



基于胶囊网络的跨领域情感分类方法

摘要

情感分析主要基于文本数据研究人们对于商品、服务、事件等对象的情感、意见或者态度。标记数据稀缺是情感分析领域面临的巨大挑战。在有监督的情感分类任务中,标记数据稀少会导致分类器的效果下降。跨领域的方法能够在一定程度上帮助解决该问题,但领域间往往存在差异。因此在利用领域适应方法进行情感分类时,分类器对目标领域的效果会变差。本文提出利用少量的目标领域标记信息来提高目标领域适应效果的思想。特别地,本文提出了一个基于胶囊网络的跨领域情感分类模型,在此基础框架上,设计了额外的胶囊网络层辅助目标领域的适应。在真实数据集上的实验结果表明,本文提出的模型效果优于以往的研究方法。

关键词

情感分类;跨领域;胶囊网络

中图分类号 TP391.1;TP183

文献标志码 A

收稿日期 2019-05-16

资助项目 国家自然科学基金(61572376, 91646206)

作者简介

林悦,男,硕士生,研究方向为自然语言处理。linyue@whu.edu.cn

钱铁云(通信作者),女,博士,教授,研究方向为自然语言处理、Web挖掘。qiy@whu.edu.cn

0 引言

随着互联网的快速发展,网络用户有更多的机会和更方便的途径去表达自己的观点。例如用户经常在社交媒体或者评论网站发表自己对一些事件、产品、服务的想法。亚马逊、京东、淘宝等购物网站,Facebook、新浪微博等社交媒体网站,烂番茄、imdb、豆瓣等影片评分网络,yelp、大众点评等餐厅评分网站,都已积累了海量的用户数据。评论信息传递了人们的各种情感倾向和观点意见,能够为他人提供重要的参考价值。比如购物时,买家可以根据其他用户的评论来判断商品是否满足自己的要求;同时,该信息也引起了商家的高度重视,因为它们能够给商家带来一定的商业价值^[1],例如,公司能够通过挖掘用户对商品的态度来制定相应的策略,进而改善自己的商品和服务来提高公司利润。如何正确理解评论中的观点态度、情感倾向,是一个极为重要也极具应用价值的研究课题。因此情感分析等领域正越来越吸引研究者的关注。

情感分析^[2],也被称为观点挖掘^[3],它基于文本数据研究人们对于商品、服务、事件等对象的情感、意见或者态度,是自然语言处理领域的一个重要分支。近年来,情感分析在很多领域中都有不同程度的应用,包括消费、医疗、旅游、金融乃至政治等领域。例如,在金融领域,可以通过股民的心情来进行股票走势的预测^[4];在政治领域,可以通过推特的信息来对选举结果进行预测^[5]等。正是其广泛的应用,情感分析已经成为最流行的研究领域之一。情感分类是情感分析中最主要、应用最广泛的任务,旨在通过训练分类器自动识别文本的情感极性,例如判断一段文本评论的态度是积极的或者消极的。

理论上,当一个领域有充足的标记数据时,有监督的学习算法往往能够很好地训练出该领域的分类器^[6]。同样地,在情感分类领域中也是如此。当领域有充足的标记数据时,情感分类器可以被训练出很好的效果。然而,在现实中感兴趣的领域往往只有少量的、不充足的标记数据,而大量的数据都是无标记的。如何训练出好的情感分类器是一个巨大的挑战。人为地对无标记数据进行标注是一件非常耗费时间、精力和金钱的工作,并且标记出的数据很难拓展到其他领域。因此,依赖于人工标注的方法并不是一个好的解决方案。目前,通过跨领域学习进行情感分类是解决标记数据不足的有效途径。跨领域学习是指利用与目标领域相关或者相近的源领域中所蕴含的丰富标记数

¹ 武汉大学 计算机学院,武汉,430072

据,提取和利用对目标领域有用的知识和信息,从而提高目标领域的情感分类任务的效果.但是,跨领域的情感分类目前仍存在很多不足.其根本原因在于领域差异性的存在:一个情感分类器经过源领域的训练后,并不能在目标领域表现出好的效果.

目前跨领域的情感分类已经得到了一定的发展.早期大部分跨领域工作是利用特征迁移的方法,需要对领域的中心词(pivots)和非中心词(non-pivots)进行人工选择.Tan等^[7]尝试选择一些广泛性的特征来进行情感分类,该特征包括在不同领域中出现较频繁并且有着相近概率分布的词.Yoshida等^[8]提出利用来自不同资源和领域的训练数据贝叶斯概率模型,该工作在建立迁移的过程中需要目标领域中大量的标记数据.该模型利用了词频的信息,并没有利用其他的文本特征.人工选择中心词的工作除了利用两个领域的词频,还包括利用源领域特征和标记间的互信息、特征和领域之间的互信息和加权似然估计比率^[9]等.

结构化一致学习(简称为SCL)是Blitzer等^[10]提出的杰出的领域适应方法.SCL方法利用了多个中心词的预测任务来推断中心词和非中心词的关系,尝试获得一个从非中心词特征空间到中心词特征空间的映射矩阵.Pan等^[11]提出了光谱特征校准(简称为SFA)方法,该方法利用中心词和非中心词的共现来找到它们的准线.具体是通过中心词校准非中心词的关系来建立不同领域间中心词特征和非中心词特征的联系,实现跨领域的情感分类.然而,上述方法都需要人工选择中心词.人工选择工作代价昂贵,并且可能会出错.

