·人工智能与模式识别·　　文章编号：1000-3428(2022)03-0069-05　　文献标志码：A　　中图分类号：TP391

融合笔画特征的胶囊网络文本分类

李冉冉，刘大明，刘　正，常高祥

（上海电力大学 计算机科学与技术学院，上海 200090）

摘　要：目前多数文本分类方法无法有效反映句子中不同单词的重要程度，且在神经网络训练过程中获得的词向量忽略了汉字本身的结构信息。构建一种GRU-ATT-Capsule混合模型，并结合CW2Vec模型训练中文词向量。对文本数据进行预处理，使用传统的词向量方法训练的词向量作为模型的第1种输入，通过CW2Vec模型训练得到的包含汉字笔画特征的中文词向量作为第2种输入，完成文本表示。利用门控循环单元分别提取2种不同输入的上下文特征并结合注意力机制学习文本中单词的重要性，将2种不同输入提取出的上下文特征进行融合，通过胶囊网络学习文本局部与全局之间的关系特征实现文本分类。在搜狗新闻数据集上的实验结果表明，GRU-ATT-Capsule混合模型相比TextCNN、BiGRU-ATT模型在测试集分类准确率上分别提高2.35和4.70个百分点，融合笔画特征的双通道输入混合模型相比单通道输入混合模型在测试集分类准确率上提高0.45个百分点，证明了GRU-ATT-Capsule混合模型能有效提取包括汉字结构在内的更多文本特征，提升文本分类效果。

关键词：词向量；笔画特征；门控循环单元；注意力机制；胶囊网络；文本分类

开放科学（资源服务）标志码（OSID）：

中文引用格式：李冉冉，刘大明，刘正，等.融合笔画特征的胶囊网络文本分类[J].计算机工程，2022，48(3)：69-73，80.

英文引用格式：LI R R，LIU D M，LIU Z，et al.Text classification using capsule network integrating stroke features[J].Computer Engineering，2022，48(3)：69-73，80.

Text Classification Using Capsule Network Integrating Stroke Features

LI Ranran，LIU Daming，LIU Zheng，CHANG Gaoxiang

（School of Computer Science and Technology，Shanghai University of Electric Power，Shanghai 200090，China）

【Abstract】Most current text classification methods have the problem that they cannot effectively reflect the importance of different words in a sentence，and the word vectors obtained in the neural network training process ignore the structural information of the Chinese characters.A GRU-ATT-Capsule hybrid model is proposed and combined with the CW2Vec model to train Chinese word vectors.First，the text data are preprocessed，and the word vector trained by the traditional word vector method is used as the first input of the model.The Chinese word vectors containing the stroke features of Chinese characters obtained by the CW2Vec model training are used as the second input to represent the text.Second，the contextual features of two different inputs are extracted through the Gated Recurrent Unit（GRU），and the attention mechanism is used to learn the importance of words in the text.The contextual features extracted from the two different inputs are fused，allowing the capsule network to learn the characteristics of the relationship between the local and global features of the text for text classification.The experimental results on the Sogou news dataset show that the GRU-ATT-Capsule hybrid model improves the classification accuracy of the test set by 2.35 and 4.70 percentage points，respectively，compared with the TextCNN and BiGRU-ATT models，and the dual-channel input hybrid model fused with stroke features compared with the single-channel input mixture model，the classification accuracy of the test set is improved by 0.45 percentage points，which proves that the GRU-ATT-Capsule hybrid model can effectively extract more text features，including Chinese character structure，and improve the text classification effect.

