

基于公共特征空间的自适应情感分类

洪文兴¹, 杞坚玮¹, 王玮玮¹, 郑晓晴¹, 翁洋²

(1. 厦门大学航空航天学院, 厦门 361005; 2. 四川大学数学学院, 成都 610064)

摘要: 针对情感分类这一项从文章或句子中得到观点态度的任务, 常规情感分类模型大多需要耗费大量人力获取标注数据。为解决某些领域缺乏标注数据, 且其他领域分类器无法在目标领域直接使用的现状, 设计了一种新颖的基于构建公共特征空间方法, 使分类模型可从有标注领域向无标注领域进行迁移适应, 减少人工标注的成本开销, 实现情感分类的领域自适应。该方法以大规模语料预训练的词向量信息作为以词为元素的特征, 在同种语言中表达情感所采用的句法结构相似这一假设前提下, 通过对领域内特有的领域特征词进行替换的方式构建有标注数据集与无标注数据集基本共有的公共特征空间, 使有标注数据集与无标注数据集实现信息共享。以此为基础借助深度学习中卷积神经网络采用不同尺寸卷积核对词语不同范围的上下文特征进行抽取学习, 进而采用半监督学习与微调学习相结合的方式从有标注数据集向未标注数据集开展领域自适应。在来自京东与携程共 5 个领域的真实电商数据集上进行实验, 分别研究了领域特征词选择方法及其词性约束对领域间适应能力的影响, 结果表明: 相较于不采用领域适应的模型, 可提升平均 2.7% 的准确率; 且在来自亚马逊电商的公开数据集实验中, 通过与现有方法进行对比, 验证了该方法的有效性。

关键词: 情感分类; 领域自适应; 半监督学习; 特征重构

中图分类号: TK448.21

文献标志码: A

文章编号: 0493-2137(2019)06-0631-07

Domain Adaptation with Common Feature Space for Sentiment Classification

Hong Wenxing¹, Qi Jianwei¹, Wang Weiwei¹, Zheng Xiaoqing¹, Weng Yang²

(1. School of Aerospace Engineering, Xiamen University, Xiamen 361005, China;

2. School of Mathematics, Sichuan University, Chengdu 610064, China)

Abstract: Sentiment classification, which extracts the opinions from sentences/documents, has been extensively studied. Most of the conventional sentiment classification models require a lot of cost to obtain the labeled data. In order to solve the problem that a trained classifier from other domain cannot be used directly on the target domain which lack labeled data, we proposed a novel domain adaptation model with reconstructing a common feature representation. This model makes the classifier from the labeled domain adapt to the unlabeled domain, reduces the cost of manual labeling and achieves the domain adaptation of sentiment classification. This model utilizes the pre-trained word vectors as the feature of the words. With the premise that the syntactic structure used to express sentiment in the same language is similar, a common feature space shared by the labeled and unlabeled data set is reconstructed by replacing the special domain words that unique to the domain. Therefore, the information sharing between the labeled and unlabeled data sets is realized. Based on this, the convolutional neural network in the model uses different size of convolution kernels to extract the context features of different range of words. With semi-supervised learning and fine-

收稿日期: 2018-10-29; 修回日期: 2018-11-13.

作者简介: 洪文兴(1980—), 男, 博士, 副教授, hwx@xmu.edu.cn.

通信作者: 翁洋, wengyang@scu.edu.cn.

基金项目: 国家重点研发计划资助项目(2018YFC0830300); 福建省科技计划资助项目(2018H0035); 厦门市科技计划资助项目(3502Z20183011); 掌数金融科技研发基金资助项目.

Supported by the National Key R&D Program of China(No. 2018YFC0830300), the Science and Technology Program of Fujian, China(No. 2018H0035), the Science and Technology Program of Xiamen, China(No. 3502Z20183011), the Fund of XMU-ZhangShu Fintech Joint Lab.

tuning learning, the model can be domain adapted from the labeled domain to the unlabeled domain. In experiments based on real data from Jingdong and Xiecheng, we separately compared the influence of different domain words selection and different POS constraints on the performance of our model, and found our model can improve the accuracy by about 2.7% compared to our baseline. In addition, we compared our model with related works on the public data from Amazon, and verified the effectiveness of our model.

