

一种基于迁移学习及多表征的微博立场分析方法

周艳芳 周 刚 鹿忠磊

(数学工程与先进计算国家重点实验室 郑州 450001)

摘 要 立场分析旨在发现用户对特定目标对象所持的观点态度。针对现有方法往往难以克服标注数据匮乏及微博文本中大量未登录词等导致的分词误差的问题,提出了基于迁移学习及字、词特征混合的立场分析方法。首先,将字、词特征输入深度神经网络,级联两者隐藏层输出,复现由分词错误引起的缺失语义信息;然后,利用与立场相关话题的辅助数据训练话题分类模型(父模型),得到更为有效的句子特征表示;接着,以父模型参数初始化立场分析模型(子模型),从辅助数据(话题分类数据)迁移知识能加强句子的语义表示能力;最后,使用有标注数据微调子模型参数并训练分类器。在 NLPCC-2016 任务 4 的语料上进行实验, F1 值达 72.2%, 优于参赛团队的最佳成绩。实验结果表明,该方法可提高立场分类性能,同时缓解分词误差带来的影响。

关键词 迁移学习, 深度学习, 立场分析, 微博

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.09.040

Approach of Stance Detection in Micro-blog Based on Transfer Learning and Multi-representation

ZHOU Yan-fang ZHOU Gang LU Zhong-lei

(State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China)

Abstract Stance detection aims to identify users' opinion towards a particular target. Aiming at the problem that existing methods are often difficult to overcome the lack of labeled data and the error caused by word segmentation of Chinese text, this paper presented a transfer learning method and a hybrid model of character-level and word-level features. Firstly, character-level and the word-level features are inputted to deep neural network and the outputs of both are concatenated to reproduce the missing semantic information caused by word segmentation. Then, a topic classification model (parent model) is trained with a large external micro-blog data to obtain the effective sentence feature representation. Next, some of parent model's parameters are used to initialize stance detection model and the knowledge transferred from auxiliary data can be used to enhance semantic representation ability of sentences. Finally, the labeled data are used to fine tune the child model and train classifiers. Experiment on NLPCC-2016 Task 4 proves that F1 value of proposed method achieves 72.2%, which is better than the best one of participating teams. The results show that this approach can improve the stance detection performance and alleviate the influence caused by word segmentation.

Keywords Transfer learning, Deep learning, Stance detection, Micro-blog

1 引言

随着 Twitter、Facebook、微博等社交平台和亚马逊、淘宝等电子商务平台的普及,网络上的评论性文本与日俱增。基于此,潜在用户可通过社交媒体评论了解大众对某一事件或产品的看法。情感分析(Sentiment Analysis, SA)是确定文本主观倾向的主要手段。然而,在更多的情况下,我们更加关注社交媒体中用户对某一特定对象的态度,例如“广大市民对烟花爆竹燃放禁令表示欢迎”。因此,立场分析(Stance Detection)应运而生。

立场分析是指自动检测文本作者对指定对象持有的支持、反对或中立的态度,该对象可以是组织机构、政府政策、

人、产品等。立场分析与情感分析之间存在实质性差异,情感分析确定文本的主观倾向,而不针对特定目标。立场分析的文本可能不包含指定对象,或者文本内容不涉及指定对象,表达较灵活,常采用讽刺、比喻和隐喻等修辞手段,人们通常会利用经验推断更广泛的语境。但对于机器来说,通过非正式词汇、语法和拼写分析文本立场具有一定的挑战性。然而,自动识别微博立场具有较大的实际应用价值。例如,其可作为传感器来衡量微博用户在政治、社会等问题上的态度,突出对特定目标实体的立场观点。

通常,立场分析研究侧重于国会辩论或在线论坛的分析^[1-4]。Rajadesingan 等^[5]研究了 Twitter 对话中的立场检测,并使用了基于转发标签(Hashtag)的传播方法。本文工作