【Key words】word vector；stroke feature；Gated Recurrent Unit（GRU）；attention mechanism；capsule network；text classification

DOI：10.19678/j.issn.1000-3428.0060560

基金项目：甘肃省自然科学基金（SKLLDJ032016021）。

作者简介：李冉冉（1996—），男，硕士研究生，主研方向为深度学习、自然语言处理；刘大明，副教授、博士；刘　正、常高祥，硕士研究生。

收稿日期：2021-01-12　　**修回日期：**2021-03-09　**E-mail：**921090271@qq.com

0 概述

文本分类是自然语言处理领域的一项基础任务，在问答系统、主题分类、信息检索、情感分析、搜索引擎等方面发挥重要的作用。传统文本分类方法大部分在向量空间模型的基础上实现，文献[1]提出一种基于TF-IDF的改进文本分类方法，文献[2]提出使用支持向量机的分类方法，文献[3]提出一种特征加权的贝叶斯文本分类方法。但是，在使用传统特征提取方法表示短文本时存在数据稀疏与特征向量维度过高的问题。

随着深度学习技术的快速发展，越来越多的学者们将深度学习模型应用到自然语言处理中[4]。文献[5]通过构建神经网络语言模型训练词向量来获得文本的分布式特征。但是，其将词视为不可分割的原子表征，忽略了词的结构信息，这一不足引起了部分学者的注意，并对利用词的结构信息的方法进行了研究。文献[6-7]在学习过程中使用部首词典提取字词特征。文献[8]提出一种使用笔画*n*-gram提取汉字的笔画序列信息的方法，该方法可以训练得到含有汉字结构信息的词向量。深度学习技术在文本分类中也得到广泛应用，文献[9]利用卷积神经网络和预先训练的词向量对输入向量进行初始化，文献[10]通过使用循环神经网络（Recurrent Neural Network，RNN）来提取全局特征，文献[11]提出使用门控循环单元（Gated Recurrent Unit，GRU）和注意力机制进行文本分类，文献[12]提出胶囊网络并将其应用在文本分类任务中，减少了卷积神经网络（Convolutional Neural Network，CNN）在池化操作中的特征丢失问题。

本文提出一种结合GRU、注意力机制和胶囊网络的混合神经网络模型GRU-ATT-Capsule。使用GRU提取2种词向量的上下文特征，结合注意力机制对GRU提取的特征重新进行权重分配，对文本内容贡献较大的信息赋予较高的权重，并将重新计算得到的特征进行融合。利用胶囊网络进行局部特征提取，从而解决CNN在池化过程中的特征丢失问题。

1 相关工作

1.1 中文词向量表示方式

传统词向量表示方式主要基于分布式假设，即上下文相似的词应具有相似的语义[13]。Word2Vec和Glove是2种具有代表性的词向量训练方式，由于高效性与有效性得到广泛应用[5,14]。

汉语词向量表示方式的研究在早期多数是参照英文词向量表示方式：先对中文进行分词，再在基于英文的方法上进行大规模语料的无监督训练。但是，该方式忽略了汉字本身的结构信息[15]。因此，汉语学者们针对用于汉语的词向量进行研究。文献[16]设计汉字和词向量联合学习模型，该模型利用汉字层次结构信息改进中文词向量的质量。文献[7]在学习过程中使用部首词典提取字词特征，实验结果表明在文本分类以及词相似度上均有所提升。文献[17]使用卷积自动编码器直接从图像中提取字符特征，以减少与生词相关的数据稀疏问题，实验结果显示利用字符结构特征进行增强会取得一定的效果。文献[8]首先对汉字的笔画特征进行建模，将词语分解为笔画序列，然后使用笔画*n*-gram进行特征提取，在文本分类任务上相比于传统方法提升了1.9%的准确率。

1.2 基于胶囊网络的文本分类

文献[18]提出胶囊网络，其思想是使用胶囊来代替CNN中的神经元，使模型可以学习对象之间的姿态信息与空间位置关系。文献[19]提出胶囊间的动态路由算法与胶囊网络结构，其在MNIST数据集上达到了当时最先进的性能。

胶囊网络与CNN的区别在于：1）使用胶囊（向量）代替神经元（标量）；2）使用动态路由算法进行低层到高层的参数更新，而不是使用池化操作，从而避免信息丢失；3）使用压缩函数代替ReLU激活函数。由于高层胶囊是由底层胶囊通过动态路由算法利用压缩函数计算得出，由多个向量神经元共同决定与整体的关系，因此使得胶囊网络可以学习到文本的局部与整体之间的关联信息。

文献[12]使用两个并行的卷积层进行特征提取，并且其中一个卷积层使用ELU激活函数进行激活，然后将提取的特征按位置对应相乘，输入到胶囊网络，在精度和训练时间上均取得了较好的效果。

