

基于 CSLSTM 网络的文本情感分类^①

庄丽榕, 叶东毅

(福州大学 数学与计算机科学学院, 福州 350108)

通讯作者: 庄丽榕, E-mail: zlr1111@126.com

摘要: 文本情感分类是自然语言处理领域的研究热点, 更是产品评价领域的重要任务. 考虑到词向量与句向量之间的语义关系和用户信息、产品信息对文本情感分类的影响, 提出余弦相似度 LSTM 网络. 该网络通过在不同语义层级中引入用户信息和产品信息的注意力机制, 并根据词向量和句向量之间的相似度初始化词层级注意力矩阵中隐层节点的权重. 在 Yelp13、Yelp14 和 IMDB 三个情感分类数据集上的实验结果表明文中方法的有效性.

关键词: 文本情感分类; 注意力机制; 用户信息; 产品信息; 语义关系; 相似度

引用格式: 庄丽榕, 叶东毅. 基于 CSLSTM 网络的文本情感分类. 计算机系统应用, 2018, 27(2): 230-235. <http://www.c-s-a.org.cn/1003-3254/6200.html>

Text Sentiment Classification Based on CSLSTM Neural Network

ZHUANG Li-Rong, YE Dong-Yi

(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China)

Abstract: Text sentiment classification is a popular subject of natural language processing and the crucial problem in product evaluation. Based on semantic relationship of word vector and sentence vector and the impact of user information, product information to text sentiment classification, Cosine Similarity Long-Short Term (CSLSTM) network is proposed. CSLSTM considers attention mechanisms of user information and product information in various semantic levels. And it involves a effective initialization method in hidden level weights of word-level attention matrix according to similarity of word vector and sentence vector. The competitive results are derived from three sentiment classification datasets, Yelp13, Yelp 14, and IMDB.

Key words: text sentiment classification; attention mechanisms; user information; product information; semantic relationship; similarity

1 引言

情感分类旨在发现主观性文档中表达的态度和情感的倾向性. 随着信息技术飞速发展, 尤其是微博、微信等社交网络的兴起, 用户可以更加方便、及时的进行信息交流和意见表达. 大量网络用户每天都会发布并传播高达上亿条的信息, 这些海量的文本信息表达了用户观点倾向和情感信息. 为了更好挖掘人们对评论事物的情感倾向, 需要对这些评论信息进行情感

分类, 根据分类的结果制定相应的策略, 因此文本情感分类具有广泛的实用价值^[1,2].

由于文本情感分类的多领域融合性以及情感信息的价值可挖掘性, 国内外学者在相关领域做了许多研究, 并取得诸多成果, 大致分为基于传统的机器学习和深度学习两个方向.

传统的机器学习方法, 通过人工设计特征, 将自然语言文本转化为结构化的特征向量提取情感特征, 并

^① 收稿时间: 2017-05-18; 修改时间: 2017-06-05; 采用时间: 2017-06-08

构造情感分类器,将待分类的文本通过分类器进行情感倾向性分类. Pang^[3]最早将机器学习方法应用到文本情感分类中,把 n-gram 和词性进行组合作为情感特征,之后分别采用朴素贝叶斯、最大熵和支持向量机(Support Vector Machine, SVM)分类器对电影评论进行情感分类.为了提高文本情感分类的精度, Kim^[4]除了考察传统的 n-gram 模型外,还引入了位置特征和评价词特征来完成文本的情感分类.徐军等人^[5]选择具有语义倾向的词语作为特征项,实验发现具有正负面情感的词语对情感分类起到决定性作用,选择这些词语作为特征可以提高文本情感分类的精度.李素科等^[6]针对监督学习分类中标注工作需要大量的人工劳动力,提出了一种基于特征聚类的半监督式情感分类方法,该方法提高了情感分类的性能,同时只需要对少量训练数据实例进行情感类别标注.