到稿日期:2017-08-10 返修日期:2017-10-04 本文受数学工程与先进计算国家重点实验室开放基金资助项目(2015A11)资助。

周艳芳(1993—),女,硕士生,主要研究方向为自然语言处理、情感分析, E-mail: lz_xd6j@163.com;周 刚(1974—),男,博士,副教授,硕士生导师,主要研究方向为在线社会网络分析、海量信息处理, E-mail: gzhongzhou@126.com(通信作者);鹿忠磊(1988—),男,硕士生,主要研究方向为自然语言处理。

的不同之处在于,在中文文本上试图通过分析单个消息文本来检测作者的立场。微博短文本与普通文本有很多不同之处,如存在未登录词(Out of Vocabulary, OOV)增多、文法不规范、数据稀疏等问题。

2016 年, NLPCC 首次组织了关于微博立场分析的测评,成绩最好的团队分别对 5 个指定对象使用独立的分类器,大多数团队使用 bi-gram、TF-IDF、词向量^[6] (Embedding Word) 和情感词典等特征。

现有方法存在以下问题:

1) 没有针对微博文本特性(噪声大、新词多、修辞手法复杂)设计算法,导致分词误差,影响后续分析效果。

2) 标注数据有限(比赛提供有标注的训练数据共 3000 条),人工标注代价大,在实际应用中难以获得大量的标注数据。数据条件严重限制了神经网络模型的复杂程度,导致语义表达不充分等问题。

为解决上述问题,本文提出基于迁移学习^[7-8] 及字、词特征混合的多表征方法。针对问题 1), 将文本字、词特征混合输入深度神经网络,通过级联字、词隐藏层输出状态,将字的特征集成到词的特征中,复现由分词错误引起的缺失语义信息,从而缓解因分词误差带来的后续影响;针对问题 2), 由于话题分类数据的标注不需要任何专业知识,相较于立场分析标注数据,标注数据极易获得,且与立场分析模型所得句子语义表示相同,只有分类器部分不同,因此可使用与立场话题相关的大量辅助数据来训练该模型,将得到的句子特征表示迁移至立场分析模型中;最后,通过有限的有标注数据微调参数,并训练分类器。在 NLPCC-2016 任务 4 数据集上的实验表明,本文所提方法的准确率优于参赛团队的最佳成绩,显示了本文所提方法可以改善立场分析效果,突破数据量限制,缓解中文分词误差带来的影响。此外,本文在实验中探索迁移学习过程中固定与微调参数的选择对立场分析结果的影响。

本文第 2 节介绍所用立场分析模型;第 3 节概述迁移学习的基础知识,并说明如何用迁移学习提高立场分析的准确率;第 4 节介绍实验设置及实验结果;最后总结全文,并指出下一步工作方向。

2 基于注意力机制的长短时记忆网络多表征立场分析模型

分词是中文自然语言处理任务中的第一步,由分词带来的误差会传递给整个深度神经网络。分词误差通常由两方面产生:1) 未登录词问题,即出现词典中没有的词,如人名、地名、机构名、一些新词等;2) 歧义问题。微博文本中存在的许多新兴网络词语均属于未登录词。为减小分词带来的影响,提高中文立场的分析效果,本文采用字、词混合特征,通过级联字、词隐藏层输出状态,将字的特征集成到词的特征中,复现由分词错误引起的缺失语义信息,减少错误语义表示。通常,微博文本中影响立场分析结果的只有部分关键词语,而不是组成句子的所有单词。因此,本文采用注意力机制,为句子

中的每个词附加权重,突出关键信息,削减冗余信息,并提出基于注意力机制的长短时记忆(Long Short-Term Memory, LSTM)多表征模型,得到句子表示向量作为分类特征。该模型由查找层、LSTM 网络与分类器 3 个部分组成,其结构如图 1 所示。