1.3 GRU和注意力机制

文献[9]将CNN应用于文本分类，其使用预先训练的词向量来初始化输入向量，首先在卷积层中利用多个不同尺度的卷积核进行特征提取，然后采用池化操作提取主要特征，最后使用softmax分类器进行分类。文献[11]使用门控循环单元学习文本序列的上下文特征，并结合注意力机制为文本中的特征学习权重分布，相比基于LSTM的方法在准确度上取得了2.2%的提升。文献[20]提出一种结合CNN、双向门控循环单元（Bidirectional Gated Recurrent Unit，BiGRU）和注意力机制的文本分类方法，其首先使用CNN提取局部短语特征，然后使用BiGRU学习上下文特征，最后通过增加注意力机制对隐藏状态进行加权计算以完成特征权重的重新分配，相比基于attention-CNN的方法与基于attention-LSTM的方法在准确率与F1值上均有所提升。

2 模型设计

本文提出一种GRU-ATT-Capsule混合模型用于文本分类。该模型由文本表示、全局特征提取和胶囊网络分类等3个模块组成，结构如图1所示。

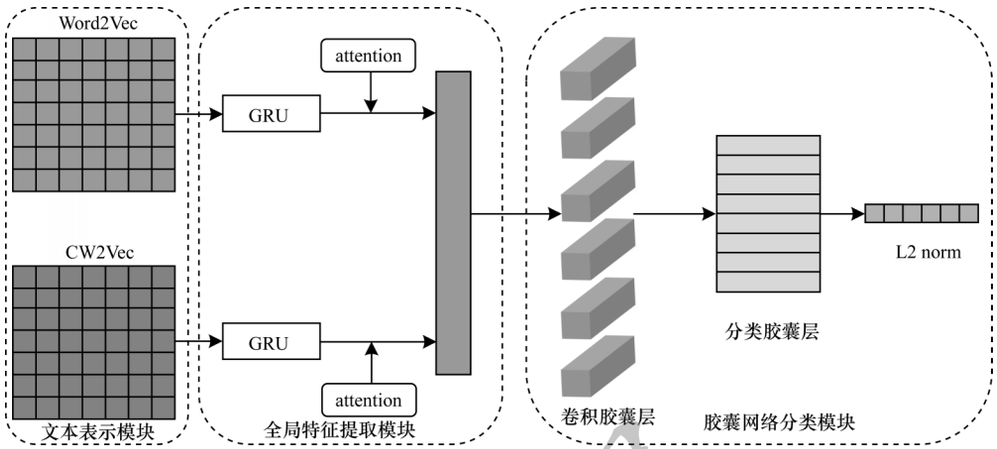

图1　GRU-ATT-Capsule混合模型结构

Fig.1　Structure of GRU-ATT-Capsule hybrid model

2.1　文本表示模块

将文本序列表示为计算机能够处理的形式,表示方法主要包括离散表示和分布式表示。离散表示将文本中的每个词用向量表示,向量的维度就是词表的大小,该方式构建简单,但是不能表示词与词之间的关系,忽略了词的上下文语义。分布式表示考虑了词的上下文语义信息并刻画了词与词之间的语义相似度,并且维度较低,主要包括Word2Vec、Glove等常用的词向量表示。由于汉语的文字系统与英语为代表的字母语言存在不同,因此为了融合汉字结构特征,采用双通道输入策略,利用基于文献[8]的笔画*n*-gram方法来提取每个字对应的笔画向量,然后使用Skip-gram模型进行词向量训练获得CW2Vec。