传统机器学习属于浅层模型,浅层模型的一个主要局限性就是需要依赖人工经验抽取样本特征,而这些特征的设计都需要专家的领域知识,耗费大量的人工成本,系统迁移性差.深度学习通过学习一种深层的非线性的网络结构来弥补这一约束,能够实现复杂函数的逼近,从样本集中抓取到数据的本质特征^[7-10].因此,采用深度学习成为近几年自然语言处理领域内的研究热点.

考虑到文本中词之间的时序信息和 RNN 循环神经网络^[11](Recurrent Neural Network, RNN)处理长文本任务会出现长期依赖的问题, Tai^[12]和 Tang^[13]等使用拥有门机制的长短期记忆模型(Long-Short Term Memory, LSTM)^[14]进行文本情感分类,首先获取句子级的文本表示,然后利用句子级表示获得更高级的语义表示作为文本的特征,提高了文本情感分类的精度.但由于大多数产品评论的情感分类将关注点更多放在文本内容上,忽略了用户和产品本身特点对情感分类的影响.考虑到用户信息和产品信息对情感分类的重要性, Tang^[15]尝试将二者的信息和神经网络结合起来,在输入层将词向量、用户向量和产品向量结合起来,然后利用 CNN 进行建模,通过 softmax 分类.但这个模型仍存在一定问题,首先,在输入层将用户向量和产品向量结合在一起导致模型的计算量较大;其次,只有词层级引入用户信息和产品信息,不够充分.于是, Chen^[16]提出了层级 LSTM 网络,通过注意力机制在不同语义级别中引入用户和产品信息,降低了模型的计

算量并且充分引入用户信息和产品信息,提高了情感分类的准确度.但 Chen 在引入注意力机制时,采取随机初始权重矩阵,忽视了词向量与句向量间的语义关系.对此,本文提出了 CSLSTM 网络(Cosine Similarity Long Short-Term Memory),通过在不同层级中引入用户信息和产品信息的注意力机制,并根据词向量和句向量的相似度初始化词层级注意力矩阵中隐层节点的权重.实验结果表明,本文所提出的 CSLSTM 网络具有较好的分类效果.

2 CSLSTM 网络的文本情感分类

本文所提出的 CSLSTM 网络情感分类主要分为以下几个部分:首先对文本分类中的一些符号进行定义;接着通过层级的 LSTM 获得文本的语义表示;之后通过注意力机制在不同语义层级引入用户信息和产品信息;然后根据词向量与句向量的语义关系对隐层节点初始权重进行赋值;最后将获得语义作为文本特征进行情感分类.

2.1 符号定义

为了方便研究,对本文中所用到的一些符号进行定义.假设用户 $u \in U$ 关于产品 $p \in P$ 发表了评论 d ,且文本 d 包含 n 个句子,则文本 d 表示如下:

$$d = (s_1, s_2, \dots, s_n) \quad (1)$$

其中, s_i 表示评论文本 d 中的第 i 个句子.

假设某个句子 s_i 包含 m 个词,则句子 s_i 表示如下:

$$s_i = (x_1^i, x_2^i, \dots, x_m^i) \quad (2)$$

其中, x_j^i 表示句子 s_i 中的第 j 个词.

2.2 情感分类模型

本文首先对句子进行建模,得到句子的语义表示,再利用句子对文本进行建模.考虑到文本中词之间的时序信息和长文本存在长期依赖问题,本文在词层级选择 LSTM 网络对句子进行建模.同样,在句子层级也选择 LSTM 网络对文本进行建模.在词层级的建模过程中,首先将词 x_j^i 映射到低维语义空间,获取词向量 $\mathbf{x}_j^i \in \mathbf{R}^d$,之后在每个时刻,给定一个输入 \mathbf{x}_j^i ,得到当前网络中的细胞状态 \mathbf{c}_j^i 和隐层状态 \mathbf{h}_j^i ,其中 \mathbf{c}_j^i 和 \mathbf{h}_j^i 的状态更新方式如下:

$$\mathbf{in}_j^i = \sigma(\mathbf{W}_{inh} \cdot \mathbf{h}_{j-1}^i + \mathbf{W}_{inx} \cdot \mathbf{x}_j^i + \mathbf{b}_{in}) \quad (3)$$

