

基于时间注意力胶囊网络的维吾尔语情感分类模型

罗涵天^{1,2,3}, 杨雅婷^{1,2,3*}, 董瑞^{1,2,3}, 马博^{1,2,3}

1.中国科学院新疆理化技术研究所, 乌鲁木齐, 830011

2.中国科学院大学, 北京, 100049

3.新疆民族语音语言信息处理实验室, 乌鲁木齐, 830011

(通讯作者邮箱: yangyt@ms.xjb.ac.cn)

摘要

维吾尔语属于稀缺资源语言, 如何在资源有限的情况下提升维吾尔语情感分类模型的性能, 是目前待解决的问题。本文针对现有维吾尔语情感分析因为泛化能力不足所导致的分类效果不佳的问题, 提出了基于时间卷积注意力胶囊网络的维吾尔语情感分类模型(TA-Cap)。本文在维吾尔语情感分类数据集中进行了实验并且从多个评价指标(准确率, 精确率, 召回率, F1值)进行评估, 实验结果表明本文提出的模型相比传统深度学习模型可以有效提升维吾尔语情感分类的各项指标。

关键词: 维吾尔语; 情感分析; 胶囊网络

Uyghur Sentiment Classification Model Based on Temporal Attention Capsule Networks

Hantian Luo^{1,2,3}, Yating Yang^{1,2,3*}, Rui Dong^{1,2,3} and Bo Ma^{1,2,3}

1.Xinjiang Technical Institute of Physics & Chemistry, Chinese Academy of Sciences, Urumqi, 830011

2.University of Chinese Academy of Sciences, Beijing, 100049

3.Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi, 830011

Abstract

Uyghur is a language with scarce resources. How to improve the performance of Uyghur sentiment classification model in the context of limited resources is a problem to be solved. Aiming at the problem of poor classification effect caused by the insufficient generalization ability of existing Uyghur sentiment analysis models, this paper proposes Uyghur sentiment classification model (TA-CAP) based on Temporal attention capsule network to solve this problem. We conducted experiments on the Uyghur sentiment analysis dataset and conducted experimental evaluation from multiple evaluation indicators (Accuracy, Precision, Recall, F1-score). The results of this experiment show that our model can effectively improve the performance of the Uyghur sentiment classification model.

Keywords: Uyghur, sentiment analysis, capsule network

1 引言

©2021 中国计算语言学大会根据《Creative Commons Attribution 4.0 International License》许可出版
*通讯作者邮箱: yangyt@ms.xjb.ac.cn

文本情感分类任务，一直是学术界研究的热点之一，其任务目标是为句子分配相应的情感标签，属于一种特殊的文本分类任务，其特殊性体现在文本情感的内在隐蔽性。研究维吾尔语情感分类对公民以及国家有着重要的意义，不仅可以帮助政府部门有效整治网络环境，使网民拥有一个干净的网络环境，还可以将此类技术应用于恶意检测，舆情分析等领域，帮助政府部门更好地掌握新疆地区动向，从而有效促进民族团结。早期的情感分析方法主要以基于规则与统计机器学习的方法为主 (Rish and others, 2001; Genkin, 2007; Joachims, 1998), Wiebe (1999) 采用朴素贝叶斯的方法来对文本的主客观性进行分类识别, Das (2007) 通过构造情感词典的方法, 使用人工标注构造的情感词典, 给予不同的词相应的情感极性, 最后通过计算文本情感词将情感极性进行累计进行判断。也有不少学者使用无监督聚类 (Wiebe and others, 2000; Turney, 2002) 的方法来对情感词进行聚类分析。此两类方法忽视了文本的语序顺序以及语义信息, 因此分类准确率不高。