Keywords: sentiment classification; domain adaptation; semi-supervised learning; feature reconstructing

随着互联网上情感文本的爆炸式增长,情感分析任务愈发受到重视.自动化的情感分类模型可以有效地帮助政府或企业监测公众反馈并收集市场舆情.

情感分类的传统方法中,研究者需要收集并整理大量的标注数据,在此基础上训练他们所关注领域的情感分类模型.然而现实生活中某些领域缺少标注数据是十分常见的现象,因此在成本有限的条件下,在这些领域训练一个可接受的分类模型是一件困难的工作.为解决这一问题,笔者采用了领域自适应的思想.研究者可在一个有高质量标注数据的领域训练分类模型,并将该模型应用在无标注数据的领域中.但要实现这一过程,有两个问题亟待解决:其一,需要确定数据中什么信息需要进行领域自适应;其二,需要知道如何将这些信息从源领域适应至目标领域.

本文提出了两个方案来分别解决上述问题.在源领域与目标领域之间重构一个公共的特征空间,该特征空间所包含的信息被两个领域所共享,即为希望进行领域适应的信息.利用卷积神经网络中不同尺寸的卷积核抽取并增强这其中有价值的信息,并通过半监督学习与神经网络微调的方式将这些信息从公共特征空间向目标领域适应.在此基础上,该模型可以逐步学习目标领域中的特征信息,并最终应用在目标领域中.

综上所述,本文实现了一个可在不同领域之间开展领域自适应过程的模型训练方法,解决了某些领域缺乏训练数据的问题.在来自京东、携程与亚马逊的电商数据集上进行了验证,实验结果证明了该方法的有效性.

1 研究现状

机器学习领域已有大量关于情感分类的研究,其中最著名是 Pang 等^[1]基于词语的 unigram、bigram 及词性特征,采用朴素贝叶斯、支持向量机与最大熵模型进行的对比分析.在 Pang 等^[2]另一项工作中将分类问题转换为排序问题,取得了良好的效果.在这些情感分类问题的研究中,可从网络博客^[3]、新闻^[4]与

社交软件^[5]上获取高质量的训练数据,然而在另一些实际领域训练数据十分匮乏.

解决方案之一是迁移学习中的领域自适应方法,其普遍方法是采用样例加权^[6],大多数以样例加权为基础的方法都有一个假设,虽然源领域与目标领域的边际概率 $P(X)$ 可能不同,但其边际条件概率 $P(Y|X)$ 是相同的^[7-9].另一些方法依赖领域中数据特征表达的转换^[10-11],这些以特征表达为基础的方法通过构建一个可观测的共同联合分布来对数据表达进行映射,从而实现领域自适应^[12-15].

情感分类中特征表达重构的直观方案是构建一个情感词典作为公共特征并在此基础上进行领域自适应^[16].结构对应学习则依靠选取支点特征,从句法结构角度寻找特征映射^[11].而基于谱特征对齐是将不同领域的词语经过聚合对齐获取共同特征,从而完成领域自适应^[17].上述方法以及本文所提出的方法均为词语级别的特征重构方法,所以在实验中将分别进行对比.

近年来,卷积神经网络模型在自然语言处理任务中也表现非常优秀,在语言建模^[18]、句子匹配^[19]、句子分类^[20]以及许多自然语言处理任务中都有不俗的表现^[21].深度学习方法在特征抽取中是无监督的过程,使得其更适合解决领域自适应问题^[22].

半监督学习是领域自适应的一种常用方法^[23-24],分类器可在一个小规模高质量标注数据集上初始化,并未从标注数据集中不断获得标注数据,通过重复训练改善模型在目标领域的表现^[25].在大规模数据集上预训练的卷积神经网络通过微调应用在新的领域已在图像识别领域得到了验证^[26].