$$\mathbf{f}_j^i = \sigma(\mathbf{W}_{fh} \cdot \mathbf{h}_{j-1}^i + \mathbf{W}_{fx} \cdot \mathbf{x}_j^i + \mathbf{b}_f) \quad (4)$$

$$\mathbf{o}_j^i = \sigma(\mathbf{W}_{oh} \cdot \mathbf{h}_{j-1}^i + \mathbf{W}_{ox} \cdot \mathbf{x}_j^i + \mathbf{b}_o) \quad (5)$$

$$\bar{\mathbf{c}}_j^i = \tanh(\mathbf{W}_{ch} \cdot \mathbf{h}_{j-1}^i + \mathbf{W}_{cx} \cdot \mathbf{x}_j^i + \mathbf{b}_c) \quad (6)$$

$$\mathbf{c}_j^i = \mathbf{f}_j^i \odot \mathbf{c}_{j-1}^i + \mathbf{i}_j^i \odot \bar{\mathbf{c}}_j^i \quad (7)$$

$$\mathbf{h}_j^i = \mathbf{o}_j^i \odot \tanh(\mathbf{c}_j^i) \quad (8)$$

其中, σ 为 sigmoid 函数; $\mathbf{i}, \mathbf{f}, \mathbf{o}$ 和 \mathbf{c} 分别为输入门、忘记门、输出门和细胞状态; \odot 表示点乘操作; $\mathbf{W}_{inh}, \mathbf{W}_{fh}, \mathbf{W}_{oh} \in R^{H*ID}, \mathbf{W}_{inx}, \mathbf{W}_{fx}, \mathbf{W}_{ox} \in R^{H*H}$ 表示权重矩阵, H 为隐层维数, ID 为输入层维数; $\mathbf{b}_{in}, \mathbf{b}_f, \mathbf{b}_o$ 为偏置因子. 其中, 权重矩阵和偏置因子需要训练, 最后将得到的隐层状态输入均值池化层获得句子语义表示 \mathbf{s}_i .

在句子层级对文本进行建模时采用与词层级的方法相同, 将 \mathbf{s}_i 输入 LSTM 网络中获得文本语义表示 \mathbf{d} .

2.3 用户和产品注意力机制

本文采用在不同语义层级通过注意力机制引入用户信息和产品信息, 以获取不同层级中对情感分类有重要影响的信息. 首先在词层级引入用户信息和产品信息得到句子层级的语义表示, 然后在句子层级引入用户信息和产品信息得到文本的语义表示.

对于同一用户或是同一个产品的评论文本中, 每个词对于所在句子中的贡献程度是不同的, 故采用均值池化层进行权重分配会损失重要词汇对语义表示的贡献率. 为了解决这个问题, 本文在词层级采用注意力机制抽取对句子语义表示具有重要贡献的词, 提高语义表示的有效性. 引入用户信息和产品信息的注意力机制计算如下:

$$\mathbf{s}_i = \sum_{j=1}^{m_i} \alpha_j^i \mathbf{h}_j^i \quad (9)$$

其中, α_j^i 衡量了第 j 个词对当前用户、产品和所在句子的重要程度, 它的计算公式如下:

$$\alpha_j^i = \frac{\exp(\text{score}(\mathbf{h}_j^i, \mathbf{u}, \mathbf{p}))}{\sum_{k=1}^{m_i} \exp(\text{score}(\mathbf{h}_k^i, \mathbf{u}, \mathbf{p}))} \quad (10)$$

其中, $\mathbf{u} \in R^{d_u}$ 表示用户向量, $\mathbf{p} \in R^{d_p}$ 表示产品向量, d_u, d_p 分别表示用户向量和产品向量的维数, $\text{score}(\mathbf{h}_j^i, \mathbf{u}, \mathbf{p})$ 为评价函数, 其计算方法如下:

$$\text{score}(\mathbf{h}_j^i, \mathbf{u}, \mathbf{p}) = \mathbf{v}_\alpha^T \tanh(\mathbf{W}_H \mathbf{h}_j^i + \mathbf{W}_U \mathbf{u} + \mathbf{W}_P \mathbf{p} + \mathbf{b}_\alpha) \quad (11)$$

