

基于卷积神经网络与多特征融合的 Twitter 情感分类方法

王汝娇, 姬东鸿

(武汉大学 计算机学院, 武汉 430072)

摘 要: 为了对社交网络平台上发表的言论和信息进行情感分类, 基于卷积神经网络和多特征融合, 提出一种情感分类方法。结合 Twitter 自身语言特性和情感字典资源设计语料特征和词典特征, 对 Twitter 文本词向量使用卷积神经网络获得对应的深度词向量特征, 将上述 3 类特征进行特征融合并采用 One-Versus-One SVM 实现情感极性的分类判别。针对 SemEval 语料的实验结果表明, 该方法取得了较好的情感分类效果, 多特征融合能够有效地提高情感分类的准确性。

关键词: 文本分类; 情感分析; 卷积神经网络; 词向量; 特征融合

中文引用格式: 王汝娇, 姬东鸿. 基于卷积神经网络与多特征融合的 Twitter 情感分类方法[J]. 计算机工程, 2018, 44(2): 210-219.

英文引用格式: WANG Rujiao, JI Donghong. Twitter Sentiment Classification Method Based on Convolutional Neural Network and Multi-feature Fusion[J]. Computer Engineering, 2018, 44(2): 210-219.

Twitter Sentiment Classification Method Based on Convolutional Neural Network and Multi-feature Fusion

WANG Rujiao, JI Donghong

(Computer School, Wuhan University, Wuhan 430072, China)

[Abstract] In order to classify the emotion for users expressions and comments on social networks, this paper presents a sentiment classification method which combines Convolutional Neural Network (CNN) and multi-feature fusion. It designs corpus features and lexicon features according to the characteristics of Twitter texts and semantic lexicons, uses the convolution neural network for the word vector of Twitter text to get the depth word vector features, combines the above three features to construct the feature fusion and uses One-Versus-One SVM to obtain the sentiment polarity classification and discrimination. Experimental results on SemEval corpus show this method performs a good result and the multi-feature fusion can efficiently improve the accuracy of sentiment classification.

[Key words] text classification; sentiment analysis; Convolutional Neural Network (CNN); word vector; feature fusion

DOI: 10.3969/j.issn.1000-3428.2018.02.037

0 概述

随着移动互联网和社交网络应用的飞速发展, 用户越来越乐于利用微博、微信、Twitter、Facebook 等社交软件在互联网中分享和传播自己对于某件事物的立场和看法, 或者在各大电商、APP 应用等平台中表达对某一产品或者服务的使用感受。如何准确高效地对这类带有感情色彩的文本进行分析处理、归纳推理已成为当前学界和商界的迫切需求, 这就大大地促进了文本情感分析技术的发展^[1]。

文本情感分析依据其处理的文本粒度不同可以分为: 文档级, 句子级, 词语级。文档级情感分析目

的是对文档表达的整体情感倾向进行判别分类, 通常重点关注话题和与话题相关的观点词和情感词。句子级情感分析是一个三元分类问题, 其重点在于判断句子是否为主观句, 并在此基础上判别主观句的情感倾向是正面、负面或中立。词语级情感分析主要包括基于词典的情感分类方法和基于语料的情感分析方法, 前者通过同义、反义以及语义层次来构建情感词, 而后者通常人工初始少量情感词作为种子, 然后通过某种统计方法或者句式分析方法在大规模语料数据中判别其他词语的情感倾向性。

对 Twitter 短文本开展情感分析研究, 通过判断 Twitter 文本作者的情感状态是积极的肯定赞赏、中

基金项目: 国家自然科学基金(61373108)。

作者简介: 王汝娇(1992—), 女, 硕士研究生, 主研方向为自然语言处理; 姬东鸿, 教授、博士、博士生导师。

收稿日期: 2017-01-09 **修回日期:** 2017-02-26 **E-mail:** kewpie1992@163.com

立观望或者消极的否定批判,获取 Twitter 舆论话题的倾向性,把握舆论的走向。但是, Twitter 短文本区别于传统文本,其长度通常只有 1 个~100 个词左右, Twitter 中包含的信息总量虽然庞大,但是单条 Twitter 短文本包含的特征少,并且 Twitter 短文本通常不遵守语法句法特性,组词造句存在很强的随意性^[2]。因此, Twitter 短文本的稀疏性和不规范性对情感分析带来了较大的影响和干扰。

本文提出一种基于卷积神经网络和多特征融合的 Twitter 情感分类方法。针对 Twitter 文本独特的语言特性,构建适用于 Twitter 文本情感分类的语料特征和词典特征。然后将 Twitter 文本词向量作为卷积神经网络的输入,通过训练神经网络获取深度词向量特征。最后,将语料特征、词典特征和深度词向量特征进行特征融合生成情感特征,使用支持向量机 SVM 对积极、中性、消极 3 类文本进行一对一的模型训练和分类,从而实现 Twitter 文本的情感分类判别。

1 相关工作

1.1 情感分析

情感分析主要任务是分析文本的主客观性,然后对主观性语句的情感倾向以及情感强度进行判别分析^[3]。目前文本情感分析主要可分为基于情感词典和基于机器学习两大类。

1) 基于情感词典方法,该类方法是利用情感词典、领域词典等人工编写的词典模板来获取最终的文本情感极性,其中最为关键的是要有一个能准确评价情感强度的情感词典。文献[4]情感词典包含 2 006 个正性词(Positive Word)和 4 783 个负性词(Negative Word),该词典不仅包含正常的用词,还包含了拼写错误、语法变形、俚语以及社交媒体标记等。文献[5]主观词典包含 2 718 个正性词和 4 912 个负性词并标示出了词性、词根以及定义了词语情感极性的强弱。NRC Hashtag 词典和 Sentiment 140 词典^[6]这 2 个词典是专门针对 Twitter 文本挖掘创建的,分别用 Hashtag 和表情符号作为某条 Twitter 是正面或负面的信号,共收集了 775 000 个有情感表达 Hashtag (如难过、生气、开心等)的 Twitter 数据,人工挖掘了与正面情感或负面情感联系紧密的 78 个种子词汇,然后计算词语偏向正面情感的互信息与偏向负面情感的互信息的差值作为该词的情感得分。General Inquirer 词典^[7]包含 1 914 个正性词和 2 293 个负性词,并依据每个词语词性和强度不同打上不同的标签值。LIWC 词典^[8]包含 70 多个类别的 2 300 个词语,该词典包含影响情感分类判别的介词、否定词和表示程度的副词等。