近年来,深度学习在计算机视觉和自然语言处理等^[12]方面都有了很大进展,例如,卷积神经网络和循环神经网络等已经成为了一种常见的网络结构.在相应领域有大量的标记数据时,深度学习的模型往往能够取得惊人的效果,并且深度学习的方法大大减少了人工的特征工程.因此,深度学习的方法很自然地引入到跨领域情感分类的研究中.然而,深度学习的方法要取得好的效果往往需要依赖于一种前提假设——训练数据和测试数据都需要有相同的数据分布.由于跨领域数据存在的差异性,成为了将深度学习方法应用在跨领域的一个障碍.

Ganin等^[13-14]提出了领域对抗训练神经网络(简称DANN),该方法利用了领域对抗的训练方法,使神经网络能够产生融合领域的分类器.该方法首

次引入对抗训练领域分类器的思想,利用梯度反转层来反转神经网络梯度方向,使其难以识别出源领域和目标领域的差别,从而达到领域融合的效果.Li等^[15-16]将记忆网络和分层的注意力机制引入到DANN中,使其能够自动辨别中心词,因此不需要人为的选择.然而,该方法不一定能够精确地捕获到中心词以及判断中心词正确的情感极性.同时,源领域和目标领域中心词特征交集较少时,该类跨领域情感分类效果将会变差.

1 领域适应的胶囊网络模型

1.1 问题分析

在本文的情感分类任务中,样本一般为文本数据,情感类别作为数据的标记,一般分为积极、中立和消极.针对本文研究的数据集特性,本文将把情感状态划分为积极和消极两种.跨领域情感分类旨在利用源领域的信息辅助提高目标领域的情感分类性能.

在跨领域的研究中,一般将领域分为源领域和目标领域,其中将源领域定义为 D_s ,目标领域定义为 D_t .在源领域里 D_s ,存在样本 $X_s^l = \{x_s^i\}_{i=1}^{N_s^l}$ 和对应标记 $\{y_s^i\}_{i=1}^{N_s^l}$.目标领域里存在着一个无标记的数据集 $X_t^u = \{x_t^i\}_{i=1}^{N_t^u}$,目标领域存在少量的标记数据定义为 $X_t^l = \{x_t^i\}_{i=1}^{N_t^l}$,其对应标记为 $\{y_t^i\}_{i=1}^{N_t^l}$.在现实中,目标领域的标记数据是相对稀缺的,而人工进行标注需要耗费大量的成本.以往的研究中,跨领域的情感分类是根据源领域的标记数据和目标领域的无标记数据训练出一个情感分类器.此类方法并未考虑到目标领域中少量标记数据对增强模型适应性的作用,从而会引起以下两个问题:第一,因为模型缺乏目标领域的标记数据,即使利用了相关的源领域的标记数据和目标领域的无标记数据,模型无法在目标领域得到充分的拟合,导致模型仍是一种欠拟合的状态,比如,同一个词在源领域中表示积极态度但在目标领域中表示消极态度,不能适应目标领域的模型将可能依赖源领域信息状态判断为积极的从而导致分类的错误;第二,当源领域和目标领域共享信息较少或者两个领域相关性较低时,模型从两个领域共享知识中能学习到帮助目标领域分类的信息将会很有限,导致跨领域分类器对目标领域的分类效果降低.

为了解决以上存在的问题,本文提出了利用目标领域的标记信息来改善模型适应能力的思想.当

模型得到目标领域标记信息后,一方面,模型将能提高对目标领域的拟合能力.比如,从目标领域中学习到一些单词、短语使得情感分类器更加适应于目标领域的语义环境.另一方面,模型在面对两个领域交集较少或者相关度较低的情况时,将能够从目标领域标记数据中学习到与目标领域直接相关的知识,从而提高模型对目标领域的适应性.而之前的模型并没有考虑加入目标领域信息的结构,直接将该信息加入到之前的模型中,因此不能充分地利用和学习该信息.

1.2 模型框架

本文的模型基于胶囊网络,并利用目标领域的标记信息辅助提高模型对目标领域的适应性.该模型简称为 target_capsule 模型.在胶囊网络模型中一般会结合卷积神经网络.卷积神经网络的卷积层通过卷积操作可以提取到 N-gram 的特征,而在进行池化操作时虽然能够保留重要的特征,也往往会丢失一部分信息.胶囊网络中的每个胶囊类型都对应于一个实体或者抽象概念,因此胶囊能够表示更丰富的抽象特征.在文本中,胶囊往往还能提取到文本的结构信息.于是在卷积神经网络的基础上,本文的模型结合了胶囊网络.为了能够学习到源领域和目标领域共享特征表示,该模型利用领域对抗训练的方法进行训练.

得到共享表示之后,该模型新增了一层胶囊网络层作为目标领域的适应结构.利用新的胶囊网络层能够更充分和高效地学习到目标领域信息,并与共享部分的模型学习的共享信息进行交互,最终提高整个分类器对目标领域的分类效果.此外,进行文本类型的数据领域适应时,除了单词层面的语义共

享,还有别的信息能够得到共享,比如单词的位置、语法、句法等信息也能够共享.该类特征进行迁移时一定程度上能补充目标领域的信息,可以提高领域适应的效果,因此我们也将利用上述信息.