其中, $\mathbf{W}_H, \mathbf{W}_U, \mathbf{W}_P$ 为权重矩阵, \mathbf{v}_α^T 为权重向量 \mathbf{v}_α 的转置, 前面所述的权重矩阵和权重向量一般采用随机

初始赋值, 但考虑到词向量和句向量之间语义关系, 本文将根据二者之间的相似度对 \mathbf{W}_H 进行初始赋值, 具体方法见 2.4 节.

评价函数将隐层状态、用户向量和产品向量结合起来, 其中用户向量和产品向量作为模型参数一起训练, 得到句子的语义表示 \mathbf{s}_i . 同样, 在句子层级也采用通过注意力机制引入用户信息和产品信息来抽取对文本语义表示有重要贡献的句子:

$$\mathbf{d} = \sum_{i=1}^n \beta_i \mathbf{h}_i \quad (12)$$

$$\beta_i = \frac{\exp(\text{score}(\mathbf{h}_i, \mathbf{u}, \mathbf{p}))}{\sum_{i=1}^n \exp(\text{score}(\mathbf{h}_i, \mathbf{u}, \mathbf{p}))} \quad (13)$$

$$\text{score}(\mathbf{h}_i, \mathbf{u}, \mathbf{p}) = \mathbf{v}_\beta^T \tanh(\mathbf{W}_{H\beta} \mathbf{h}_i + \mathbf{W}_{U\beta} \mathbf{u} + \mathbf{W}_{P\beta} \mathbf{p} + \mathbf{b}_\beta) \quad (14)$$

2.4 注意力权重矩阵初始化策略

注意力权重矩阵一般简单采用随机初始赋值, 忽略掉了词向量和句向量之间的语义关系. 对此, 本文采用根据二者之间的相似度对权重矩阵中的 \mathbf{W}_H 进行初始化. 此处所指的句向量 $\bar{\mathbf{s}}_i$ 与句子的语义表示 \mathbf{s}_i 不同, $\bar{\mathbf{s}}_i$ 是由句子中所有的词向量求平均值得到的向量表示:

$$\bar{\mathbf{s}}_i = \text{avg}(\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_m^i) \quad (15)$$

采用余弦相似度衡量词向量和句向量二者之间的相似度:

$$\text{sim}(\mathbf{x}_j^i, \bar{\mathbf{s}}_i) = \cos \langle \mathbf{x}_j^i, \bar{\mathbf{s}}_i \rangle = \frac{\mathbf{x}_j^i \cdot \bar{\mathbf{s}}_i}{\|\mathbf{x}_j^i\| \cdot \|\bar{\mathbf{s}}_i\|} \quad (16)$$

最后, 将相似度进行归一化, 得到权重 \mathbf{W}_H 初始值:

$$\mathbf{w}_{Hj}^i = \frac{\text{sim}(\mathbf{x}_j^i, \bar{\mathbf{s}}_i)}{\sum_{k=1}^m \text{sim}(\mathbf{x}_k^i, \bar{\mathbf{s}}_i)} \quad (17)$$

2.5 情感分类

将获得的文本语义表示 \mathbf{d} 进一步抽象以挖掘更深层次的语义, 在 LSTM 层后设置一个全连接层, 将 \mathbf{d} 映射到 C 类别空间, 得到定长向量 $\hat{\mathbf{d}}$, 具体计算如下:

$$\hat{\mathbf{d}} = \tanh(\mathbf{W}_c \mathbf{d} + \mathbf{b}) \quad (18)$$

其中, \mathbf{W}_c 为全连接层的权重矩阵, \mathbf{b} 为偏置因子.

