

中文网络评论的产品特征提取及情感倾向判定^①

任远远, 王卫平

(中国科学技术大学 管理学院, 合肥 230026)

摘要: 随着互联网和电子商务的发展, 用户在购买或使用商品之后会在网络站点上发表对产品的评论, 大量的产品评论中所包含的丰富信息, 可以为生产厂商和用户重要的决策依据. 基于文本的语义和语言分析, 提出了从产品评论中提取用户关注的产品特征的方法, 并根据用户的关注程度对产品特征进行排序; 同时, 根据观点词的极性值判定用户对产品特征的情感倾向以及情感倾向强度. 本研究采用从互联网上获得的针对笔记本电脑的产品评论作为实验对象, 实验结果初步证明该方法具有良好的准确率和召回率.

关键词: 网络评论; 产品特征; 情感倾向; 观点挖掘

Extracting Product Features and Determine Sentiment Orientation from Chinese Online Reviews

REN Yuan-Yuan, WANG Wei-Ping

(Faculty of Management, University of Science and Technology of China, Hefei 230026, China)

Abstract: With the development of the Internet and e-commerce, users can write their opinions about the product on the website after they buy it or use it. There are a large number of product reviews contain abundant information which can help manufacturers and users to make important decisions. Based on semantic and linguistic analysis, this paper proposes a method which can extract product features from the online reviews and rank them based on customer attention strength. At the same time, the sentiment orientation users have on the feature is determined according to the polarity scores of opinion words. The online reviews about notebook Computers are used as experimental data. Experimental results show that the method achieved good precision and recall rate.

Key words: online reviews; product features; sentiment orientation; opinion mining

1 引言

随着互联网和电子商务的飞速发展, 网络已经成为收集用户对于产品观点的重要来源. 用户在购买或使用商品之后会在网络上发表对产品的评论, 这些网络评论会成为影响生产厂商和用户决策行为的重要因素之一. 目前, 越来越多的消费者选择网上购物, 无论是在实体店还是在网上购物, 用户都会浏览大量的网络评论, 了解产品的优势与劣势, 进而决定自己的购买行为. 同样, 生产厂商通过分析产品评论, 可以更好的了解客户需求, 找出用户最关注的产品特征以及用户对这些特征的态度, 对于用户认可的产品特征加以宣传, 而让用户不满意的特征加以改进. 但是,

生产厂商和用户无法快速的从大量的非结构化数据中找到自己想要的信息. 为了了解一个产品, 他们需要浏览大量网站, 通过人工阅读的方式来获取信息, 这样不仅消耗时间而且容易得到错误的信息. 因此, 需要借助一定的技术手段对大量信息进行自动处理, 使生产厂商和用户可以直观的看到他们所需要的信息. 在这样的需求下, 以有效获取网络用户评论信息为目标的非结构化数据分析技术——“观点挖掘”成为了研究热点. 借助于观点挖掘工具, 可以得到用户对于整个产品以及他们所关注的产品特征的态度, 从而为生产厂商和潜在消费者的决策提供信息支持.

挖掘产品特征和情感分析是观点挖掘的两个主要

^① 收稿时间:2014-03-05;收到修改稿时间:2014-04-10

研究方向. 近年来, 大量的学者对这两方面进行研究并取得了一定的成果. 产品特征挖掘是指从大量的网络评论中自动获取用户所关注的特征, 也是分析用户对于产品特征情感倾向的前提. 对于特征挖掘的研究已经取得了一些成果. Popescu 等利用 KnowItAll 系统计算 PMI 值, 然后进行贝叶斯分类, 提取候选特征^[1]. Kobayash 等提出利用产品、产品特征和观点词之间的共现模式的半自动化方法提取产品特征和观点词^[2]. Hu 和 Liu 提出利用非监督型的关联规则分类方法提取英文评论中的产品特征, 取得了较好的结果^[3]. 李实等将这种方法引入到中文网络客户评论中, 并结合汉语的特点对抽取结果进行筛选, 也取得了较好的效果^[4]. Liu 和 Hu 利用自然语言处理和监督模式发现方法从特殊模式的评论中挖掘产品特征^[5]. Miao 和 Li 等将半结构化评论中挖掘出的结果作为先验知识, 使用聚类方法挖掘非结构话评论中的产品特征和观点^[6]. Zhai 和 Liu 等提出了在朴素贝叶斯分类的 EM 算法之上加入两个软约束将具有相同含义的特征合并的方法^[7].