基于深度学习的方法有效提高了情感分析任务的分类效果, 如基于RNNs (Bengio et al., 1994; Cho et al., 2014a; Hochreiter and Schmidhuber, 1997) 的方法可用很好地从隐藏层中获取上下文信息, 但是RNN(Recurrent Neural Network)不能并行计算, 当序列较长的时候, 训练相对较为困难, 这无疑增加了训练的时长, 而且当文本较长的时候容易出现长距离依赖的问题导致文本信息丢失。基于CNN(Convolutional Neural Network) (LeCun et al., 1989; Kim, 2014) 的方法可以有效提取长文本序列的信息特征, 并且在数据量大的时候依然可以保证较高的准确率, 但是基于CNN的模型不能很好地记忆语序信息, 因此会造成部分语义信息的丢失。近几年维吾尔语情感分类模型是以基于RNN的方法为主 (王树恒 et al., 2017; 帕丽旦 et al., 2019) 但是基于RNN的方法在文本序列较长的时候容易造成语义信息的丢失, 并且模型的泛化能力不足, 这对维吾尔语的语言特性来说并不友好, 维吾尔语是一种较为复杂的语言形态, 并且是一种资源稀缺语言, 因此需要模型有较强的泛化能力以及特征抽取能力。针对此问题, 本文提出了TA-Cap模型, 并在维吾尔语数据集中进行了详细实验, 结果证明本文所提出的模型相比常用模型方法在分类效果上有显著的提升, 除此之外, 本文还探究了加入维吾尔语情感特征对实验结果的影响。

2 相关工作

维吾尔语属于阿尔泰语系, 是一种黏着语并且构词形态丰富, 这无疑增加了维吾尔语在自然语言处理 (Natural Language Processing, NLP) 任务中的难度。现有的维吾尔语情感分类方法主要有基于词典与机器学习相结合的方法 (Tuerhongtai et al.,) 以及基于深度学习的方法, 目前基于深度学习的方法已占据主流。王树恒 (2017) 提出了基于Bi-LSTM和词嵌入的维吾尔语情感分析模型, 较于SVM (Joachims, 1998), RNN以及CNN模型, 分类的效果更好。帕丽旦 (2019) 融合了改进后的词性特征向量以及韵律短语向量, 提出了基于Bi-RNN的情感分类模型, 有效提升了模型的分类效果。

Bai等人 (2018) 重新定义了序列模型(RNN), 将卷积作为序列模型的起点, 提出了时间卷积网络模型(Temporal Convolutional Network, TCN), 并在大量通用任务中做了实验, 并证实了它的优越性能。TCN可以获取语序信息并且可以保证信息不易被丢失, 这可以很好地解决了RNNs模型训练时间长以及长距离依赖的问题, 同时也解决了CNN模型无法获取语序信息的不足。

胶囊网络 (Sabour et al., 2017) 由深度学习之父hinton提出, 最早应用于图像分类任务并取得了显著的效果, 随后有学者将其应用于NLP领域, 并取得了相关成就 (Yang et al., 2018; Wang et al., 2018; Aly et al., 2019; Vu et al., 2019; Chen et al., 2020)。宏观来看, 胶囊网络就是通过使用一种无监督聚类算法, 将不同的底层语义胶囊实体进行组合从而衍生出不同的语义信息。胶囊网络已被证实通过其独有的外推机制 (Zhao et al., 2019), 可以在训练过程中外推生成未见语句, 可以将低层语义实体映射到高层语义空间当中, 衍生出多种语义, 因此可以使得模型具有较强的泛化能力以及鲁棒性。Wang (Wang et al., 2018) 提出了RNN-Capsule模型, 并将其应用于英文文本情感分类任务, 取得了较好的效果。目前, 基于胶囊网络的结构模型还尚未在维吾尔语自然语言处理任务中使用, 为探究其有效性, 本文在维吾尔语情感分类数据集进行了实验。且本文提出的方法不同于传统的胶囊网络算法, 是在基于动态路由的胶囊网络模型中融入了多头注意力机制 (Vaswani et al., 2017) 来增加模型对多个语义信息的关注能力, 从而进一步提升文本的泛化能力。

