

基于静态路由分组胶囊网络的文本分类模型

朱海景, 余 谅, 盛钟松, 陈贵强, 王 争  
(四川大学计算机学院, 成都 610065)

**摘 要:** 近年来,随着胶囊网络的广泛研究,其在图像、语言等领域取得了重大进展.但胶囊网络存在参数多、训练时间长的缺点.分组反馈路由机制是一种称为分组路由的监督路由策略,该策略将胶囊平均地分成若干组,胶囊局部共享转换权重,从而减少路由参数和计算复杂度,在图像分类领域取得较好效果.本文将胶囊分组方法运用于文本分类任务中,再引入胶囊压缩、静态路由机制,提出了一种新的文本分类模型 CapsNet-GSR.该模型通过胶囊分组在提取文本局部信息的同时减少参数,利用胶囊压缩和静态路由机制,进一步提高胶囊质量、降低参数数量.在 20 news 文本分类数据集上的实验证明,其在参数数量和训练时间上有明显减少.在 AG's news、TREC 和 20 news 数据集上的实验表明,该模型在准确率上也有所提高.

**关键词:** 文本分类; 胶囊网络; 路由机制

**中图分类号:** TP391      **文献标识码:** A      **DOI:** 10.19907/j.0490-6756.2021.062001

Text classification model based on static routing and grouped capsule network

ZHU Hai-Jing, YU Liang, SHENG Zhong-Song, CHEN Gui-Qiang, WANG Zheng  
(College of Computer Science, Sichuan University, Chengdu 610065, China)

**Abstract:** The use of capsule network in the fields of image and language have progressed tremendously over the past few years, with the extensive research of the capsule network. However, the capsule network has the disadvantages of too many parameters and long training time. The group feedback routing mechanism is a supervised routing strategy called group-routing. The strategy divides the capsules into several groups evenly, and the capsules locally share the conversion weights, thereby reducing routing parameters and computational complexity, and achieving better results in image classification. In this paper, a new text classification model CapsNet-GSR(CapsNet-Grouped capsule based on Static Routing) is proposed based on the capsule grouping method, and the capsule compression and static routing mechanism are also introduced. The model uses capsule grouping to extract local information of the text while reducing parameters. In addition, it uses capsule compression and static routing mechanisms to further improve the quality of the capsule and reduce the number of parameters. The experiment on the 20 news text classification dataset proves that the number of parameters and training time of the proposed model decrease obviously. The experiment results on AG's news, TREC, and 20 news datasets show that the accuracy of the proposed model is also improved.

**Keywords:** Text classification; Capsule network; Routing mechanism

收稿日期: 2021-03-31  
基金项目: 国家自然科学基金(61806135)  
作者简介: 朱海景(1988—), 男, 江西南康人, 硕士研究生, CCF 会员, 研究方向为自然语言处理. E-mail: zhuhaifengjing@163.com  
通讯作者: 余谅. E-mail: yuliang@scu.edu.cn

## 1 引言

文本分类是自然语言处理中至关重要的一步,其目的是利用非结构化的文本数据来归纳信息<sup>[1]</sup>.早期的文本分类方法是通过人工设计特征,然后用机器学习方法进行分类.但随着深度学习的深入研究,神经网络等被大量用于文本分类,并且能够自动获取文本特征,分类效果更好.

胶囊网络是一种新兴的神经网络结构,该结构的特点是能在小数据集中提升分类准确率.胶囊网络模型是 2017 年由 Sabour 等<sup>[2]</sup>在 NIPS 上首次提出的神经网络模型,类似于卷积神经网络,在图像领域获得了很好的效果. Ding 等<sup>[3]</sup>采用分组反馈胶囊网络,利用分组胶囊获取局部信息并共享权重矩阵以减少参数,用反馈方式的网络模型预测胶囊来提高应对复杂数据环境的适应能力. Zhao 等<sup>[4]</sup>将胶囊网络引入到文本分类领域,也取得了不错的性能.近年来,随着对胶囊网络研究的不断深入,其改进方法也越来越多,但主要集中在两个方面:(1) 是提升胶囊质量;(2) 是改进路由规则减少网络参数<sup>[5]</sup>. 本文通过对胶囊进行分组,以提取更有价值的文本信息,减少冗余信息,达到提高胶囊质量的目的;同时,利用静态路由机制<sup>[6]</sup>,去除动态路由中的耦合系数<sup>[2]</sup>;采用胶囊压缩<sup>[7]</sup>操作,减少了网络参数.为此我们提出了 CapsNet-GSR 文本分类模型.