图 1 为 target_capsule 模型的结构示意.该模型分两个阶段进行训练.在上面蓝色框的部分为模型第一阶段的训练,下面绿色框为模型第二阶段的训练.参与第一阶段训练的模型参数能通过蓝色框中的卷积层和胶囊网络层学习到共享信息.参与第二阶段训练的模型参数只有绿色框中的卷积层和胶囊层,该层能学习到目标领域的独有信息.最后,通过将两阶段最后的两层胶囊层进行交互,可以得到完整的 target_capsule 模型.

本文采用两个阶段的训练方式,其目的是为了保证模型在学习共享表示时达到最优,以及在学习目标领域表示时也能达到最优,模型的两个模块能够充分地学习到对应的表示,最终使两阶段的总和最优.

1.3 模型细节

下文开始详细介绍 target_capsule 模型中各个模块的结构.

1.3.1 词向量表示

输入层中接收的数据是文本评论,每条文本评论是由单词组成的.输入层中的数据为 n 个单词, n 为文本长度.假若每个单词以 one-hot 表示将导致过于离散,反而会丢失一部分信息.因此本文使用词嵌入的方法.词嵌入的方法是将每个单词从 one-hot 的表示形式映射到一个低维的连续空间.单词在映射后的空间中的向量表示称为该单词的词向量.词向量矩阵表示全部文本中出现的单词的词向量集合,

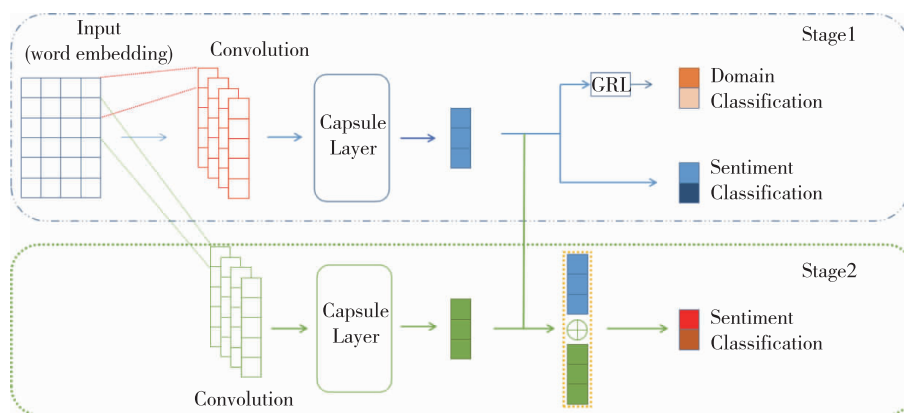


图 1 target_capsule 模型结构

Fig. 1 The framework of target_capsule

在本文中表示为 $\mathbf{L} \in \mathbf{R}^{d \times |V|}$, 其中 d 为词向量的维度, $|V|$ 为词汇表大小. 假设句子的表示为 $s = \{w_1, w_2, \dots, w_n\}$, 其中 w_i 表示为第 i 个单词的词向量, 词向量大小为 $d \times 1$, 则句子 s 的大小为 $n \times d$.

1.3.2 卷积层

在卷积层一般采用多种不同的卷积核进行卷积操作. 每个卷积核都对应着特定的 N-gram 特征. 假设卷积核参数 $\mathbf{W} \in \mathbf{R}^{h \times d}$, 其中 h 为卷积的窗口长度, 即需要卷积的单词数量, d 为词向量长度. 根据式(1)的计算可以得到该卷积核提取的特征:

$$c_i = f(\mathbf{W} \cdot \mathbf{w}_{i:i+h-1} + \mathbf{b}), \quad (1)$$

其中 \mathbf{b} 为偏置项, f 为非线性函数. 卷积操作会对输入的词向量矩阵以大小为 h 的窗口进行滑动遍历, 即在词组子集的集合中依次进行卷积计算, 得到 $n - h + 1$ 个特征. 将所有特征组合在一起形成一个新的集合, 称为特征图表示, 如式(2)所示:

$$\mathbf{c} = [c_1, c_2, \dots, c_{n-h+1}], \quad (2)$$

其中 $\mathbf{c} \in \mathbf{R}^{(n-h+1) \times 1}$. 该特征图表示的是单个卷积核在文档上进行卷积操作得出的结果. 当存在 m 个卷积核时, 将能够得到 m 个特征图表示.

为了能够得到维度大小相同的特征图, 本文使用的卷积核窗口大小 h 是一个固定的值. 因此, m 个特征图的大小是一致的. 将 m 个特征图拼接在一起得到特征图矩阵 $\mathbf{z} = [\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_m]$, 其中 $\mathbf{z} \in \mathbf{R}^{(n-h+1) \times m}$.