3 TA-Cap模型

模型的输入为预处理后的维吾尔语文本序列，输出为情感分类的类别标签，TA-Cap模型由三个部分组成：第一个部分为词嵌入模块，使用训练后200维的CBOW (Mikolov et al., 2013)词向量作为Word Embedding初始化；第二个部分为时间卷积模块，将初始化后的Word Embedding送入到TCN中进行序列建模，通过设置扩张率(d)以获取不同层次的语义信息，其中扩张率按 2^n 逐层增长，TCN序列无长距离依赖问题，可以保证语义信息以及语序信息不易被丢失。第三个部分为注意力胶囊模块，此模块将序列映射到多维度语义空间，并通过胶囊网络将多维语义的实体到进行组合去生成更高的语义实体；最后的部分为全连接层，并经过softmax函数得到情感分类输出。本文所提出的模型结构图如Figure 1所示：

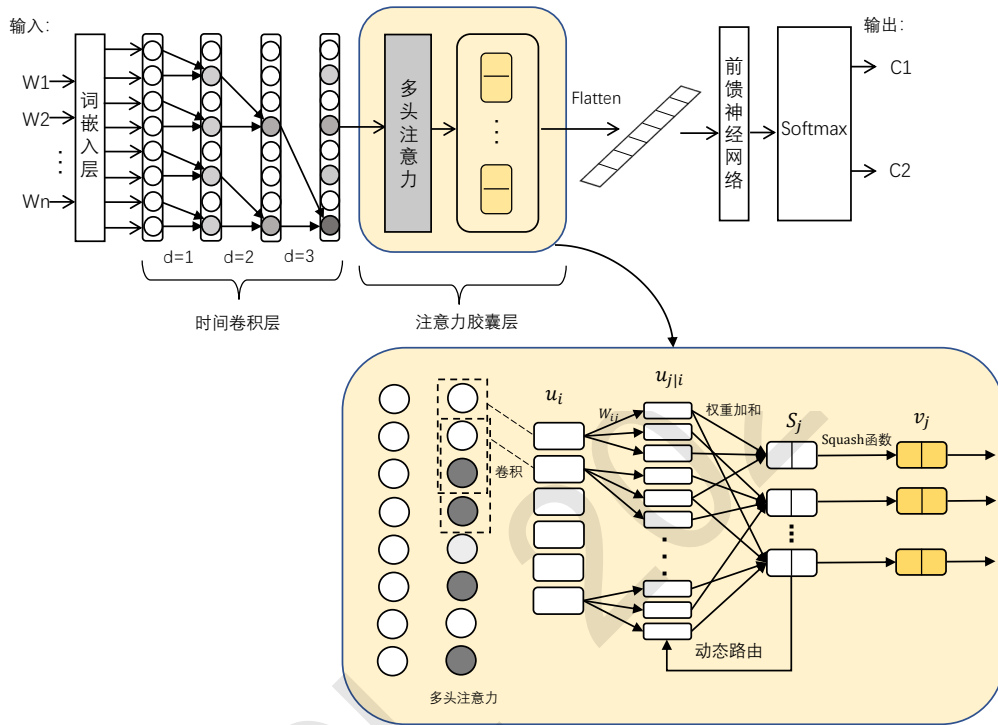


Figure 1: TA-Cap模型结构图

3.1 时间卷积层

时间卷积网络(TCN)由空洞因果卷积以及残差链接 (He et al., 2016)组成，因果卷积 (Long et al., 2015)是一种受严格的时间约束的序列模型，每时刻的输入只和当前输入和前一时刻的输入有关，这无疑增加了运算的时间且无法获得较远距离的信息，空洞卷积 (Oord et al., 2016)可以扩大文本的感受野，并且理论上可以获取到无限远的距离并且可以并行计算，从而提高训练效率，缩短训练时间。除此之外，本文还加入了残差链接来保证当网络较深时，模型可以达到较好的性能。给定输入 x ，残差网络定义为：

$$y(x) = ReLU(\gamma(x) + x) \tag{1}$$

其中 $\gamma(x)$ 为残差模块，并且本文使用ReLU函数 (Nair and Hinton, 2010)作为输出的激活函数。

3.2 注意力胶囊层

在注意力胶囊层中，输入为经TCN建模的文本向量，其中向量维度不变，经多头注意力机制的过程如Figure 2:

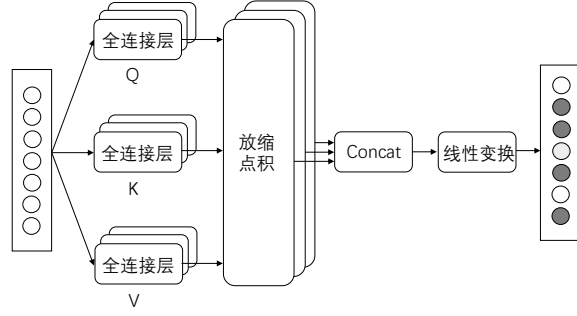


Figure 2: 多头注意力计算过程

其中放缩点积的计算公式如下:

$$Attention(Q, K, V) = softmax\left(\frac{Q * K^T}{\sqrt{d_k}}\right)V \quad (2)$$

其中Q, K, V分别代表了查询向量(Query), 键向量(Key), 值向量(Value), $\sqrt{d_k}$ 为缩放因子, 用来调节点积的大小。多头注意力(Multi-Attention)完善了Attention机制的功能, 赋予了模型关注到多个不同位置信息的能力, h头注意力机制可以得到h个多头并得到h组Q, K, V向量, 最后, 将输出的h组向量进行拼接。本文使用的是三头注意力机制, 在Concat过程中计算公式如下:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (3)$$

其中:

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (4)$$

$head_1$ 到 $head_h$ 表示h个头的自注意力点积, W^O 表示与Concat后的矩阵做线性运算的矩阵, 本文将线性变换后的序列经卷积操作后送入到胶囊层中, 在胶囊层中本文使用的是基于动态路由的方法, 基于动态路由的算法过程如下所示:

Algorithm 1 动态路由算法

Input:

输入为低胶囊层向量 u_i , 通过仿射变换得到的预测向量为 $\hat{u}_{j|i} = W_{ij}u_i$

Output:

输出为包含概率信息的高层胶囊实体 v_j

- 1: 对于所有l层的胶囊i和(l+1)层的胶囊j中, 初始化 $b_{ij} = 0$
 - 2: 初始化 $n = 0$
 - 3: while $n < N$:
 - 4: $c_{ij} = softmax(b_{ij})$
 - 5: $s_j = \sum_i c_{ij} \hat{u}_{j|i}$
 - 6: $v_j = squash(s_j)$
 - 7: $b_{ij} = \hat{u}_{j|i} \cdot v_j + b_{ij}$
 - 8: $n++$
 - 9: **return** v_j ;
-

其中 s_j 表示胶囊实体, v_j 表示输出胶囊实体的概率信息, 输入向量 u_i 通过仿射变换乘以权重矩阵 W_{ij} 得到预测向量 $\hat{u}_{j|i}$ 。 $W_{ij} \in R^{d_i \times d_j}$, d_i 表示低层胶囊i的的维度, d_j 表示高层胶囊j的维度, c_{ij} 为耦合系数, 等价于路由过程中的权重系数, 设置 b_{ij} 初始值为0, 且 b_{ij} 是随着预测向量 $\hat{u}_{j|i}$ 和输出向量 v_j 不断更新的。其中squash函数为非线性压缩函数, 可以获得输出胶囊实体的概率信息:

$$v_j = squash(s_j) = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (5)$$

3.3 融合维吾尔语情感特征

在文本中，情感词能较好的体现文本的情感信息，在文本向量中加入情感特征会对实验的结果造成影响，为进一步探究维吾尔语情感词特征对模型实验结果的影响，本文使用自建的维吾尔语情感词典（包含维吾尔语正向词语和负向词语），在文本中抽取相应的情感词，作为情感特征将其词嵌入（Word Embedding）融合（Concat）到原文本的词向量中。融合后的文本词向量作为模型的输入，实验过程如figure 3所示：

