

基于标签传播的情感词典构建方法

张 璞, 王俊霞, 王英豪

(重庆邮电大学 计算机科学与技术学院, 重庆 400065)

摘 要: 传统情感词典构建方法存在依赖语义知识库、覆盖率有限、领域适应性差等问题。为此, 提出一种利用语料库来构建情感词典的方法。该方法选取情感种子词, 在语料上训练 Word2Vec 词向量来选取与种子词相似度高的词语作为候选情感词, 并在语料上分析与种子词具有连词关系的词语作为候选情感词。通过种子词和候选情感词之间的相似度构建语义关联图, 使用标签传播算法计算情感词的极性, 从而构建情感词典。实验结果表明, 与基线方法相比, 该方法能获得较高的准确率和较好的鲁棒性。

关键词: 情感分析; 情感词典构建; 词向量; 连词关系; 标签传播

中文引用格式: 张 璞, 王俊霞, 王英豪. 基于标签传播的情感词典构建方法[J]. 计算机工程, 2018, 44(5): 168-173.

英文引用格式: ZHANG Pu, WANG Junxia, WANG Yinghao. Sentiment Lexicon Construction Method Based on Label Propagation[J]. Computer Engineering, 2018, 44(5): 168-173.

Sentiment Lexicon Construction Method Based on Label Propagation

ZHANG Pu, WANG Junxia, WANG Yinghao

(College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

[Abstract] Traditional sentiment lexicon construction methods have problems such as relying on semantic knowledge base, limited coverage and poor domain adaptability. Aiming at these problems, this paper proposes a method to construct sentiment lexicon based on label propagation. The method firstly selects some sentiment seed words manually, and then uses Word2Vec to train word embeddings on corpus and treats the words which have high similarities with seed words as candidate sentiment words, the method also finds out the words which have conjunctive relations with seed words, and treats them as candidate sentiment words. By constructing a semantic association graph through the similarities between seed words and candidate sentiment words, the method uses label propagation algorithm to identify polarities of candidate sentiment words and construct sentiment lexicon. Experimental results show that the method can obtain higher accuracy and better robustness compared with baseline method.

[Key words] sentiment analysis; sentiment lexicon construction; word vector; conjunction relationship; label propagation

DOI: 10.19678/j.issn.1000-3428.0046862

0 概述

随着电子商务的快速发展, 各电商网站产生了带有情感色彩的海量评论文本, 文本情感分析技术也得到越来越多的应用。通过对产品评论文本进行情感倾向性的自动分析, 不仅可以为消费者节省大量时间和精力, 提供商品属性粒度级别的购买依据, 还可以使商家更加了解消费者对产品的情感倾向以及关心的问题, 由此明确产品的优势和不足, 对用户需求和产品改进方向作出有效反应, 提高企业竞

争力^[1]。

文本情感分析领域中, 情感词典的构建是重要研究内容之一。情感词典作为一种重要的情感资源, 在词语、短语、属性、句子及篇章级等不同粒度的情感分析任务中起着重要的作用^[2]。

通用情感词典如 General Inquirer (GI)、SentiWordNet 等主要以手工或半自动方式构建而成, 难以覆盖不同领域的情感词, 其领域适应性及可靠性受到限制^[2-3]。因此, 许多研究关注于情感词典的自动构建。其中一类主要方法是利用语义知识库

基金项目: 教育部人文社会科学研究青年基金(17YJCZH247); 重庆市教委人文社会科学研究“社交媒体背景下的产品评论挖掘及应用研究”项目(17SKG055); 重庆市教委科技项目(KJ1600440)。

作者简介: 张 璞(1977—), 男, 副教授、博士, 主研方向为文本挖掘、自然语言处理; 王俊霞、王英豪, 硕士研究生。

收稿日期: 2017-04-18 **修回日期:** 2017-05-18 **E-mail:** 1263038889@qq.com

(WordNet、HowNet 等)来进行情感词典的扩充^[3-6]。这类方法主要以一组已知极性的词语作为种子集,通过利用知识库中词语的同义、反义、上下位、词语和义原的关联等语义关系来对未知词的情感倾向进行判定,可以方便快捷地扩充情感词典,但也存在依赖于语义知识库、词典覆盖率有限、领域适应性差等问题。针对上述问题,本文利用语料库来进行情感词典的构建,提出一种基于标签传播的情感词典构建方法。该方法通过在语料上利用 Word2Vec 训练词向量,以及分析词语之间的连词关系等 2 种方式来对情感种子扩充候选情感词,并根据扩充词和种子词之间的相似性构建图,最后再利用标签传播算法来计算词汇的情感极性,得到情感词典。