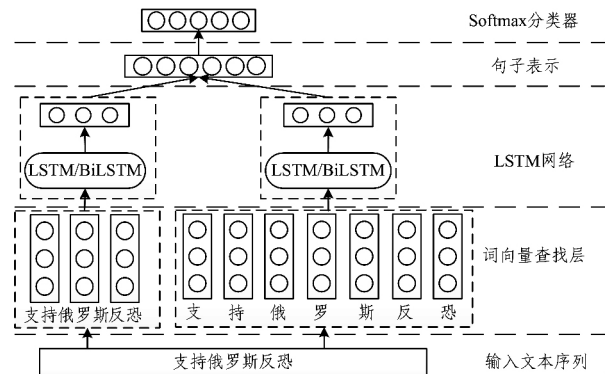


图 1 基于注意力机制的多表征模型

Fig. 1 Multi-representation model based on attention mechanism

2.1 查找层

词向量是将文本特征转换为连续的稠密向量表示,获得对应词语的语义特征。查找层是该模型的第一层,将输入文本序列中的每个词映射为“ N ”维向量。给定有限词典 D , D 从训练语料中提取,未知词被映射为 $\langle unk \rangle$ 。查找表是一个 $N \times |D|$ 的矩阵 M ,单词 w_i 的词向量为 Ew_i ,其为 M 中索引为 i 的列向量。因此,查找层可将输入单词序列 $\{w_1, w_2, \dots, w_n\}$ 转换为一系列词向量 $\{Ew_1, Ew_2, \dots, Ew_n\}$ 。查找表分为词向量查找表与字向量查找表。预训练的词向量可改善模型的训练结果^[9],因此,我们使用 word2vec¹⁾ 工具训练词向量,从大量微博数据及论坛数据中训练字和词的向量表示。

2.2 LSTM 网络

循环神经网络(Recurrent Neural Networks, RNN)是常规前馈神经网络(Feedforward Neural Network, FNN)的扩展,可处理变长输入序列^[10]。但在实际训练中,RNN 存在梯度爆炸和梯度消失的问题^[11],对长文本的表示效果不佳,因此 Hochreiter 等^[12] 提出了改进模型——长短时记忆(LSTM)模型。该循环神经网络通过在隐藏层增加 3 种“门”结构来控制细胞状态,关键架构是具有门控访问的存储单元,用于存储和获取信息,非常适合解决长序列建模问题。LSTM 模型的结构如图 2 所示。

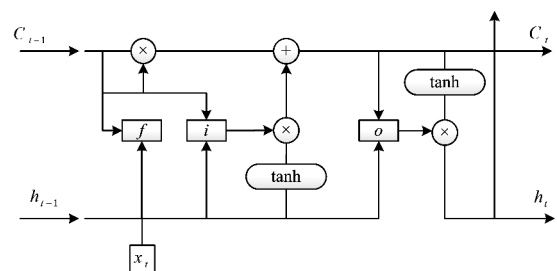


图 2 LSTM 模型的隐藏层单元

Fig. 2 Hidden layer unit of LSTM model

¹⁾ <http://code.google.com/p/word2vec/>

给定输入序列 (x_1, x_2, \dots, x_T) , LSTM 计算 h 序列 (h_1, h_2, \dots, h_T) 和 C 序列 (C_1, C_2, \dots, C_T) 。令 h_{t-1} 和 C_{t-1} 分别表示 $t-1$ 时刻的隐藏状态和存储单元, LSTM 网络根据输入 x_t 更新隐藏状态和存储单元的值, 如下所示:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$\tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \quad (4)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t \quad (5)$$

$$h_t = o_t \otimes \tanh C_t \quad (6)$$

其中, i_t, f_t, o_t 分别是输入门、遗忘门和输出门, σ 是 sigmoid 函数, \otimes 表示对应元素做乘法, 矩阵 W 、矩阵 U 和向量 b 是网络的参数。

LSTM 只进行单向编码。进一步, 对于时间步长 t , 其隐藏层输出仅包含 1 至 $t-1$ 时刻的信息, 即上文信息。而下文信息对整个语义的刻画同样重要。为更好地利用上下文信息, 且基于 BiRNN 模型^[13-14]在语音识别领域的成功应用, 本文提出使用 BiLSTM 模型, 该类模型可利用历史和将来的所有可用输入信息进行训练, 获得更加全面且准确的语义向量表示。

