

情感分析综述

1. 情感分析的描述

情感分析或观点挖掘是对人们对产品、服务、组织、个人、问题、事件、话题及其属性的观点、情感、情绪、评价和态度的计算研究。该领域的开始和快速发展与社交媒体的发展相一致，如评论、论坛、博客、微博、推特和社交网络，因为这是人类历史上第一次拥有如此海量的以数字形式记录的观点数据。早在2000年，情感分析就成为NLP中最活跃的研究领域之一。它在数据挖掘、Web挖掘、文本挖掘和信息检索方面得到了广泛的研究。实际上，因其对商业和社会的整体重要性，它已经从计算机科学扩展到管理学和社会学，如营销、金融、政治学、传播学、健康科学，甚至历史学。

文本情感分析又称意见挖掘，是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。按照处理文本的粒度不同，情感分析可分为词语级、短语级、句子级、篇章级以及多篇章级等几个研究层次；按照处理文本的类别不同，可分为基于新闻评论的情感分析和基于产品评论的情感分析。文本情感分析可归纳为3项主要任务，即情感信息抽取、情感信息分类以及情感信息的检索与归纳。

2. 情感分析研究现状

随着社交网络的不断发展，人们更愿意通过微博，博客社区来表达自己的观点，发表对热点事件的评论，从而使通过微博，博客，影评以及产品评价等来了解社交网络用户的情感倾向得到了学术界的广泛关注。根据微博数据进行情感分析是一个具有挑战性的任务，近年来引发了学者极大的兴趣。目前，情感分析的主要研究方法还是一些基于机器学习的传统算法，例如，SVM，信息熵，CRF等。这些方法归纳起来有3类：有监督学习，无监督学习和半监督学习。而当前大多数基于有监督学习的研究都取得了不错的成绩，但是由于有监督学习依赖于大量人工标注的数据，使得基于有监督学习的系统需要付出很高的标注代价。相反的，无监督学习不需要人工标注数据训练模型，是降低标注代价的解决方案，但由于其完全依赖算法学习结果，往往效果不佳，难以达到实际要求。而半监督学习则是采取综合利用少量已标注样本和大量未标注样本来提高学习性能的机器学习方法，它兼顾了人工标注成本和学习效果，被视为一种折中方案。已有研究所采用的方法大多数都基于词袋模型，而这种模型无法捕获到很多有关情感倾向性分析的语言现象特征。例如，“反法西斯联盟击溃了法西斯”和“法西斯击溃了反法西斯联盟”这两个词组拥有相同的词袋模型表示方法，而前一个带有积极的感情色彩，后一个带有消极的感情色彩。除此之外，还有很多研究者使用人工标注的数据（情感词典及句法分析等），虽然采用这些方法可以有效的提高情感分析的准确性，但由于需要较多的人工标注数据，从而限制了这些方法在其他领域以及跨语言的推广。本文综合文本情感分析领域的研究成果以及已有的中文微博情感分析方法，将中文微博的情感分析算法进行归纳。

3. 情感分析的主要研究方法

3.1 基于RAE的情感极性转移模型^[3]-半监督

该模型基于自编码算法围绕半监督学习方法提出一种新的分析中文微博情感倾向的深度学习算法。该模型首先将文本数据转为低维实数向量表示,建立表示文本特征的矩阵,然后将其作为基于R A E的情感极性转移模型的输入,最后使用L B F G S算法多次迭代生成最终的模型。该模型可对低维实数向量表示的文本进行情感分类输出其情感极性。

训练R A E情感极性转移模型

输入: 训练语料 S 及其对应标签 t

输出: $\theta = \{w(1), w(2), w(3), A, b(2), b(3)\}$

1) 使用高斯分不初始化训练语料的词向量表示;