1 情感词典的构建

1.1 总体流程

本文方法的整体思路如下:首先选取一定数量积极极性和消极极性的情感种子词,然后使用 Word2Vec 在语料上训练词向量。如果词语 a 和种子词 b 的相似性超过阈值,则词语 a 和 b 之间有一条边;此外,通过在语料中分析种子词的连词关系,如果词语 c 和种子词 b 具有连词关系,则词语 c 和 b 之间有一条边。最后得到一个图,再通过标签传播算法在图上进行极性传播,最终确定所有扩充得到的候选情感词的极性。总体流程如图 1 所示。

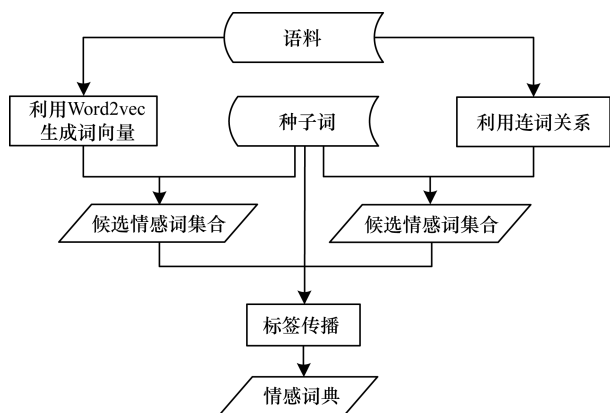


图1 本文方法的总体流程

1.2 种子词的选取

在情感词典构建中,通常选用具有明显情感倾向的词语作为种子词。本文选择 SentiWordNet^[7] 中部分主观性强的词语作为种子词。SentiWordNet 中对各种词性的情感词打有分值,本文选择分数高且常见的情感词作为种子词,将极性为积极的情感词放入积极种子词集合(“Positive”集合,简称“Pos”集合)中,将极性为消极的情感词放在消极种子词集合(“Negative”集合,简称“Neg”集合)中。所选取的部分种子词如表 1 所示。

表1 部分种子词集合

“Pos”集合	“Neg”集合
good	bad
happiness	sad
love	unfortunate
nice	worst

1.3 候选情感词的 Word2Vec 抽取方法

Word2Vec 是一款将词语表征为词向量(也称词嵌入)的高效工具。词向量是词语的分布式表示(Distributed Representation)方式,其基本思想是利用深度学习技术,通过在语料上训练后,将每个词映射成多维的实数向量,使得语义相近的词间的距离也比较接近。Word2Vec 中包含了 2 种训练模型,分别是连续词袋模型(Continuous Bag-Of-Words, CBOW)和 skip-gram 模型。其中,CBOW 模型通过上下文来预测当前词,skip-gram 模型则通过当前词来预测上下文。相比于词语的独热表示方式(One-hot Representation),词向量不但能更好地表达词语间的语义关系,也避免了维数灾难,因而被应用于情感新词的发现^[8]、情感词典的自动构建^[9]等工作中。受上述工作的启发,本文也采用 Word2Vec 工具来进行情感词的扩充,通过 Word2Vec 在语料上训练词向量模型,得到每个词语的词向量。再将训练语料中与种子词相似度大于设定阈值的词语抽取出来作为候选情感词,添加到候选情感词集合中。在计算词语间相似度时,本文选用余弦相似度方法来进行计算。

1.4 候选情感词的连词关系抽取方法

本文所用连词关系主要有并列关系和转折关系这 2 种。一般而言,评论文本中所出现的并列关系连词不会改变连续出现的情感词极性,而转折关系连词会使文本中前后出现的情感词极性发生变化。具有并列关系的评论文本如例 1)、例 2),具有转折关系的评论文本如例 3)、例 4):例 1)“The cellphone is beautiful and durable”;例 2)“The phone is neither cheap nor beautiful”;例 3)“The phone’s performance is good, but the price is too expensive”;例 4)“For the price, this is a cheap, yet somewhat protective option”。