2 模型与方法

为实现上述两个解决方案,笔者构建了以卷积神经网络为基础的分类器,设计了基于特征表达重构的领域自适应方法.模型结构将在第 2.1 节阐述,第 2.2 节阐述重构特征表达的过程,在第 2.3 节以伪代码的形式阐述半监督学习过程.

2.1 模型架构

本文将句子级别的情感分类任务看作词语级别序列分类问题,借鉴 Kim 等^[20]提出的多尺寸卷积核模型,如图 1 所示,在其基础上添加一层隐藏层作为半监督学习的特征映射。

如图所示,词典中第 i 个词语 w_i 被嵌入为一个 K 维的向量 v_i 。每个句子被表达为词向量构成的序列,通过截长补短使每个句子序列的长度保持一致。

在卷积神经网络中,设计了多种尺寸 s 的卷积核。卷积核的尺寸代表着取词窗口的宽度,例如当 $s=3$ 时,卷积核从每 3 个词中抽取 1 个特征。模型中有 h 种卷积核尺寸,每种尺寸有 f 个卷积核,通过卷积可得到 $h \times f$ 个特征。在池化过程中,采用了最大池化的操作^[21],即每个特征序列中保留最显著的特征。

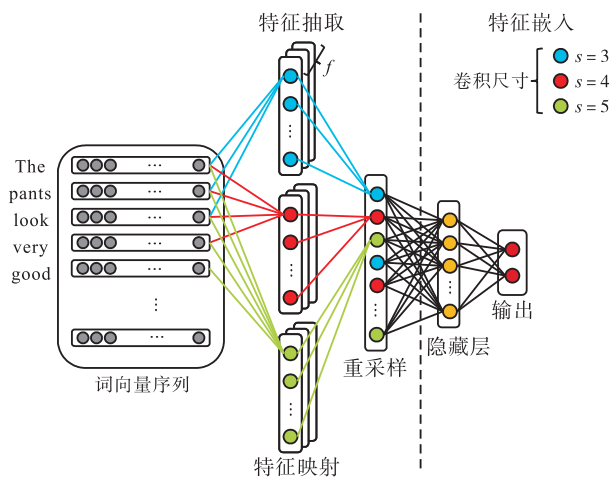


图 1 分类模型架构

Fig.1 Classification model architecture

上述卷积与池化的过程可称为特征抽取过程,其后的隐藏层作为特征映射层。模型依照句子的特征向量对其进行分类。

2.2 公共特征空间

不同领域的语料可看作是从一个公共词典里基于不同概率分布进行采样。笔者认为在同一种语言中,构建公共特征空间的基础在于表达情感时所采用的句法是类似的,它们的不同之处主要在于领域特征词汇的组成。领域特征词汇是个相对的概念,当选定源领域和目标领域后,一些词语在目标领域中出现非常频繁而在源领域中几乎不出现,即为目标领域中的特征词汇,反之亦然。表 1 中列出了服装评论与酒店评论中的两个例子,其中“裤子”与“面料”即为服装评论领域相对于酒店评论领域的特征词汇。如上文所述,这两条例句在表达对所消费商品的认可时,所使用的句法结构是相似的,即“名词-动词-形容词”的形式。