2) 基于机器学习方法,其重点在于选取合适的特征来表征文本,从人工标注情感类别的文本中提

取特征,用于分类器的训练和构建,最终利用机器学习方法获得文本的情感极性。文献[9]通过提取文本的词性、否定词等特征,使用有监督方法进行文本情感分类。文献[10]提出一种基于情感词图模型的方法,将 PageRank 算法引入情感词权重中,使用 SVM 进行微博语料的情感分析。文献[11]通过对微博层次结构的多策略方法的特征进行对比分析研究,重点分析了表情符号的规则方法、情感词典的规则方法以及基于 SVM 的层次结构的多策略方法。文献[12]提出一种基于 Bootstrapping 的半监督情感分析方法,通过自学习来扩大情感类别标记的规模,使用迭代学习提高 SVM 情感分类的准确性。但上述两大类方法并非完全独立,基于情感词典的机器学习情感分类方法也是研究热点之一。

Twitter 微博属于短文本,针对其开展的句子级情感倾向性分析的最终目的是将 Twitter 文本的情感极性分为正向、负向和中性 3 大类。目前,国内外的研究学者已经在 Twitter 情感分析领域做出了诸多探索,主要是分析整个句子中的主客观性,并结合微博所表现的总体情感倾向性进行情感分类。近年来涌现了大量对 Twitter 文本情感分析的研究,文献[13]提到了 Twitter 情感分析面临的一些困难:数据的稀疏性,非标准词汇的使用,语法缺乏规范性以及 Twitter 的多语言特性。研究者设计了 2 种不同的思路解决上述这些困难:一种是基于 Twitter 本身的内容,另一种是使用 Twitter 的上下文。基于内容的模型使用每个 Twitter 的字符 n-gram 特征图表示,语言的独立性和语法、句法的错误变得不相关,因此分析效果更好。文献[14]则是构建了一个基于 Twitter 领域的词法器,然后利用 N-gram 词组模型作为情感特征并通过动态神经网络 DAN2 对 Twitter 消息进行分类。文献[15]针对含有 Hashtag 的 Twitter 进行分析,提出了 Hashtag 图模型,将 Hashtags 之间的共现关系和 Hashtag 的字面意思并入到分类框架中,从而辅助极性判别。针对 Twitter 文本特性构建了一系列文本特征,包括一元词、二元词、否定文本个数、情感词数量等,最后采用 lib-SVM 进行情感分类^[6]。文献[16]设计了一个 Twitter 情感分析系统,能抓取 Twitter 中与总统选举有关的社会舆论信息并进行情感性分析。

1.2 深度学习

上述这类基于浅层机器学习的方法已在情感分类中得到广泛应用,但这类基于传统机器学习方法较为依赖构造分类器的情感词典特征以及分类器模型的参数调优。为解决这一问题,目前已有研究学者将深度学习运用到自然语言处理领域中,深度学习模型能够更加高效地完成特征提取任务,并且提取出的特征能够更好地反映出文本语义信息。

深度学习作为一种新兴的神经网络学习算法,开启了机器学习二次革命的浪潮。深度学习依据学习方法的不同可以分为深层无监督学习和深层监督学习。1)深层无监督学习可以看作是多个浅层无监督学习器的叠加,输入是数据的原始特征,输出是数据新的表达,模型中的任意一层输出均可以看作是数据的特征表达,代表方法包括限制玻尔兹曼机 Deep Boltzmann Machines、自编码器 Autoencodes 等;2)深层监督学习方法输入也是数据的原始特征,但输出是数据的 label 标签,与传统的浅层监督方法最大的不同在于深层监督学习方法在学习迭代过程中会生成很多数据的中间表达或者特征,因此深层监督学习方法中间层的学习或者特征提取过程可以看作是一个特征自学习或者特征表达的过程,代表方法包括深信度网络(Deep Belief Network, DBN)、深度神经网络(Deep Neural Network, DNN)、卷积神经网络(Convolutional Neural Network, CNN)等。

文献[17]采用卷积神经网络 CNN 实现了一个多层结构的神经网络学习算法, CNN 利用空间相对关系减少参数数目,从而提高反向传播 BP 算法的训练性能。它的权值共享网络结构使之更类似于生物神经网络,减少了权值的数量,降低了网络模型的复杂度。近年来, CNN 被用于自然语言处理领域和情感分析问题上并取得了良好的效果^[18]。文献[19]提出了一个可以同时利用基于特定任务调整的词向量和静态的预训练词向量的卷积神经网络模型,用于句子级别的文本分类任务。文献[20]改进了 CNN 的结构,进行了两次卷积操作,并提出了动态 K-max pooling 池化策略,返回每一个 feature maps 中最大的 K 个值。文献[21]受图像有 RGB、CMYK 多通道的启发,将句子视为图像,将一个 d 维的词向量看成一个有 d 个通道的像素,在卷积层中,每一个卷积核作用于维度为: filter 长度与词向量维度的乘积即 $P \times |V|$ 的 region 上,为了降低需要学习的参数数量,将 P 个词向量相加,从而将参数数量由 $P \times |V|$ 降低为 $|V|$ 。不同于上述论文中只使用单词的词向量表示,文献[22]同时使用了字符级的向量,将词向量和字符向量合并在一起作为其向量表示。文献[23]针对句子匹配的问题,提出了一种基于 CNN 的句子建模网络,模型解决句子输入长度不一致的方法是规定最大的可输入句子长,然后长度不够的进行补 0 填充。文献[24]通过选择使用词性标注、Word embedding 等方法增强模型性能,并且在卷积、池化以及句子表征方面进行局部相似度计算来提升性能。文献[25]提出一种使用 Siamese 框架的 CNN 模型,并且采用多粒度的交互特征,一方面提高了模

型的分类性能,另一方面增强了模型的鲁棒性。文献[26]构建了一个深度卷积神经网络模型。针对训练集用 word2vec 训练出每个词的词向量,然后把它们输入 CNN 中并作为网络参数一起参与迭代训练过程。最后由 CNN 中最后一层 Softmax 层进行分类输出情感极性。

此外,深度学习中递归自编码方法也广泛应用在情感分析中。文献[27]提出一种基于递归自编码方法对句子级文本进行情感分析,使用了无监督 RAE 方法学习句子组合机构,采用有监督方法训练带情感标签的数据通过分类判别获取情感极性。文献[28]设计了一种基于 RAE 的微博情感分析方法,通过将低维实数向量表示的文本作为 RAE 情感极性转移模型的输入,最终获得情感分类判别的极性。