与 LSTM 相比, BiLSTM 使用两个单独的隐藏层来读取两个方向的输入: 正向和反向。正向以原始顺序(1 到 t)读取输入; 反向按照相反的顺序(t 到 1)读取输入。本文使用 BiLSTM 生成句子向量表示时, 通过级联方式综合表示语义。

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (7)$$

从而, 隐藏状态既包含正向信息也包含反向信息, 即包含上下文信息。通过双向模型, LSTM 在较长序列上的记忆表现可得到有效提高。

2.3 注意力机制

注意力机制(Attention Mechanism)最早被用于神经机器翻译(Neural Machine Translation, NMT), 在生成单个目标语言单词时, 仅与源语言句子序列中的个别词相关。与绝大多数词无关或弱相关。因此, Bahadanau 等^[15]提出使用 Attention 机制为每个目标语言词动态生成源语言词的上下文向量, 而非采用整个源语言句子序列的定长向量。

对短文本的立场分析同样如此, 为将模型的“注意力”放在影响立场分析的关键词语上, 本文采用 Attention 机制, 基于上下文语义动态地为输入序列中的每个词赋予权重。图 3 为 Attention 机制原理图。

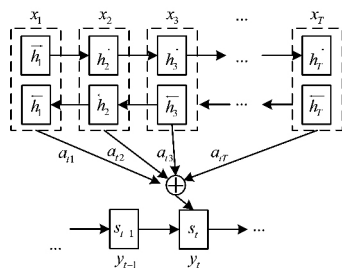


图3 注意力机制原理图

Fig. 3 Schematic diagram of attention mechanism

$$u_t = \tanh(W_u h_t + b_u) \quad (8)$$

$$a_t = \frac{\exp(u_t^T u_w)}{\sum_t \exp(u_t^T u_w)} \quad (9)$$

$$v = \sum_t a_t h_t \quad (10)$$

本文使用 Attention 机制获得句子组成词对该句子含义表示的贡献程度, 且通过整合这些重要词语形成句子表示。隐藏层 h_t 通过单层神经网络后可得到 u_t , 然后测量 u_t 与 u_w 之间的相似度, 得到归一化的权重 a_t , 最后将计算得到的权重相加和结果作为句子向量表示。

2.4 分类器

基于以上工作, 得到待处理文本的向量表示。给定待分类类别 C , 通过线性计算将句子向量映射为维度为 C 的实值向量, 再使用 Softmax 将这些实值映射为对应的概率, 计算公式如下:

$$P_i = \frac{\exp(x_i)}{\sum_{j=1}^C \exp(x_j)} \quad (11)$$

其中, $\sum_{i=1}^C P_i = 1$ 。

本文训练数据中的待处理文本均有标准标签。采用交叉熵作为损失函数, 使用随机梯度下降(SGD)法更新参数。

$$loss = - \sum_{s \in T} \sum_{i=1}^C P^s(s) \cdot \log(P_i(s)) \quad (12)$$

其中, T 是训练语料库, s 代表中文短文, C 是类别数, $P^s(s)$ 为标准标签, $P(s)$ 为预测标签概率。

3 迁移学习

定义 1^[7] 给定源域数据 D^S 和源域任务 T^S 、目标域数据 D^T 和目标域任务 T^T , 迁移学习就是研究如何利用源域数据和源域任务来帮助改善目标域数据的学习效果。一般地, $D_S \neq D_T, T_S \neq T_T$ 。

迁移学习是一种特殊的机器学习方法, 它放宽了传统机器学习中的两个基本假设: 1) 用于学习的训练样本与新的测试样本满足独立同分布的条件; 2) 必须有足够可利用的训练样本才能学习得到一个良好的分类模型。其目的是将已有知识进行迁移, 用于解决目标领域中仅有少量有标注样本数据甚至没有标注样本数据的问题。