本文选用的并列关系连词有:and, neither...nor, either...or, as well as, not only...but also...等。转折关系连词有:but, yet, however, still, while, on the contrary 等。为了利用词语之间的连词关系抽取候选情感词,传统方法通常采用词汇窗口技术。考虑到有些具有连词关系(尤其是转折关系)的词语的距离比较远,使用词汇窗口技术来抽取候选情感词时会由于窗口大小设置不当而失效,因此,本文使用 Stanford Parser^[10]来对语料进行依存句法分析,得到词语之间的依存关系,再将和种子词具有连词依存

关系的词语抽取出来作为候选情感词。Stanford Parser 能够依据连词依存关系来有效地抽取出候选情感词。例如,对例 3) 所示的评论而言,“good”和“expensive”这 2 个词分别位于句子的不同子句中,距离比较远。通过 Stanford Parser 对评论句子进行依存句法分析后,可以得到以下依存关系结果,“det (phone-2, The-1), nmod_poss (performance-4, phone-2), case (phone-2, 's-3), nsubj (good-6, performance-4), cop (good-6, is-5), root (ROOT-0, good-6), cc (good-6, but-8), det (price-10, the-9), nsubj (expensive-13, price-10), cop (expensive-13, is-11), advmod (expensive-13, too-12), conj_but (good-6, expensive-13)”。其中,每条依存关系分析结果提供的信息主要由两部分构成:关系名称(如 det 代表冠词修饰关系;conj_but 表示转折关系;nsubj 代表主谓关系),2 个参与关系的词语及它们分别在句中的位置。通过 conj_but (good-6, expensive-13) 这一依存关系结果可知“good”和“expensive”具有转折关系,因此,当“good”为种子词时,则可以根据 conj_but 依存关系来抽取出候选情感词“expensive”。类似地,对例 1)、例 2)、例 4) 中的评论句子,可以分别利用 conj_and、conj_nor、conj_yet 等连词依存关系来抽取候选情感词。

当评论文本中的词语和种子词具有并列关系时,将该词语加入到对应候选情感词集合中,并记录词语在集合中的出现次数。当集合中已有该词时,则将其次数加 1,且标记为并列关系;当评论文本中的词语和种子词具有转折关系时,将该词语也加入到对应候选情感词集合中。若集合中已有该词时,则将其次数加 1,且标记为转折关系。如此在评论文本中对选取的种子词利用连词关系不断扩充候选情感词,直到再没有可以扩充的词语为止。假设“beautiful”和“good”都是种子词,而例 1) 中“beautiful”和“durable”具有并列关系,所以“durable”被抽取出来;例 3) 中“good”和“expensive”具有转折关系,所以“expensive”被抽取出来。

在扩充得到的候选情感词集合中,通过一个种子词扩充得到的具有相同连词关系的词语数量来计算种子词和该词语的相似度。如种子词“good”扩充到 3 个具有并列关系的词语:“durable”“beautiful”“cheap”,其数量分别为 3、4、3,则这 3 个词语与种子词“good”的相似度分别为 $3/(3+4+3)$, $4/(3+4+3)$, $3/(3+4+3)$ 。

1.5 利用标签传播算法的情感词典构建

1.5.1 标签传播算法

标签传播算法 (Label Propagation Algorithm, LPA) 是一种基于图的半监督学习方法,目前已在 Web 页面挖掘^[11]、社区发现^[12]、人物社会关系抽取^[13]、情感词典构建^[14]等诸多领域得到运用。标签

传播算法根据结点之间的相似度关系构建关系图,在图完成初始化后,图中结点可以分为已知标签的结点和未知标签的结点,LPA 的基本思想是通过已知标签的结点来预测未知标签结点的标签信息。结点之间的标签主要通过相似度来进行传播。在标签传播的过程中,未知标签的结点通过相邻已知标签的结点信息来更新自己的标签。若相邻结点与它的相似度越大,则对其标签影响的权重越大,也就更容易进行标签的传播。