然而,单纯使用深度学习方法处理情感分类问题会忽略现有情感分类领域积累的大量情感词典、知识库、人工特征等资源。此外, Twitter 文本是一种口语化的非规范文本,受字数限制,通常就是简短的一句话,甚至是若干单词的组合,缺乏必要的上下文语境关联,导致单纯地使用词向量模型来表示文本特征效果可能不太理想。现阶段情感分类领域积累的大量情感词典、知识库、人工特征等资源,其中关于情感词、否定词、标点符号等这些短词精心设计的特征在情感分析中也取得不错的效果。本文将这类特征与深度词向量特征进行融合,以实现 Twitter 短文本的情感分类。如何合理科学地设计深度卷积神经网络,将深度学习技术与已有的情感分类资源进行有效融合,提高情感分类准确性是本文研究的切入点和突破口。

2 本文方法描述

本文提出一种基于卷积神经网络和特征融合的 Twitter 情感分类方法,首先预先对 Twitter 文本进行预处理,将 Twitter 文本转化为统一规则的规范化文本,并结合 Twitter 文本的语言特性,构建 Twitter 文本情感分类的语料特征和词典特征。然后使用 GloVe 获取 Twitter 文本的词向量表示,并将文本词向量特征输入本文设计的卷积神经网络中,通过网络模型训练获取深度词向量特征。最终用于情感分类特征为语料特征、词典特征和深度词向量进行特征融合的结果,最后将融合特征使用 One-Versus-One SVM 支持向量机进行 3 类情感极性的一对一模型训练和分类。

2.1 Twitter 文本预处理

由于 Twitter 文本并不像传统媒体发布的文本那样规范正式,这种非正式文本通常很短不超过

140 字,多数都是一句话、一个标题,所使用的语言有其自己的特色,内容包含有创意的拼写和标点、表情符号、俚语、新词、URL、特定术语缩写、用户名以及 Twitter 中特有的主题标签 #Hashtag 等,这些非正式文本给情感分类带来了极大挑战。

为了去除不必要的噪声干扰,本文预先对 Twitter 文本进行预处理,具体规则如表 1 所示。针对 8 种非规范对象进行规范化转换,并在转换后的文本对象后添加对应的标记,从而将原始 Twitter 文本转化为统一规则的规范化文本。

表 1 Twitter 文本规范化预处理规则

非规范对象	标记	规范化转换示例
用户名	<USER>	"@ elifyk1" => "<USER> @ elifyk1"
URL 链接	<URL>	"http://bit.ly/XQEhJU" => "<URL> http://bit.ly/XQEhJU"
数字	<NUMBER>	"32" => "<NUMBER> 32"
主题标签	<HASHTAG>	"#NBA Final" => "<HASHTAG> #NBA Final"
大写文本	<ALLCAPS>	"LAKERS DICK" => "<ALLCAPS> LAKERS DICK"
连续标点	<REPEAT>	"!!!" => "<REPEAT> !!!"
刻意延长词	<ELONG>	"wayyyy" => "<ELONG> wayyyy"
表情符号	<EMOTICON>	":D" => "<EMOTICON> :D"

经过规范化处理后,得到相对规范化的 Twitter 文本,在此基础上提取语料特征、词典特征以及深度词向量特征。

2.2 语料特征和词典特征提取

Twitter 文本受到长度限制,通常内容简单直观,整体表达内容单一,导致 Twitter 文本中情感表达的词语和短语往往就代表该文本的倾向性。例如:"They may have a SuperBowl in Dallas, but Dallas ain't winning a SuperBowl.", 其中"SuperBowl"和"Dallas"代表文本的主题内容,而"but"和"ain't"就代表了文本的情感倾向。针对 Twitter 短文本的独特语言特性,本文设计了 Twitter 文本的语料特征和词典特征用于对其进行情感分类。

如表 2 所示,使用 6 种语料特征,总计 30 维:

1) 标签的数量:该标签使用 Twitter 词性标注工具包"ark-tweet-nlp",其中定义了 24 种标签,包括名词、表情符号、用户名等。

2) Hashtag 主题标签数量:Hashtag 是 Twitter 中用来标注线索话题的标签,如 #NBA Final 则代表这条 Twitter 与 NBA 篮球总决赛有关。

3) Twitter 中否定上下文的数量:否定上下文的定义是以否定词(no, wouldn't 等)开始,以标点符号“,”“.”“:”“!”“?”为结尾。

4) Twitter 中纯大写的词数量,如 GOOD 等。

5) Twitter 中连续出现的标点符号序列的个数以及最长序列的长度,如“!!!”“!?”“?!?”“??”等。

6) Twitter 中是否出现刻意延长词:如 looove、sooo 等。

表 2 Twitter 短文本的语料特征

特征描述	特征维度
标签数量	24 维
Hashtag 主题标签数量	1 维
否定上下文数量	1 维
纯大写词数量	1 维
连续标点符号个数和长度	2 维
是否出现刻意延长词	1 维

除语料特征外,还选取了 4 个情感字典,分别计算对应的 4 种情感分数,得到总计 16 维的词典特征。

对于 Liu Bing 词典和 MPQA 主观词典,计算如下特征:

- 1) 计算肯定上下文中包含 token 的正面分数;
- 2) 计算肯定上下文中包含 token 的负面分数;
- 3) 计算否定上下文中包含 token 的正面分数;
- 4) 计算否定上下文中包含 token 的负面分数。

否定上下文的定义与语料特征 3) 中定义一致,除去否定上下文的剩余部分即为肯定上下文。其中,对于 Bing Liu's Lexicons 词典中的情感词,正面 token 给予 +1 的情感分数,负面 token 给予 -1 的情感分数。对于 MPQA Subjectivity Lexicon 词典中的情感词,由于该词典考虑了情感极性的强弱,因此给予弱情感词的情感分数分别为 +1/-1,强情感词的情感分数分别为 +2/-2。

对于 NRC Hashtag 和 Sentiment 140 词典,计算如下特征:

- 1) 情感分数不为 0 的 token 总数;
- 2) 语句中所有 token 的情感分数之和;
- 3) 语句中 token 的情感分数最大值;
- 4) 语句中最后一个 token 的情感分数。

将 30 维语料特征和 16 维词典特征组合在一起获得本文构建的 46 维特征。

由于该 46 维的特征维度间差异较大,为避免分类中某一维特征支配整个分类过程对结果过分干扰,对 46 维语料特征和词典特征进行线性归一化,归一化公式如下:

$$Feature_{\text{norm}} = \frac{Feature - Feature_{\min}}{Feature_{\max} - Feature_{\min}}$$

其中, $Feature_{\text{norm}}$ 为某一维特征归一化后的结果; $Feature$ 为某一维的原始特征; $Feature_{\max}$ 和 $Feature_{\min}$ 分别为某一维原始特征中的最大值和最小值,通过归一化将 46 维特征进行等比例缩放。