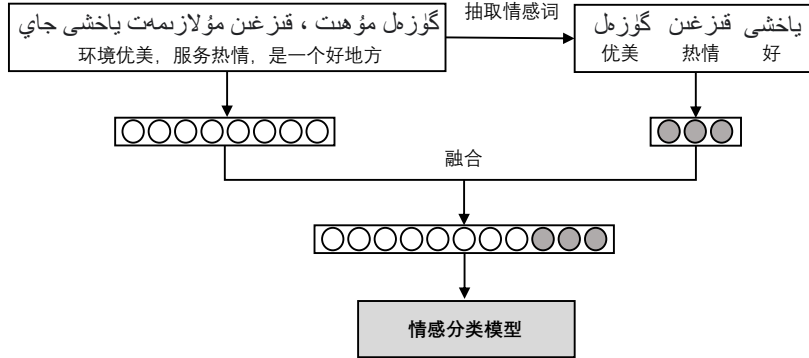


Figure 3: 融合维吾尔语情感特征的实验过程

4 实验与结果分析

4.1 语料获取

维吾尔语属于稀缺资源语言，目前无公开的维吾尔语情感分析任务的语料库，因此难以提供语料资源供研究人员进行研究。本文使用中国科学院计算技术研究所谭松波老师收集的酒店评论情感分类数据集 (Tan and Zhang, 2008) 中的6000条（3000正向情感，3000负向情感）平衡语料，通过维汉机器翻译系统进行翻译最后通过人工筛选加修正的方式来对文本语料进行修改调整，其方式不仅可以保证语料的质量，而且可以减少人工标注语料的成本，其不足之处为语料中存在少量语句不通顺的情况，增加了人工筛选和修改的时间成本。本次实验使用的维汉机器翻译系统由中国科学院新疆理化技术研究所多语种信息技术研究室研发，曾多次在CCMT维汉机器翻译评测中取得优异成绩，目前被CCMT指定为维汉机器翻译评测任务语料提供单位。

本文使用10%(600条)的语料作为测试训练集，90%(5400条)的语料作为训练集，其中在训练集中抽取10%的数据作为交叉验证集。在翻译获得的数据集中，本文对语料进行了分词、去停用词、去非文本词等预处理操作。除此之外，在天山网维语版等维吾尔语网站爬取收集了200w条维吾尔语新闻语料，并使用CBOW模型进行训练，将训练后的200维(dim=200)的词向量用于初始化Word Embedding。

4.2 实验环境

在本文中用于实验的环境如Table 1所示。

实验环境	环境配置
GPU型号	Tesla K80
操作系统	Ubuntu 18.04.1
深度学习框架	tensorflow 1.15.0
编程语言	python 3.7.4

Table 1: 环境说明

4.3 评价指标

在本文中，采用国际通用标准：精确率(Precision)、召回率(Recall)、F1分数(F1-score)以及准确率(Accuracy)来综合判断模型的好坏，除此之外本文还记录了每个模型在训练过程中所

花费的时间。其中准确率为预测正确的结果占所有样本的比例,精确率为预测为正的样本中实际上有多少是真正的样本,召回率为实际样本中为正的样本中被预测成正样本的概率, F1分数则为综合了精确率和召回率的结果,各指标计算公式如下:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1 = \frac{(2 \times Precision \times Recall)}{Precision + Recall} \quad (9)$$

4.4 对比模型

在实验中,本文实现了常用于分类任务及情感分析任务所使用的深度学习模型作为对比实验模型,包括基于RNN的模型:LSTM (Hochreiter and Schmidhuber, 1997)、Bi-LSTM、GRU (Cho et al., 2014b)、Bi-GRU,基于TCN (Bai et al., 2018)的模型,基于CNN的模型:TextCNN (Kim, 2014)以及基于胶囊网络的模型: Capsule-B(CNN-Capsule) (Yang et al., 2018)、RNN-Capsule (Wang et al., 2018)。

4.5 参数设定

在训练过程中权重矩阵服从glorot均匀分布,并采用Adam优化器作为模型优化器,本文使用200维词向量初始化词嵌入矩阵,采用3头注意力机制,胶囊网络模块中的胶囊个数为10,胶囊维度为16,动态路由次数N=3;除此之外本文训练使用的各参数设置如Table 2所示。