迁移学习中迁移的“知识”有很多形式, 如“有标注的数据”“模型参数”等。在基于深度学习的自然语言处理任务中, 最大的障碍是在训练大模型的海量参数过程缺乏所需的海量数据。深度学习取得的成果一部分就在于大数据的支撑, 在面对某一领域的具体问题时, 通常可能无法得到构建模型所需规模的数据。然而, 在一个模型的训练任务中, 针对某种类型数据获得的特征可直接应用于同一领域的不同任务中。

Yosinski 等^[16]在基于深度学习的图像分类任务中, 使用迁移学习特征初始化可提升模型的准确率。在自然语言处理领域中, 在无标注的数据上学习的句子表示已被证明可用于各种分类和语义相似度任务中^[17-19]。Weston 等^[20]使用推特主题标签(hashtag)分类任务来改进基于下游任务的推荐系统的句子表示。

在立场分析任务中,使用深度学习模型可获得文本语义,且结构复杂的神经网络模型可获得更准确的语义表示,但建模能力受训练数据量大小的限制。相关话题分类的数据标注不需要任何专业知识及主观判断,且标注数据较易获得,因此本文提出基于迁移学习的思想,将外部辅助数据获取的知识迁移至立场分析模型,使其更准确地建模该文本所表达的立场。首先通过获取相关话题的辅助数据(与5个立场分析任务相同的话题)来训练话题分类模型(父模型),利用父模型参数(即“知识”)初始化立场分析模型(子模型),再利用有标签的数据微调子模型,即可得到立场分类器。具体的迁移过程如图4所示。

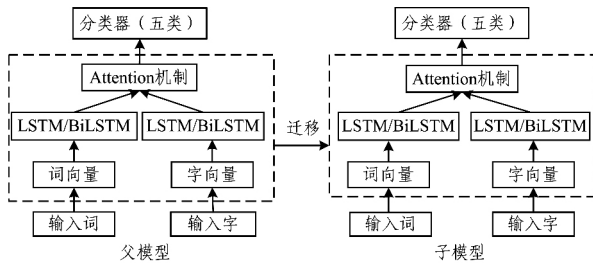


图4 迁移学习流程图

Fig. 4 Flowchart of transfer learning

父模型中5个模块(词向量、字符向量、两部分LSTM/BiLSTM与Attention机制)的参数分别向子模型的对应模块迁移。父、子模型只在分类器部分不同,从输入序列到分类器之前(即图4中虚线框所示部分)均可看作文本的特征提取过程,得到文本句子的特征表示。两个分类任务相关但不相同,符合迁移条件。将外部大量相关数据训练得到的父模型参数迁移至子模型,可得到高质量的句子向量表示,为准确分析立场打下基础。在子模型中,利用有标签数据微调模型,可改善立场分析效果。

4 实验

本文实验内容分为4部分:1)验证迁移学习在立场分析中的有效性;2)探索立场分析模型的结构对分析效果的影响;3)验证字、词特征混合的多表征输入对句子语义建模能力的影响;4)探索迁移学习过程中固定和微调参数的选择对准确率的影响。

4.1 实验数据的预处理与参数的设置

本实验所使用的实验数据为NLPC-2016任务4提供的微博立场观点分析数据。

任务:有监督学习,给定有标签训练数据3000条,测试数据1000条,包含5个话题:“深圳禁摩限电”“开放二胎”“俄罗斯叙利亚反恐行动”“iPhone SE”“春节放鞭炮”。分析文本中对指定目标的立场观点。

此外,本文额外爬取了8万条与这5个话题相关的数据,用以训练话题分类模型。

数据预处理过程包括数据清洗和分词。数据清洗过程包括去除一些较短句子和标点。分词采用jieba分词工具。构建词表与字表,预训练使用word2vec分别训练100维的词向

