DOI：10.11830/ISSN.1000-5013.201804046

对抗长短时记忆网络的跨语言文本情感分类方法

党莉，陈锻生，张洪博

（华侨大学 计算机科学与技术学院，福建 厦门 361021）

摘要： 针对文本情感分类任务中，有情感标注的语料在不同语言中的不均衡问题，结合深度学习和迁移学习，提出一种基于对抗长短时记忆网络（ALSTM）的跨语言文本情感分类方法. 设置双语各自独立的特征提取网络和共享特征提取网络，把获取到的特征拼接输入到分类器进行分类. 在共享特征提取网络中，设置语言分类器，运用对抗思想优化模型，通过投票法决定文本最终的情感极性. 实验表明：该方法可以取得跨语言文本情感分类任务更高的准确度.

关键词： 文本情感；跨语言；对抗；长短时记忆网络；共享特征

中图分类号： TP 183；TP 391.1 **文献标志码：** A **文章编号：** 1000-5013(2019)02-0251-06

Cross-Lingual Sentiment Classification Method Based on Adversarial Long Short Term Memory Network

DANG Li, CHEN Duansheng, ZHANG Hongbo

(College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

Abstract: This paper proposes a cross-lingual sentiment classification method based on adversarial long short term memory (ALSTM) network, which aims at the problem of text sentiment classification in the disparity of emotionally annotated corpus in different languages, combined with deep learning and transfer learning. Bilingual feature extraction networks and a shared feature extraction network are set up, and then the extracted features are merged for classification. In the shared feature extraction network, a language classifier is set up. Using the adversarial idea to optimize the model, and the final polarity of the text depends on the voting results. Experiments show that cross-lingual sentiment classification can achieve higher accuracy by this method.

Keywords: sentiment of the text; cross-lingual; adversarial; long short term memory network; shared features

电子商务行业蓬勃发展，在各种交易平台都会找到各种商品的评价. 如何从这些海量数据找到其背后的潜在价值，成为亟需解决的问题. 由于中文较其他语言起步较晚，缺乏高质量的语料资源，人工标注又需要投入巨大的人力物力，这在一定程度上阻碍了中文情感分类的研究. 跨语言情感分析是利用一种语言的丰富情感资源协助或提高另一种语言的情感分析[1]. 在跨语言情感分类任务中，最常用的是机器翻译的方法[2]，但是机器翻译的方法会出现翻译误差. Wan 等[3]采用半监督的方法弥补翻译损失. 此外，还有基于双语词典和平行语料的方法[4]，但是，这些双语资源在现实中都很难获取. 近年来，深度学

收稿日期： 2018-04-14

通信作者： 陈锻生（1959-），教授，博士，主要从事数字图像处理与模式识别的研究. E-mail：dschen@hqu.edu.cn.

基金项目： 国家自然科学基金资助项目（61502182）；福建省科技计划重点项目（2015H0025）

习在自然语言处理领域的应用越来越广泛[5-6]. Zhou 等[7]提出一种基于去噪自动编码机(DAE)的双语情感词嵌入算法,通过语义学习和情感学习阶段获得两个视图的共同表示,准确率达到 80.68%. Zhou 等[8]提出双语语义和情感特征表示(BSSR)算法,准确率达到 82.24%. Zhou 等[9]提出一种基于注意力机制的长短时记忆网络(Attention-based LSTM),利用词级注意力和句子级注意力进行优化,将分类准确度提高到 82.40%. Ben-David 等[10]提出,一个良好的特征表示,应该是域分类器分不出此特征来自源领域还是目标领域. 目前,利用对抗思想,在计算机视觉的图片生成[11]和领域适配[12]方面取得了很好的效果. 跨语言问题也是同样道理,一个好的迁移特征应该使语言分类器分不清特征来源于源语言还是目标语言. 因此,本文提出一种基于对抗长短时记忆网络(ALSTM)的跨语言文本情感分类方法.