2.3 Twitter 文本词向量表示方法

在使用卷积神经网络处理自然语言问题时,通常以矩阵表示的句子,采用一个低维的实数向量来表示词语以避免数据维数灾难。Twitter 文本情感分析的输入是短文本,可以看作是一串单词序列的组合,那么矩阵的每一行对应于一个分词元素或者一个单词,矩阵中每一行表示一个单词的向量,这些向量都是 word embeddings 的形式,例如 Word2Vec 和 GloVe 等。这样就将文本数据用低维度实数向量表示,然后可作为卷积神经网络的输入。

本文采用了斯坦福开源工具 GloVe (Global Vector for Word Representation) 获得文本词向量表示^[29]。它提供了一个用大量 Twitter 语料预先训练好的词向量 glove.twitter.27B, 里面包含有 25 维、50 维、100 维以及 200 维 4 种不同维度的词向量。

词向量可将每一个词语映射到 d 维实数向量空间上,假如词向量空间 Ω 的大小为 $|\Omega| \times d$, $|\Omega|$ 是词表的大小,代表着词向量中包含的词语个数, Ω 中每一行表示一个单词的 d 维的词向量。例如,输入的一条 Twitter 语料 E 可以单词序列 $[e_1, e_2, \dots,$

$e_{|E|}]$ 表示,其中, $|E|$ 表示语料 E 中包含的词语个数,即给定句子 E , 包含 $|E|$ 个词语 $e_i, 1 \leq i \leq |E|$ 。将语料 E 转换为词向量矩阵时, $\forall e_i \in E, 1 \leq i \leq |E|$, 在 Ω 中搜索 e_i 的词向量,若存在则选取它对应的实数向量 θ_i 表示,若不在 Ω 中则将其向量置为 0, 即 $\theta_i = 0$ 。通过将各个词语的词向量堆放在一起,最终语料 E 可转换为大小为 $|E| \times d$ 的词向量 S 特征矩阵表示,即:

$$[e_1, e_2, \dots, e_{|E|}] \Rightarrow \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{|E|} \end{bmatrix}$$

词向量特征矩阵 S 的每一行代表输入的 Twitter 文本中一个单词的词向量。

2.4 卷积神经网络模型

如图 1 所示,本文构建的卷积神经网络模型由卷积层、池化层、全连接层和 softmax 层组成,每一层的输出均为下一层的输入。

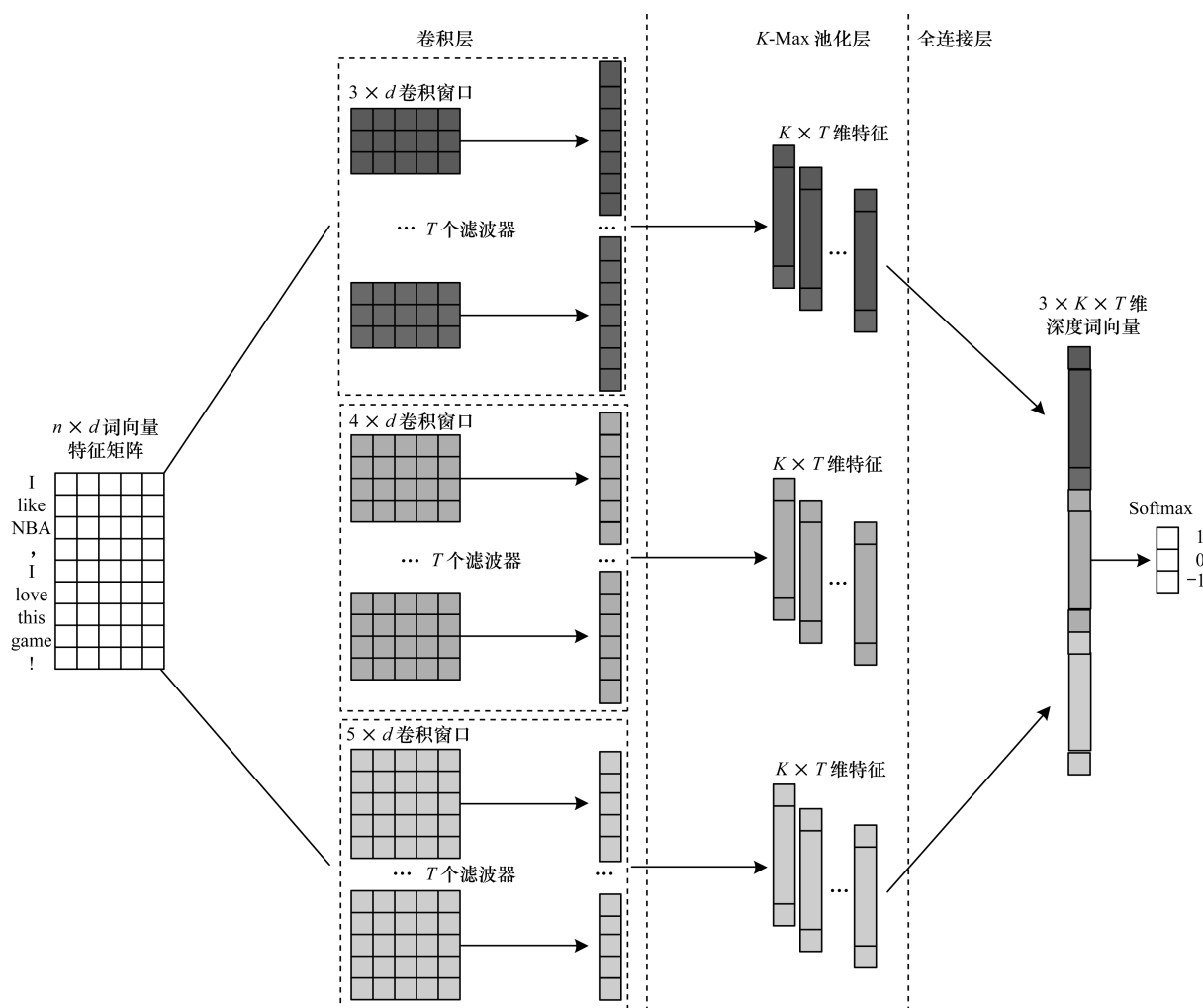


图 1 深度词向量特征提取算法

利用该模型获取情感分类的深度词向量特征,卷积神经网络构造方法如下:

1) 卷积层

卷积层接收大小为 $n \times d$ 的 Twitter 短文本的词向量特征矩阵 S , 其中 S 包含 n 个单词, 词向量维度为 d , 矩阵 S 每一行表示一个单词的 d 维词向量。选取大小为 $m \times d$ 的滤波器 F 来对特征矩阵 S 进行卷积操作, 从而获得新的特征值 c_i , 其中 i 表示第 i 个特征值, 则:

$$c_i = (F * S)_i = f(w \times S_{i:i+m-1} + b)$$

其中, f 是一个非线性卷积核函数, m 表示卷积计算中滑动窗口的大小, w 为大小 $m \times d$ 的滤波器 F , b 是一个偏置项, $S_{i:i+m-1}$ 表示从 S 的第 i 行至第 $i+m-1$ 行抽取的局部特征矩阵。这个卷积窗口被应用到所有可用的局部特征矩阵 $S_{1:m}, S_{2:m+1}, \dots, S_{n-m+1:n}$ 中, 最终生成了特征向量 C 。

$$C = [c_1, c_2, \dots, c_{n-m+1}]$$

2) 池化层

由于每条 Twitter 文本的单词长度不确定, 词向量特征矩阵 S 大小也不确定, 最终导致卷积层输出特征维度不固定, 因此需要对获得的特征向量 C 进行池化操作。

池化操作采用 K-Max Pooling 方法, 对于某个卷积核获得的若干特征值, 只保留其特征值中得分在前 K 的特征值, 并且保留这 K 个特征在卷积后的先后顺序。

$$\hat{c} = \max(C, K)$$

当 $K=1$ 时, K-Max Pooling 方法与 Max Pooling 方法一致, 即取得最大值。

通过池化 Pooling 层可以固定全连接层的神经元个数, 也固定了全连接层输出特征的长度, 自然地解决了 Twitter 文本情感分析中句子长度不一致干扰特征提取的问题。

3) 全连接层

上述过程描述了卷积窗口为 $m \times d$ 的特征提取过程, 而在整个卷积神经网络中使用了多个不同大小卷积核。本文使用了 3 种不同尺寸的卷积核, 分别为 $m=3, m=4, m=5$ 。每个尺寸卷积核中包含 T 个滤波器, 每个滤波器产生的特征中又池化获得 K 个特征。因此, 所有特征在全连接层汇聚为最终 $3 \times K \times T$ 维的深度词向量特征 V :

$$V = \begin{bmatrix} \hat{c}_{1 \cdot 3 \times d}^1 & \dots & \hat{c}_{1 \cdot 3 \times d}^K & \dots & \hat{c}_{T \cdot 3 \times d}^1 & \dots & \hat{c}_{T \cdot 3 \times d}^K \\ \hat{c}_{1 \cdot 4 \times d}^1 & \dots & \hat{c}_{1 \cdot 4 \times d}^K & \dots & \hat{c}_{T \cdot 4 \times d}^1 & \dots & \hat{c}_{T \cdot 4 \times d}^K \\ \hat{c}_{1 \cdot 5 \times d}^1 & \dots & \hat{c}_{1 \cdot 5 \times d}^K & \dots & \hat{c}_{T \cdot 5 \times d}^1 & \dots & \hat{c}_{T \cdot 5 \times d}^K \end{bmatrix}$$

其中, $\hat{c}_{T \cdot 3 \times d}^1$ 表示大小为 $3 \times d$ 的卷积核在第 T 个滤波器所形成的特征经 K-Max 池化生成最大的特征值, $\hat{c}_{1 \cdot 5 \times d}^K$ 表示大小为 $5 \times d$ 的卷积核在第 1 个滤波器所形成的特征经过 K-Max 池化生成第 K 大的特征值。通过将不同尺寸卷积核、不同滤波器生成的特征值进行组合获得情感分类特征 V 。

4) Softmax 层

全连接层输出的特征 V 利用 Softmax 输出情感极性 3 分类 (正面、负面、中立) 判别的结果, 并依据训练 Twitter 文本数据中的真实标签, 采用反向传播算法对卷积神经网络中的参数进行梯度更新:

$$P(\text{result} | V, W_x, b_x) = \text{softmax}_x(W_x \cdot V + b_x)$$

其中, softmax 为分类器, result 表示情感极性判别的结果 $\text{result} \in \{+1, 0, -1\}$, 下标 x 代表第 x 轮的梯度更新。

5) 情感分类模型的训练和判别

当整个 CNN 网络梯度更新收敛后, 获得最终用于情感分类的深度词向量特征。将语料特征、词典特征和深度词向量特征组合在一起作为情感分类的特征, 并采用支持向量机 SVM 对训练和测试样本进行模型训练和分类判别, 其中: 情感分类特征 = [语料特征; 词典特征; 深度词向量特征]。

由于情感分类面对的是一个 3 类分类的问题, 因此本文使用 One-Versus-One SVM 进行分类, 在积极、中性、消极中任意 2 类之间设计一个 SVM, 总计设计 3 个 SVM 分类器, 当对一个未知语句 Sample 进行情感类别分类时, 使用 3 个分类器进行分类判决投票, 得票最多的情感类别即是该语句的情感类别。具体分类投票算法如下:

$S_{\text{Negative}}, S_{\text{Neutral}}, S_{\text{Positive}}$ 为消极、中性、积极 3 类 Twitter 情感语句的训练样本, 训练生成 3 个分类模型:

$$\text{Model}_1 = \text{Train}(S_{\text{Negative}}, S_{\text{Neutral}})$$

$$\text{Model}_2 = \text{Train}(S_{\text{Neutral}}, S_{\text{Positive}})$$

$$\text{Model}_3 = \text{Train}(S_{\text{Positive}}, S_{\text{Negative}})$$

初始化 $\text{Negative} = \text{Neutral} = \text{Positive} = 0$, 如果 $\text{Classifier}(\text{Model}_1, \text{Sample}) = \text{Negative}$, 则 $\text{Negative} = \text{Negative} + 1$, 否则 $\text{Neutral} = \text{Neutral} + 1$; 如果 $\text{Classifier}(\text{Model}_2, \text{Sample}) = \text{Neutral}$, 则 $\text{Neutral} = \text{Neutral} + 1$, 否则 $\text{Positive} = \text{Positive} + 1$; 如果 $\text{Classifier}(\text{Model}_3, \text{Sample}) = \text{Positive}$, 则 $\text{Positive} = \text{Positive} + 1$, 否则 $\text{Negative} = \text{Negative} + 1$; 最终 Sample 的情感类别 Label 为: $\text{Label} = \max(\text{Negative}, \text{Neutral}, \text{Positive})$ 。