情感分析主要包括两个方面, 一方面是判定评论语句的情感倾向, 即正面评论或负面评论; 另一方面主要是判定用户对产品特征的情感倾向以及情感倾向强度. 很多学者已经对情感分析进行了一定的研究. 朱嫣岚等基于 HowNet, 根据刘群提出的计算语义相似度和语义相关场的方法^[8]计算词语的语义倾向, 准确率达到了 80% 以上^[9]. 王振宁等提出了基于 HowNet 和 PMI 的词语情感极性计算方法, 比传统方法的准确率提高了 5%^[10]. Hatzivassiloglo 和 McKeown 采用对数线性模型分析用连接词连接在一起的形容词的情感倾向^[11]. Esuli 等创建了 SentiWordNet, 计算词语在不同的同义词组中的客观程度值、褒义程度值以及贬义程度值^[12]. Hu 和 Liu 首先利用 WordNet^[13]中形容词的同义词组和反义词组判断形容词的情感倾向, 然后利用评论语句中形容词的情感倾向判断评论语句的情感倾向^[14]. Ding 和 Liu 等提出了一个全面的基于词典的观点挖掘方法, 该方法不仅考虑了基于上下文的观点词的极性, 同时也考虑了一些对极性有影响的特殊词语和短语, 取得了较好的效果^[15]. Pang 等以电影评论作为数据源, 测试了贝叶斯分类、最大熵分类和支持向量机三种机器学习的方法对评论进行情感倾向分类^[16]. Turney 通过计算评论语句与“excellent”和“poor”之间的 PMI 值来判断该评论语句的极性^[17]. Tong 通过跟踪

用户对电影的在线讨论, 显示出正面的情感信息和负面的情感信息随时间变化的图^[18].

基于文本的语义和语言分析, 本文将观点挖掘的任务分为以下两个方面:

- (1) 提取用户关注的产品特征, 并根据用户的关注程度进行排序.
- (2) 判定用户对其所关注的产品特征的情感倾向(正面或负面)以及情感倾向强度.

2 总体架构

本文提出的观点挖掘的任务总体架构如图 1:

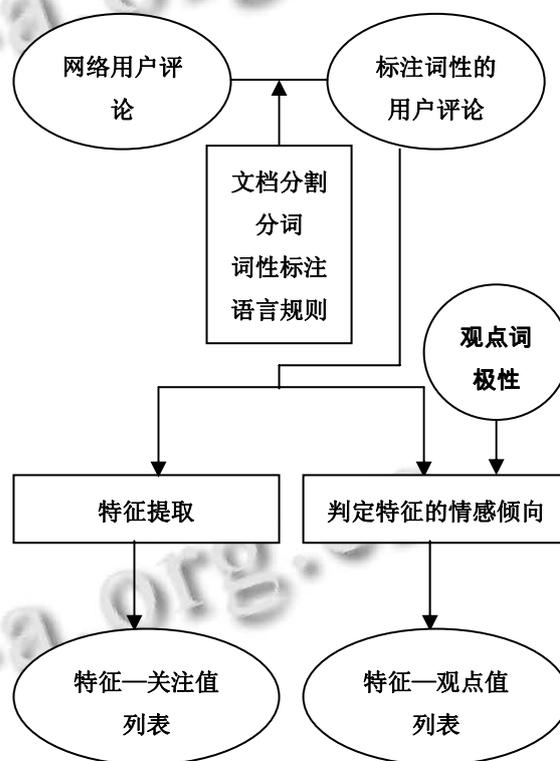


图 1 总体架构

具体来说, 分为以下三个步骤:

- (1) 网络评论的预处理: 首先要对文档进行分割, 接着进行分词和词性标注, 最后根据语言规则完善词性标注结果, 得到标注词性的网络用户评论;
- (2) 特征提取: 利用特征提取算法提取出用户关注的产品特征, 并按照用户关注度对它们进行排序, 得到特征—关注值对;
- (3) 判定用户对特征的情感倾向: 利用观点词的极性和极性强度判定用户对特征的情感倾向和情感倾