为了得到情感类别估计值, 本文采用 softmax 函数进行情感分类, 获得文本的情感分布:

$$p_c = \frac{\exp(\hat{d}_c)}{\sum_{k=1}^C \exp(\hat{d}_k)} \quad (19)$$

其中, C 情表示情感类别数目, p_c 表示预测文本 d 属于类别 c 的概率, 之后将 d 归类到概率最大的那个类别中.

本文采用交叉熵损失 (Categorical Cross-Entropy Loss) 作为损失函数, 模型的训练目标是最小化真实类别与预测类别间的交叉熵损失, 其中交叉熵损失函数表示如下:

$$L = - \sum_{d \in D} \sum_{c=1}^C p_c^g(d) \cdot \log(p_c(d)) \quad (20)$$

其中, D 表示整个训练集, p_c^g 表示文本 d 的真实类别是否为 c , 即当 d 的真实类别为 c 时, $p_c^g=1$, 否则 $p_c^g=0$.

3 情感分类实验

为了验证本文所提出方法的有效性, 本文利用 Tang¹⁵ 提供的评论文本数据进行文本情感分类实验.

3.1 实验数据集

本文采用 Yelp13、Yelp14 和 IMDB 三个数据集, 三个数据集中已包含了用户信息和产品信息. 表 1 为三个数据集的信息摘要. 实验中, 将每个数据集按 8:1:1 的比例分配训练集、验证集和测试集.

表 1 数据集分布情况

Dataset	train size	dev. size	test size	users	products
Yelp13	62522	7773	8671	1613	1633
Yelp14	183019	22745	25399	4818	4194
IMDB	67426	8381	9112	1310	1635
Dataset	classes	docs/user	docs/product	sens/Doc	words/sen
Yelp13	5	48.42	48.36	10.89	17.38
Yelp14	5	47.97	55.11	11.41	17.26
IMDB	10	64.82	51.94	16.08	24.54

注: train size 为训练集大小, dev. size 为验证集大小, test size 为测试集大小, users 为用户数量, products 为产品数量, classes 为类别数目, docs/user 为平均每个用户发表的评论文本数, product 为平均每个产品包含的评论文本数, sens/doc 为平均每篇文本包含的句子数, words/sen 为平均每个句子包含的单词数

Yelp 2013 和 Yelp 2014 分别为 Yelp 数据库 2013 和 2014 的评论, 两者的评级范围从 1 颗星到 5 颗星, 星数越高代表用户对产品的越满意.

IMDB 是由 84919 条 IMDB 电影网站上的评论组成的数据集, 它的评分范围从 1 分到 10 分 (只能选择整数分值), 分数越高代表用户对电影的满意程度越就

越高.

3.2 评价指标

本文采用正确率和均方根误差作为评价标准, 其中正确率衡量整体情感分类性能, 均方根误差反映预测情感类别 (评分等级) 和真实情感类别 (评分等级) 的偏离程度. 两种评价标准的公式如下:

$$Accuracy = \frac{T}{N} \quad (21)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (gr_i - pr_i)^2}{N}} \quad (22)$$

其中, T 为正确分类的文本数, N 为整个数据集的文本数, gr_i 和 pr_i 分别为真实情感类别 (评分等级) 和预测的情感类别 (评分等级).

3.3 实验设置

本文采用 SkipGram 算法生成词向量, 并设置词向量的维度为 200, 其它参数均采用默认值. 同时将用户向量和产品向量初始化为随机向量, 维度设置为 200 维. LSTM 网络中隐藏节点数和细胞状态数也设置为 200, 并采用 AdaDelta 进行参数优化.

3.4 基准实验

为了验证本文所提方法的有效性, 本文选择以下算法进行对比, 以下算法大致可以分为两类: 不考虑用户、产品信息和考虑用户、产品信息, 其中不考虑用户、产品信息的算法如下:

1) CBOW: 对所有词向量进行求和后输入 Softmax 进行分类.

2) Majority: 采用多种情感策略的启发式算法.

3) Trigram: 分别采用 Unigram、Bigram、Trigram 训练支持向量机 (SVM).