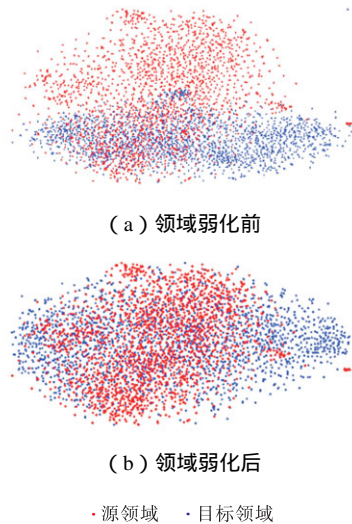
表 1 领域特征词汇示例

Tab.1 Example of domain words

领域	句子	特征词汇
服装评论	这条裤子看着不错,面料很舒服	裤子,面料
酒店评论	这家酒店的床感觉很舒服,位置也不错	床,酒店,位置

笔者将领域特征词汇定义为:在源领域中出现频次与目标领域中出现频次之比大于特征词汇阈值 h_w 的词汇。这样的定义可简单地理解为在一个领域中大量出现的词汇,通常含有有价值的信息,而在另一个领域中几乎不出现,则证明该词汇所含有的信息是前者所特有的,排除了每个领域中都大量出现的助词性词汇。每个句子中的词汇可看作是该样本的一个特征,若将在源领域训练的模型直接应用到目标领域上,大量的独有特征会误导模型分类效果。

因此笔者提出了一种方法,首先选取无意义的符号替换原有句子中的领域特征词,对领域信息先进行弱化,待模型训练完成再强化。在领域弱化的过程中,通过对特征词的替换,可构建一个公共的特征空间。为更直观地观察领域弱化的作用,图 2(a)、2(b)中利用 T-SNE 算法将源领域与目标领域的高维词向量空间转变为二维分布^[27]。在可视化图形中可发现,领域弱化前源领域与目标领域的向量空间重合度较小,经过领域弱化两个领域的向量空间基本重合,即领域弱化过程使源领域与目标领域的向量空间更加相似。



(a) 领域弱化前

(b) 领域弱化后

· 源领域 · 目标领域

图 2 领域弱化前后的特征分布变化

Fig.2 Changes in the distribution of features before and after the weakening of domain

借助预训练的词向量模型,通过计算词向量序列间的余弦距离可判断两个句子的相似性。如表 2 所示,经过领域弱化后两个句子趋于相同,因此可利用该公共特征空间构建训练数据集,以此初始化分类模

型,并逐步学习目标领域的独有特征。

表 2 领域弱化示例

Tab.2 Example of weakening of domain

领域	句子	相似度
服装评论	这条裤子看着不错,面料很舒服	0.879
酒店评论	这家酒店的床感觉很舒服,位置也不错	
弱化后	这条 X 看起来不错,X 很舒服 这家 X 的 X 感觉很舒服,X 也不错	0.932

该过程的关键点在于如何选取领域特征词。本文中采用基于统计词频的方式衡量词语在领域语料中的独特程度,若词语在源领域中出现频次与其在目标领域中出现频次的比值大于实验设定的领域词发现阈值,即认为该词语是源领域的特征词,反之亦然。为更精确选取特征词,本文采用词性约束的方式对特征词加以限制,分别以最为常见的名词性约束与形容词性约束选择特征词。在添加词性约束时,词语在满足频次比值高于阈值,且属于目标词性的条件下,被选取为特征词。

2.3 半监督学习过程

综上所述,基于公共特征空间可构建弱化后的源领域数据集,记为 S_w ,弱化后的目标领域数据集,记为 T_w ,原本的目标领域数据集记为 T ,新样本置信度阈值为 h_s 。如下用伪代码阐述半监督学习的过程。

Semi-supervised Learning(S_w, T_w, T, h_s)

输入 :weakened source dataset S_w ,weakened target dataset T_w ,original target dataset T ,threshold h_s

输出 :classifier for target dataset C

1. 初始化 C
2. Train C on S_w
3. for sentence_w ,sentence in T_w, T :
4. label ,conf = Predict(C ,sentence_w)
5. if conf > h_s
6. add(sentence ,label)to D
7. remove sentence from T
8. Fine-tune C
9. while D increase
10. Train C on D
11. for sentence in T
12. label ,conf = Predict(C ,sentence)
13. if conf > h_s
14. add(sentence ,label)to D
15. remove sentence from T
16. return C

在 1、2 行,初始化模型并利用经过弱化的源领域进行训练。3~7 行模型对目标领域经过弱化的句子进行分类标注,置信度高于阈值的样本则将其未经弱化句子与标签一同加入新的训练集 D 中,同时从目标领域数据集中去除该样本。8 行对模型进行微调初始化,即固定神经网络中特征抽取部分神经节点参