1 基于 ALSTM 的跨语言文本情感分类

1.1 数据预处理

采用基于 FoolNLTK 的中文分词和基于 Glove 模型[13]的词向量进行预训练. 基于 FoolNLTK 分词工具(https://github.com/rockyzhengwu/FoolNLTK)进行优化,构建常用的网络用语词表作为中文分词的辅助词表,实现更准确的中文分词,为后续词向量的预训练提供更好的数据来源. 同时,采用基于 Glove 模型的词向量预训练方法,通过预训练的方式引入额外的语料库,其中,包括大量中文无标注数据. Glove 模型综合运用词的全局统计信息和局部统计信息生成语言模型和词的向量化表示. 由于词向量是依据大量的无标注语料生成的,所以,能比单纯地使用标注语料进行情感文本分类接触到更多的数据. 实验中,采用的词向量维数为 300.

1.2 网络框架

跨语言文本情感分类网络研究框架,如图 1 所示. 图 1 中,x^s 和 x^t 分别为源语言和目标语言的输入. 该网络主要包括特征提取和分类预测两部分. 特征提取部分主要包括 3 个网络:源语言特征提取网络 F_p^s,目标语言特征提取网络 F_p^t 和共享特征提取网络 F_c. 分类预测部分主要包括 2 个分类器:文本情感极性分类器 C_p 和语言分类器 C_l. 其中,文本情感极性分类器用来预测 3 种方式融合得到的特征极性;语言分类器用来预测共享特征提取网络提取的特征来源于源语言还是目标语言. 文本情感极性分类器最小化文本极性分类损失,语言分类器最大化语言分类损失,使语言分类器最大程度分不清特征来源. 通过这种对抗训练,使网络参数得以优化,以便于获取双语不变特征.

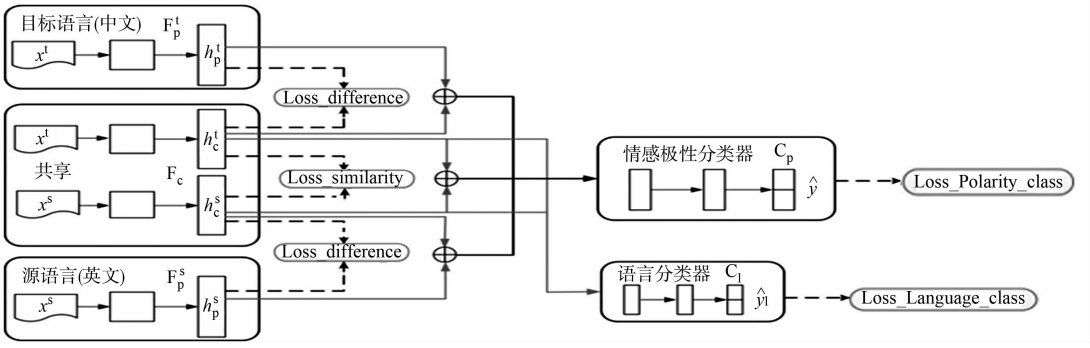


图 1 跨语言文本情感分类框架图

Fig. 1 Cross-lingual sentiment classification framework

在特征提取部分,采用加入注意力机制的长短时记忆网络(LSTM)[14]. 传统的 LSTM 网络中,每个输入的词语都赋予了相同的权重. 采取基于注意力机制的 LSTM 网络有两方面原因:一是机器翻译的方法不可避免地引入噪音,通过注意力机制可以减小这些噪音的影响;二是注意到不同的词语对语句的极性贡献大小不同. 注意力机制通过保留 LSTM 对输入序列的中间输出结果,经过 Softmax 进行归一化得到符合概率分布取值区间的注意力分配概率分布数值,然后,训练模型对输入进行选择性学习.

通过一个单层的神经网络,计算每个时刻隐藏状态所占权重,即

$$\beta_i = V_a \tanh(h_i W_a + b_a), \qquad \alpha_i = e^{\beta_i} / \sum_{k=1}^{t} e^{\beta_k}.$$

加权求和，得到最终的隐藏状态，即

$$h = \sum_{i=1}^{t} h_i \alpha_i .$$

特征提取网络，如图 2 所示. 通过机器翻译工具，把源语言的训练数据 x^s 翻译成目标语言的训练数据 x^t. 源语言特征提取网络 F_p^s 获取到源语言特征 h_p^s，目标语言特征提取网络 F_p^t 获取到目标语言特征 h_p^t，共享特征提取网络 F_c 获取到源语言特征 h_c^s 和目标语言特征 h_c^t. 然后，将源语言特征和源语言特征拼接，目标语言特征和目标语言特征拼接，源语言特征和目标语言特征拼接，分别输入极性分类器 C_p 进行分类. 最终的分类结果取决于 3 种融合特征的投票结果.