本文的主要贡献有三个方面:(1) 将分组胶囊网络应用到文本分类中,验证了该方法在文本分类领域的有效性;(2) 采用胶囊压缩和静态路由机制减少网络模型的参数;(3) 通过对比实验,突出本文方法在模型参数和训练时间上具有明显优势.

## 2 相关研究

文本分类是指用计算机对文本数据集(或其他实体)按照一定的分类标准或体系进行自动分类标记的过程<sup>[8-9]</sup>. 文本分类始于 20 世纪 50 年代,利用的是专家规则,但该方法需要对相关领域有深入研究才能写出合适的规则,因此发展受限. 80 年代出现了专家系统,即利用专家掌握的知识构造分类系统<sup>[10]</sup>,比如 CONSTRUE 系统<sup>[11]</sup>. 90 年代有了支持向量机,便产生了基于机器学习的文本分类模型. 2010 年后随着深度学习的广泛运用,基于神经网络的文本分类模型表现突出,并广泛应用于学术界和工业界.

### 2.1 文本分类模型

从文本分类的发展来看,文本分类模型可分为传统的机器学习模型和深度学习模型.

传统的机器学习模型有朴素贝叶斯(NB)<sup>[12]</sup>、k-最近邻(KNN)<sup>[13]</sup>、决策树(DT)<sup>[14]</sup>、支持向量机(SVM)<sup>[15]</sup>等.与早期基于规则的方法相比,此类模型在准确性和稳定性上更甚一筹,但此类模型仍需进行特征选择与提取的过程,且容易丢失文本内的前后位置关系或语境,导致单词语义理解不准确.

深度学习模型是基于神经网络的模型,如卷积神经网络(CNN)<sup>[16-17]</sup>、循环神经网络(RNN)<sup>[18]</sup>、长短期记忆网络(LSTM)<sup>[19]</sup>和胶囊网络(CapsNet)<sup>[4,6-7,20-21]</sup>等.与传统的机器学习模型相比,深度学习模型不需要人为选择和提取特征,即可自动获取有用的语义信息.由此可见,文本分类采用深度神经网络更智能,而且利用深度神经网络,可以挖掘出更多更深层次的文本数据信息,以提高算法的分类准确度、鲁棒性和泛化能力.

### 2.2 基于胶囊网络的文本分类方法

胶囊网络主要由卷积层、初级胶囊层、数字胶囊层、全连接层等组成.胶囊层又由许多胶囊组成,其中每个胶囊又包含多个神经元.每个神经元表示图像中特定实例的属性,如姿态、大小、位置、方向、纹理等.每个胶囊即为一个向量,向量的长度表示实体存在的概率,长度越长,则存在的概率越大.因此胶囊层的胶囊输入是向量,输出也是向量.

胶囊网络中的胶囊包含丰富的空间位置等信息,相邻节点之间具有强相关性,而且能够保留原始数据中的底层细节信息,这些特性正好契合了文本数据中上下文之间存在的联系性和自然顺序性,能够很好地提取单词的语义信息,有利于文本的正确分类.另一方面,胶囊网络的结构较浅,对小数据样本集友好.但胶囊网络自 2017 年被提出动态路由机制后才得到广泛关注,因此将胶囊网络运用在文本分类领域的研究并不多.