量与字向量。模型参数的设置如表1所列。

表1 参数设置

Table 1 Setting of parameters

参数	设置	实验范围
最大文本长度	50	40~100
词向量维数	100	50,100,200
字符向量维数	100	50,100,200
隐藏层单元数	400	200,300,400
轮数(Epoch size)	10	10
批处理大小	16	8,16,32
丢弃率(dropout)	0.5	0.4,0.5,0.7
学习率	0.01	0.01

以Tensorflow为实验平台,实验硬件环境为Intel core i7处理器,16GB内存,操作系统为64位Ubuntu 14.04 LTS。为提高训练效率,在此基础上额外增加了一块GTX 980的GPU加速卡。

$$F_{AVG} = \frac{F_{FAVOR} + F_{AGAINST}}{2}$$

$$F_{FAVOR} = \frac{2 * P_{FAVOR} * R_{FAVOR}}{P_{FAVOR} + R_{FAVOR}}$$

$$F_{AGAINST} = \frac{2 * P_{AGAINST} * R_{AGAINST}}{P_{AGAINST} + R_{AGAINST}}$$

该比赛以 F_{FAVOR} 与 $F_{AGAINST}$ 的平均值作为评测标准。

4.2 实验内容

为完成3个实验内容,本文设置以下对比实验。

1)NBOW模型:使用TF-IDF作为单词特征。

2)LSTM模型:分别使用随机初始化词向量WR、预训练的词向量WE、预训练的字向量CE以及预训练的词向量与字向量(隐藏层输出状态级联)同时训练LSTM模型,未使用迁移学习。

3)BiLSTM模型:与LSTM的实验设置相同,增加一组迁移学习实验T。

4)固定与微调参数设置:在BiLSTM迁移学习实验的基础上,同时微调Attention机制与分类器参数。

词表由训练集与测试集数据生成,以上实验词表、分类器Softmax以及其他参数的设置均相同。

4.3 实验结果及分析

实验结果如表2所列。可以看出,使用预训练后的向量可提高分类性能,单独使用预训练后的词向量与字向量的效果较好,且使用预训练后的词向量与字向量混合训练的效果最好,表现出能缓解分词误差带来的影响能力;在模型选择方面,同等参数设置下,BiLSTM模型的性能优于LSTM模型与NBOW模型,表明BiLSTM模型对句子语义特征的建模能力更好。相比于非迁移学习方法,使用迁移学习后的结果明显优于同等参数设置下其他模型的结果,且在支持与反对两个类别上的效果都有提高,证明了迁移学习的有效性。在迁移学习过程中,仅微调分类器参数的效果较同时调整Attention机制的效果略差,说明在话题分类模型中影响分类性能的关键词与在立场分析中影响分类性能的关键词不同,这与我们的常规认识一致。

表2 测评结果
Table 2 Test results

方法	总评		
	FFAVOR	FAGAINST	FAVG
NBoW	0.498	0.520	0.509
LSTM+WR	0.613	0.643	0.628
LSTM+CE	0.632	0.655	0.643
LSTM+WE	0.653	0.679	0.666
LSTM+WE+CE	0.689	0.693	0.691
BiLSTM+WR	0.620	0.645	0.633
BiLSTM+CE	0.632	0.667	0.650
BiLSTM+WE	0.655	0.682	0.668
BiLSTM+WE+CE	0.692	0.707	0.700
BiLSTM+WE+CE+T (微调分类器)	0.706	0.729	0.718
BiLSTM+WE+CE+T (微调 Attention 机制与分类器)	0.711	0.732	0.722

结束语 本文提出一种基于迁移学习及字、词特征混合的多表征微博立场分析方法,构建基于注意力机制的 LSTM 多表征模型,通过联合字、词特征重构由中文分词引起的缺失语义信息,加强句子的语义表示,对缓解分词误差带来的影响具有一定的效果。针对立场分析标注匮乏的问题,引入大量外部相关话题辅助数据来训练话题分类模型(父模型),利用父模型参数初始化立场分类模型(子模型),并使用有标注的数据微调子模型,训练立场分类器,得到立场分析结果。在 NLPCC-2016 任务 4 提供的语料库上进行测试,结果验证了本文方法具有更好的句子语义特征建模能力,基于该句子特征向量的立场分析性能较高;并且在迁移学习过程中,同时微调 Attention 机制与分类器参数的效果更好。