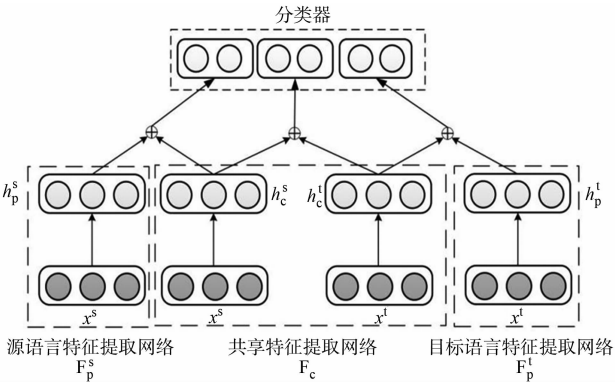


图 2　特征提取网络

Fig. 2　Feature extraction networks

整个网络的损失为

$$\text{losses} = \text{loss\_en\_en} + \text{loss\_cn\_cn} + \text{loss\_en\_cn}.$$

上式中：loss_en_en＝loss_Polarity_class＋α×loss_difference＋loss_en 表示对源语言特征拼接后的极性分类预测损失；loss_cn_cn＝loss_Polarity_class＋β×loss_difference＋loss_cn 表示对目标语言特征拼接后的极性分类预测损失；loss_en_cn＝loss_Polarity_class＋loss_similarity＋loss_Language_class 表示对源语言和目标语言特征拼接后的极性分类预测损失. 其中，loss_Polarity_class 为极性分类误差；loss_difference 为独立特征提取网络获取的特征和通过共享特征提取网络获取的特征之间的距离，文中采用欧氏距离；loss_en 为英文分类损失；loss_cn 为中文分类损失；loss_similarity 为通过共享特征提取网络获取的特征 h_c^s 和 h_c^t 之间的距离；loss_Language_class 为通过语言分类器的语言类别分类损失；α，β 为超参数，取 0.1.

1.3　领域对抗训练

设置共享特征提取网络，使源语言的分布和目标语言的分布尽可能地接近，以便于网络学习到双语的不变性特征. 对共享特征提取网络获取的源语言特征 h_c^s 和目标语言特征 h_c^t，采用最大均值差异损失(MMD)[15]衡量 2 个分布的相似性. 表达式为

$$\text{loss\_similarity} = \frac{1}{(N^s)^2}\left(\sum_{i,j=0}^{N^s} k(h_{c,i}^s, h_{c,j}^s) - \frac{2}{N^s N^t}\sum_{i,j=0}^{N^s,N^t} k(h_{c,i}^s, h_{c,j}^t) + \frac{1}{(N^t)^2}\sum_{i,j=0}^{N^t} k(h_{c,i}^t, h_{c,j}^s)\right).$$

上式中：$k(\cdot)$为映射，用于把原变量映射到高维空间中，实验采用高斯核函数.

设置语言分类器，最大程度模糊两个分布，让分类器分不清特征来源于源语言还是目标语言. 对共享特征提取网络提取到的源语言特征 h_c^s 添加标签(0,1)，目标语言特征 h_c^t 添加标签(1,0)，依次输入到语言分类器 C_l 中. 定义 $L(y,\hat{y})$，其中，y 为数据的原始标签，$\hat{y}$ 为域分类器预测的标签.