1.5.2 情感词典的构建

首先,将 1.3 节和 1.4 节中扩充所得到的候选情感词和种子词都作为图中的结点,图的构造方法如下:如果词语 a 通过 Word2Vec 方法可以扩充得到词语 b ,那么 a 和 b 之间有一条边,权重为词语 a 和 b 的相似度;如果词语 a 和词语 c 具有连词关系,则 a 和 c 之间有一条边,权重为词语 a 和 c 的相似度。因此,所有抽取出的候选情感词和种子词被抽象为一张图,如图 2 所示。

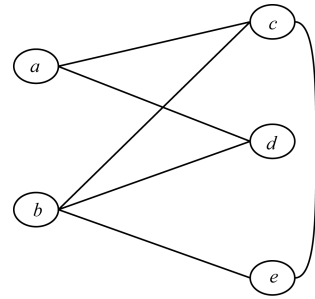


图 2 图模型结构

假设图中共有 m 个结点,则可构建一个 m 维的相似度概率转移矩阵。计算方法如下:

$$T[i][j] = \frac{SIM(w_i, w_j)}{\sum_{j=0}^{m-1} SIM(w_i, w_j)} \quad (1)$$

其中, $T[i][j]$ 表示词语 i 到词语 j 的相似度转移概率, $SIM(w_i, w_j)$ 表示词语 i 和 j 的相似度。

在图 2 中,假设 a 和 b 分别是积极种子词和消极种子词,极性分别记为 $+1$ 和 -1 。其余词语的极性未知,记为 0 。则词语 $a \sim e$ 初始情感极性向量 V 如下:

$$V = \begin{bmatrix} +1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (2)$$

然后利用相似度概率转移矩阵和词语的初始情感极性进行不断地迭代,得到每个未知极性的词语极性。计算方法如下:

$$PO[i] = \sum_{j=0}^{m-1} T[j][i] \times V[j] \quad (3)$$

其中, $PO[i]$ 表示迭代后结点 i 的情感极性, $T[j][i]$

表示结点 j 到结点 i 的相似度矩阵中的转移概率, $V[j]$ 表示迭代前结点 j 的初始情感极性。在每轮迭代中,种子词的极性保持不变。经过不断迭代,直到图中词语的极性不再发生任何变化为止。迭代过程结束后,将极性的绝对值大于某一阈值的词语作为最终情感词,并得到未知标签的词语极性,扩充得到情感词典。

1.6 本文方法的算法描述

本文提出了基于标签传播的情感词典构建方法,算法描述如算法1所示。

算法1 情感词典构建算法

输入 评论语料 (Review Corpus, RC), 情感种子词 SD , 利用 Word2Vec 扩充词语的阈值 wv , 存放情感词的集合 U , 确定是否是积极情感词的阈值 p_{pos} , 确定是否是消极情感词的阈值 p_{neg}

输出 积极情感词典 (SL_{pos}) 和消极情感词典 (SL_{neg})

1) 初始化情感词典, 令 $SL_{pos} = \emptyset, SL_{neg} = \emptyset$ 。

2) 对语料 RC 进行预处理。

3) 对每个种子词 $SD_i \in SD$ 。

(1) 利用 Word2Vec 在语料 RC 上训练词向量, 若 RC 中的词语 SW_i 和种子词 SD_i 的相似性大于阈值 wv , 则 $U = U \cup SD_i \cup SW_i$, 并记录词语之间的相似度。

(2) 对语料 RC 进行连词关系和依存句法分析, 若 RC 中的词语 SW_j 和种子词 SD_i 具有连词关系, 则 $U = U \cup SD_i \cup SW_j$ 。

(3) 按 1.4 节所叙述方法来计算步骤 (2) 中词语 SW_j 和种子词 SD_i 的相似度。

4) 通过步骤 3 得到所有扩充的候选情感词以及词语之间的相似度矩阵后, 再根据式 (1) 得到词语之间的相似度概率转移矩阵, 并构建情感词之间的语义关联图。

5) 在整个图中, 根据式 (3) 以及词语的初始情感极性 V 计算未知极性的词语极性 PO 。

6) 令 $V = PO$ 。

7) 重复步骤 5)、步骤 6), 直至整个图中词语的极性不再发生任何变化。

8) 得到最终词语的极性向量 PO 。 PO 中每个值代表每个词语的情感倾向。如果词语 SW_k 的极性 $PO_k > 0$ 且 $|PO_k| > p_{pos}$, 则 $SL_{pos} = SL_{pos} \cup SW_k$, 如果词语 SW_k 的极性 $PO_k < 0$ 且 $|PO_k| > p_{neg}$, 则 $SL_{neg} = SL_{neg} \cup SW_k$ 。