3 实验与分析

3.1 Twitter 情感分析数据集

如表 3 所示,本文选用了 SemEval 发布的关于 Twitter 情感分析任务所提供的公开语料数据集来验证本文方法的有效性。将 SemEval2013、SemEval2014 及 SemEval2015 数据集内所有 Twitter 语料合成在一起,去除重复和标注自相矛盾的 Twitter 短文本,最终获得 9 396 条已标注情感极性的短文本,其中情感极性判定积极的有 3 574 条,消极的为 1 344 条,中性的为 4 478 条。由于 3 个类别的样本数目存在差异,这种数量的不平衡性会导致分类器训练的过程中严重偏向训练样本多的类别,对分类器性能造成较大的损失。因此,本文选取训练集时保证了 3 种类别的训练样本数量一致。

表 3 SemEval 语料集合样例

情感类别	语料样例
积极	1) 007 in Detroit wed august 17th I' ll be dancing live!!!! D town stand up!!! Get ready its going down!
	2) Keen for the eagles game today. #goeagles
	3) Our new dance instructor for # ClubFitness has toured w/ Beyonce!!! You know our Beyonce themed night is going to be CRAZY!!!
中性	1) Im not watching them because the Red Sox vs Yankees game is on tonight for who's in first.
	2) I stay up til 6:30 every morning just to watch Between the Lions.
	3) If Cundiff would've made the Field Goal Sunday, I could've been eatin' a free 6in sub at Subway for lunch!
消极	1) And whilst listening to Dashboard with a hangover, I may as well slit my wrists.
	2) I never had a good time, I sat by my bedside. With papers and poetry about Estella.
	3) OKC Thunder center Kendrick Perkins may miss the preseason due to a wrist injury.

如表 4 所示,实验中随机抽取积极、中性、消极样本各 800 条,总计 2 400 条语料组成训练集;将剩余 6 996 条语料作为测试语料。其中制作 2 个测试集:测试集 1 为全部剩余的 6 996 条语料均作为测试集。测试集 2 为在 6 996 条测试集语料中积极、中性、消极样本各抽取相同的 544 条。

表 4 实验数据集语料分布

样本类型	训练集	测试集 1	测试集 2
积极样本数量	800	2 774	544
中性样本数量	800	544	544
消极样本数量	800	3 678	544
总数量	2 400	6 996	1 632

3.2 评价指标

本文选择精确率 P 、召回率 R 以及 F 值作为评价指标。精确率 P 表示情感分类后,实际属于某个类别的文本数量与情感分类判断属于某个类别的文本数量的比值: $P = a/(a + c)$ 其中, a 、 b 、 c 和 d 的定义如表 5 所示。

表 5 情感分类混淆矩阵

判断属性	实际上属于某个情感类别的文本	实际上不属于某个情感类别的文本
判断属于某个情感类别的文本	a	c
判断不属于某个情感类别的文本	b	d

召回率 R 表示情感分类后,情感分类判断属于某个类别的文本数量与实际属于某个类别的文本数量的比值:

$$R = a/(a + b)$$

为了综合考虑精确率和召回率,使用两者的加权调和平均数 F 来衡量分类效果:

$$F = \frac{2 \times P \times R}{P + R}$$

由于本文设定的情感类别包括消极、积极和中性 3 个类别,因此采用了平均精确率 av_P 、平均召回率 av_R 、平均 av_F 值作为最终衡量情感分类方法的性能的评价指标,即先对不同类别单独计算 P 、 R 和 F 值,再取算术平均值。

3.3 实验模型参数设置

卷积神经网络中参数和函数设置如表 6 所示。使用 ReLU 作为激活函数,采用 Adadelta 方法作为梯度更新规则。而参数 T 和 d 则采用单因子变量法实验确立最优参数。实验采用表 4 中训练集和测试集 2 种样本,使用本文方法测试在不同滤波器数 T 和词向量维度 d 下的 av_F 值。

表 6 实验参数设置

参数	参数/函数名称	参数/函数设置
—	激活函数	ReLU (Rectified Linear Units)
—	优化方法	Adadelta 方法
m	滤波器窗口大小	3, 4, 5
x	神经网络迭代次数	50 次
K	K-max 最大池化值	2
T	滤波器个数	需对比实验 (50, 80, 100, 120)
d	词向量维度	需对比实验 (25, 50, 100, 200)

如表 7 所示,随着词向量维度 d 的增加,情感分类的效果有明显提升;而当滤波器的数量 T 增多

时,情感分类 F 值有所提升,但当 T 达到 100 个后,对情感分类的提升作用较小。所以,在卷积神经网络训练复杂度和分类器训练时间没有显著增加的前提下,最终选择 100 个滤波器数量、200 维词向量,深度词向量特征为 $3 \times 2 \times 100$ 总计 600 维。在此条件下,本文设计的情感分类方法准确性能达到 0.704。

表 7 不同滤波器数 T 和词向量维度 d 下的情感分类结果

d 值	$T = 50$	$T = 80$	$T = 100$	$T = 120$
20	0.528	0.559	0.629	0.640
50	0.550	0.551	0.640	0.634
100	0.559	0.627	0.684	0.689
200	0.638	0.667	0.704	0.702

表 8 给出了本文方法针对测试集 2 中 1 632 条 Twitter 语料的分类结果,可以看出本文方法训练出来的情感分类器在消极和积极语料中表现较好, F 值分别为 0.743 和 0.726,针对中性 Twitter 语料的 F 值为 0.643。

表 8 本文方法的预测准确性

情感类别	精确度 P	召回率 R	加权调和平均数 F
消极 Negative	0.761	0.726	0.743
中性 Neutral	0.612	0.676	0.643
积极 Positive	0.741	0.711	0.726

3.4 对比实验

为验证本文提出的基于卷积神经网络和特征融合的 Twitter 情感分类方法的有效性,将本文方法与 N-gram 词袋模型、UNITN 模型进行对比实验,具体实验如下:

- 1) 方法 1:采用 uni-gram 特征,使用 SVM 分类;
- 2) 方法 2:采用 uni-gram 和 bi-gram 特征,使用 SVM 分类;
- 3) 方法 3:采用 uni-gram、bi-gram 和 tri-gram 特征,使用 SVM 分类;
- 4) UNITN-CNN:采用 word2vec 训练的词向量,并迭代训练,最后使用 CNN 的 Softmax 进行分类;
- 5) Corpus-Lexicons-SVM:采用本文提出 46 维的语料特征和词典特征,使用 One-Versus-One SVM 分类;
- 6) GloVe-CNN-SVM:采用本文提出的深度词向量特征,使用 One-Versus-One SVM 分类;
- 7) 本文方法:采用本文提出的语料特征、词典特征和深度词向量融合情感特征,使用 One-Versus-One SVM 分类。