参数	参数值
Learning_rate	0.001
hidden_layer	64
dropout	0.5
batch_size	64

Table 2: 参数设置

4.6 实验结果

在保证参数一致的情况下,实验结果如Table 3所示。

模型	Accuracy(%)	Precision(%)	Recall(%)	F1(%)	Train-Time
LSTM	88.50	87.85	89.86	88.85	13min29s
Bi-LSTM	88.33	91.69	86.73	89.13	21min50s
GRU	88.83	89.46	89.17	89.31	11min28s
Bi-GRU	87.67	87.22	88.92	88.06	18min47s
TCN	89.16	90.09	89.24	89.67	59.9s
TextCNN	89.50	91.37	88.81	90.07	58.5s
Capsule-B	90.00	90.73	90.15	90.44	1min28s
RNN-Capsule	90.16	92.01	89.44	90.70	18min29s
TA-Cap	91.67	93.92	90.46	92.16	2min58s

Table 3: 各模型实验结果对比

实验结果表明本文所提出的模型在准确率、精确率、召回率、F1分数相比传统深度学习模型以及基于胶囊网络的模型都有了明显提升,在训练时长方面,TA-Cap模型训练时长较短且

模型的测试效果较好，虽时间没有达到最短，但是在训练时长和精度方面达到了一个最佳的平衡，并且模型具有较强的泛化性能以及鲁棒性。

原因分析：1.基于RNN的模型可以获取序列的语义信息，但RNN有长距离依赖的问题，容易造成信息的丢失，而且RNNs无法并行计算，因此训练时间较长；2.基于CNN的模型能获取到关键字的信息，但CNN无法较好地获取到上下文的语序信息，因此也会造成信息的丢失导致模型效果欠佳；3.TCN与胶囊网络的结合在维吾尔语情感分类任务中效果优于CNN和RNN与胶囊网络的结合（Capsule-B和RNN-Capsule），其原因为TCN通过空洞卷积可以很好地获取语序信息以及语义信息，并且不易造成语义信息的丢失，因此在胶囊层的时候可以保证语义的准确性，从而有效提升了模型的泛化性能以及鲁棒性。

4.7 消融实验

本文为探究TA-Cap模型各模块的性能表现，进行了消融实验，实验结果表明在TCN结构中融入多头注意力机制与胶囊网络可以显著提升模型的综合表现力，实验结果如Table 4所示。

模型	Accuracy(%)	Precision(%)	Recall(%)	F1(%)	Train-Time
TCN	89.16	90.09	89.24	89.67	59.9s
Capsule	90.17	90.41	89.84	90.12	54.1s
TCN+Capsule	90.83	92.01	90.57	91.28	1min15s
TCN+Multi-Att+Capsule(TA-Cap)	91.67	93.92	90.46	92.16	2min58s

Table 4: 各模型实验结果对比

实验结果说明，本文所提出的模型综合表现力较好，原因在于胶囊网络可以提升模型的泛化性能以及鲁棒而且多头注意力机制具有关注到多个语义空间的能力，从而可以进一步提升模型的特征抽取能力。

4.8 实验参数对实验结果的影响

4.8.1 胶囊个数

考虑到胶囊的个数对于实验结果的影响比较大，胶囊的个数体现了不同的实体特征，是模型泛化能力的体现，针对不同的语言需要选择合适的胶囊个数来达到最优效果，在其他参数变量不变的情况下，本文设置了不同的胶囊个数来进行测试，并使用准确率(Accuracy)以及F1-分数(F1-score)来对模型进行测试，F1分数为精确率和召回率的加权平均结合准确率可以更好地衡量模型的好坏，Figure 4 为胶囊个数对实验测试准确率的影响。

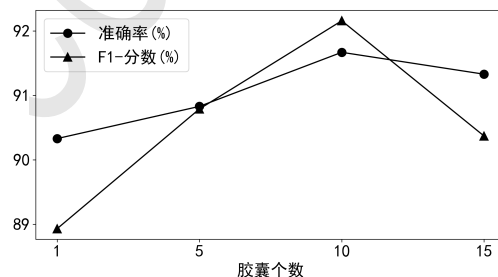


Figure 4: 胶囊个数对实验结果的影响

从Figure 4 中可以看出随着胶囊个数的增加，模型的测试准确率和F1-分数逐渐上升，在胶囊的个数等于10的时候，准确率和F1-分达到最高点。

4.8.2 TCN层数

TCN可以获取到序列的顺序信息，并且保证信息不易被丢失，层数的多少表示模型可获取到前序序列信息量的多少，在保证其他参数不变的情况下，本文针对层数进行了五组实验，实验结果如图5所示。