向强度, 得到特征—观点值对。

3 特征抽取及其情感倾向的判定

3.1 评论文本预处理

3.1.1 文档分割

由于网络客户评论都是非结构化的数据, 不便于以后的挖掘任务, 所以需要将 these 数据进行分割, 以句子为分割单位, 得到易于挖掘的句子级网络客户评论。

3.1.2 分词与词性标注

文档分割之后, 利用中国科学院计算机所编写的中文分词软件 ICTCLAS(<http://www.ictclas.org/>)对网络用户评论进行分词和词性标注。

3.1.3 语言规则

由于产品特征主要表现为名词和名词词组, 因此, 特征提取转化成对名词和名词词组的提取。由于 ICTCLAS 在词性标注的过程中不会直接标注出名词短语(除了专有名词以外), 所以本文提出了四条语言规则对词性标注后的评论进行完善:

(1) 生产厂商提供名词规则: 生产厂商都会提供一些产品的特征, 供用户参考, 但是其中的一些产品特征不会被 ICTCLAS 标注为名词, 所以需要将 these 产品特征导入 ICTCLAS 用户词典中, 从而将它们的词性标注为 “n”。如例 1 中的 “显卡”。

例 1: 显卡很不错, 玩游戏速度流畅。

(2) 字符名词规则: 由英文字母或英文字母加数字构成的字符串虽然不会被标注为名词, 但具有名词性质, 很有可能是产品特征或其简写形式, 所以将它的词性标注为 “n”。如例 2 中的 “CPU”、“WIN7”。

例 2: CPU 很好, 正版 WIN7 很棒, 散热不太好, 总的来说, 我很喜欢。

(3) 名词词组规则: 连续两个或连续两个以上名词直接相连构成的名词词组很可能是一个产品特征, 因为分解成单个名词之后无法准确表达整个词组的含义, 所以只保留最后一个名词的词性标注。如例 3 中的 “键盘手感”由两个连续名词 “键盘”和 “手感”组成。

例 3: 键盘手感好, 散热还是不错的。

(4) “的”字连接名词规则: 两个名词之间有 “的”字相连, 则该短语很可能是一个产品特征, 所以取消第一个名词和 “的”字之后的词性标注。如例 4 中

的 “风扇的声音”由两个 “的”字相连的名词 “风扇”和 “声音”组成

例 4: 风扇的声音太大, 让我感觉很烦躁。

3.2 特征提取

网络评论中的产品特征可以分为两种: 一种是显性特征, 是用直观的文字描述的, 如例 5 中的 “外观”; 另一种是隐性特征, 是隐含在语意中的, 如例 5 中的 “携带方便”隐含的特征是 “重量”。由于大多数的特征都是显性特征, 并且提取隐性特征有一定的难度, 所以本文的研究主要是提取显性特征。

例 5: 外观很漂亮, 携带也很方便。

一般来说, 如果用户经常表达观点对某些名词和名词短语, 那么这些名词和名词短语很有可能是重要的产品特征, 而用来表达观点的词语被称为观点词。人们经常用形容词来表达对产品特征的观点, 所以, 本文将形容词作为观点词。基于上述思想, 本文提出了特征提取算法。该算法包括四个步骤:

(1) 提取网络评论中所有的名词和名词短语, 对每个名词和名词短语赋予一个关注值。关注值是指形容名词或名词短语的观点词的个数, 初始值设为 0;

(2) 找出每条评论语句中的形容词, 形容词通常是来形容离它最近的名词或名词短语, 所以将离它最近的名词或名词短语的关注值加 1;

(3) 对具有相同含义的名词或名词短语进行合并, 因为不同的名词或名词短语可能描述的是同一个产品特征, 例如 “价钱”和 “价位”都是指 “价格”这一特征, 合并后得到候选特征及其关注值;

(4) 根据关注值对候选特征排序, 关注值大于设置的特定阈值的候选特征即为用户关心的产品特征。

该特征提取算法的伪代码形式如下:

R : 网络产品评论。

S : 网络评论语句, s_i 为第 i 条评论语句。

A : 形容词, a_j 为第 j 个形容词。

N : 名词或名词短语, n_k 为第 k 个名词或名词短语。

NIS : 名词和名词短语的关注值, n_{is_k} 为名词或名词短语 n_k 的关注值。

$NNIS$: 名词和名词短语以及它们对应的关注值。

$FFIS$: 用户关注的特征以及它们对应的关注值。

t : 特征提取阈值

Input: $R, NNIS$

for each n_k in $NNIS$

```

     $nis_k = 0$ 
end for
for each  $s_i$  in  $R$ 
    for each  $a_j$  in  $s_i$ 
        find the closest  $n_k$ 
         $nis_k ++$ 
    end for
end for
combine  $n_k$  in  $NNIS$ 
sort  $n_k$  in  $NNIS$  by  $nis_k$ 
for each  $n_k$  in  $NNIS$ 
    if  $nis_k > t$  then add  $n_k$  and  $nis_k$  into  $FFIS$ 
end for
Output:  $FFIS$ 

```

3.3 判定用户对特征的情感倾向

从网络评论中提取出了用户关注的产品特征, 总结基于这些产品特征的观点, 依靠情感分析技术就可以得出用户对这些产品特征的情感倾向和情感倾向强度。

3.3.1 观点词的褒贬程度计算

观点词的褒贬程度值由观点词与基准词的语义关联的紧密程度决定。基准词是指褒贬态度非常明显、强烈、具有代表性的词语。与褒义基准词联系越密切, 则观点词的褒义倾向越强烈。反之, 与贬义基准词联系越密切, 则观点词的贬义倾向越强烈。本文词语相似度的计算采用夏天提出的方法^[19]。基于上述思想, 得出计算观点词褒贬程度值的公式:

$$O(w) = \max_{i=1}^m \frac{S(P_i, w_k)}{m} - \max_{j=1}^m \frac{S(N_j, w_k)}{m} \quad (k=1..t)$$

其中, P_i 表示褒义基准词 i , N_j 表示贬义基准词 j , w_k 表示观点词 w 的第 k 个义原, $S(P_i, w_k)$ 表示义原 w_k 与褒义基准词 P_i 的相似度, $S(N_j, w_k)$ 表示义原 w_k 与贬义基准词 N_j 的相似度, m 表示基准词的对数, $O(w)$ 是位于 ± 1 之间的实数, 以 0 作为默认阈值, 大于阈值, 则为褒义, 并且值越大, 褒义程度越强; 反之, 为贬义, 并且值越小, 贬义程度越强。

3.3.2 特征观点值的计算

形容词用来形容离它最近的名词或名词短语, 所以将形容某一个名词或名词短语的所有形容词的褒贬

程度值相加, 并取平均值, 即得到该名词或名词短语的观点值。观点值用来说明用户对该名词或名词短语的情感倾向和情感倾向强度, 它是位于 ± 1 之间的实数, 以 0 作为默认阈值, 大于阈值, 说明用户对该特征的情感倾向是正面的, 并且值越大, 正面程度越强; 反之, 说明用户对该特征的情感倾向是负面的, 并且值越小, 负面程度越强。该算法包括六个步骤:

(1) 提取网络评论中的所有形容词, 计算每个形容词的褒贬程度值;

(2) 为提取出的每个名词或名词短语赋予一个情感值, 情感值表示描述该名词或名词短语的所有形容词的褒贬程度值的加总, 初始值设为 0;

(3) 建立否定词表, 包括“不”、“没有”等否定词。如果观点词的左边有否定词出现, 则该观点词的褒贬程度值乘以 -1;

(4) 找出每条评论中的形容词, 将离它最近的名词或名词短语的情感值加上该形容词的褒贬程度值;

(5) 将具有相同含义的名词或名词短语合并, 得到候选特征及其情感值;

(6) 根据提取出的用户关心的产品特征, 用它的情感值除以它的关注值, 即得到它的观点值。

该方法的伪代码形式如下:

R : 网络产品评论.

S : 网络评论语句, s_i 为第 i 条评论语句.

A : 形容词, a_j 为第 j 个形容词.

AS : 形容词的褒贬程度值, as_j 为形容词 a_j 的褒贬程度值.

AAS : 形容词以及它们对应的褒贬程度值.

N : 名词或名词短语, n_k 为第 k 个名词或名词短语.

$NOOS$: 名词和名词短语的情感值, $noos_k$ 为名词或名词短语 n_k 的情感值.