在特征提取网络和语言分类器之间，采用梯度反转层(GRL)[16]. 在前向传播期间，GRL 作为一种恒等交换；在后向传播过程中，GRL 从后面的层获得梯度并改变其符号，即将其乘以－1，然后，将其传递到前一层. GRL 是对抗性的. 一方面，优化网络以增强语言分类器区分特征来源于源语言还是目标语言的能力；另一方面，梯度反转层使判别特征表示来自哪种语言的能力被降低. 交叉熵损失为

$$\text{loss\_Language\_class} = \sum_{i=0}^{N_s+N_t} \{y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\}.$$

2　实验结果与分析

2.1　实验数据

采用第二届自然语言处理与中文计算会议中跨语言情感分析评测任务的公开数据集. 其中，源语言为英文，目标语言为中文. 数据主要来自亚马逊的中英文商品评论，覆盖 book，dvd，music 3 个领域. 3

个领域的训练数据均包含积极、消极比例为1∶1的4 000条有标注的英文训练数据和4 000条无标注中文测试数据,以及积极和消极的比例不平衡的大量中文无标注数据.实验数据集,如表1所示.

表1 实验数据集

Tab.1 Experimental data set 条

数据集	领域		
	book	dvd	music
英文(有标注训练集)	4 000	4 000	4 000
中文(无标注测试集)	4 000	4 000	4 000
中文(无标注训练集)	44 113	17 815	29 678

将4 000条英文数据作为源语言训练集,同时,利用谷歌在线翻译工具,将4 000条英文数据进行翻译,得到对应的中文数据作为目标语言训练集.将4 000条中文数据作为测试集,并将4 000条中文数据进行翻译得到的对应英文数据作为源语言测试集.同时,将大量无标注中文数据作为前期词向量预训练的部分语料.采用Google在线翻译工具(https://translate.google.cn),此工具被认为是现阶段最好的机器翻译系统.采用的编程语言为Python,编辑器为PyCharm,网络框架为Tensorflow.采用Ubuntu16.04 64位操作系统,主机内存为64 GB,处理器是Intel® Xeon(R) CPU E5-2630.

2.2 结果分析

为了检测实验结果,采用两部分的基准模型作为实验结果的参照.第一部分是与当前在跨语言文本情感分类任务上比较先进的方法作对比;第二部分是实验所采用方法作自身对比.

当前比较先进的方法主要包括逻辑回归(LR),支持向量机(SVM),DAE[7],BSSR[8]和Attention-based-LSTM[9]等方法.

为了说明文中系统框架下提出方法的有效性,将实验的分支部分独立进行结果验证,具体流程包括如下4种.

1) Sh.仅利用共享特征提取网络对源语言和目标语言进行特征提取,然后,拼接输入分类器分类.

2) Sh-Ad.利用共享特征提取网络对源语言和目标语言进行特征提取,同时,加入语言分类器,进行对抗训练.

3) Pr-Sh.同时设置独立的网络特征提取网络和共享特征提取网络、源语言和源语言特征拼接、目标语言和目标语言拼接、源语言和目标语言拼接,输入到分类器进行分类.

4) ALSTM.文中的最终做法是设置3个特征提取网络,同时,加入语言分类器,运用对抗思想优化模型,通过投票法决定文本最终的情感极性.

实验中,词向量的维度取300维;batch_size大小为200;训练集的dropout率设置为0.5,防止过拟合;学习率衰减权重为0.95;迭代次数为40次,直至准确度不再提升.两部分的基准模型与文中方法的对比结果,如表2所示.