2) while 不收敛 do

$\nabla J=0$

 forall $\langle s, t \rangle \in S$ do

 使用贪心算法生成句子二叉树结构

 计算 $\nabla J_i = \partial J(s, t) / \partial \Theta$

 更新 $\nabla J = \nabla J + \nabla J_i$

 endfor

 更新 $\Theta = N1 \nabla J + \lambda \theta$

endwhile

该实验数据采用C O A E 2 0 1 4 微博数据集完整版作为测试集,进行完全封闭 实验,不采用任何评测外数据集,并与其他应用于该评测数据集的结果做比较。

COAE2014 数据集样例

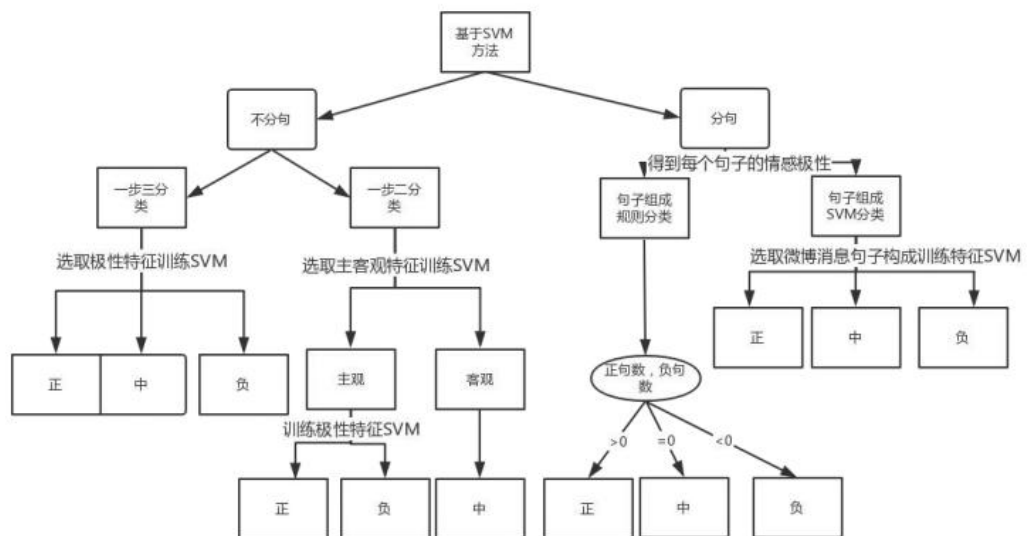
积极	消极
奥迪 Q3SUV, 不错	保险公司很无语啊!
我的三星被我摔了好几次, 摔坏了好几个壳, 手机还是完好无损哈。	我同意, 外资的保险公司似乎缺少了狼性。但友邦有待考量。
蒙牛那个点滴的广告还不错	三星要是能领导行业那才见鬼了

模型对比结果

系统	准确率	召回率	F-Score
RecursiveAutoEncoder	0.7875	0.7843	0.7787
R A E 的情感极性转移模型	0.8058	0.8059	0.7979

3.2 基于机器学习的微博情感分类^[7]-监督学习

机器学习分类方法主要是应用机器学习模型,如支持向量机,朴素贝叶斯,最大熵等,通过对训练集的特征进行学习,构造模型,从而应用于对测试集的分类判断。采用支持向量机(SVM)和信息增益(IG),以及使用三种机器学习算法,三种特征选取算法以及三种特征项权重计算方法对微博进行了情感分类的实证研究。研究发现,综合考虑三种因素,采用支持向量机(SVM)和信息增益(IG),以 TF-IDF (TermFrequency-InverseDocumentFrequency)作为特征项权重,三者结合对微博的情感分类效果最好。TF-IDF (TermFrequency-InverseDocumentFrequency)作为特征项权重,三者结合对微博的情感分类效果最好。



3.3 基于 SVM 和 CRF 多特征组合的微博情感分析^[2]-监督学习

实验使用的是 COAE2014 中任务 4 所提供的语料,包含 2174 个训练语料和 5000 个测试语料。

语料的统计结果

语料	正向语料个数	负向语料个数	总数
训练语料	993	1181	2174
测试语料	2656	2344	5000

3.3.1 基于 SVM 多特征组合的微博情感分析

首先对微博进行预处理,然后使用 NLPIR2013 分词工具进行分词以及词性标注,抽取文本中的特征,分别在 SVM 模型和 CRF 模型上进行实验。为了训练出一个较好的 SVM 模型,选取了五类文本特征,包括词性,情感词,否定词,程度副词及特殊符号。