假定文本序列包含n个词,将其One-hot表示为$\boldsymbol{x}_i=\left(x_1,x_2,\cdots,x_n\right),i\in(0,n)$作为输入。首先,利用训练好的CW2Vec模型获取的词向量作为词嵌入层a的权重$\boldsymbol{W}_1^{p\times V}$,并随机初始化词嵌入层$b$的权重为$\boldsymbol{W}_2^{p\times V}$,将输入$\boldsymbol{x}_i$分别与输入权重$\boldsymbol{W}_1$、$\boldsymbol{W}_2$进行运算,具体过程如式(1)所示:

$$\boldsymbol{W}_i=\boldsymbol{x}_i\boldsymbol{W}_k^{p\times V} \tag{1}$$

然后,得到词向量矩阵$\boldsymbol{W}_a^{n\times V}=(\boldsymbol{W}_1^1,\boldsymbol{W}_2^1,\cdots,\boldsymbol{W}_i^1)$和$\boldsymbol{W}_b^{n\times V}=(\boldsymbol{W}_1^2,\boldsymbol{W}_2^2,\cdots,\boldsymbol{W}_i^2)$,并将其作为后续模型的输入。

2.2　全局特征提取模块

2.2.1　GRU

循环神经网络是一种专门用于处理序列数据的神经网络,通过网络中的循环结构记录数据的历史信息,即当前隐藏层单元的输出不仅与当前时刻的输入有关,还依赖上一时刻的输出。RNN还可以处理变长序列。本文采用门控循环单元提取文本特征,它是LSTM的一种变体,在结构上与LSTM相似,不同点在于GRU简化了LSTM的门控结构。LSTM网络由输入门、输出门、遗忘门和记忆单元来实现历史信息的更新和保留。GRU将LSTM中的输入门和遗忘门合并,称为更新门$\boldsymbol{z}_t$。此外,重置门为$\boldsymbol{r}_t$,记忆单元为$\boldsymbol{h}_t$。更新门控制当前时间步的信息中保留多少先前时刻的历史信息。重置门决定了当前时刻的输入与前一时间步的信息之间的依赖程度。这两个门控向量决定了哪些信息最终能作为GRU的输出并被保存,GRU结构如图2所示。

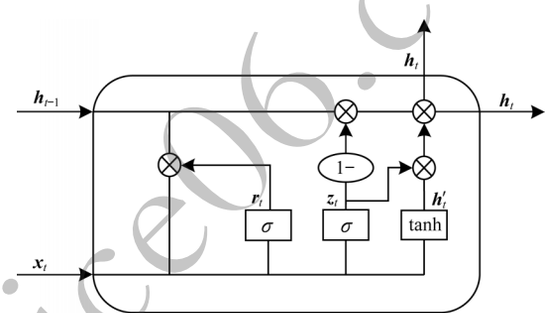

图2　GRU结构

Fig.2　Structure of GRU

GRU单元在时刻t的更新过程如式(2)~式(5)所示:

$$\boldsymbol{r}_t=\sigma\left(\boldsymbol{W}_r\boldsymbol{x}_t+\boldsymbol{U}_r\boldsymbol{h}_{t-1}+\boldsymbol{b}_r\right) \tag{2}$$

$$\boldsymbol{z}_t=\sigma\left(\boldsymbol{W}_z\boldsymbol{x}_t+\boldsymbol{U}_z\boldsymbol{h}_{t-1}+\boldsymbol{b}_z\right) \tag{3}$$

$$\boldsymbol{h}'_t=\tanh\left(\boldsymbol{W}_c\boldsymbol{x}_t+\boldsymbol{U}_c\left(\boldsymbol{r}_t\odot\boldsymbol{h}_{t-1}\right)+\boldsymbol{b}_c\right) \tag{4}$$

$$\boldsymbol{h}_t=\boldsymbol{z}_t\odot\boldsymbol{h}'_t+\left(1-\boldsymbol{z}_t\right)\odot\boldsymbol{h}_{t-1} \tag{5}$$

其中:$\boldsymbol{x}_t$为当前时刻的输入;$\boldsymbol{h}_{t-1}$为前一时刻GRU的输出;$\boldsymbol{h}_t$为当前时刻GRU的输出;σ是sigmoid激活函数;$\boldsymbol{W}_r$、$\boldsymbol{W}_z$、$\boldsymbol{W}_c$、$\boldsymbol{U}_r$、$\boldsymbol{U}_z$、$\boldsymbol{U}_c$为权重矩阵;$\boldsymbol{b}_r$、$\boldsymbol{b}_z$、$\boldsymbol{b}_c$为偏置项;$\odot$表示向量间的点乘运算。