1.3.3 胶囊层

在特征图矩阵 \mathbf{z} 中共有 m 个列向量, 每个列向量表示的是一类不同的特征. 传统的卷积神经网络用池化的方法把向量压缩为标量, 从而导致信息丢失, 而在胶囊层中往往需要保留着向量信息. 本文将向量转化为更加方便使用的形态. 因为使用的卷积核大小是一致的, 因此在每个向量的相同维度上是提取同一种特征属性. 对相同维度的特征值再进行一次卷积, 其中使用的卷积核的大小为 $\mathbf{R}^{1 \times (n-h+1)}$. 通过使用该尺寸的卷积核对特征图矩阵进行卷积, 将每个列向量转化新的特征. 假设使用的胶囊层有 J 个胶囊单元, 其中每个胶囊的维度为 d_j , 则需要使用 $J \times d_j$ 个卷积核来产生新特征. 根据卷积公式(1)可以计算出新特征图矩阵, 其大小为 $\mathbf{R}^{m \times (J \times d_j)}$. 将其形状从二维矩阵转化为三维矩阵, 则其大小转化为 $\mathbf{R}^{m \times J \times d_j}$. 通过上述的卷积操作和变换, 原本的二维特征图矩阵 $\mathbf{z} \in \mathbf{R}^{(n-h+1) \times m}$ 变形为新的三维特征图 $\mathbf{R}^{m \times J \times d_j}$ 的形式. 接下来利用文献[17]中的动态路由

算法对该特征图进行计算, 得到 J 个胶囊单元向量, 其中每个胶囊的维度为 d_j .

1.3.4 全连接层与情感分类

为了将所有单元向量进行线性变换并完成从胶囊层到全连接层的连接, 我们将二维的胶囊矩阵 $\mathbf{R}^{J \times d_j}$ 的每一个行向量依次拼接从而形成一个一维向量 $\mathbf{J} \in \mathbf{R}^{(J \times d_j) \times 1}$. 通过一个线性变换可以将向量表示 \mathbf{J} 映射为一个长度为 $|C|$ 的向量, 其中 $|C|$ 表示情感分类任务中情感倾向的数量. 最后通过 softmax 函数计算出每个类别出现的概率, 如式(3)所示:

$$\hat{y}_i = \text{softmax}(\mathbf{Wc} \cdot \mathbf{J} + \mathbf{bc}), \quad (3)$$

\hat{y}_i 表示第 i 个样本属于某种情感倾向的概率. 在此基础上进一步定义情感分类损失函数如式(4)所示:

$$L_{\text{sen}}(H(X_s^l; \theta_{\text{stl}})) = -\frac{1}{N_s^l} \sum_{i=1}^{N_s^l} y_i \ln \hat{y}_i + (1 - y_i) \ln(1 - \hat{y}_i), \quad (4)$$

其中, 情感标记 $y_i \in \{0, 1\}$, 分别表示积极和消极两种情感倾向; $H(\cdot; \theta)$ 函数表示样本到向量的映射过程, 即情感分类器, 其中 θ_{stl} 表示该网络部分中出现的全部参数. 通过优化该损失函数能够训练出参数 θ_{stl} .

1.3.5 领域分类

为了使得该网络能够得到源领域和领域的共享表示, 本文采取了对两个领域进行对抗训练的方式, 使得模型不能识别数据来自哪个领域, 从而使得两个领域表示无差异. 该模型应用梯度反转层到领域分类的对抗训练中. 在前向传播中梯度反转层作为一个普通的传递过程, 而在反向传播中, 梯度反转层使后一层传向前一层的梯度的方向反转, 式(5)和式(6)分别表示其前向和反向的传播过程:

$$Q_\lambda(x) = x, \quad (5)$$

$$\frac{\partial Q_\lambda(x)}{\partial x} = -\lambda I. \quad (6)$$

在数学上, 可以用 $f(Q_\lambda(\cdot; \theta_{\text{stl}}); \theta_D)$ 来表示领域分类器, 其中 θ_D 表示领域分类的参数. 通过梯度反转层, 一方面, 它能够通过训练 θ_D 来提高领域分类的效果; 另一方面, 它能够训练 θ_{stl} 使得领域分类效果下降, 而最终情感分类任务的参数与 θ_D 无关, 只与 θ_{stl} 相关, 即 θ_{stl} 能够学习到领域共享的信息. 领域分类的损失函数如式(7)所示:

$$L_{\text{dom}} = -\frac{1}{N_s^l + N_t^l} \sum_{i=1}^{N_s^l + N_t^l} d_i \ln \hat{d}_i + (1 - d_i) \ln(1 - \hat{d}_i), \quad (7)$$

其中,领域标记 $d_i \in \{0, 1\}$, 分别表示两个领域. \hat{d}_i 为第 i 个样本属于其中一个领域的概率.

图 1 蓝色框中最终的损失函数为情感分类的损失与领域分类损失之和, 计算如式(8)所示:

$$L_{\text{stage1}} = L_{\text{sen}}(H(X_s^l; \theta_{\text{stl}})) + L_{\text{dom}}. \quad (8)$$

因此, 根据式(8)能够完成该模型第一阶段的训练.

1.3.6 目标领域适应

第一阶段的模型经过训练之后得到的是一个能利用目标领域和源领域共享信息进行分类的模型. 在此基础上本文新增加了一层胶囊网络层作为目标领域的适应结构. 新增的胶囊层通过目标领域的标记信息进行训练, 因此该层能够学习到目标领域独有的特征表示.

在目标领域适应阶段, 首先进行的是卷积层的卷积操作, 并采用与阶段一卷积层中同样尺寸大小和数目的卷积核. 通过对输入的词向量进行卷积操作, 我们得到大小相同的特征图矩阵 $z' = [C'_1, C'_2, \dots, C'_m]$, 其中 $z' \in \mathbf{R}^{(n-h+1) \times m}$.