Zhao 等<sup>[4]</sup>在动态路由胶囊网络的基础上,提出了一个用于文本分类的胶囊网络.通过在胶囊层中增加共享矩阵和非共享矩阵,提高了胶囊内局部信息的多样性;通过引入孤立类别、修正连接强度等方法减少噪声. Zhao 等<sup>[20]</sup>针对多标签输出数据集,通过引入自适应优化器、胶囊压缩和部分路由来提高胶囊网络的可靠性可拓展性,提出了可靠可扩展的胶囊网络模型. Kim 等<sup>[6]</sup>提出运用一种简单的静态路由方法来降低动态路由的计算复杂度.

Ren 等<sup>[21]</sup>利用组合编码胶囊构造词嵌入来减少词嵌入参数, 并采用 K 均值路由算法来提高网络模型的稳定性和鲁棒性.

### 3 本文模型

本文模型框架如图 1 所示. 模型由卷积层、分组胶囊层、压缩胶囊层、文本胶囊层等组成. 其中压缩胶囊层到文本胶囊层之间采用静态路由方法进行传递.

#### 3.1 卷积层

首先, 我们在嵌入文本上采用卷积过滤器进行卷积操作来提取特征. 模型的输入为  $X \in R^{L \times V}$  的文本句子. 其中,  $L$  为句子的长度;  $V$  为词嵌入大小 (词向量维度).  $X_i \in R^V$  为一个输入句子中第  $i$  个词

的词向量.  $W^a \in R^{k \times V}$  为卷积过滤器. 其中,  $k$  为过滤器的宽度, 本文采用 3 种过滤器  $W^a$  以提取不同的特征, 其宽度分别为 (3, 4, 5). 然后将该过滤器与句子局部区域  $X_{i:i+k-1} \in R^{k \times V}$  进行卷积以生成一个特征.

$$m_i = f(W^a \circ X_{i:i+k-1} + b_0) \quad (1)$$

其中, “ $\circ$ ” 表示为对应元素乘运算;  $b_0$  为偏置项;  $f(\cdot)$  为 ReLU 激活函数. 利用过滤器以步长为 1 滑过整个句子文档后得到所有的  $m_i$  组成一个特征图  $\mathbf{m}(m_1, \dots, m_{L-k+1})$ . 因此, 当有  $A$  个过滤器  $W^a$  ( $a=1, \dots, A$ ) 就可以得到  $A$  个特征图:

$$\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_A] \in R^{(L-k+1) \times A} \quad (2)$$

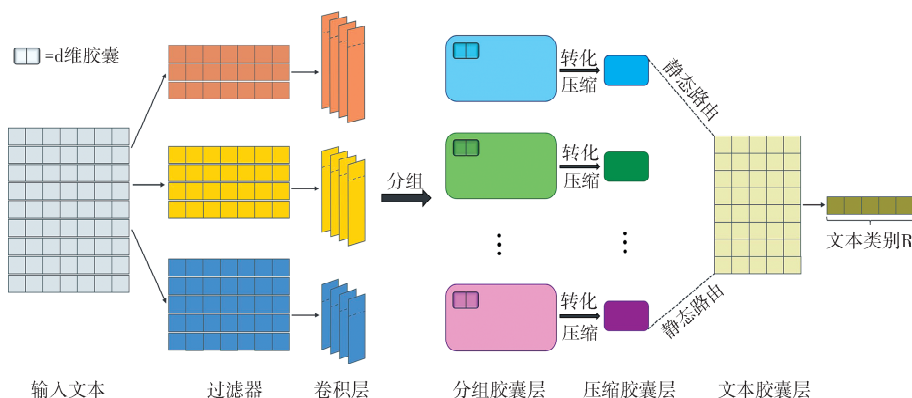


图 1 本文模型示意图  
Fig. 1 Schematic diagram of model

#### 3.2 分组胶囊层

分组胶囊层采用 Ding 等<sup>[3]</sup>提出的胶囊分组方法. 每个胶囊组包含若干个  $d$  维胶囊, 而每个胶囊是由特征图上不同通道相同位置的元素组成, 同一通道的不同位置表示不同的胶囊. 该胶囊示意图如图 2 所示. 我们将胶囊均匀地分为若干组, 具体分组情况将在第 4.2 节中讨论.