通过文本表示模块,长度为n的语句$\boldsymbol{S}$可分别表示如下:

$$\boldsymbol{S}_a=\left(\boldsymbol{z}_1^1,\boldsymbol{z}_2^1,\cdots,\boldsymbol{z}_i^1\right),\boldsymbol{z}_i^1\in\boldsymbol{W}_a^{n\times V}$$

$$\boldsymbol{S}_b=\left(\boldsymbol{z}_1^2,\boldsymbol{z}_2^2,\cdots,\boldsymbol{z}_i^2\right),\boldsymbol{z}_i^2\in\boldsymbol{W}_b^{n\times V},i\in n$$

其中:$\boldsymbol{z}_i^1$、$\boldsymbol{z}_i^2$分别表示语句$\boldsymbol{S}$中基于CW2Vec和Word2Vec的词向量。

将文本表示$\boldsymbol{S}_a$、$\boldsymbol{S}_b$输入GRU中获得上下文特征

$\boldsymbol{h}_t^1$、$\boldsymbol{h}_t^2$用于后续的注意力计算。

2.2.2 注意力机制

注意力机制是一种权重分配机制，对于重要的语义信息分配较多的注意力。在文本分类任务中，句子中不同的词对分类效果的影响是不同的，本文引入注意力机制识别文本中的重要信息。将GRU的隐藏层输出$\boldsymbol{h}_t$作为输入，得到：

$$\boldsymbol{u}_i=\tanh\left(\boldsymbol{W}_a\boldsymbol{h}_t+\boldsymbol{b}_a\right) \tag{6}$$

$$\boldsymbol{a}_i=\text{softmax}\left(\boldsymbol{u}_i\right) \tag{7}$$

$$\boldsymbol{v}_i=\sum_i \boldsymbol{a}_i\boldsymbol{h}_i \tag{8}$$

其中：$\boldsymbol{W}_a$为权重矩阵；$\boldsymbol{b}_a$为偏置；$\boldsymbol{u}_i$为$\boldsymbol{h}_i$的隐层表示；权重$\boldsymbol{a}_i$为经过softmax函数得到的每个词的标准化权重，然后通过加权计算得到注意力机制的输出向量$\boldsymbol{v}_i$。

将GRU的输出$\boldsymbol{h}_t^1$、$\boldsymbol{h}_t^2$首先使用注意力机制重新分配词的权重，获得输出向量$\boldsymbol{v}_i^1$、$\boldsymbol{v}_i^2$，然后将其相加获得融合汉字结构特征的上下文特征$\boldsymbol{v}_i'$作为胶囊网络的输入。

2.3 胶囊网络分类模块

CNN的池化操作会丢失大量的特征信息使得文本特征大幅减少，其中神经元为标量。在胶囊网络中使用胶囊（即向量）来代替CNN中的标量神经元。胶囊网络的输入输出向量表示为特定实体类别的属性，通过使用动态路由算法在训练过程中迭代调整低层胶囊与高层胶囊的权重c_{ij}，低层胶囊根据多次迭代后的权重通过压缩函数共同来决定某个高层胶囊表示，并将通过动态路由得到的所有高层胶囊进行拼接得到最终的高阶胶囊表示。

卷积胶囊层的输入为融合层的输出$\boldsymbol{v}_i'$，输出$\boldsymbol{u}_i'$作为分类胶囊层的输入。动态路由通过多次迭代路由来调整耦合系数c_{ij}确定合适的权重：

$$c_{ij}=\frac{\exp\left(b_{ij}\right)}{\sum_k \exp\left(b_{ik}\right)} \tag{9}$$