在该阶段的胶囊层中, 保持与阶段一中胶囊层相同的超参数, 即 J 个胶囊单元和每个单元的维度大小为 d_j . 同样地, 使用 $J \times d_j$ 个大小为 $\mathbf{R}^{1 \times (n-h+1)}$ 的卷积核对特征图矩阵再一次卷积. 在得到新特征图矩阵后, 将其变换为三维形式的特征图 $\mathbf{R}^{m \times J \times d_j}$. 利用文献[17]中动态路由算法来计算可以得到新的胶囊矩阵 $\mathbf{J}' \in \mathbf{R}^{J \times d_j}$. 每个向量的每一维度和其组合都代表着不同的性质和抽象概念, 比如单词语义、单词情感倾向、语法、词的位置等信息. 为了能够对目标领域有更好的适应, 该模型通过交互共享的胶囊层和目标领域适应的胶囊层, 使其在多种层面能够适应目标领域. 将该层的 $\mathbf{J}' \in \mathbf{R}^{J \times d_j}$ 转化为 $\mathbf{J}' \in \mathbf{R}^{(J \times d_j) \times 1}$ 的一维向量, 该向量与阶段一中的 $\mathbf{J} \in \mathbf{R}^{(J \times d_j) \times 1}$ 向量进行拼接, 形成新的向量 $\mathbf{J}'' \in \mathbf{R}^{(2 \times J \times d_j) \times 1}$. 该向量为两个领域共享特征和目标领域独有特征进行交互后的特征表示. 最后用一个线性变换将拼接后的文档向量映射成一个长度为 $|C|$ 的向量, 其中 $|C|$ 表示情感分类任务中情感倾向的数量. 通过 softmax 函数计算出每个类别出现的概率, 计算如式(9)所示:

$$\hat{y}' = \text{softmax}(\mathbf{W}_i \mathbf{J}'' + \mathbf{b}t), \quad (9)$$

\hat{y}'_i 表示第 i 个样本属于某种情感倾向的概率, 因此能够得到情感分类损失函数, 计算如式(10)所示:

$$L_{\text{stage2}}(H(X_i^l; z')) =$$

$$-\frac{1}{N_i^l} \sum_{i=1}^{N_i^l} y_i \ln \hat{y}'_i + (1 - y_i) \ln(1 - \hat{y}'_i). \quad (10)$$

2 实验分析

2.1 数据集

本文实验采用的是与 Blitzer 等^[10]实验一致的数据集, 数据集由 Amazon 的评论数据组成. 该数据集从 2007 年以来一直被跨领域情感分类的研究广泛地使用. 该数据集由以下 5 个领域构成: Books (B)、DVD (D)、Electronics (E)、Kitchen (K) 以及 Video (V). 在 5 个领域中, 每个领域都有 6 000 条有标记的评论数据, 其中 3 000 条评论数据的态度倾向为积极的 (positive) 以及 3 000 条评论数据是消极的 (negative). 同时, 每个领域中都存在一定数量的无标记评论数据, 无标记评论数据的具体数量如下: 在 Books 领域中有 9 750 条, DVD 领域中有 11 843 条, Electronics 领域中有 17 009 条, Kitchen 领域中有 13 856 条, Video 领域中有 30 180 条. 该数据集的分布如表 1 所示.

在该数据集上, 本文建立了 20 个跨领域的情感分类任务, 分别是: B→D, B→E, B→K, B→V, D→B, D→E, D→K, D→V, E→B, E→D, E→K, E→V, K→B, K→D, K→E, K→V, V→B, V→D, V→E, V→K. 其中 B 表示 Books 领域, D 表示 DVD 领域, E 表示 Electronics 领域, K 表示 Kitchen 领域以及 V 表示 Video 领域. 箭头表示跨领域的方向, 每个跨领域任务通过箭头连接并由源领域指向目标领域. 对每个任务 A→B, 本文从源领域 A 中随机选了 2 800 条积极的评论数据和 2 800 条消极的评论数据作为训练集数据, 剩下的数据作为验证集数据. 为了模拟现实中目标领域少量标记数据的情况, 从目标领域 B 中的 6 000 条数据随机划分 1/10, 即 600 条的标记数据作为已知的目标领域标记数据一起加入训练集中. 领域 B 中剩余的 5 400 条数据则作为测试集数据.

表 1 Amazon 数据集的统计信息

Table 1 Statistics of Amazon review dataset

Domain	Train	Test	Unlab
Books (B)	5 600	400	9 750
DVD (D)	5 600	400	11 843
Electronics (E)	5 600	400	17 009
Kitchen (K)	5 600	400	13 856
Video (V)	5 600	400	30 180

2.2 评价指标

本文的情感分类任务是将评论分类为积极的态度和消极的态度.因此,该任务为二分类的问题.

二分类任务中,预测标记与真实标记存在着4种情况,包括:

1) TP(True Positive):预测标记为积极评论,并且正确标记为积极评论;

2) FP(False Positive):预测标记为积极评论,但是正确标记为消极评论;

3) FN(False Negative):预测标记为消极评论,但是正确标记为积极评论;

4) TN(True Negative):预测标记为消极评论,并且正确标记为消极评论.