上述 7 种不同情感分类方法在表 4 中 2 个测试集下对比实验结果如表 9 所示。

表 9 不同方法的分类准确性对比

方法	测试集 1			测试集 2		
	av_P	av_R	av_F	av_P	av_R	av_F
方法 1	0.563	0.512	0.536	0.517	0.526	0.522
方法 2	0.551	0.542	0.546	0.550	0.545	0.547
方法 3	0.598	0.581	0.590	0.600	0.591	0.596
UNITN-CNN	0.686	0.689	0.688	0.686	0.695	0.690
Corpus-Lexicons-SVM	0.581	0.589	0.585	0.598	0.604	0.601
GloVe-CNN-SVM	0.676	0.689	0.683	0.686	0.691	0.688
本文方法	0.703	0.708	0.706	0.706	0.702	0.704

通过对比方法 1、方法 2 和方法 3 发现:引入 bi-gram 和 tri-gram 特征能显著提升情感分类的准确性,但是随着 N-gram 模型 N 值变大,会导致模型训练难度增大。

通过对比 N-gram 词袋模型和 UNITN-CNN 发现:卷积神经网络使用不同尺寸的多个卷积核来学习文本词向量特征,并且通过池化层来固定全连接层的神经元个数,降低了输出特征的维数,避免维数灾难的出现,并且也一定程度地提升了检测准确率。

通过对比 UNITN-CNN 和本文方法发现:本文方法在采用 GloVe 词向量作为原始特征并使用 One-Versus-One SVM 分类器取代 CNN 的 softmax 进行模型训练和情感分类时能够提高情感分类精确度。

本文设计特征融合方法(对比 Corpus-Lexicons-SVM、GloVe-CNN-SVM 与本文特征融合方法):这种特征融合在保证深度词向量特征反映文本上下文语义关系和其中蕴含的情感信息表达的基础之上,又采用如情感词得分、延长词、连续标点等特征增强了对原词向量的情感信息表达。实验结果证明,将语料特征、词典特征与深度词向量特征进行特征融合能够有效地提高情感分类的准确性。

此外,从 SemEval 数据集中抽选了 6 条 Twitter 语句,如表 10 所示,分别给出了它们采用方法 3、UNITN-CNN 以及本文方法的情感分类结果。其中人工标记“-1”代表该条语句被标记为消极、“0”代表中性、“1”代表积极。

1) 样例 1 中包含了“great”“awesome”等明显的正面极性词,样例 3 中出现了“hate”“bad”“no”明显的负面极性词,样例 5 为不带情感色彩的陈述句,所以 3 种方法都做出了准确的情感判别。

2) 样例 2 不是情感词单独起作用,而是通过词语的序列表达出逻辑关系。CNN 处理此类样例时,

由于出现了“not”“didn't”等大量的负面词,从而被误分类为消极的情感倾向。而本文方法中的深度词向量特征反映了词语间的序列关系,“may not”是否定词,“but”表示转折关系然后接着“amazing”是正面极性词,因而被判定为积极的情感倾向。深度词向量特征与语料特征、词典特征的融合纠正了仅靠情感词判别导致的分类错误,在考虑词语的情感极性的同时也兼顾了词语序列表达的逻辑,从而做出正确的分类。

表 10 测试集实验结果样例

序号	SemEval 数据样例	方法 3	UNITN-CNN	本文方法	人工标记
1	Today's a great day to splurge on an awesome bottle of Hollywood Tans lotion with our Treat Your Tuesday daily special! Call 866-970-8267...	1	1	1	1
2	I may not be a fan of the Spurs, but they play some amazing basketball.	1	-1	1	1
3	I hate watching movies like Friday after next on VH1, BET, MTV, & etc. because they take out the bad language - _ - . It's no fun !	-1	-1	-1	-1
4	I could have had the BEST DAY EVER!.. if I had went to tha Nicki Minaj concert in AC tonight...	1	1	1	-1
5	I'm going to the ravens game in philly this Thursday #getthefucoutofhere	0	0	0	0
6	They spontaneously sang "There's a Great Big Beautiful Tomorrow" out loud in the grocery store today.	1	1	1	0

综上所述,对于包含虚拟语气、反语以及一些既定事实陈述句的无情感转述容易影响情感分类的准确性,针对这一类语句的情感分类研究也是后续研究的重点方向。

4 结束语

本文提出了一种基于卷积神经网络和多特征融合的 Twitter 情感分类方法,根据 Twitter 文本的独特性,构建了适用于 Twitter 文本情感分类的语料特征和词典特征。在此基础上,通过卷积神经网络获取 Twitter 文本深度词向量特征,并将语料特征、词典特征和深度词向量特征融合为情感分类特征,最终使用 One-Versus-One SVM 进行分类,实验结果显示本文方法能够有效地对 Twitter 文本进行情感分类。下一步将针对一些复杂语法结构的短文本进行研究,从而进一步提高 Twitter 文本情感分类的准确性。

参考文献

- [1] 陆文星,王燕飞. 中文文本情感分析研究综述[J]. 计算机应用研究,2012,29(6):2014-2017.
- [2] KOULOUMPI E, WILSON T, MOORE J D. Twitter Sentiment Analysis: The Good the Bad and The Omg! [C]// Proceedings of International Conference on Weblogs & Social Media. Berlin, Germany: Springer, 2011: 538-541.

3) 样例 4 从词语的极性来说“best”是正向极性词,词语序列关系也没有出现转折,但该样例是一种虚拟语气的表达,其实表达的是一种消极的后悔情绪,所以 3 种方法都将其误分类为积极的情感倾向。

4) 样例 6 陈述了他们在杂货店里自发的大声唱歌这样一个事件,并没有任何情感倾向性。然而在语句中出现“great”“beautiful”明显的正向极性词(实际为歌词),因此系统将其误判为积极的情感倾向。

- [3] 赵妍妍,秦兵,刘挺. 文本情感分析[J]. 软件学报, 2010,21(8):1834-1848.
- [4] HU Mingqing, LIU Bing. Mining and Summarizing Customer Reviews[C]//Proceedings of the 10th ACM International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2004:168-177.
- [5] WILSON T, WIEBE J, HOFFMANN P. Recognizing Contextual Polarity in Phrase-level Sentiment Analysis[J]. International Journal of Computer Applications, 2005, 7(5): 347-354.
- [6] MOHAMMAD S M, KIRITCHENKO S, ZHU Xiaodan. NRC-Canada: Building the State-of-The-Art in Sentiment Analysis of Tweets [C]//Proceedings of the 7th International Workshop on Semantic Evaluation. Berlin, Germany: Springer, 2013:1-9.
- [7] The General Inquirer Home Page[EB/OL]. (2006-01-31). <http://www.wjh.harvard.edu/~inquirer>.
- [8] Linguistic Inquiry and Word Count[EB/OL]. (2016-01-20). <http://liwc.wpengine.com>.
- [9] PANG Bo, LEE L. Opinion Mining and Sentiment Analysis[J]. Foundations and Trends in Information Retrieval, 2008, 2(1-2):1-135.
- [10] 黄挺,姬东鸿. 基于图模型和多分类器的微博情感倾向性分析[J]. 计算机工程, 2015, 41(4):171-175.
- [11] 谢丽星,周明,孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取[J]. 中文信息学报, 2012, 26(1):73-83.
- [12] ZHU Shaojie, XU Bing, ZHENG Dequan, et al. Chinese Microblog Sentiment Analysis Based on Semi-supervised Learning[M]. Berlin, Germany: Springer, 2013.