NLPCC-2016 任务 4 中 B 题主要针对无监督的立场分析技术进行研究,在实际应用中也存在无标注数据的立场分析情况。因此,下一步将对无监督的立场分析技术进行研究。

参考文献

- [1] THOMAS M, PANG B, LEE L. Get out the vote: determining support or opposition from congressional floor-debate transcripts[C]//Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2006:327-335.
- [2] SOMASUNDARAN S, WIEBE J. Recognizing Stances in Online Debates[C]//Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. 2009:116-124.
- [3] MURAKAMI A, RAYMOND R. Support or oppose?: classifying positions in online debates from reply activities and opinion expressions[C]//International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, 2010:869-875.
- [4] WALKER M A, ANAND P, ABBOTT R, et al. Stance classification using dialogic properties of persuasion[C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2013:592-596.
- [5] RAJADESINGAN A, LIU H. Identifying Users with Opposing Opinions in Twitter Debates[M]//Social Computing, Behavior-

al-Cultural Modeling and Prediction. Springer International Publishing, 2016:153-160.

- [6] MIKOLOV T, LE Q V, SUTSKEVER I. Exploiting Similarities among Languages for Machine Translation[J]. arXiv preprint arXiv:1309.4168. 2013.
 - [7] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Data Engineering, 2010, 22(10):1345-1359.
 - [8] SCHÖLKOPF B, PLATT J, HOFMANN T. Analysis of Representations for Domain Adaptation[C]//International Conference on Neural Information Processing Systems. MIT Press, 2006:137-144.
 - [9] ERHAN D, BENGIO Y, COURVILLE A C, et al. Why Does Unsupervised Pre-training Help Deep Learning? [J]. Journal of Machine Learning Research, 2010, 11(3):625-660.
 - [10] GRAVES A. Supervised Sequence Labelling with Recurrent Neural Networks[OL]. <http://mediatum.ub.tum.de/doc/673554/file.pdf>.
 - [11] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 2002, 5(2):157-166.
 - [12] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735.
 - [13] SCHUSTER M, PALIWAL K K. Bidirectional re-current neural networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11):2673-2681.
 - [14] GRAVES A, JAITLEY N, MOHAMED A R. Hybrid speech recognition with Deep Bidirectional LSTM[C]//Automatic Speech Recognition and Understanding. IEEE, 2014:273-278.
 - [15] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[C]//3rd International Conference on Learning Representation. 2015.
 - [16] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks? [J]. EprintArxiv, 2014, 27:3320-3328.
 - [17] KIROS R, ZHU Y, SALAKHUTDINOV R, et al. Skip-Thought Vectors[OL]. <http://www.cs.toronto.edu/~zemel/documents/skipThought.pdf>.
 - [18] DAI A M, LE Q V. Semi-supervised Sequence Learning[C]//International Conference on Neural Information Processing Systems. 2015:3079-3087.
 - [19] HILL F, CHO K, KORHONEN A. Learning Distributed Representations of Sentences from Unlabelled Data[C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016:1367-1377.
 - [20] WESTON J, CHOPRA S, ADAMS K. # TagSpace: Semantic Embeddings from Hashtags [C]//Conference on Empirical Methods in Natural Language Processing. 2014:1822-1827.
 - [21] YANG W, SONG J J, TANG J Q. A Study on the Classification Approach for Chinese MicroBlog Subjective and Objective Sentences[J]. Journal of Chongqing Institute of Technology, 2013, 27(1):51-56. (in Chinses)
- 杨武, 宋静静, 唐继强. 中文微博情感分析中主客观句分类方法[J]. 重庆理工大学学报(自然科学), 2013, 27(1):51-56.