根据所得 TP、FP、FN 和 TN 的数目(分别记为 n_{TP} 、 n_{FP} 、 n_{FN} 、 n_{TN}),可以用式(11)、(12)、(13)、(14)分别计算出精确率(P)、召回率(R)、F1分数(F_1)与准确率(A).

$$P = \frac{n_{TP}}{n_{TP} + n_{FP}}, \quad (11)$$

$$R = \frac{n_{TP}}{n_{TP} + n_{FN}}, \quad (12)$$

$$F_1 = \frac{2PR}{P + R}, \quad (13)$$

$$A = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{TN} + n_{FP} + n_{FN}}. \quad (14)$$

为了能够和现有的工作做比较,本文采用了准确率 A (Accuracy)作为第一个评价指标.与精确率 P (Precision)和召回率 R (Recall)相比,F1分数(F1-Score)指标更加均衡、代表性更强.因此本文还采用了 F1 分数作为第二个评价指标.

2.3 基线方法

为了验证和比较本文模型的性能,本文选择了从2016到2018年中具有代表性的3个模型作为基线方法,接下来具体介绍这3个基线方法:

1) DANN 模型是文献[13]提出的模型.该方法利用了领域对抗的训练方法来使神经网络产生融合领域分类器表示.该模型提出了用梯度反转层对领域进行分类,即通过反转神经网络梯度方向,从而使其无法识别出源领域和目标领域的差别,达到领域共享的效果.该方法首次利用了对抗的思想进行跨领域情感分类,因此是对抗训练方法中经典的基线方法.

2) AMN 模型是文献[15]提出的模型.该方法主

要利用了对抗的训练和深度记忆网络模型来学习特征表示.其中深度记忆网络模型能够利用一个外部记忆单元存放信息,同时结合了注意力机制来提取对句子影响程度较大的单词,并且其注意力分布能够经过多层反复计算得到.AMN 模型通过两个记忆力网络共享参数以及注意力分布,并利用对抗训练的方法使得该网络能够学习到两个领域的共享特征.

3) P-Net 模型是文献[16]提出的一个模型,该模型是 AMN 模型的一种改进,它同时使用了分层的注意力机制和对抗训练的方法.比起 AMN 模型,该模型通过分层的注意力机制,能够更准确地提取到对评论影响最重要的句子并捕获对句子影响程度最大的单词,最终通过对抗训练的方法使模型学习到源领域和目标领域的共享特征.该文献中另外提到的扩充模型由于使用了额外工作提供的其他非共享特征,因此不作为本文的基线方法.

对所有的基线方法均按照原文献中设置的超参数进行训练.

2.4 实验结果分析

表2为所有模型在20个任务上的准确率(Accuracy)结果,表3为所有模型在20个任务上的F1分数(F1-score)结果,为了表格方便显示,将本文的模型简称为 T-Caps.综合表2和表3的结果,可以看出本文提出的模型在跨领域的情感分类任务中的表现都要比基线方法优秀.

在基线方法中,DANN 是一个最基础的模型,它虽然采用了对抗训练的思想,但是其网络结构单一,无法更有效地提取到共享的信息.AMN 和 P-Net 模型除了采用对抗训练的思想,还分别将记忆网络和分层注意力机制结合到模型中,使得它们的模型能够更加有效地提取一些共享的核心词.

从表2和表3可以看出,模型 AMN 和模型 P-Net 结果都要比 DANN 优秀,因此设计对任务更有效的共享架构是提升性能的途径之一.虽然基线方法也加入了目标领域标记数据进行训练,但是效果均比本文的模型差.因为在它们的结构中,缺乏针对目标领域信息作出适应的结构,从而无法判断领域标记信息是否能提高其对目标领域的适应效果.例如,当同一个共享信息在两个领域中表现出矛盾的情感态度时,由于不能确保对目标领域的分类有帮助,因而影响了分类器的效果.

表2 准确率结果
Table 2 Accuracy result

分类任务	DANN ^[13]	AMN ^[15]	P-Net ^[16]	T-Caps (本文)
B→D	0.826 8	0.836 7	0.859 9	0.867 5
B→E	0.755 9	0.799 1	0.836 9	0.841 1
B→K	0.772 1	0.826 4	0.848 2	0.850 5
B→V	0.824 9	0.836 8	0.856 2	0.861 0
D→B	0.799 7	0.836 1	0.857 1	0.864 7
D→E	0.755 9	0.795 5	0.838 8	0.845 5
D→K	0.773 8	0.807 8	0.847 7	0.852 7
D→V	0.851 0	0.864 5	0.876 1	0.881 1
E→B	0.727 3	0.767 5	0.822 1	0.828 3
E→D	0.754 4	0.797 3	0.812 9	0.818 9
E→K	0.836 1	0.869 6	0.888 5	0.894 4
E→V	0.763 6	0.813 1	0.833 7	0.836 2
K→B	0.734 4	0.781 9	0.827 0	0.829 3
K→D	0.745 7	0.786 4	0.821 0	0.826 1
K→E	0.846 8	0.857 4	0.866 5	0.871 8
K→V	0.756 1	0.812 6	0.845 7	0.849 8
V→B	0.793 2	0.827 6	0.856 6	0.861 9
V→D	0.834 0	0.861 1	0.863 7	0.868 4
V→E	0.750 4	0.789 7	0.834 9	0.837 1
V→K	0.745 5	0.802 6	0.848 0	0.853 1
平均	0.782 4	0.818 5	0.847 1	0.852 0