- [13] AISOPOS F, PAPADAKIS G, VARVARIGOU T. Sentiment Analysis of Social Media Content Using N-Gram Graphs [C]//Proceedings of the 3rd ACM SIGMM International Workshop on Social Media. New York, USA: ACM Press, 2011: 9-14.
- [14] GHIASSI M, SKINNER J, ZIMBRA D. Twitter Brand Sentiment Analysis: A Hybrid System Using N-gram Analysis and Dynamic Artificial Neural Network [J]. Expert Systems with Applications, 2013, 40(16): 6266-6282.
- [15] JIANG Long, YU Mo, ZHOU Ming, et al. Target-dependent Twitter Sentiment Classification [C]//Proceedings of Annual Meeting of the Association for Computational Linguistics. Oregon, USA: Association for Computational Linguistics, 2011: 151-160.
- [16] WANG Hao, CAN Dogan, KAZEMZADEH A, et al. A System for Real-time Twitter Sentiment Analysis of 2012 U. S. Presidential Election Cycle [C]//Proceedings of ACL 2012 System Demonstrations. Jeju Island, Korea: Association for Computational Linguistics, 2012: 115-120.
- [17] LECUN Y, KAVUKCUOGLU K, FARABET C C. Convolutional Networks and Applications in Vision [C]//Proceedings of IEEE International Symposium on Circuits and Systems. Washington D. C., USA: IEEE Press, 2010: 253-256.
- [18] 吴 轲. 基于深度学习的中文自然语言处理 [D]. 南京: 东南大学, 2014.
- [19] KIM Y. Convolutional Neural Networks for Sentence Classification [EB/OL]. (2014-09-03). <https://arxiv.org/abs/1408.5882>.
- [20] KALCHBRENNER N, GREFFENSTETTE E, BLUNSOM P. A Convolutional Neural Network for Modelling Sentences [EB/OL]. (2014-09-08). <https://arxiv.org/abs/1404.2188>.
- [21] JOHNSON R, ZHANG T. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks [EB/OL]. (2015-03-26). <https://arxiv.org/abs/1412.1058>.
- [22] DOS SANTOS C N, GATTI M. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts [C]//Proceedings of International Conference on Computational Linguistics. New York, USA: ACM Press, 2014: 69-78.
- [23] HU Baotian, LU Zhengdong, LI Hang, et al. Convolutional Neural Network Architectures for Matching Natural Language Sentences [C]//Proceedings of International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2015: 2042-2050.
- [24] HE Hua, KEVIN G, LIN Jimmy. Multi-perspective Sentence Similarity Modeling with Convolutional Neural Networks [C]//Proceedings of Conference on Empirical Methods in Natural Language Processing. Berlin, Germany: Springer, 2015: 1576-1586.
- [25] YIN Wenpeng, HINRICH S. Convolutional Neural Network for Paraphrase Identification [C]//Proceedings of International Conference on Neural Information Processing. Berlin, Germany: Springer, 2015: 901-911.
- [26] SEVERYN A, MOSCHITTI A. UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification [C]//Proceedings of the 9th International Workshop on Semantic Evaluation. Berlin, Germany: Springer, 2015: 464-469.
- [27] SOCHER R, PENNINGTON J, HUANG E H, et al. Semi-supervised Recursive Autoencoders for Predicting Sentiment Distributions [C]//Proceedings of Conference on Empirical Methods in Natural Language Processing. New York, USA: ACM Press, 2011: 151-161.
- [28] 梁 军, 柴玉梅, 原慧斌, 等. 基于深度学习的微博情感分析 [J]. 中文信息学报, 2014, 28(5): 155-161.
- [29] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global Vectors for Word Representation [C]//Proceedings of Conference on Empirical Methods in Natural Language Processing. Berlin, Germany: Springer, 2014: 1532-1543.

编辑 顾逸斐

(上接第209页)

- [12] WENG Jianshu, LIM E P, JIANG Jing, et al. TwitterRank: Finding Topic-sensitive Influential Twitterers [C]//Proceedings of ACM International Conference on Web Search & Data Mining. New York, USA: ACM Press, 2010: 261-270.
- [13] 王晨旭, 管晓宏, 秦 涛, 等. 微博消息传播中意见领袖影响力建模研究 [J]. 软件学报, 2015, 26(6): 1473-1485.
- [14] 吴 渝, 马璐璐, 林 茂, 等. 基于用户影响力的意见领袖发现算法 [J]. 小型微型计算机系统, 2015, 36(3): 561-565.
- [15] ASUR S, HUBERMAN B A, SZABO G, et al. Trends in Social Media: Persistence and Decay [EB/OL]. (2011-02-05). http://www.hpl.hp.com/research/scl/papers/trends/trends_web.pdf.
- [16] CHA M, HADDADI H, BENEVENUTO F, et al. Measuring User Influence in Twitter: The Million Follower Fallacy [C]//Proceedings of the 4th International Conference on Weblogs and Social Media. Washington D. C., USA: [s. n.], 2010.
- [17] KWAK H, LEE C, PARK H, et al. What is Twitter, A Social Network or A News Media? [C]//Proceedings of International Conference on World Wide Web. New York, USA: ACM Press, 2010: 591-600.
- [18] 许树柏. 层次分析法原理 [M]. 天津: 天津大学出版社, 1988.
- [19] SONG Xiaodan, CHI Yun, HINO K, et al. Identifying Opinion Leaders in the Blogosphere [C]//Proceedings of the 16th ACM Conference on Information and Knowledge Management. New York, USA: ACM Press, 2007: 971-974.
- [20] MIAO Qingliang, ZHANG Shu, MENG Yao, et al. Domain-sensitive Opinion Leader Mining from Online Review Communities [C]//Proceedings of the 22nd International Conference on World Wide Web Companion. New York, USA: ACM Press, 2013: 187-188.

编辑 顾逸斐