数,将特征映射部分节点设为可训练状态。9~15 行在新训练集 D 的基础上重复训练模型,并不断从 T 中获取新的高置信度数据,不断扩充模型训练集。当 D 不再扩增时即停止训练过程,此时的模型即为由源领域自适应至目标领域的分类器。

3 实验与讨论

在构建分类器与半监督自适应过程的基础上,本节通过一系列对比实验验证该方法的有效性。

第 3.1 节简要阐述实验所使用的数据集与部分超参数设置。第 3.2 节展示实验结果,而第 3.3 节笔者对实验结果进行了讨论,第 3.4 节中笔者在公开数据集上实验该方法并与现有方法进行比较。

3.1 数据集与参数设定

表 3 中列举了实验中所使用的数据集概况。

表 3 数据集概况

Tab.3 Information of data sets

数据集	领域	样本量	词汇量	平均长度
京东	服装	10 000	8 643	17
	水果	10 000	10 891	20
	掌上电脑	10 000	13 486	23
	洗发露	10 000	10 850	21
携程	酒店	10 000	30 311	58

上述数据均来自京东与携程电商平台,参照通用电商评论标注方式,其标签以评论附加的星数确定,4、5 星为正样本,1、2 星为负样本,3 星默认为不具有情感极性的评价故而舍弃,数据集中每个领域正负样本各 5 000 条。实验中采用十折交叉验证的方式对模型进行验证,最终结果为各次实验的平均表现。

卷积核尺寸设置为 $s = \{3, 4, 5\}$,每个尺寸设置卷积核数 $f = 100$,即共计 300 个卷积核。隐藏层节点数量为 100,网络节点中 dropout 概率设为 0.4^[28],且添加 0.001 的 L2 正则约束以避免过拟合。在训练集中随机抽取 10% 数据作为验证集,在训练过程中以验证集上模型的表现调整学习速率。词嵌入所使用的预训练词向量是以中文维基百科所训练的 50 维向量^[29]。

3.2 实验结果

为确定模型训练过程中的若干超参数,如领域词发现阈值、新样本置信度阈值、领域词的词性约束,笔者以服装领域评论与水果领域评论进行了一系列先导实验。

如图 3 所示,在多组领域词发现阈值的实验中,模型在阈值为 10 时达到最好效果。

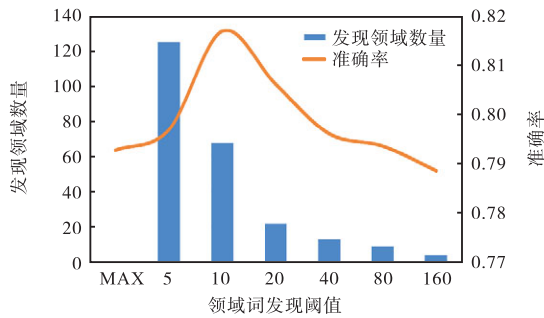


图3 模型在不同领域词发现阈值下的表现

Fig.3 Model performance under different domain word discovery thresholds

图4中,在不同置信度阈值的作用下,模型的表现 在 0.999 达到最优。

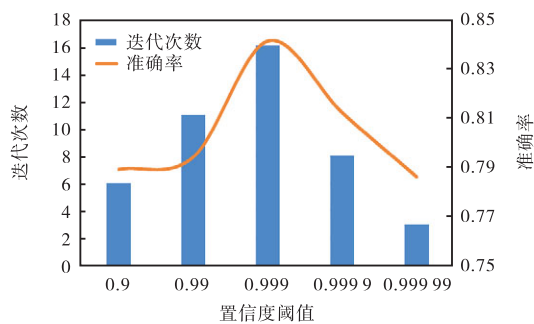


图4 模型在不同新样本置信度阈值下的表现

Fig.4 Model performance under different sample confidence thresholds

在正式对比实验中,以无领域自适应训练的结果为基线,以在目标领域内训练的结果为最优表现,在表4中展示了不同词性约束下模型的准确率表现情况,均为十折交叉验证平均结果。