SVM 模型的特征类型及其含义

特征类型	含义
词性	形容词、动词、名词、副词、介词、叹词的个数
情感词	正向情感词和负向情感词的个数、强度及情感得分
否定词	情感词之前是否出现否定词
程度副词	情感词之前是否出现程度副词
特殊符号	问号和感叹号的个数

SVM 模型中不同特征组合的实验结果

实验	特征组合	正面语料正确率/%	负面语料正确率/%	整体正确率/%
1	词性	56.40	68.43	62.04
2	情感词	88.14	82.59	85.54
3	词性+情感词	88.37	87.16	87.80
4	词性+情感词+否定词	87.95	89.59	88.72
5	词性+情感词+否定词+程度副词	87.84	88.87	88.32
6	词性+情感词+否定词+特殊符号	88.79	88.25	88.54

实验结果显示,当使用词性,情感词和否定词的组合特征时效果最好,正确率达到 88.72%。其中情感词的作用最大,使正确率提高了 25.76%;其次是词性,使正确率提高了 2.46%;否定词也对情感分析起到一定作用,使正确率提高了 0.92%。但是,加入程度副词和特殊符号作为特征后,正确率略微降低,说明程度副词和特殊符号并不适合作为 SVM 模型的特征。

3.3.2 基于 CRF 多特征组合的微博情感分析

由于 CRF 一般用于序列标注任务中,而文本情感分析是要判断整个句子的情感倾向,并不是一个典型的标注任务。为了将情感分析转换成标注问题,本文将文本的极性对应到文本中每个词语的极性,通过标注每个词语的极性来判断文本的极性。换句话说,如果一条微博的极性为正向,则将微博中每个词语的极性都标注为正向;如果一条微博的极性为负向,则将微博中每个词语的极性都标注为负向。反之,如果一条微博中每个词语的极性都被预测成正向,则微博的极性就被判定为正向;如果一条微博中每个词语都被预测成负向,则微博的极性就被判定为负向。

CRF 模型的特征类型及其含义

特征类型	值
词本身	
词	n, v, a, d, nr, ...
词性	1: 正向, -1: 负向, 0: 中性
情感词极性	1: 是否定词, 0: 不是否定词
否定词	0.5~2.0: 程度副词的强度值, 0: 不是程度副词
程度副词	2: 问号, 1: 感叹号, 0: 不是问号或感叹号
特殊符号	