其中：i表示底层；j表示高层。

路由算法利用式（10）中的变换矩阵$\boldsymbol{W}_{ij}$将低层胶囊$\boldsymbol{u}_i'$进行转换用于获得预测向量$\hat{\boldsymbol{u}}_{j|i}$，通过式（11）获得来自底层胶囊的预测向量$\hat{\boldsymbol{u}}_{j|i}$的加权和$\boldsymbol{s}_j$。为使胶囊的模长表示对应类别的分类概率，通过压缩函数式（12）进行归一化得到高层胶囊$\boldsymbol{v}_j$。利用式（13）计算高层胶囊$\boldsymbol{v}_j$与底层预测向量$\hat{\boldsymbol{u}}_{j|i}$的点积，当低层预测向量$\hat{\boldsymbol{u}}_{j|i}$与输出胶囊$\boldsymbol{v}_j$方向趋向一致时，增加对应的耦合系数$c_{ij}$。通过多次迭代路由算法对耦合系数进行调整得到修正后的高层胶囊v_j。

$$\hat{\boldsymbol{u}}_{j|i}=\boldsymbol{W}_{ij}\boldsymbol{u}_i' \tag{10}$$

$$\boldsymbol{s}_j=\sum_i c_{ij}\hat{\boldsymbol{u}}_{j|i} \tag{11}$$

$$\boldsymbol{v}_j=\frac{\left\|\boldsymbol{s}_j\right\|^2}{1+\left\|\boldsymbol{s}_j\right\|^2}\cdot\frac{\boldsymbol{s}_j}{\left\|\boldsymbol{s}_j\right\|} \tag{12}$$

$$b_{ij}=b_{ij}+\hat{\boldsymbol{u}}_{j|i}\cdot\boldsymbol{v}_i \tag{13}$$

3 实验与结果分析

3.1 实验环境

为验证GRU-ATT-Capsule混合模型在文本分类中的效果，采用基于TensorFlow的Keras进行开发，编程语言为Python3.6。数据集被随机打乱。服务器配置如下：CPU为Intel Core i5-8500主频3.0 GHz，内存32 GB，硬盘1 TB，操作系统为Ubuntu16.04，GPU为GeForce RTX 2080 SUPER。

3.2 实验数据

为评估GRU-ATT-Capsule混合模型的有效性，采用搜狗实验室的中文新闻数据集，该数据集包含419 715条数据，可被标出类别的共有14类，去掉4类样本数不足的数据，保留其中的10类作为分类文本分解。每类数据选择2 000条文本，训练集、验证集和测试集的划分比例为8∶1∶1。数据集分布如表1所示。

表1 数据集样本类别分布

Table 1 Dataset sample category distribution

序号	类别	总样本数	训练集样本数	验证集样本数	测试集样本数
1	汽车	20 000	16 000	2 000	2 000
2	财经	20 000	16 000	2 000	2 000
3	IT	20 000	16 000	2 000	2 000
4	健康	20 000	16 000	2 000	2 000
5	旅游	20 000	16 000	2 000	2 000
6	教育	20 000	16 000	2 000	2 000
7	军事	20 000	16 000	2 000	2 000
8	文化	20 000	16 000	2 000	2 000
9	娱乐	20 000	16 000	2 000	2 000
10	体育	20 000	16 000	2 000	2 000

实验的预处理部分首先将文本数据进行转码，然后使用Jieba分词工具对每条数据进行分词，再加上对应的标签。

3.3 评价指标与实验设置

实验采用准确率作为评价标准，准确率计算公式如下：

$$A_{\text{Acc}}=\frac{T_{\text{TP}}+T_{\text{TN}}}{T_{\text{TP}}+F_{\text{FN}}+F_{\text{FP}}+T_{\text{TN}}} \tag{14}$$

其中：T_{TP}为真正例；T_{TN}真负例；F_{FP}为假正例；F_{FN}为假负例。

使用搜狗语料SogouCA训练CW2Vec词向量作为第1个通道的词向量。窗口n设置为5，词向量维度设置为200，最低词频设为10，用于过滤词频较低的词。在训练过程中，使用Adam优化器，将学习率设置为0.001，并且在每个epoch将学习率衰减0.99来降低学习率。GRU单元数量设置为64。底层胶囊数量设置为6，文献[19]使用了1 152个胶囊进行图像分类，之所以进行这种操作，是因为该方法生成