表4 模型在不同词性约束下的准确率表现

Tab.4 Model accuracy under different constraints

实验	准确率			
	C-F	C-H	C-P	C-S
无自适应	0.861	0.845	0.743	0.836
无约束	0.866	0.852	0.765	0.848
名词约束	0.887	0.861	0.758	0.867
形容词约束	0.793	0.856	0.746	0.838
目标领域内训练	0.902	0.884	0.917	0.916
实验	F-H	F-P	F-S	H-P
无自适应	0.756	0.702	0.839	0.764
无约束	0.781	0.753	0.872	0.761
名词约束	0.802	0.754	0.907	0.768
形容词约束	0.758	0.711	0.842	0.763
目标领域内训练	0.884	0.917	0.916	0.917
实验	H-S	P-S		
无自适应	0.723	0.763		
无约束	0.741	0.761		
名词约束	0.743	0.762		
形容词约束	0.722	0.758		
目标领域内训练	0.916	0.916		

表中 C、F、H、P、S 分别代表服装、水果、酒店、掌上电脑、洗发露各个领域,C-F 即为以服装为源领域、F 为目标领域进行实验.以红色字体标出了每组实验中效果最好的约束条件,从结果中可以看出,以名词为约束条件所训练的模型取得了最佳的效果.与无自适应的实验过程相对比,在名词约束下的特征变换模型准确率平均提升 2.7%,证明该模型能从一定程度上改善跨领域的自适应过程。

3.3 讨论

在第 1 个先导实验中,通过比较模型在不同领域词发现阈值下的表现可看出,随着阈值增大,领域词数量逐步减少,模型准确率先增大后减小.在阈值较小时,过多领域词被替换导致模型没有学习足够的特征;而当阈值较大时,由于几乎没有领域词被替换,此时类似无自适应过程。

第 2 个先导实验中,随着新样本置信度阈值的增大,迭代次数与模型准确率均先增大后减小.当阈值较小时,对新样本的选取较为宽松,目标数据集中的样本快速进入新训练数据集,此时训练集中数据质量较低导致模型准确度较差;而当阈值较大时,训练数据集很难获取新的训练样本,导致训练集较早停止扩增,模型迭代训练次数较少,故而准确率较低。

在对比实验中,对比不同组实验的结果可发现下列现象。

掌上电脑领域的模型表现均不甚理想.笔者认为原因在于该领域与其他日常生活领域相差较远,用户重叠较少,电子产品的评价主要针对其性能而非给人的感受,不适用第 2.2 节所述在表达情感时所用句法大致相近的前提,故而模型训练结果较差。

服装、酒店、水果领域的自适应效果较好.笔者认为这源于这些领域均侧重于商品给人的直观感受,感受的表述上大致相近。

名词性约束在大多数领域取得了较好的结果.笔者认为这得益于在预训练词向量的基础上,相近的形容词类属性特征的聚合程度更好,所以模型训练效果更好。

综上,该方法所需的成本较小,在训练情感分类器的同时能得到领域的特征词典.神经网络的表现与训练数据量有直接关系,笔者相信该模型在大规模数据集下有取得更好表现的潜力。

该实验过程运行于 GTX1080 8G GPU,每次实验的时间开销约为 2h.实验中模型通常需要在半监督学习过程迭代 10 次以上才能取得较好的效果。

3.4 公开数据集实验

该实验数据集来自亚马逊电商评论数据^[11],涵

盖书籍、DVD 影像、电器与厨房用具 4 个领域。其中数据处理方式同上,评分 4、5 星为正样本,1、2 星为负样本,每个领域分别包含正负样本各 1 000 条。

该实验中采用基于谷歌新闻数据集预训练的 300 维英文词向量模型进行向量嵌入,网络结构与超参数设置均采用上述中文数据集实验中最优条件,以名词性约束抽取领域特征词。