CRF 模型中不同特征组合的实验结果

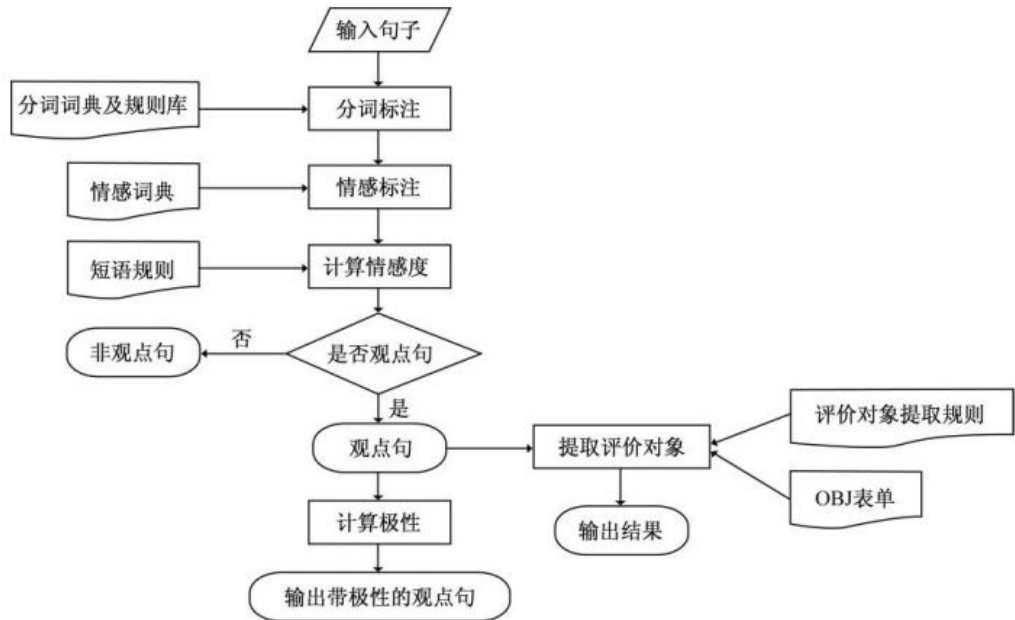
	正面语料 正确率/%	负面语料 正确率/%	整体正确 率/%
1 词	96.16	50.64	74.82
2 词+词性	95.78	53.11	75.78
3 词+词性+情感极性	97.81	74.78	87.02
4 情感极性	88.47	88.31	88.40
5 情感极性+否定词	90.47	89.97	90.24
6 情感极性+否定词+程度副词	90.40	90.36	90.38
7 情感极性+否定词+程度副词+特殊符号	90.51	90.36	90.44
8 所有特征	97.59	76.07	87.50

实验结果显示, 当使用情感极性, 否定词, 程度副词和特殊符号的组合特征时效果最好, 正确率达到 90.44%。其中情感词极性的作用最大, 仅使用情感词极性作为特征时, 正确率就达到 88.40%; 其次是否定词, 使正确率提高了 1.84%; 程度副词和特殊符号的作用较小, 加入特征组合后使正确率提高了 0.2%。笔者还发现, 传统方法中经常使用的词和词性特征并不适合作为 CRF 模型的特征, 仅使用情感极性作为特征时正确率为 88.40%, 加入词和词性特征后, 正确率变为 87.02%, 降低了 1.48%, 而且使 CRF 模型对正面语料和负面语料的效果不平衡, 前者高达 97.81%, 后者仅为 74.78%。

3.4 基于情感短语词典以及语义规则的情感分析系统 CUCsas[1]-半监督

情感分析及评价对象抽取, 目前主要有两种基本方法. 一种是基于句型, 句法, 语义的词典及规则方法, 一种是基于标注语料的机器学习方法. 我们认为情感分析, 分析的是语言的意义, 要想在对语言进行深层理解的基础上完成情感分析, 规则的方法更有优势, 所以主要采取的是基于语义的词典加规则的方法, 统计方法只在建立词典确立词条属性, 建立规则库确定规则优先级时发挥作用. 针对话题型微博, 我们主要采用了加大情感因子颗粒度, 构建基于短语的情感词典, 通过短语规则确定句子极性, 重点研究否定形式, 建立基于话题的 OBJ 表单等策略. 情感短语词典突破了以词为主要收录对象的局限, 将明确表达观点的情感短语甚至短句都收录其中, 如“不得好死, 不咋的, 扯犊子, 就好了, 瞎折腾个屁, 不是找抽吗, 什么玩意儿, 没一个好东西, 好不到哪儿去, 以为自己是谁啊, 可以说脏话

