值的增加, 越靠近核心层节点的准确率越来越高. 而随着层级的增加, 召回率 R 呈现下降趋势. 因为层级越高, 准确率越来越高, 但剩下的节点数却越来越少, 就会导致召回率越来越低. 为了使四类产品的 P 、 R 、 F 值的实验结果对比更加清晰, 我们分别用如表 2 至表 5 所示.

表2 洗面奶数据分析结果

网络	特征词			观点词		
	P	R	F	P	R	F
无权网络	0.83	0.48	0.61	0.7	0.7	0.7
加权网络	0.71	0.91	0.8	0.89	0.8	0.84

表3 羽毛球拍数据分析结果

网络	特征词			观点词		
	P	R	F	P	R	F
无权网络	0.78	0.48	0.6	0.4	0.67	0.5
加权网络	0.65	0.92	0.76	0.67	0.67	0.67

通过对上述实验结果进行对比分析我们发现在这四组数据中, 利用 B-核算法提取特征词的准确率普遍

高于 BW-核算法. 这是由于在候选特征集中, 错误的特征词出现的频率也很高, 所以就导致利用加权网络提取特征词的准确率比无权网络低.

表4 乐视手机数据分析结果

网络	特征词			观点词		
	<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>
无权网络	0.8	0.4	0.57	0.8	0.8	0.8
加权网络	0.7	0.78	0.74	0.82	0.9	0.86

表5 华为手机数据分析结果

网络	特征词			观点词		
	<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>
无权网络	0.87	0.58	0.7	0.8	0.92	0.86
加权网络	0.67	0.89	0.76	0.92	0.85	0.88

从表2至表5中我们还可以看出利用无权网络提取特征词的召回率普遍较低. 这是因为当一个特征词被多个观点词修饰时, 这个特征词是真正特征词的概率很高, 但这并不代表真正的特征词都会有多个观点词修饰. 例如在华为手机评论文本中, 真正的特征词“屏幕”可以被真正的观点词“大”、“好”以及“清晰”修饰, 但真正的特征词如“像素”却只能用观点词“高”或“低”修饰. 因为加权网络考虑了频次, 出现次数越多是真正的特征词的概率越大. 所以在加权网络中提取特征词的召回率高于无权网络. 但通过 *F* 值的比较我们发现, 无论是哪一类产品评论文本的分析结果, 加权网络的 *F* 值均高于无权的网络. 所以实验结果表明, BW-核算法的性能要优于 B-核算法, 即在加权特征-观点对二分网络中更有利于高频特征观点词的提取.

5 结论

本文针对一个具体的网络, 即对特征-观点对二分网络做了详细分析. 将二分网络节点重要性排序研究引入进高频特征观点词提取研究当中. 首先提出了 B-核算法, 即将节点的度值作为节点重要性排序的度量值. 后针对无权网络中算法的缺陷改进了算法, 提出了 BW-核算法, 该算法是将节点的权值作为节点重要性排序的度量值. 通过实验发现, 两种算法在实际操作中都取得了很好效果.

将复杂网络中节点重要性排序引入特征观点词挖掘研究当中, 不仅是意见挖掘领域的一大创新, 更是扩大了复杂网络在实际中的应用. 二分网络是复杂网络中一种特殊的网络模式, 二分网络中两类节点的连边

与单顶点网络中节点的连边相比, 有更多的意义. 所以接下来我们将对两类节点之间的连接边做进一步研究, 将复杂网络更好地应用于提取特征词和观点词的研究当中.

参考文献

- Zhao WX, Jiang J, Yan HF, *et al.* Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. Proceedings of 2010 Conference on Empirical Methods in Natural Language Processing. Cambridge, MA, USA. 2010. 56–65.
- Hu MQ, Liu B. Mining and summarizing customer reviews. Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, WA, USA. 2004. 168–177. [doi: 10.1145/1014052.1014073]
- Popescu AM, Etzioni O. Extracting product features and opinions from reviews. In: Kao A, Poteet SR, eds. Natural Language Processing and Text Mining. London, UK: Springer. 2007. 9–28.
- 李实, 叶强, 李一军, 等. 中文网络客户评论的产品特征挖掘方法研究. 理科学学报, 2009, 12(2): 142–152. [doi: 10.3321/j.issn:1007-9807.2009.02.015]
- 马柏樟, 颜志军. 基于潜在狄利特雷分布模型的网络评论产品特征抽取方法. 计算机集成制造系统, 2014, 20(1): 96–103.
- Qiu G, Liu B, Bu JJ, *et al.* Opinion word expansion and target extraction through double propagation. Computational Linguistics, 2011, 37(1): 9–27. [doi: 10.1162/coli_a_00034]
- Hai Z, Chang KY, Cong G, *et al.* An association-based unified framework for mining features and opinion words. ACM Transactions on Intelligent Systems and Technology, 2015, 6(2): 26.
- 孙晓, 唐陈意. 基于层叠模型细粒度情感要素抽取及倾向分析. 模式识别与人工智能, 2015, 28(6): 513–520.
- 刘臣, 韩林, 李丹丹, 等. 基于汉语组块产品特征——观点对提取与情感分析研究. 计算机应用研究, 2017, 34(10): 2942–2945. [doi: 10.3969/j.issn.1001-3695.2017.10.014]
- 刘通, 张聪, 吴鸣远. 在线评论中基于边界平均信息熵的产品特征提取算法. 系统工程理论与实践, 2016, 36(9): 2416–2423.
- Jin W, Ho HH. A novel lexicalized hmm-based learning framework for web opinion mining. Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Quebec, Canada. 2009. 465–472.
- 李志义, 王冕, 赵鹏武. 基于条件随机场模型的“评价特征-评价词”对抽取研究. 情报学报, 2017, 36(4): 411–421. [doi:

- 10.3772/j.issn.1000-0135.2017.04.010]
- 13 Titov I, McDonald R. Modeling online reviews with multi-grain topic models. Proceedings of the 17th International Conference on World Wide Web. Beijing, China. 2008. 111–120.
- 14 彭云, 万常选, 江腾蛟, 等. 基于语义约束 LDA 的商品特征和情感词提取. 软件学报, 2017, 28(3): 676–693. [doi: 10.13328/j.cnki.jos.005154]
- 15 Kamal A, Abulaish M, Anwar T. Mining feature-opinion pairs and their reliability scores from web opinion sources. Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics. Craiova, Romania. 2012. 15. [doi: 10.1145/2254129.2254150]
- 16 郝亚辉. 产品评论特征及观点抽取研究. 情报学报, 2014, 33(3): 326–336. [doi: 10.3772/j.issn.1000-0135.2014.03.011]
- 17 Liu K, Xu LH, Zhao J. Extracting opinion targets and opinion words from online reviews with graph co-ranking. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, MD, USA. 2014. 314–324.
- 18 Zhang L, Liu B, Lim SH, *et al.* Extracting and ranking product features in opinion documents. Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Beijing, China. 2010. 1462–1470.
- 19 吴亚晶, 张鹏, 狄增如, 等. 二分网络研究. 复杂系统与复杂性科学, 2010, 7(1): 1–12. [doi: 10.3969/j.issn.1672-3813.2010.01.001]