$NNOOS$: 名词和名词短语以及它们对应的情感值.

F : 产品特征, f_m 为第 m 个产品特征.

FOS : 特征的观点值, fos_m 为特征 f_m 的观点值.

$FOOS$: 特征的情感值, $foos_m$ 为特征 f_m 的情感值.

FIS : 特征的关注值, fis_m 为特征 f_m 的关注值.

$FFIS$: 用户关注的特征以及它们对应的关注值.

$FFOS$: 用户关注的特征以及它们对应的观点值.

IW : 否定词, iw_t 为第 t 个否定词.

Input: $R, AAS, IW, NNOOS, FFIS$

for each n_k in $NNOOS$

```

    noosk = 0
end for
for each si in R
    for each aj in si
        if iwi in left context then asj = asj * -1
        find the closest nk
        noosk = noosk + asj
    end for
end for
combine nk in NNOOS
for each fm in FFIS
    find fm in NNOOS
    foosm = noosk
    fosm = foosm / fism
    add fm and fosm in FFOS
end for
Output: FFOS

```

3.4 算法的局限性

通过使用特征提取算法和判定用户对特征情感倾向的算法,不仅可以提取出用户关注的产品特征以及用户对这些特征的情感倾向,而且可以将用户的关注程度以及用户对它们的情感倾向强度进行量化,使用户可以更直观的看到他们想要的信息。但是,这两个算法也存在一定的局限性,它们适用于评论格式比较规范的产品评论,对于评论格式比较随意的产品评论,不会取得较好的结果。

表 1 不同阈值下的特征提取结果对比

阈值	人工标注特征数	挖掘特征数	真正特征数	查准率 Precision	查全率 Recall	F-Score
1	51	100	51	51.00(%)	100.00(%)	67.55(%)
2	51	63	47	74.60(%)	92.16(%)	82.46(%)
3	51	46	41	89.13(%)	80.39(%)	84.54(%)
4	51	39	37	94.87(%)	72.55(%)	82.22(%)

表 2 特征提取结果(按关注程度排序)

排名	特征	关注值	排名	特征	关注值
1	Y460	260	6	键盘	92
2	外观	223	7	价格	84
3	散热	217	8	显卡	79
4	音响	140	9	速度	72
5	配置	93	10	性价比	71

从表 1 中的数据可以看出:当阈值为 3 时,得到了很好的查准率和查全率,查准率与查全率的调和平均

4 实验数据分析

对于本文提出的评论文本预处理、特征提取以及判定用户对特征的情感倾向的方法均使用 JAVA 语言构造。

4.1 语料数据

本文从天极网(<http://www.yesky.com/>)上提取联想 Y460A-IFI(白)(E)笔记本电脑的网络评论作为实验语料。该网站中总共包含对该笔记本电脑的 276 条评论,按句子分割后得到 597 条评论语句。用人工标注的方法对用户关注的产品特征进行标注和识别,将相同含义的特征合并后得到了 51 个产品特征,这些特征将作为参照特征对实验结果进行评估。

4.2 实验结果

4.2.1 特征提取结果

根据特征提取算法得到候选特征之后,需要设置阈值提取用户关注的产品特征。特征提取的结果评估指标采用查准率(Precision)、查全率(Recall)以及查准率与查全率的调和平均值(F-Score)。查准率(Precision)用来评估提取出的特征的精确度;查全率(Recall)用来评估提取出的特征的完整度;虽然查准率和查全率在理论上不相关,而且这两方面又是相互矛盾的,追求查准率的提升往往是以降低查全率为代价,反之亦然,所以实际挖掘中就需要寻求两者的一种平衡,达到整体性能的最佳状态,那就是查准率与查全率的调和平均值(F-Score)。

值(F-Score)也达到了最高。表 2 给出依照关注程度排名前 10 位的产品特征挖掘结果。

根据表 1 和表 2 可以看出:与之前的一些特征提取算法^[1-6,14,15]相比,本文没有提取出所有的产品特征,只是提取用户关注的产品特征,而且将它们关注程度进行量化,给用户一个直观的展示;同时,特征提取的结果也取得了较好的查准率和查全率。

4.2.2 用户对特征的情感倾向值

根据本文提出的判定用户对产品特征的情感倾向的算法, 得出产品特征的观点值. 表 3 给出依照关注程度排名前 10 位的产品特征的观点值.