吗,这是人的行为吗,良心让狗吃了”等等.当然,一些明确表达观点的口语词,网络用语和脏话也必须收录到情感词典中,如“得瑟,傻逼,狗逼,操蛋,犯贱,傻B,欠揍,找死”等,这样包含这些情感短语和情感词的观点句就能够被准确地识别出来,从而在不降低准确率的前提下,提高系统对观点句的召回率.我们构建了一个基于情感短语词典以及语义规则的情感分析系统 CUCsas. 系统包括分词标注模块,短语和句子规则解释器,评价对象提取规则解释器和一个 OBJ 模块.其中的分词标注模块与一般的不同,为做情感分析,在原有分词词典 和规则库的基础上,又加进了几部与情感分析相关的词典,主要的三部是:一部用 户词典 Usrsas,专门储存情感分析需要切分出来的词和短语并赋予词性;一部添 加了情感短语,带有“po(褒),ne(贬)”标记的情感词典 Dicsas,用于给待分析语 料中的情感因子进行标记;一部褒贬词词典 Dicsvl,Dicsas 中的所有词语在这里 都要根据语感人工赋给一个从 1 至-1 之间的情感值,用于计算情感因子的情感 度.短语和句子规则解释器对短语和句子规则进行解释,实现对句子的情感度计 算;评价对象提取规则解释器对评价对象提取规则进行解释,实现一般情况下评 价对象的提取.OBJ 模块是对评价对象提取规则的补充,实现对复杂情况的处理。



以中国计算机学会(CCF)2012 年主办的第一届中文微博情感分析评测提供的 评测语料为对象,对 20 篇话题型微博语料进行了观点句识别,观点句极性识别和 观点句评价对象提取的实验。

情感分析系统 CUCsas 对话题型微博语料的分析结果

实验（评测）项目	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
观点句识别	0.756	0.812	0.783	0.757	0.797	0.773
观点句极性识别	0.844	0.685	0.756	0.842	0.672	0.745
评价对象识别（严格）	0.303	0.275	0.288	0.305	0.265	0.278
评价对象识别（宽松）	0.387	0.356	0.371	0.391	0.344	0.359

4. 情感分析总结

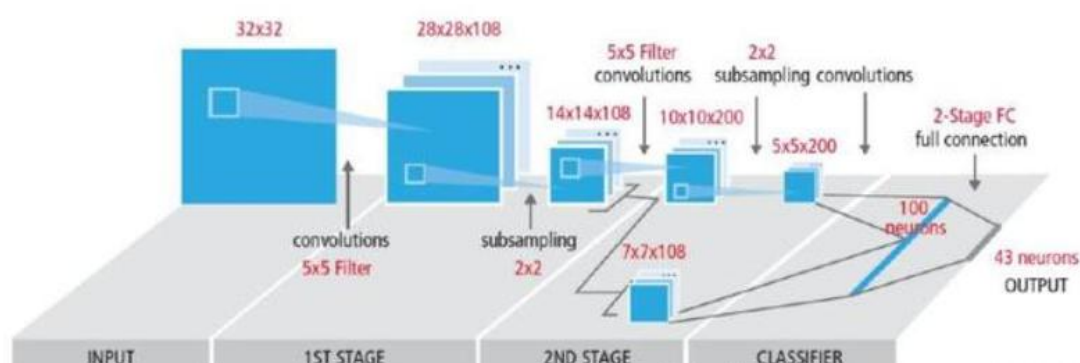
4.1 情感分析应用技术

4.1.1 卷积神经网络

CNN 非常擅长处理图像数据，如下图所示，它们一般由若干个卷积和池化操作组成，通常跟随着一个或多个全连接层。

相比于全连接神经网络，卷积网络每一个单元都只会和上一层部分单元相连接。一般每个卷积层的单元都可以组织成一个三维张量，即矩阵沿第三个方向增加一维数据。例如 Cifar-10 数据集的输入层就可以组织成 $32 \times 32 \times 3$ 的三维张量，其中 32×32 代表图片的尺寸或像素数量，而 3 代表 RGB 三色通道。