如表 5 所示,模型在书籍与 DVD 影像、电器与厨房用具领域间的适应效果较好,这一现象与中文数据集中属性词相近的领域在经过领域弱化后更加相似是一致的。

表 5 模型在英文数据集的准确率表现

Tab.5 Model accuracy on English data set

源/目标领域	准确率			
	Book	DVD	Electronics	Kitchen
Book	0.827	0.811	0.772	0.781
DVD	0.782	0.843	0.743	0.823
Electronics	0.764	0.786	0.876	0.843
Kitchen	0.649	0.763	0.812	0.882

同时,参照前期研究者的实验方法,选择其中一个领域作为目标领域,其余 3 个领域作为源领域进行实验,可得到模型准确率如表 6 所示。

表 6 模型准确率对比

Tab.6 Comparison of model accuracy

方法	准确率			
	Book	DVD	Electronics	Kitchen
无自适应	0.690 2	0.724 8	0.734 7	0.792 6
SCL-MI ^[11]	0.745 6	0.763 0	0.789 3	0.820 6
SFA ^[17]	0.777 3	0.763 1	0.753 0	0.814 8
SST ^[16]	0.763 2	0.782 6	0.838 6	0.851 8
本文模型	0.784 5	0.801 4	0.821 5	0.866 7
目标领域内训练	0.832 9	0.846 4	0.883 1	0.891 2

可发现在拥有更大训练集的情况下,模型的准确率得到进一步提高。笔者分析认为原因在于一方面,源数据集更加具有普适性,在领域弱化过程中与目标领域会更加相似;另一方面,神经网络在大规模数据集上能更有效率地抽取特征。与现有方法相比,该方法所需的开销更小,且能取得较好的效果。

4 结 语

本文中,笔者致力于研究多领域中情感分类问题的领域自适应。设计了一套基于公共特征空间重构的方法进行领域自适应。以卷积神经网络抽取特征,半监督学习过程进行特征映射,从而节约大量人力解决了某些领域无标注数据、无法训练有效分类器的问题。实验结果表明,该方法在多领域情感分类问题中

能取得良好的表现。

参考文献:

- [1] Pang B, Lee L, Vaithyanathan S. Thumbs up? : Sentiment classification using machine learning techniques[C]// Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. Philadelphia, USA, 2002: 79-86.
- [2] Pang B, Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales[C]// Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Ann Arbor, USA, 2005: 115-124.
- [3] Yang Changhua, Lin K, Chen H. Emotion classification using web blog corpora[C]// IEEE/WIC/ACM International Conference on Web Intelligence. California, USA, 2007: 275-278.
- [4] Read J. Using emoticons to reduce dependency in machine learning techniques for sentiment classification [C]// Proceedings of the ACL Student Research Workshop. Ann Arbor, USA, 2005: 43-48.
- [5] Hassan A, Abbasi A, Zeng D. Twitter sentiment analysis: A bootstrap ensemble framework[C]// 2013 International Conference on Social Computing. Washington, DC, USA, 2013: 357-364.
- [6] Jiang Jing, Zhai Chengxiang. Instance weighting for domain adaptation in nlp[C]// Proceedings of the 45th Annual Meeting on Association for Computational Linguistics. Prague, Czech Republic, 2007: 264-271.
- [7] Dai Wenyuan, Yang Qiang, Xue Guirong, et al. Boosting for transfer learning[C]// Proceedings of the 24th International Conference on Machine Learning. Corvallis, USA, 2007: 193-200.
- [8] Huang Jiayuan, Gretton A, Borgwardt K M, et al. Correcting sample selection bias by unlabeled data[C]// Advances in Neural Information Processing Systems. Vancouver, Canada, 2007: 601-608.
- [9] Sugiyama M, Nakajima S, Kashima H, et al. Direct importance estimation with model selection and its application to covariate shift adaptation[C]// Advances in Neural Information Processing Systems. Vancouver, Canada, 2008: 1433-1440.
- [10] Bendavid S, Blitzer J, Crammer K, et al. Analysis of representations for domain adaptation[C]// Advances in Neural Information Processing Systems. Vancouver, Canada, 2007: 137-144.