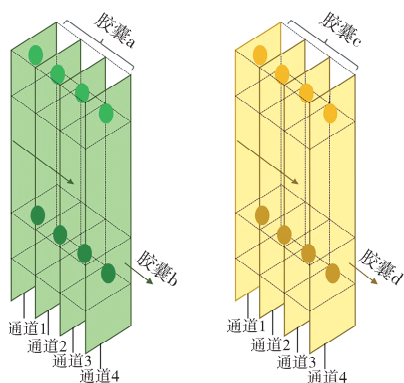


图 2 四维胶囊示意图  
Fig. 2 Schematic diagram of 4-dimensional capsule

胶囊只有与其相关的胶囊之间才有利用价值, 因此可以利用余弦相似度来衡量一对胶囊之间的相似度. 我们希望不同组胶囊间的相似度尽可能小, 同一组的相似度尽可能大, 并通过最小化分组损失<sup>[3]</sup>来达到分组的目的, 具体实现方法如下.

设  $\{G_t\} \ t=1 \sim T$ ,  $t$  表示胶囊组, 每组有  $M$  个胶囊  $\{v_j\} \ j=1 \sim M$ , 各组胶囊  $G_t$  的平均值可定义为

$$\mu_t = \frac{1}{M} \sum_{v \in G_t} v, \quad (t = 1, \dots, T) \quad (3)$$

一对胶囊的相似度定义如下.

$$\sum_t = \sum_{v_1, v_2 \in G_t} \frac{(v_1 - \mu_t) \cdot (v_2 - \mu_t)}{|v_1 - \mu_t| |v_2 - \mu_t|}, \quad (v_1 \neq v_2) \quad (4)$$

其中,  $(v_1 - \mu_t) \cdot (v_2 - \mu_t)$  表示向量  $(v_1 - \mu_t)$  和向量  $(v_2 - \mu_t)$  的内积, 然后可以得到所有胶囊类内相似度为

$$S_\omega = \sum_{t=1}^T \sum_t =$$

$$\sum_{t=1}^T \sum_{v_1, v_2 \in G_t} \frac{(v_1 - \mu_t) \cdot (v_2 - \mu_t)}{|v_1 - \mu_t| |v_2 - \mu_t|}, (v_1 \neq v_2) \quad (5)$$

所有胶囊类间相似度为

$$S_b = \sum_{t_1, t_2=1}^T \frac{(\mu_{t_1} - \mu) \cdot (\mu_{t_2} - \mu)}{|\mu_{t_1} - \mu| |\mu_{t_2} - \mu|}, (t_1 \neq t_2) \quad (6)$$

其中,  $\mu_{t_1}$  表示第  $t_1$  个胶囊组的平均值, 可用式(3)计算得出,  $\mu$  表示所有胶囊的平均值. 最后, 分组损失可定义为

$$L_{\text{Group}} = S_b / S_w \quad (7)$$

通过最小化分组损失, 确保胶囊能够被正确地分成若干个组, 即将同一组的胶囊拉近, 将不同组的胶囊推开.

为了进一步提取胶囊组的局部特征, 需对胶囊组进行整形转换, 然后每个胶囊组内部局部共享相同的转换矩阵. 假设胶囊共分为  $D$  组, 每组有  $m_t$  个  $d$  维胶囊, 首先将每组胶囊整形为  $R^{l_t \times h_t \times d}$ , 其中,  $l_t \times h_t = m_t$ . 然后使用  $d_h$  个 3D 卷积过滤器  $W^b \in R^{c \times c \times d}$  进行转化, 其中,  $c$  为过滤器  $W^b$  的宽度, 进而得到转化胶囊  $\hat{v}_t \in R^{\hat{l}_t \times \hat{h}_t \times d_h}$ . 其中  $\hat{l}_t$  和  $\hat{h}_t$  由下式求得.

$$\hat{l}_t = \frac{l_t - c + 2 \times \text{填充}}{\text{步长}} + 1 \quad (8)$$

$$\hat{h}_t = \frac{h_t - c + 2 \times \text{填充}}{\text{步长}} + 1 \quad (9)$$

最后, 将转化胶囊  $\hat{v}_t$  整形成  $v_t = R^{\hat{m}_t \times d_h}$ , 其中,  $\hat{m}_t = \hat{l}_t \times \hat{h}_t$ . 分组胶囊层通过转化后, 每组各得到  $\hat{m}_t$  个  $d_h$  维转化胶囊  $v_{ij}$ . 其转化过程如图 3 所示.