的特征图包含的有效信息不足。高层胶囊数量设置为10，对应不同类别，胶囊的模长代表所属类别的概率。为验证GRU-ATT-Capsule混合模型的有效性，将TextCNN模型[9]和BiGRU-ATT模型[11]作为基线进行对比分析。

3.4　结果分析

从训练集、验证集和测试集出发比较TextCNN、BiGRU-ATT和GRU-ATT-Capsule混合模型的文本分类效果，实验结果如表2所示。

表2　3种神经网络模型的分类准确率对比

Table 2　Comparison of classification accuracy of three neural network models　%

数据集	TextCNN	BiGRU-ATT	GRU-ATT-Capsule
训练集	98.44	99.22	99.63
验证集	86.15	82.55	89.25
测试集	85.20	82.80	87.55

3种模型的epoch设置为40，在训练集上的准确率均达到98%以上，其中基于GRU-ATT-Capsule混合模型的文本分类准确率最高，达到99.63%。在验证集上的分类准确率均达到82%以上，仍是基于GRU-ATT-Capsule混合模型的文本分类准确率最高，达到89.25%。对于测试集，3种模型在测试集上的分类准确率明显低于训练集和验证集，但是本文GRU-ATT-Capsule混合模型的准确率仍为最高，达到87.55%。由此可见，在文本分类任务中，本文提出的GRU-ATT-Capsule混合模型的分类效果要优于TextCNN模型和BiGRU-ATT模型。

由于卷积神经网络的池化操作会导致特征信息丢失，因此采用动态路由机制的胶囊网络用于获得合适的权重，将低层胶囊的特征信息传递到高层胶囊中，从而更好地捕捉文本序列中的特征信息。为进一步证明本文融合的笔画特征对文本分类的有效性，设置模型对比实验进行分类性能验证。将GRU-ATT-Capsule混合模型的输入分别设置如下：

1）Single：采用单通道输入，使用随机初始化的Word2Vec词向量作为输入，在训练过程中通过模型的反向传播进行不断调整。

2）Dual：采用双通道输入，一个通道采用包含笔画特征的CW2Vec词向量作为输入，另一个通道采用随机初始化的Word2Vec词向量作为输入，在训练过程中可以通过模型的反向传播进行不断调整。

2种输入方式下的GRU-ATT-Capsule混合模型分类准确率对比结果如表3所示。由表3可知，本文提出的融合笔画特征的GRU-ATT-Capsule混合模型在验证集和测试集上的分类准确率相比基于单通道输入的GRU-ATT-Capsule混合模型分别提高了0.15和0.45个百分点，由此说明本文所提出的融合笔画特征的GRU-ATT-Capsule混合模型相比基于单通道输入的GRU-ATT-Capsule混合模型能够提取包括汉字结构在内的更多的文本特征，从而提高文本分类效果。

表3　2种输入方式下的GRU-ATT-Capsule混合模型分类准确率对比

Table 3　Comparison of classification accuracy of GRU-ATT-Capsule hybrid model with two input modes　%

数据集	Single	Dual
训练集	99.67	99.63
验证集	89.10	89.25
测试集	87.10	87.55

4　结束语

本文提出一种融合笔画特征的胶囊网络文本分类方法。通过2种不同的词向量来表示文本内容，补充文本的汉字结构特征。使用GRU模型提取全局文本特征，在GRU后引入注意力机制，并利用注意力机制对词的权重进行调整进一步提取文本序列中的关键信息。融合2种表示结果作为胶囊网络的输入，通过胶囊网络解决了卷积神经网络池化操作的信息丢失问题。实验结果证明了GRU-ATT-Capsule混合模型的有效性。后续将对GRU-ATT-Capsule混合模型的词向量融合方式进行改进，进一步提高文本分类准确率。

参考文献

[1] 叶雪梅，毛雪岷，夏锦春，等. 文本分类TF-IDF算法的改进研究[J]. 计算机工程与应用，2019，55(2)：104-109，161.
YE X M，MAO X M，XIA J C，et al. Improved approach to TF-IDF algorithm in text classification[J]. Computer Engineering and Applications，2019，55(2)：104-109，161.（in Chinese）

[2] WANG Z Q，SUN X，ZHANG D X，et al. An optimal SVM-based text classification algorithm[C]//Proceedings of 2006 International Conference on Machine Learning and Cybernetics. Washington D. C.，USA：IEEE Press，2006：1378-1381.