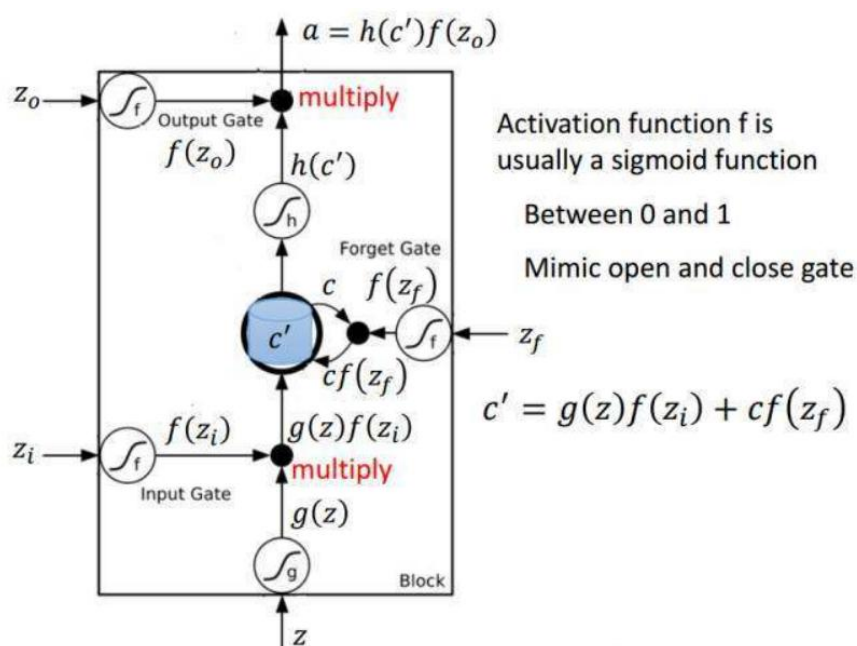
卷积神经网络中最重要的就是卷积层，卷积层试图将神经网络中的每一小块进行更加深入的分析，从而得出抽象程度更高的特征。一般来说通过卷积层处理的神经元结点矩阵会变得更深，即神经元的组织在第三个维度上会增加。



4.1.2 LSTM 网络

下面简要地介绍 LSTM 单元选择记忆或遗忘的具体处理流程。

以下是 LSTM 单元的详细结构，其中 Z 为输入部分， Z_i 、 Z_o 和 Z_f 分别为控制三个门的值，即它们会通过激活函数 f 对输入信息进行筛选。一般激活函数可以选择为 Sigmoid 函数，因为它的输出值为 0 到 1，即表示这三个门被打开的程度。



4.1.3 RNN 与注意力机制

一般来说,我们可能会认为双向 RNN 与 LSTM 就能处理数据中的长期依赖性。但是在实践中,时序数据的长期依赖性问题仍然很难处理。因此,Bahdanau 等人提出了注意力机制。

神经网络中的注意力机制受到人类视觉中注意力的启发,即人类视觉注意力能够聚焦到图像的特定区域,并在这个区域有非常高的分辨率,而在其它区域有较低的分辨率。在自然语言处理中,注意力机制允许模型根据输入文本以及它到目前为止已经生成的隐藏状态来学习要注意什么,而不像标准 RNN 与 LSTM 那样将全部原文本编码成固定长度的向量。

4.2 总结

随着互联网的飞速发展,人们通过互联网进行交流,参与评论增加,文本情感分析对于客户体验,市场研究,客户洞察,数字分析和媒体评测等变得越来越重要。目前在文本情感分析处理上应用到神经网络以及跨语言多任务学习,利用对抗生成网络解决数据不足的问题。文本情感分析是一个新兴的研究方向,而且前期工作都是以长文本为研究重点。近年来由于 Facebook,Teitter,新浪微博等短文本社交平台流行,对短文本进行情感分析的研究越来越多。但是对于短文本的情感分析研究主要集中在通过情感分析挖掘其存在的商业价值,且对情感的分析类别研究比较宽泛,缺乏对某方面情感的进一步研究,如通过互联网用户产生的文本信息判断其情绪消极程度,从而进一步判断其患有抑郁症的可能性。接下来,对互联网用户文本情感进行细化研究成为一个值得探索的方向。