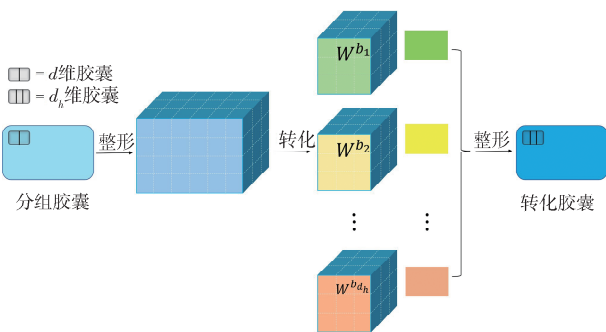


图 3 一个分组胶囊的转化过程

Fig. 3 The transformation process of a grouped capsule

### 3.3 压缩胶囊层

由于胶囊的数量与输入文档的大小成比例的增大, 当输入的文档或句子较长时, 对 GPU 内存有很大的要求. 因此在转化胶囊传递至压缩胶囊层

时, 我们对转化胶囊进行一个压缩操作, 以合并相似胶囊, 去除噪声胶囊. 每个压缩胶囊  $u_q$  通过使用每组所有转化胶囊  $v_{ij}$  的加权求和来计算<sup>[7]</sup>.

$$u_q = \sum_{j=1}^J b_{jq} v_{ij}, q \in [1, Q] \quad (10)$$

其中, 参数  $b_{jq}$  可通过监督学习得到, 其利用均匀分布  $U(-0.01, 0.01)$  采样进行初始化;  $J$  为每组转化胶囊数量;  $Q$  为压缩胶囊数量;  $j, q$  分别为转化胶囊、压缩胶囊索引.

### 3.4 文本胶囊层

对于图像分类领域, 低层实体的空间位置等信息对分类结果有很大的影响, 因此动态路由机制表现出很好的效果. 然而在自然语言处理中, 文本和情感的表达方式有很大的自由, 句子中单词的前后顺序有时对分类结果影响较小. 例如, “This movie is rubbish, I don’t like it!” 与 “I don’t like this movie, it is rubbish!”, 虽然单词的位置发生变化, 但其表达的意思是一样的. 因此从这个角度来看, 运用静态路由机制对分类结果影响不大, 而且能够降低计算复杂度, 其具体实现方法如式(11)和式(12)所示.

$$s_r = \sum_q W_{qr} u_q \quad (11)$$

$$v_r = \frac{\|s_r\|^2}{1 + \|s_r\|^2} \frac{s_r}{\|s_r\|} \quad (12)$$

其中,  $s_r$  为文本胶囊层胶囊  $r$  的所有输入向量;  $W_{qr} = [R \times Q]$  为由压缩胶囊层到文本胶囊层的权重矩阵, 可通过学习得到;  $R$  为文本胶囊的数量.  $v_r$  为文本胶囊层中胶囊  $r$  的输出向量, 式(12)即采用非线性挤压函数<sup>[2]</sup>得到胶囊  $r$  的输出向量. 由此可见, 静态路由取消了动态路由中的耦合系数, 而是直接由压缩胶囊经权重矩阵, 利用非线性挤压函数得到文本胶囊, 此过程大幅缩短训练时间, 减少了计算开销.