[3] ZHANG L G，JIANG L X，LI C Q，et al. Two feature weighting approaches for naive Bayes text classifiers[J]. Knowledge-Based Systems，2016，100：137-144.

[4] OTTER D W，MEDINA J R，KALITA J K. A survey of the usages of deep learning for natural language processing[J]. IEEE Transactions on Neural Networks and Learning Systems，2021，32(2)：604-624.

[5] MIKOLOV T，CHEN K，CORRADO G，et al. Efficient estimation of word representations in vector space[EB/OL]. [2020-12-15]. https://arxiv. org/pdf/1301. 3781. pdf.

[6] SUN Y M，LIN L，YANG N，et al. Radical-enhanced Chinese character embedding[C]//Proceedings of 2014 International Conference on Neural Information Processing. Berlin，Germany：Springer，2014：279-286.

[7] 武娇，洪彩凤，顾永春，等. 基于类邻域字典的线性回归文本分类[J]. 计算机工程，2021，47(8)：93-99，108.
WU J，HONG C F，GU Y C，et al. Linear regression text classification based on class-wise nearest neighbor dictionary[J]. Computer Engineering，2021，47(8)：93-99，108.（in Chinese）

（下转第80页）

（上接第73页）

[8] CAO S, LU W, ZHOU J, et al. CW2Vec: learning Chinese word embeddings with stroke n-gram information[EB/OL]. [2020-12-15]. https://ojs. aaai. org/index. php/AAAI/article/view/12029.

[9] KIM Y. Convolutional neural networks for sentence classification[EB/OL]. [2020-12-15]. https://arxiv. org/abs/1408. 5882.

[10] TIAN Z X, RONG W G, SHI L B, et al. Attention aware bidirectional gated recurrent unit based framework for sentiment analysis [C]//Proceedings of International Conference on Knowledge Science, Engineering and Management. Berlin, Germany: Springer, 2018: 67-68.

[11] DU C S, HUANG L. Text classification research with attention-based recurrent neural networks[J]. International Journal of Computers Communications & Control, 2018, 13(1): 50-60.

[12] KIM J, JANG S, PARK E, et al. Text classification using capsules[J]. Neurocomputing, 2020, 376: 214-221.

[13] HARRIS Z S. Distributional structure [M]. Berlin, Germany: Springer, 1970.

[14] PENNINGTON J, SOCHER R, MANNING C. Glove: global vectors for word representation[C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2014: 1532-1534.

[15] 赵浩新，俞敬松，林杰. 基于笔画中文字向量模型设计与研究[J]. 中文信息学报，2019，33(5)：17-23.
ZHAO H X, YU J S, LIN J. Design and research on Chinese word embedding model based on strokes[J]. Journal of Chinese Information Processing, 2019, 33(5): 17-23. (in Chinese)

[16] CHEN X, XU L, LIU Z, et al. Joint learning of character and word embeddings[C]//Proceedings of International Joint Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2015: 1236-1242.

[17] SU T R, LEE H Y. Learning Chinese word representations from glyphs of characters[EB/OL]. [2020-12-15]. https://arxiv. org/abs/1708. 04755.

[18] HINTON G E, KRIZHEVSKY A, WANG S D. Transforming auto-encoders[C]//Proceedings of International Conference on Artificial Neural Networks. Berlin, Germany: Springer, 2011: 44-51.

[19] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[EB/OL]. [2020-12-15]. https://arxiv. org/abs/1710. 09829.

[20] 王丽亚，刘昌辉，蔡敦波，等. CNN-BiGRU网络中引入注意力机制的中文文本情感分析[J]. 计算机应用，2019，39(10)：2841-2846.
WANG L Y, LIU C H, CAI D B, et al. Chinese text sentiment analysis based on CNN-BiGRU network with attention mechanism[J]. Journal of Computer Applications, 2019, 39(10): 2841-2846. (in Chinese)

编辑 陆燕菲