2 实验设置与基线实验

2.1 实验设置

语料: 本文语料选用斯坦福大学所提供的亚马逊评论语料^[15], 该语料中有书籍、手机、衣服、电子产品以及电影等不同领域的评论, 本文选择其中的

手机领域评论语料作为实验语料, 共 194 185 条。

利用 Word2Vec 在语料上训练词向量时, 由于 skip-gram 模型具有高效的特点^[16], 本文选择 skip-gram 来训练词向量, 并在训练过程中采用 Negative Sampling 方法。相应参数设置如下: 词向量维度值为 200, 上下文窗口取值为 8。

评价标准: 类似于文献[17]的评估方法, 本文利用人工评判方法对情感词典进行评估, 具体做法是选取情感词典中一定数量的词语, 人工判断它们的极性是否正确, 通过这些情感词的正确率来衡量所构建的情感词典性能。

2.2 基线实验

为了验证本文提出的情感词典构建方法的有效性, 选取了以下方法作为基线方法:

1) 基于 WordNet 的情感词典构建方法 (A Method of Constructing Sentiment Lexicon Based on WordNet, WN): 直接利用 WordNet 语义知识库对情感种子词进行同义词的迭代抽取, 得到扩充的情感词典。

2) 基于连词关系的情感词典构建方法 (Method of Constructing Sentiment Lexicon Based on Conjunctive Relations, CR): 分析语料, 迭代抽取和种子词具有连词关系的词语。和种子词具有并列关系的词语极性和种子词极性相同, 和种子词具有转折关系的词语极性和种子词极性相反, 得到扩充的情感词典。

3) 基于 Word2Vec 的情感词典构建方法 (Method of Constructing Sentiment Lexicon Based on Word2Vec, W2V): 利用 Word2Vec 在语料上训练词向量, 然后迭代计算语料中词语和情感种子词之间的语义相似度, 相似度大于某个阈值的词语和该种子词的极性相同, 以此得到扩充的情感词典。

4) 基于连词关系和标签传播的情感词典构建方法 (Method of Constructing Sentiment Lexicon Based on Conjunctive Relations and Label Propagation, CRLP): 利用连词关系抽取候选情感词后, 然后利用标签传播算法确定候选情感词的极性以及构建情感词典。

5) 基于 Word2Vec 和标签传播的情感词典构建方法 (A Method of Constructing Sentiment Lexicon Based on Word2Vec and Label Propagation, W2VLP): 利用 Word2Vec 在语料上训练词向量后, 然后计算语料中词语和情感种子词之间的语义相似度, 将相似度大于某个阈值的词语抽取出来作为候选情感词, 最后利用标签传播算法确定候选情感词的极性以及情感词典。

3 实验结果与分析

为了方便, 将本文提出的方法称为 CR&W2V-LP 方法。分别选取 20 个、30 个、50 个种子词进行

实验。其中,确定是否是积极情感词的阈值 p_{pos} 设为 0.01,确定是否是消极情感词的阈值 p_{neg} 设为 0.02。人工判断所构建情感词典的正确率,实验结果分别如表 2~表 4 所示。

表 2 种子词为 20 个的实验结果 %

实验方法	积极情感词典的正确率	消极情感词典的正确率	正确率
WN	81.6	60.0	70.8
CR	79.0	62.8	70.9
W2V	74.0	76.6	75.3
CRLP	83.3	62.5	72.9
W2VLP	84.0	68.0	76.0
CR&W2V-LP	88.0	68.0	78.0

表 3 种子词为 30 个的实验结果 %

实验方法	积极情感词典的正确率	消极情感词典的正确率	正确率
WN	80.0	52.0	66.0
CR	80.0	62.0	71.0
W2V	66.0	79.6	72.8
CRLP	83.3	66.7	75.0
W2VLP	84.0	70.0	77.0
CR&W2V-LP	87.0	70.0	78.5