4) TextFeature: 抽取语义特征和情感特征采用支持向量机进行训练.

5) AvgWordvec: 对所有词向量求平均值作为文本向量表示, 采用支持向量机进行训练.

6) SSWE: 生成情感词向量后利用支持向量机进行训练.

7) LSTM: 生成词向量输入 LSTM 网络进行训练.

8) RNTN+RNN: 经过 RNTN 网络得到句子语义表示, 之后输入 RNN 网络得到文本语义表示.

9) B-CLSTM: 基于 B-LSTM 网络添加缓存机制,

并将记忆元分为多组,对不同组采用不同的遗忘率。

其中考虑用户信息和产品信息的实验如下:

10) Trigram +UPF: 在算法 3) 的基础上添加用户信息和产品信息。

11) TextFeature++UPF: 在算法 4) 的基础上添加用户信息和产品信息。

12) JMARS: 基于协同过滤和主题模型引入用户信息和产品信息。

13) UPNN: 在输入层引入用户信息和产品信息, 并采用 CNN 网络进行训练。

14) NSC+UPA: 在不同层级通过注意力机制引入了用户信息和产品信息。

本实验所采用的数据集与文献[15-17]的数据集相同,因此算法 1)-13) 直接使用文献[15-17]中的结果数据,其中算法 14) 使用作者提供代码进行实验。实验结果如表 2 所示。

表 2 实验结果对比

Models	Yelp2013		Yelp2014		IMDB	
	Acc.(%)	RMSE	Acc.(%)	RMSE	Acc.(%)	RMSE
CBOW	54.5	0.840	56.8	0.787	34.8	1.69
Majority	44.1	1.060	39.2	1.097	19.6	2.495
Trigram	56.9	0.814	57.7	0.804	39.9	1.783
TextFeature	55.6	0.845	57.2	0.800	40.2	1.793
AvgWordvec	52.6	0.898	53.0	0.893	30.4	1.985
SSWE	54.9	0.849	55.7	0.851	31.2	1.973
LSTM	53.9	0.810	56.3	0.769	37.8	1.612
RNTN+RNN	57.4	0.804	58.2	0.821	40.0	1.764
B-CLSTM	59.8	0.741	61.9	0.704	46.2	1.453
Trigram+UPF	57.0	0.803	57.6	0.789	40.4	1.764
TextFeature+UPF	56.1	1.822	57.9	0.791	40.2	1.774
JMARS	N/A	0.985	N/A	0.999	N/A	1.773
UPNN	59.6	0.784	60.8	0.764	43.5	1.602
NSC+UPA	64.7	0.692	66.6	0.658	51.4	1.317
CSLSTM	64.9	0.686	67.1	0.658	51.1	1.302

注: 由于实验JMARS中的评分取值为实数值,而类别中的实际取值为整数值,因此正确率不可获得。

3.5 结果比较

对比 CSLSTM 和算法 1)-9), 可以发现考虑用户信息和产品信息可以提高文本情感分类的精度。其中, 对比 CSLSTM 与算法 1) 可见, 添加用户信息和产品信息 CSLSTM 在三个数据集上的正确率都得到了 10% 以上的提高; 对比 CSLSTM 与算法 7), 两者同样采用 LSTM 网络进行文本分类, 但本文的 CSLSTM 在分类中还考虑了用户信息和产品信息, 所以

CSLSTM 的分类效果相比与算法 7) 在各个数据集上的正确率都提高了 10% 以上, 其中在数据集 IMDB 上的正确率更是得到了 13.3% 的提高。CSLSTM 对比算法 3) 和 10) 也可以发现, 用户信息和产品信息在分类中占据一定作用, 在 Yelp2013 和 IMDB 数据集上添加用户信息和产品信息提高了情感分类的效果。从 CSLSTM 和算法 13) 结果可以发现, 在不同语义层级通过注意力机制引入用户信息和产品信息的效果好于在输入层引入用户信息和产品信息。表明多层级引入用户信息和产品信息对情感分类具有重要意义。对比 CSLSTM 和算法 14) 可以发现, 在 Yelp13 数据集中, CSLSTM 的准确率和均方根误差均高于算法 14); 在 Yelp14 数据集中, CSLSTM 的准确率高于算法 14); 在 IMDB 数据集中, CSLSTM 的均方根误差好于算法 14), 在三个数据集上的结果说明了考虑词向量和句向量之间的语义关系有利于文本情感分类的效果, 表明了本文所提出方法的有效性。