表2 实验结果对比

Tab.2 Comparison of experimental results

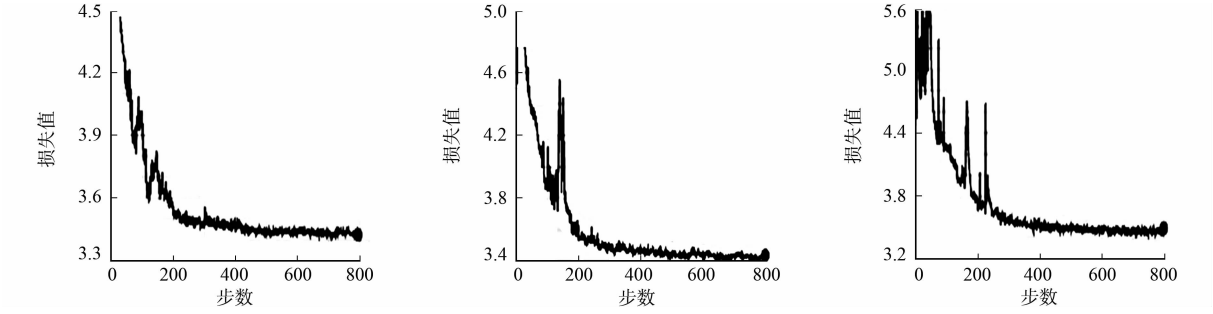
分类方法	准确度/%			平均准确度/%
	book	dvd	music	
LR	76.60	79.50	75.50	77.20
SVM	79.60	80.20	78.50	79.43
DAE	81.05	81.60	79.40	80.68
BSSR	82.15	83.03	81.55	82.24
Attention-based-LSTM	82.10	83.70	81.30	82.40
Sh	79.85	80.35	78.70	79.63
Sh-Ad	80.50	81.20	79.40	80.37
Pr-Sh	82.20	83.30	81.50	82.33
ALSTM	83.10	83.85	82.70	83.22

由表2可知:传统的机器学习方法LR和SVM的平均准确度都低于80%;而DAE降噪自动编码机方法的准确度相对传统方法有所提升;BSSR方法相比DAE方法的准确度提升了1.56%;Attention-based-LSTM方法较BSSR方法的准确度提升了0.16%;文中提出的ASLTM模型比Attention-based-LSTM方法的准确度提升了0.82%.实验结果证明了文中方法的有效性.

通过对实验自身模型的层层剖析可以看出:加入对抗训练的Sh-Ad方法比Sh方法的准确度提升了0.74%;加入对抗训练的ASLTM比Pr-Sh提升了0.89%.这说明通过对抗训练,源语言和目标语言之间进行了交互学习,这种知识迁移让源语言和目标语言之间的联系更紧密,达到了知识迁移的效果.而Pr-Sh方法较Sh-Ad方法的准确度有1.96%的提升,这是因为该方法既可以获取到双语的不变特征,又能保留各自的独的特征,说明设置独立特征提取网络和共享特征提取网络的有效性.最终的ALSTM方法得益于这两个方面的共同作用

力，从而提高了分类准确率.

在 book，dvd，music 3 个领域上的损失函数函数值变化，如图 3 所示.



(a) book　　(b) dvd　　(c) music

图 3　损失函数数值

Fig. 3　Value of loss function

由图 3 可知：在 book，dvd，music 3 个领域上的损失函数函数值随着训练步数的增加，整体呈递减趋势，直至最终收敛，证明了网络的可训练性和有效性.

为了验证不同大小数据集对实验结果的影响，选取了大、中、小 3 种规模的数据集，分别进行对比实验. 实验中的准确度包括在整个测试集上的准确度(η_{tot})、在积极数据集上的准确度(η_P)和在消极数据集上的准确度(η_N). 各数据集上的准确度，如表 3 所示.

表 3　各数据集上的准确度

Tab. 3　Accuracy results on each data set

参数	大规模			中规模			小规模		
	book	dvd	music	book	dvd	music	book	dvd	music
训练集/条	4 000	4 000	4 000	2 000	2 000	2 000	1 000	1 000	1 000
测试集/条	4 000	4 000	4 000	4 000	4 000	4 000	4 000	4 000	4 000
η_{tot}/%	83.10	83.85	82.70	76.45	76.50	76.20	69.65	70.10	70.35
η_P/%	84.00	84.40	81.10	75.10	77.15	76.35	67.30	71.85	72.10
η_N/%	82.20	83.30	84.30	77.80	75.85	76.05	72.00	68.35	68.60

由表 3 可知：训练集数据的大小对实验结果影响很大，当训练集的数量减小时，测试准确度降低；同时，积极和消极 2 个子数据集的准确度在整个测试集准确度的合理范围内波动，其准确度的主要影响因素是数据集的质量，说明网络处于一个相对稳定的状态.