表3 F1-score 结果
Table 3 F1-score result

分类任务	DANN ^[13]	AMN ^[15]	P-Net ^[16]	T-Caps (本文)
B→D	0.831 0	0.841 3	0.864 5	0.872 4
B→E	0.768 4	0.812 2	0.851 1	0.855 0
B→K	0.781 3	0.836 3	0.858 5	0.860 9
B→V	0.829 7	0.841 5	0.861 6	0.865 9
D→B	0.803 2	0.839 9	0.860 7	0.868 5
D→E	0.772 8	0.813 0	0.856 9	0.864 0
D→K	0.787 4	0.822 0	0.862 9	0.867 5
D→V	0.854 4	0.867 6	0.879 6	0.884 6
E→B	0.734 1	0.774 3	0.829 7	0.835 6
E→D	0.763 2	0.806 1	0.821 8	0.827 8
E→K	0.835 8	0.869 3	0.888 4	0.894 4
E→V	0.772 5	0.822 1	0.842 8	0.845 4
K→B	0.740 3	0.788 6	0.833 6	0.836 6
K→D	0.764 3	0.806 3	0.841 5	0.847 5
K→E	0.852 2	0.862 5	0.872 6	0.877 6
K→V	0.784 5	0.842 0	0.877 0	0.880 9
V→B	0.795 4	0.829 8	0.859 3	0.863 9
V→D	0.835 9	0.863 2	0.865 9	0.871 2
V→E	0.767 2	0.807 5	0.853 5	0.855 6
V→K	0.755 5	0.813 5	0.859 2	0.864 9
平均	0.791 5	0.827 9	0.857 1	0.862 0

对比上述模型,本文模型具有以下优点:一方面,本文的模型在第一阶段通过胶囊网络方法能提取更多类型的共享特征,使得更多的信息得到迁移;另一方面,在第二阶段,结合目标领域信息的网络结构使得本文模型更加适应目标领域情景.此外,分阶段的训练保证了对共享领域和目标领域都能够达到最优,即对信息充分学习.因此最终本文的模型比基线方法的效果都更加优越.

2.5 参数分析

本文选取了两种参数进行分析.第一种参数属于模型外的参数,与加入的目标领域有标记数据集的大小有关;第二种参数属于模型内的参数,与模型的结构相关.

2.5.1 模型外参数分析

在前面实验设置时,本文选择将 10% 的目标领域标记数据加到训练集来模拟少量目标领域标记数据的情景.在实际中,已有的目标领域数据的数量或者比例大小是未知的,但该比例对模型分类效果有一定的影响.因此,本文探索了加入不同比例的目标领域标记数据对 target_capsule 模型效果的影响.

本文总共进行了 20 个跨领域任务,本节选取了其中两个代表性的任务,分别是 B→D 以及 K→E,前者任务中两个领域较为相关,后者任务中两个领域较不相关,同时保持模型其他参数不变.本文将目标领域数据集中的标记数据的比例从 5% 到 50% 逐渐递增.考虑到过多标记数据与现实中目标领域标记数据不足的情况不符,所以将 50% 作为一个上限值.实验的结果如图 2 所示.

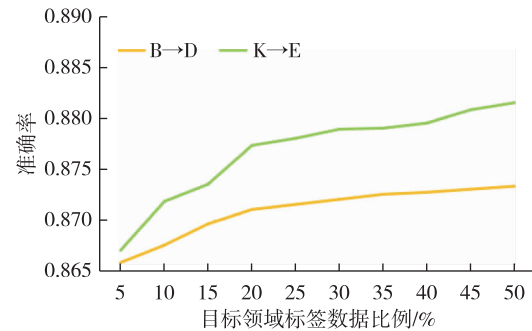


图2 目标领域标记数据比例对准确率的影响

Fig. 2 The impact of target label data ratio on accuracy

由图 2 可知,模型的效果随着目标领域标记数据比例的增加而提升,表明了目标领域标记数据增加的时候会对模型效果产生正向的影响,与现实认知相符合.随着数据的增加,模型性能提升的幅度开

始减小,因为 target_capsule 模型捕获的信息具有普遍性,比如语法结构等信息在大多句子中都能提取到.当数据增加时,该类信息重复出现的次数增多,使得模型的提升有限.此外,加入目标领域标记信息时,模型在 $K \rightarrow E$ 任务中的表现比 $B \rightarrow D$ 任务优秀.其原因在于:在 $K \rightarrow E$ 任务中,两个任务之间的相关性较低,两个领域共享的知识较少,因此加入目标领域的标记数据可以更大程度地提高模型对目标领域分类的效果.

2.5.2 模型内参数分析

接下来分析模型网络结构本身的超参数.在本模型中,卷积核个数为模型内最主要的超参数,因此本文选择卷积层中的卷积核个数来分析其变化对模型效果的影响.与第一种参数实验分析相同的原因,本文同样选择了任务 $B \rightarrow D$ 和任务 $K \rightarrow E$ 的数据作为代表性任务进行该部分的实验参数分析.

保持其他参数不变,并将卷积核个数的情况分为 16、32、64、128 和 256 个.根据图 3 可知,卷积核个数大约在 64 时,模型达到最优的效果.这是因为在卷积核个数较少时,不能充分提取到所需要的特征,使得模型不能提取到充分的特征,从而导致模型的效果下降.另一方面,当卷积核个数过多时,容易导致训练参数过多使得模型欠拟合.