表 4 种子词为 50 个的实验结果 %

实验方法	积极情感词典的正确率	消极情感词典的正确率	正确率
WN	62.0	40.0	51.0
CR	82.0	66.0	74.0
W2V	70.0	74.0	72.0
CRLP	84.2	66.7	75.5
W2VLP	86.0	68.0	77.0
CR&W2V-LP	86.8	70.6	78.7

从表 2~表 4 可以看出,在基线实验中,WN 方法所得情感词典的正确率的实验结果最差。原因在于 WordNet 是一个人工构建的语义知识库词典,在迭代扩充过程中所引入的噪声词比较多,并且覆盖有限,对于不在 WordNet 中的情感词扩充不到,因而有相应局限。

相比于依赖语义知识库的 WN 方法而言,基于语料库的 CR 方法和 W2V 方法的实验结果均取得了较好效果,说明了基于语料库的情感词典构建方法的优越性。进一步,发现在表 2~表 4 中,W2VLP 方法的性能均好于 W2V 方法,CRLP 方法的性能也优于 CR 方法,这说明了标签传播算法的有效性。

相比于基线方法,本文提出的 CR&W2V-LP 方法在表 2~表 4 中的正确率均取得了最高值,相对于 CRLP 和 W2VLP 这 2 个性能较好的基线方法分别获得了 1.5%~5.1% 间的性能提升,可见本文方法对于情感词典的构建有较大的优势。下面对原因进行分析。

在 W2V 和 W2VLP 方法中,对于语料中和种子词距离较远的情感词会由于相似度较小而被过滤掉。CR 和 CRLP 方法虽然可以将语料中和种子词具有连词关系的词语抽取出来作为情感词,但是对于其他和种子词没有连词关系的情感词却会被忽略。如“The cellphone is beautiful and durable, I love it very much”这一评论中,若“beautiful”作为种子词,利用 CR 或者 CRLP 方法可以抽取出来和它并列的情感词“durable”,但是情感词“love”则因为和种子词没有连词关系而被忽略掉。因此,从上述分析可以看出,基于 Word2Vec 的方法和基于连词关系的方法各有优缺点。

本文方法则因为结合了 Word2Vec 和连词关系这 2 种方法来抽取候选情感词,因而,能较好地抽取各自方法所扩充不到的情感词。此外,标签传播算法可以将种子词的情感极性通过边向相邻结点传播。在图中,词语的极性不仅受到近距离种子词的影响,也受到其他种子词极性的影响。通过多次迭代,标签在不断地传播,可以对极性标注不正确的情感词进行及时地更新,使最终扩充得到的情感词典的正确率更高。综上所述,本文方法所构建的情感词典质量更高,在性能和正确率上具有更大优势。

此外,本文方法还具有良好的鲁棒性。图 3 直观地展示了基线方法以及本文方法随种子词个数变化的正确率变化情况,从图 3 可以看出,本文方法较为稳定,在种子词数量变化的情况下性能均优于其他方法。

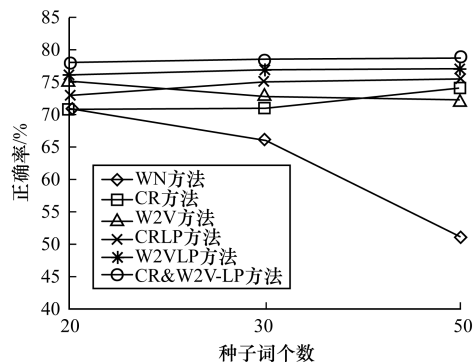


图 3 随种子词个数变化的实验结果

4 结束语

本文提出一种基于标签传播算法的情感词典构建方法。该方法首先选取情感种子词,然后通过 Word2Vec 和连词关系等 2 种方式在语料中找出候选情感词,构建语义关联图,使用标签传播算法更新候选情感词的极性,最终构建得到情感词典。实验结果表明,该方法优于基线方法,具有良好的性能及鲁棒性。下一步将对本文方法在其他领域语料上进行实验,并考虑将语言学知识融入标签传播算法中,提高情感词典构建方法的准确率。