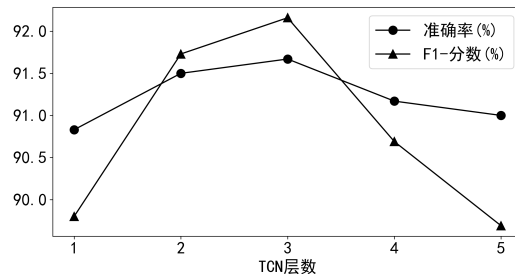


Figure 5: TCN层数对实验结果的影响

其中第n层分别对应的扩张率 $d = 2^{n-1}$ ，从表中可以看出，随着层数的增加，模型的测试准确率和F1-分数逐渐上升，当层数为3，即 $d=4$ 的时候，模型测试准确率和F1-分数达到最高点，当层数超过3层的时候，准确率以及F1-分数逐步下降。

4.9 加入维吾尔语情感特征对实验结果的影响

文本的情感词可以体现文本整体的情感倾向，考虑到文本情感词特征易对实验结果造成影响，本文在多个模型上测试了维吾尔语情感特征对模型测试准确率以及F1-分数的影响。实验结果如figure 6和7所示：

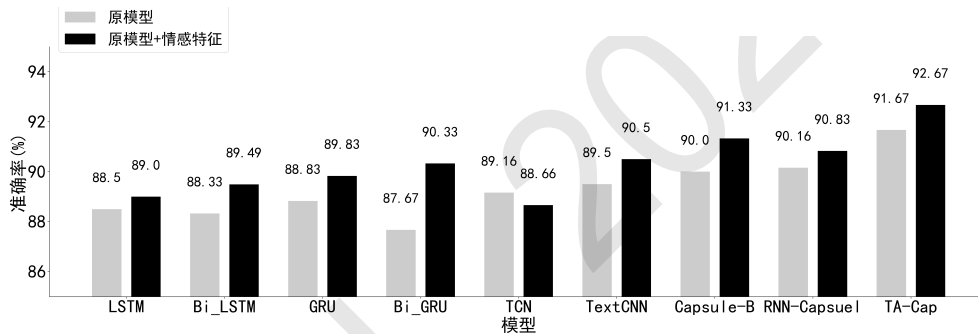


Figure 6: 模型测试的准确率

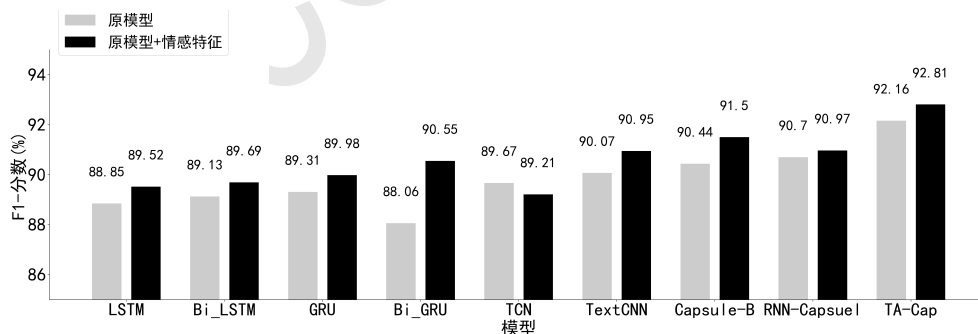


Figure 7: 模型测试的F1-分数

实验结果表明，融入维吾尔语情感特征可以提升模型的分类效果，准确率平均提升0.98个百分点，F1分数平均提升0.93个百分点，本文提出的模型(TA-Cap)准确率以及F1-值分别提升1个百分点至92.67%和0.68个百分点至92.81%，将维吾尔语情感特征融入到原本本向量中可以有效提升模型对于情感信息的关注度，从而可以提升模型的分类效果。