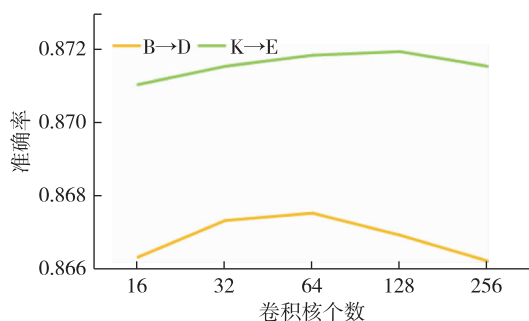


图3 卷积核个数对准确率的影响

Fig. 3 The impact of number of convolution kernels on accuracy

3 结束语

在跨领域的情感分类研究中,存在着分类器对目标领域适应能力不足的问题.本文提出了一种利用目标领域的标记信息来改善模型适应效果的思想.之前的模型缺乏针对目标领域信息应用的结构,因此不能充分利用该信息.本文提出了一个基于胶囊网络的模型,在此基础框架上,新增加了额外一层胶囊网络层作为目标领域的适应的结构,使得模型能够充分利用目标领域的信息.从实验的结果来看,

本文提出的模型效果优于以往的研究方法,有一定的可行性和价值.今后的研究重点将放在如何更高效地结合目标领域的信息,使得模型对目标领域的适应能力有进一步的提升.

参考文献

References

- [1] Anderson M, Magruder J. Learning from the crowd: regression discontinuity estimates of the effects of an online review database [J]. The Economic Journal, 2012, 122 (563): 957-989
- [2] Pang B, Lee L. Opinion mining and sentiment analysis [J]. Foundations and Trends in Information Retrieval, 2008, 2(1/2): 1-135
- [3] Liu B. Sentiment analysis and opinion mining [J]. Synthesis Lectures on Human Language Technologies, 2012, 5(1): 1-167
- [4] Devitt A, Ahmad K. Sentiment polarity identification in financial news: a cohesion-based approach [C] // ACL 2007, Proceedings of the Meeting of the Association for Computational Linguistics, Prague, Czech Republic, 2007
- [5] Kim S M, Hovy E H. Crystal: analyzing predictive opinions on the web [C] // EMNLP-CoNLL 2007, Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Prague, Czech Republic. DBLP, 2007: 1056-1064
- [6] Wang S, Manning C D. Baselines and bigrams: simple, good sentiment and topic classification [C] // Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Association for Computational Linguistics, 2012
- [7] Tan S B, Cheng X Q, Wang Y F, et al. Adapting naive bayes to domain adaptation for sentiment analysis [M] // Tan S B, Cheng X Q, Wang Y F, et al. eds. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009: 337-349. DOI: 10.1007/978-3-642-00958-7_31
- [8] Yoshida Y, Hirao T, Iwata T, et al. Transfer learning for multiple-domain sentiment analysis: identifying domain dependent/independent word polarity [C] // Twenty-Fifth AAAI Conference on Artificial Intelligence, 2011
- [9] Yu J, Jiang J. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification [C] // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: 236-246
- [10] Blitzer J, Dredze M, Pereira F. Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification [C] // Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, 2007: 440-447
- [11] Pan S J, Ni X, Sun J T, et al. Cross-domain sentiment classification via spectral feature alignment [C] // Proceedings of the 19th International Conference on World Wide Web. ACM, 2010: 751-760

- [12] Goodfellow I, Bengio Y, Courville A. Deep learning [M]. Cambridge, MA: MIT Press, 2016
- [13] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks [J]. The Journal of Machine Learning Research, 2016, 17(1): 2096-2030
- [14] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation [J]. arXiv Preprint arXiv: 1409.7495, 2014
- [15] Li Z, Zhang Y, Wei Y, et al. End-to-end adversarial memory network for cross-domain sentiment classification [C] // IJCAI, 2017: 2237-2243
- [16] Li Z, Wei Y, Zhang Y, et al. Hierarchical attention transfer network for cross-domain sentiment classification [C] // Thirty-Second AAAI Conference on Artificial Intelligence, 2018
- [17] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules [C] // Advances in Neural Information Processing Systems, 2017: 3856-3866

Cross-domain sentiment classification by capsule network

LIN Yue¹ QIAN Tiejun¹

¹ School of Computer Science, Wuhan University, Wuhan 430072

Abstract Sentiment analysis aims to extract users' sentiments and opinions about or their attitude towards a specific product, service, or event. The lack of labeled data is a significant challenge in sentiment analysis and will deteriorate the performance of the classifier in a supervised sentiment analysis task. The cross-domain approach has been shown to be effective in addressing this problem. However, the inherent difference between the source and target domains will make it difficult for the classifier to be adaptive to the target domain. In this paper, we propose a novel method to use the available labeled data, however few they may be, in the target domain to enhance the domain adaptation. Specifically, we present a cross-domain sentiment classification model using the capsule network. Based on this architecture, we design extra capsule layers for domain adaptation. Extensive experiments with real-world datasets prove that our proposed model outperforms baselines by a large margin.

Key words sentiment classification; domain adaptation; capsule network