3　结束语

针对跨语言文本情感分类任务，提出一种基于对抗长短时记忆网络的跨语言情感分类方法. 通过设置独立的特征提取网络和共享的特征提取网络，获取到双语各自的独立特征和共享特征. 同时，设置语言分类器，通过对抗训练使源语言特征和目标语言特征在空间分布上尽可能接近，以获得双语的不变特征. 较之以前的研究方法，这种方法既保留了双语各自的特征，又获得了双语之间的不变特征，加强了双语之间的交互学习，减小了语义鸿沟，达到了较好的迁移效果. 实验结果也证明了此方法的有效性.

参考文献：

[1] NAKOV P，RITTER A，ROSENTHAL S，*et al*. SemEval-2016 task 4：Sentiment analysis in twitter[C]// International Workshop on Semantic Evaluation. San Diego：[s. n.]，2016：1-18. DOI：10.18653/v1/S16-1001.

[2] WAN Xiaojun. Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis [C]// Conference on Empirical Methods in Natural Language Processing. Hawaii：DBLP，2008：553-561. DOI：10.3115/1613715.1613783.

[3] WAN Xiaojun. Co-training for cross-lingual sentiment classification[C]// Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the

AFNLP. Singapore:DBLP,2009:235-243. DOI:10.3115/1687878.1687913.

[4] LU Bin, TAN Chenhao, CARDIE C, *et al*. Joint bilingual sentiment classification with unlabeled parallel corpora[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Oregon:DBLP,2011:320-330.

[5] TANG Xuewei, WAN Xiaojun. Learning bilingual embedding model for cross-language sentiment classification[C]// IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). Warsaw:IEEE Press,2014:134-141. DOI:10.1109/WI-IAT.2014.90.

[6] 方圆.跨语言文本情感分类技术研究[D].厦门:华侨大学,2015.

[7] ZHOU Huiwei, CHEN Long, SHI Fulin, *et al*. Learning bilingual sentiment word embeddings for cross-language sentiment classification[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing:[s.n.],2015:430-440. DOI:10.3115/v1/P15-1042.

[8] ZHOU Huiwei, YANG Yunlong, LIU Zhuang, *et al*. Jointly learning bilingual sentiment and semantic representations for cross-language sentiment classification[C]// China Conference on Information Retrieval. Shanghai: Springer,2017:149-160. DOI:10.1007/978-3-319-68699-8_12.

[9] ZHOU Xinjie, WAN Xiaojun, XIAO Jiangguo. Attention-based LSTM network for cross-lingual sentiment classification[C]// Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics,2016:247-256. DOI:10.18653/v1/D16-1024.

[10] BEN-DAVID S, BLITZER J, CRAMMER K, *et al*. Analysis of representations for domain adaptation[C]// International Conference on Neural Information Processing Systems. Vancouver:DBLP,2006:137-144.

[11] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, *et al*. Generative adversarial nets[C]// Advances in Neural Information Processing Systems. Montreal:MIT Press,2014:2672-2680.

[12] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation[C]// Proceedings of the 32nd International Conference on Machine Learning. Lille:arXiv,2014:1180-1189.

[13] PENNINGTON J, SOCHER R, MANNING C. GloVe: Global vectors for word representation[C]// Conference on Empiricial Methods in Natural Language Processing. Doha:[s.n.],2014:1532-1543. DOI:10.3115/v1/D14-1162.

[14] TAI K S, SOCHER R, MANNING C D. Improved semantic representations from tree-structured long short-term memory networks[J]. Computer Science,2015,5(1):36. DOI:10.3115/v1/P15-1150.

[15] IYER A, NATH S, SARAWAGI S. Maximum mean discrepancy for class ratio estimation: Convergence bounds and kernel selection[C]// Proceedings of the 31st International Conference on Machine Learning. Beijing: JMLR org,2014:1-530.

[16] GANIN Y, USTINOVA E, AJAKAN H, *et al*. Domain-adversarial training of neural networks[J]. Journal of Machine Learning Research,2016,17(1):2096-2030. DOI:10.1007/978-3-319-58347-1_10.

(责任编辑: 黄晓楠 **英文审校:** 吴逢铁)