表 3 产品特征的观点值(按关注程度排序)

排名	特征	观点值	排名	特征	观点值
1	Y460	0.6273	6	键盘	-0.6355
2	外观	0.2396	7	价格	0.0732
3	散热	0.0564	8	显卡	0.5293
4	音响	0.5026	9	速度	0.4632
5	配置	0.7073	10	性价比	0.5445

根据表 3 可以看出: 与之前的一些判定情感倾向的方法^[14-17]相比, 本文没有判定评论语句的情感倾向, 而是判定用户对产品特征的情感倾向, 而且将情感倾向程度进行量化, 这样就可以使用户更直观的看到其他用户对这些产品特征的满意程度, 从而做出决策.

5 结论

本文采用基于语义和语言分析的产品特征挖掘算法从大量的网络评论中提取出用户关注的产品特征, 并按照特征的关注程度排名, 同时结合情感分析技术, 得出用户对这些特征的情感倾向和情感倾向强度. 通过使用本文的挖掘方法进行的实验, 取得了较好的效果. 生产厂商和消费者可以根据本文的挖掘方法挖掘出有用的信息, 为其提供决策支持.

参考文献

- Popescu A, Etzioni O. Extracting product features and opinions from reviews. Proc. of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP). 2005. 339-346.
- Kobayashi N, Inui K, Matsumoto Y, et al. Collecting evaluative expressions for opinion extraction. Proc. of The 1st International Joint Conference on Natural Language Processing. Sanya City. Hainan Island, China. 2004. 584-589.
- Hu M, Liu B. Mining opinion features in customer reviews. 19th National Conf. on Artificial Intelligence (AAAI'04). 2004. 755-760.
- 李实, 叶强, 李一军, Law R. 中文网络客户评论的产品特征挖掘方法研究. 管理科学学报, 2009, 12(2): 142-152.
- Liu B, Hu M, Cheng J. Opinion observer: Analyzing and comparing opinions on the web. Proc. of the 14th International Conference on World Wide Web, WWW'05. 2005. 342-351.
- Miao Q, Li Q, Dai R. An integration strategy for mining product features and opinions. Proc. of the 17th ACM Conference on Information and Knowledge Management, CIKM'08. 2008. 1369-1370.
- Zhai Z, Liu B, Xu H, Jia P. Grouping product features using semi-supervised learning with soft-constraints. Proc. of the 23rd International Conference on Computational Linguistics. COLING'10. 2010. 1272-1280.
- 刘群, 李素建. 基于《知网》的词汇语义相似度的计算. 第三届汉语词汇语义学研讨会, 台北, 2002.
- 朱嫣岚, 闵锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算. 中文信息学报, 2005, 20(1): 14-20.
- 王振宇, 吴泽衡, 胡方涛. 基于 HowNet 和 PMI 的词语情感极性计算. 计算机工程, 2012, 38(15): 187-193.
- Hatzivassiloglou V, McKeown KR. Predicting the semantic orientation of adjectives. Proc. of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, ACL'98. 1997. 174-181.
- Esuli A, Sebastiani F. SentiWordNet: A publicly available lexical resource for opinion mining. Proc. of 5th Conference on Language Resources and Evaluation. Genova, Italy. 2006. 417-422.
- Miller G, Beckwith R, Fellbaum C, Gross D, Miller K. Introduction to WordNet: An on-line lexical database. International Journal of Lexicography (special issue), 1990, 3(4): 235-312.
- Hu M, Liu B. Mining and summarizing customer reviews. Proc. of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'04. 2004. 168-177.
- Ding X, Liu B, Yu P S. A holistic lexicon-based approach to opinion mining. Proc. of the International Conference on Web Search and Web Data Mining, WSDM'08. 2008. 231-240.
- Pang B, Lee L, Vaithyanathan S. 2002. Thumbs up? Sentiment classification using machine learning techniques. Proc. of EMNLP 2002.
- Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. ACL'02. 2002.
- Tong R. An operational system for detecting and tracking opinions in on-line discussion. SIGIR 2001 Workshop on Operational Text Classification. 2001.
- 夏天. 汉语词语语义相似度计算研究. 计算机工程, 2007, 33(6): 190-194.