参考文献

- [1] 杜嘉忠,徐 健,刘 颖.网络商品评论的特征-情感词本体构建与情感分析方法研究[J].现代图书情报技术,2014,30(5):74-82.
- [2] 王 科,夏 睿.情感词典自动构建方法综述[J].自动化学报,2016,42(4):495-511.
- [3] 谢松县,刘 博,王 挺.应用语义关系自动构建情感词典[J].国防科技大学学报,2014(3):111-115.
- [4] KIM S M, HOVY E. Determining the sentiment of opinions [C]//Proceedings of the 20th International Conference on Computational Linguistics. Geneva, Switzerland: Association for Computational Linguistics, 2004:1367-1373.
- [5] HASSAN A, RADEV D. Identifying text polarity using random walks [C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, Sweden: Association for Computational Linguistics, 2010:395-403.
- [6] 周咏梅,杨佳能,阳爱民.面向文本情感分析的中文情感词典构建方法[J].山东大学学报(工学版),2013,43(6):27-33.
- [7] BACCIANELLA S, ESULI A, SEBASTIANI F. SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining [C]//Proceedings of International Conference on Language Resources and Evaluation. Valletta, Malta: European Languages Resources Association, 2010:83-90.
- [8] 杨 阳,刘龙飞,魏现辉,等.基于词向量的情感新词发现方法[J].山东大学学报(理学版),2014,49(11):51-58.
- [9] 杨小平,张中夏,王 良,等.基于 Word2Vec 的情感词典自动构建与优化[J].计算机科学,2015,44(1):42-47.
- [10] StanfordParser[EB/OL]. [2016-11-21]. <https://nlp.stanford.edu/software/lex-parser.shtml>.
- [11] 张乃洲,曹 薇,李石君.一种基于结点密度分割和标签传播的 Web 页面挖掘方法[J].计算机学报,2015,38(2):349-364.
- [12] 马千里,张俊浩.一种局部强化的多标签传播社区发现算法[J].计算机工程,2014,40(6):171-174.
- [13] 刘锦文,许 静,张利萍,等.基于标签传播和主动学习的人物社会关系抽取[J].计算机工程,2017,43(2):234-240.
- [14] 李寿山,李逸薇,黄居仁,等.基于双语信息和标签传播算法的中文情感词典构建方法[J].中文信息学报,2013,27(6):75-81.
- [15] MCAULEY J, TARGETT C, SHI Q, et al. Image-based recommendations on styles and substitutes [C]//Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA: ACM Press, 2015:43-52.
- [16] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [C]//Proceedings of the International Conference on Learning Representations. Scottsdale, Arizona, USA: ACM Press, 2013:1-12.
- [17] NEVIAROUSKAYA A, PRENDINGER H, ISHIZUKA M. SentiFul: a lexicon for sentiment analysis [J]. IEEE Transactions on Affective Computing, 2011, 2(1):22-36.
- [7] 李 聪,梁昌勇,董 珂.基于项目类别相似性的协同过滤推荐算法[J].合肥工业大学学报(自然科学版),2008,31(3):360-363.
- [8] 李小慧.基于 Jaccard 项目类别相似性的个性化推荐算法研究[D].长沙:中南大学,2010.
- [9] 窦羚源,王新华,孙 克.融合标签特征和时间上下文的协同过滤推荐算法[J].小型微型计算机系统,2016,37(1):48-52.
- [10] 王洪明.基于本体和标签的用户偏好提取系统的设计与实现[D].北京:北京邮电大学,2011.
- [11] 翟延冬,王康平,张东娜,等.一种基于 WordNet 的短文本语义相似性算法[J].电子学报,2012,40(3):617-620.
- [12] 张 科.基于《知网》义原空间的文本相似度计算研究与实现[D].重庆:重庆大学,2013.
- [13] 宋万鹏.短文本相似度计算在用户交互式问答系统中的应用[D].合肥:中国科学技术大学,2010.
- [14] SHIRAKAWA M, NAKAYAMA K, HARA T, et al. Wikipedia-based semantic similarity measurements for noisy short texts using extended naive Bayes [J]. IEEE Transactions on Emerging Topics in Computing, 2015, 3(2):1.
- [15] MARTINEZ G J. An overview of textual semantic similarity measures based on web intelligence [J]. Artificial Intelligence Review, 2014, 42(4):935-943.
- [16] 王小林,王 东,杨思春,等.基于《知网》的词语语义相似度算法[J].计算机工程,2014,40(12):177-181.

编辑 刘 冰

编辑 刘 冰