5. 参考文献

- [1] 侯敏, 腾永林, 李雪燕. 话题型微博语言特点及其情感分析策略研究[J]. 语言文字应用, 2013.
- [2] 李婷婷, 姬东鸿. 基于 SVM 和 CRF 多特征组合的微博情感分析[J]. 计算机应用研究, 2015.
- [3] 梁军, 柴玉梅, 原慧斌, 咎红英, 刘铭. 基于深度学习的微博情感分析[J]. 中文信息学报, 2014.
- [4] 刘爽, 赵景秀, 杨红亚, 徐冠华. 文本情感分析综述[J]. 软件导刊, 2018.
- [5] 何炎祥, 孙松涛, 牛菲菲, 李飞. 用于微博情感分析的一种情感语义增强的深度学习模型[J]. 计算机学报, 2018.
- [6] 何炎祥, 孙松涛, 牛菲菲, 李飞. 用于微博情感分析的一种情感语义增强的深度学习模型[J]. 计算机学报, 2017.
- [7] 周胜臣, 瞿文婷, 石英子, 施询之, 孙韵辰. 中文微博情感分析研究综述[J]. 计算机学报, 2013.
- [8] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- [9] Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2(3):27.
- [10] Huimin Chen, Maosong Sun, Cunchao Tu, Yankai Lin, and Zhiyuan Liu. 2016a. Neural sentiment classification with user and product

attention. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1650 – 1659.

[11] Xilun Chen, Yu Sun, Ben Athiwaratkun, Claire Cardie, and Kilian Weinberger. 2016b. Adversarial deep averaging networks for cross-lingual sentiment classification. arXiv preprint arXiv:1606.01614.

[12] Alexis Conneau, Guillaume Lample, Marc’Aurelio Ranzato, Ludovic Denoyer, and Herve J ’egou. 2017. Word translation without parallel data. arXiv preprint arXiv:1710.04087.

[13] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532 – 1543.

[14] Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. 2017. Adversarial multi-task learning for text classification. arXiv preprint arXiv:1704.05742.

[15] Zhou Zhao, Hanqing Lu, Deng Cai, Xiaofei He, and Yueting Zhuang. 2017. Microblog sentiment classification via recurrent random walk network learning. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, pages 3532 – 3538. AAAI Press.

[16] Fangzhao Wu, Yangqiu Song, and Yongfeng Huang. 2015. Microblog sentiment classification with contextual knowledge regularization. In AAAI, pages 2332 – 2338.

[17] Fangzhao Wu and Yongfeng Huang. 2016. Personalized microblog sentiment classification via multi task learning. In AAAI, pages 3059 – 3065.

[18] Duyu Tang, Bing Qin, and Ting Liu. 2015. Learning semantic representations of users and products for document level sentiment classification. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), volume 1, pages 1014 – 1023.

[19] Xiao Sun, Fei Gao, Chengcheng Li, and Fuji Ren. 2015. Chinese microblog sentiment classification based on convolution neural network with content extension method. In Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on, pages 408 – 414. IEEE.

[20] Kaisong Song, Shi Feng, Wei Gao, Daling Wang, Ge Yu, and Kam-Fai Wong. 2015b. Personalized sentiment classification based on latent individuality of microblog users. In IJCAI, pages 2277 – 2283.

[21] Kaisong Song, Shi Feng, Wei Gao, Daling Wang, Ling Chen, and Chengqi Zhang. 2015a. Build emotion lexicon from microblogs by combining effects of seed words and emoticons in a heterogeneous graph. In Proceedings of the 26th ACM Conference on Hypertext & Social Media,

pages 283 – 292. ACM.

[22] Kaisong Song, Ling Chen, Wei Gao, Shi Feng, Daling Wang, and Chengqi Zhang. 2016. Persentiment: A personalized sentiment classification system for microblog users. In Proceedings of the 25th International Conference Companion on World Wide Web, pages 255 – 258. ACM.

[23] Vibhu Jawa and Varun Hasija. 2015. A sentiment and interest based approach for product recommendation. In Modelling and Simulation (UKSim), 2015 17th UKSim-AMSS International Conference on, pages 75 – 80. IEEE.

[24] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, pages 55 – 60.