4 结语

针对文本情感分类问题, 考虑词向量与句向量之间的语义关系和用户、产品对情感分类的影响, 本文提出基于 CSLSTM 网络的文本情感分类方法。在不同语义层级引入用户信息和产品信息的注意力机制, 并根据词向量和句向量的相似度初始化词层级注意力矩阵中隐层节点的权重。实验结果表明, 相比多数基准实验, 本文方法具有较好的分类效果, 较好地利用了词向量与句向量的语义关系和用户信息、产品信息。在未来的工作中, 将在文本方法上考虑在词向量中融入更多的情感信息, 对词向量的训练部分进行扩展, 比如在 word2vec 训练词向量的过程中利用情感词典刻画情感词, 生成具有情感信息的词向量, 以提高情感分类的精度。

参考文献

- 唐慧丰, 谭松波, 程学旗. 基于监督学习的中文情感分类技术比较研究. 中文信息学报, 2007, 21(6): 88-94, 108.
- 赵妍妍, 秦兵, 刘挺. 文本情感分析. 软件学报, 2010, 21(8): 1834-1848.
- Pang B, Lee L, Vaithyanathan S. Thumbs up?: Sentiment classification using machine learning techniques. Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA,

- USA. 2002. 79–86.
- 4 Kim SM, Hovy E. Automatic identification of pro and con reasons in online reviews. Proceedings of the COLING/ACL on Main Conference, Poster Sessions. Sydney, Australia. 2006. 483–490.
- 5 徐军, 丁宇新, 王晓龙. 使用机器学习方法进行新闻的情感自动分类. 中文信息学报, 2007, 21(6): 95–100.
- 6 李素科, 蒋严冰. 基于情感特征聚类的半监督情感分类. 计算机研究与发展, 2013, 50(12): 2570–2577. [doi: [10.7544/issn1000-1239.2013.20130878](https://doi.org/10.7544/issn1000-1239.2013.20130878)]
- 7 Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science, 2006, 313(5786): 504–507. [doi: [10.1126/science.1127647](https://doi.org/10.1126/science.1127647)]
- 8 张庆庆, 刘西林. 基于深度信念网络的文本情感分类研究. 西北工业大学学报(社会科学版), 2016, 36(1): 62–66.
- 9 Bengio Y, Delalleau O. On the expressive power of deep architectures. Proceedings of the 14th International Conference on Discovery Science. Espoo, Finland. 2011. 18–36.
- 10 Zhou D, Bousquet O, Lal TN, *et al.* Learning with local and global consistency. Advances in Neural Information Processing Systems 16. Vancouver, CB, Canada. 2003: 321–328.
- 11 İrsoy O, Cardie C. Opinion mining with deep recurrent neural networks. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar. 2014. 720–728.
- 12 Tai KS, Socher R, Manning CD. Improved semantic representations from tree-structured long short-term memory networks. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China. 2015. 1556–1566.
- 13 Tang DY, Qin B, Liu T, *et al.* Document modeling with gated recurrent neural network for sentiment classification. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal. 2015. 1422–1432.
- 14 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)]
- 15 Tang DY, Qin B, Liu T. Learning semantic representations of users and products for document level sentiment classification. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China. 2015. 1014–1023.
- 16 Chen HM, Sun MS, Tu CC, *et al.* Neural sentiment classification with user and product attention. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, TX, USA. 2016. 1650–1659.
- 17 Xu JC, Chen DL, Qiu XP, *et al.* Cached long short-term memory neural networks for document-level sentiment classification. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, TX, USA. 2016. 1660–1669.