5 总结

维吾尔语属于黏着语系，构词形态丰富，现有的维吾尔语情感分类模型尚处于起步阶段，需要研究的内容还有许多。本文提出了一种用于维吾尔语情感分类任务的新模型(TA-Cap)，其目的是为解决目前维吾尔语情感分类模型因泛化能力不足导致的分类效果不佳的问题，经过详细的实验评估，本文验证了TA-Cap模型的有效性，相比传统的深度学习方法，本文提出的模型方法可以显著提升维吾尔语情感分类效果以及性能。

致谢

本文工作得到中国科学院青年创新促进会资助项目(2017472号)；中国科学院西部之光人才培养计划A类资助项目(2017-XBQNXZ-A-005)；国家自然科学基金资助项目(U2003303)；国家重点研发计划项目(2018YFC0823404)；中国科学院西部青年学者A类资助项目(2019-XBQNXZ-A-004)；新疆高层次引进人才项目(新人社函[2017]699号)的资助。

参考文献

- Rami Aly, Steffen Remus, and Chris Biemann. 2019. Hierarchical multi-label classification of text with capsule networks. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 323–330.
- Shaojie Bai, J Zico Kolter, and Vladlen Koltun. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.
- Yoshua Bengio, Patrice Simard, and Paolo Frasconi. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166.
- Boli Chen, Xin Huang, Lin Xiao, and Liping Jing. 2020. Hyperbolic capsule networks for multi-label classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3115–3124.
- Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014a. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.
- Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014b. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- Sanjiv R Das and Mike Y Chen. 2007. Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management science*, 53(9):1375–1388.
- BM Genkin. 2007. Economics and sociology of labor. *Norma*.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Thorsten Joachims. 1998. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
- Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. 1989. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.
- Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *ICML*.
- Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.
- Irina Rish et al. 2001. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46.
- Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. 2017. Dynamic routing between capsules. In *Advances in neural information processing systems*, pages 3856–3866.
- Songbo Tan and Jin Zhang. 2008. An empirical study of sentiment analysis for chinese documents. *Expert Systems with applications*, 34(4):2622–2629.
- Rexidanmu Tuerhongtai, Wushour Silamu, Yierxiati Tuerhong, et al. 词典与机器学习方法相结合的维吾尔语文本情感分析. *中文信息学报*, 31(1):177–183.
- Peter D Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- Thanh Vu, Tu Dinh Nguyen, Dat Quoc Nguyen, Dinh Phung, et al. 2019. A capsule network-based embedding model for knowledge graph completion and search personalization. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2180–2189.
- Yequan Wang, Aixin Sun, Jialong Han, Ying Liu, and Xiaoyan Zhu. 2018. Sentiment analysis by capsules. In *Proceedings of the 2018 world wide web conference*, pages 1165–1174.
- Janyce Wiebe et al. 2000. Learning subjective adjectives from corpora. *Aaai/iaai*, 20(0):0.
- Janyce Wiebe, Rebecca Bruce, and Thomas P O’Hara. 1999. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics*, pages 246–253.
- Min Yang, Wei Zhao, Jianbo Ye, Zeyang Lei, Zhou Zhao, and Soufei Zhang. 2018. Investigating capsule networks with dynamic routing for text classification. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 3110–3119.
- Wei Zhao, Haiyun Peng, Steffen Eger, Erik Cambria, and Min Yang. 2019. Towards scalable and reliable capsule networks for challenging nlp applications. *arXiv preprint arXiv:1906.02829*.
- 帕丽旦, 杨文忠, 吾守尔, 斯拉木, et al. 2019. 基于birnn 的维吾尔语情感韵律短语注意力模型. *电子科技大学学报*, 48(1):88–95.
- 王树恒, 吐尔根, 艾山, 吾买尔, et al. 2017. 基于blstm 的维吾尔语文本情感分析. *计算机工程与设计*, 38(10):2879–2886.