- [11] Blitzer J , Dredze M , Pereira F. Biographies , bollywood , boom-boxes and blenders : Domain adaptation for sentiment classification[C]// Proceedings of the 45th Annual Meeting on Association for Computational Linguistics. Prague , Czech Republic , 2007 : 440-447.
- [12] Ando R K , Zhang T. A framework for learning predictive structures from multiple tasks and unlabeled data[J]. Journal of Machine Learning Research , 2005(6) : 1817-1853.
- [13] Argyriou A , Evgeniou T , Pontil M. Multi-task feature learning[C]//Advances in Neural Information Processing Systems. Vancouver , Canada , 2007 : 41-48.
- [14] Raina R , Battle A , Lee H , et al. Self-taught learning : Transfer learning from unlabeled data[C]//Proceedings of the 24th International Conference on Machine Learning. Corvallis , USA , 2007 : 759-766.
- [15] Pan S J , Kwok J T , Yang Q. Transfer learning via dimensionality reduction[C]//AAAI Conference on Artificial Intelligence. Chicago , USA , 2008 : 677-682.
- [16] Bollegala D , Weir D , Carroll J. Cross-domain sentiment classification using a sentiment sensitive thesaurus[J]. IEEE Transactions on Knowledge and Data Engineering , 2013 , 25(8) : 1719-1731.
- [17] Pan S J , Ni Xiaochuan , Sun J T , et al. Cross-domain sentiment classification via spectral feature alignment [C]//Proceedings of the 19th International Conference on World Wide Web. Montreal , Canada , 2010 : 751-760.
- [18] Kalchbrenner N , Grefenstette E , Blunsom P. A convolutional neural network for modelling sentences [C]//the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore , USA , 2014 : 655-665.
- [19] Hu Baotian , Lu Zhengdong , Li Hang , et al. Convolutional neural network architectures for matching natural language sentences[C]//Advances in Neural Information Processing Systems. Montreal , Canada , 2014 : 2042-2050.
- [20] Kim Y. Convolutional neural networks for sentence classification[J]. Empirical Methods in Natural Language Processing , 2014 : 1746-1751.
- [21] Collobert R , Weston J , Bottou L , et al. , Natural language processing(almost)from scratch[J]. Journal of Machine Learning Research , 2011 , 12(8) : 2493-2537.
- [22] Glorot X , Bordes A , Bengio Y. Domain adaptation for large-scale sentiment classification : A deep learning approach[C]//Proceedings of the 28th International Conference on Machine Learning. Bellevue , USA , 2011 : 513-520.
- [23] Blum A , Mitchell T M. Combining labeled and unlabeled data with co-training[C]// Proceedings of the 11th Annual Conference on Computational Learning Theory. Madison , USA , 1998 : 92-100.
- [24] Joachims T. Transductive inference for text classification using support vector machines[C]//16th International Conference on Machine Learning. San Francisco , USA , 1999 : 200-209.
- [25] Nigam K , McCallum A , Thrun S , et al. Text classification from labeled and unlabeled documents using em[J]. Machine Learning , 2000 , 39 : 103-134.
- [26] Tajbakhsh N , Shin J Y , Gurudu S R , et al. Convolutional neural networks for medical image analysis : Full training or fine tuning?[J]. IEEE Transactions on Medical Imaging , 2016 , 35(5) : 1299-1312.
- [27] Van der Maaten L , Hinton G. Visualizing data using t-sne[J]. Journal of Machine Learning Research , 2008 , 9(11) : 2579-2605.
- [28] Srivastava N , Hinton G E , Krizhevsky A , et al. Dropout : A simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research , 2014 , 15(1) : 1929-1958.
- [29] Mikolov T , Sutskever I , Chen K , et al. Distributed representations of words and phrases and their compositionality[C]//Advances in Neural Information Processing Systems. Lake Tahoe , USA , 2013 : 3111-3119.

(责任编辑 :王晓燕)