文本胶囊层中的胶囊最终被展平成胶囊列表, 对应着文本分类的各个类别. 其中胶囊  $r$  的长度  $\|v_r\|$  表示属于文本类别  $r$  的概率, 当且仅当类别标签  $r$  的文本被馈送到网络中时, 类别标签  $r$  的胶囊  $v_r$  具有最大有效长度. 本文采用边际损失<sup>[2]</sup>作为分类损失函数, 边际损失  $L_r$  定义如式(13)所示. 其中, 文本的类别数为  $R$ , 如果文本的类别标签为  $r$ , 则  $T_r = 1, m^+ = 0.9$  表示胶囊的长度  $\|v_r\|$  至少为 0.9 才能被激活,  $m^- = 0.1$  表示如果类别标签不是  $r$ , 则胶囊的长度  $\|v_r\|$  应小于 0.1,  $\lambda$  是设置为 0.5 的超参数.



$$L_r = \sum_{r=1}^R (T_r \max (0, m^+ - \|v_r\|)^2 + \lambda (1 - T_r) \max (0, \|v_r\| - m^-)^2) \tag{13}$$

4 实验结果与分析

4.1 实验环境与评价指标

本文在 5 个文本分类数据集上进行了对比实验,囊括了新闻分类、评论分类、问题分类、观点分类等文本分类任务,其简要介绍如表 1 所示.表中“ $T_r$ ”为训练集样本数,“ $T_s$ ”为测试集样本数,“ $Cl_s$ ”为数据集类别数,“ $L_{avg}$ ”为句子平均长度.

本实验均由 PyTorch 实现,在 Ubuntu 16.04 LTS 操作系统上,Python3.6 环境下,使用 NVIDIA GeForce RTX 2070 显卡,CPU 为 AMD Ryzen5 2600X 处理器,运行内存 16 G.采用预训练好的 300 维 Word2vec<sup>[22]</sup>词向量初始化嵌入向量,批量大小除 AG’s news 为 50 外其他均为 25.为了能将胶囊平均地分成若干个组,且兼顾文本数据集的最大句子平均长度,我们将输入文本长度  $L$  调整为 387.使用 Adam 优化器,初始学习率为 0.001.过滤器 $W^a$ 的宽度分别为 3、4、5,步长为 1,填充为 0,各 128 个.过滤器 $W^b$ 的宽度为 3,步长、填充各为 1,每个胶囊组有 64 个.

表 1 文本数据集简介  
Tab. 1 Introduction to text datasets

Dataset	$T_r$	$T_s$	$Cl_s$	$L_{avg}$	Classification Task
AG’snews <sup>①</sup>	108 000	7 600	4	45	News categorization
MR(2005) <sup>②</sup>	8 635	1 067	2	22	Review classification
TREC <sup>③</sup>	5 500	500	6	9	Question categorization
MPQA <sup>④</sup>	8 587	1 067	2	3	Opinion classification
20 news <sup>⑤</sup>	10 182	7 532	20	315	News categorization

注:①[http://groups.di.unipi.it/~gulli/AG\\_corpus\\_of\\_news\\_articles.html](http://groups.di.unipi.it/~gulli/AG_corpus_of_news_articles.html). ②<http://www.cs.cornell.edu/people/pabo/movie-review-data/>. ③<http://cogcomp.org/Data/QA/QC/>. ④<http://mpqa.cs.pitt.edu>. ⑤<http://scikit-learn.org/0.19/datasets/twenty-newsgroups.html>

本文采用准确率作为评价指标,准确率越高,模型效果越好.对于模型参数和模型训练时间,参数越少,训练时间越少,模型越好越稳定.

4.2 胶囊分组实验

本文将胶囊平均地分为若干组,每个组可以被视为具有相似语义信息的胶囊的集合.我们根据 20 news 文本数据集的分类准确率的高低来衡量胶囊组分组效果的好坏.如表 2 所示,我们将胶囊

手动地分为  $D=1、2、4、8、16、32、64$  组时,对应的参数数量分别为 0.69 M、0.76 M、0.91 M、1.20 M、1.79 M、2.97 M、5.33 M,准确率分别为 79.01%、82.58%、84.66%、85.93%、86.85%、87.21%、85.07%,由此可知,随着分组的增加,参数数量也随之增加,这是因为有更多的权重在各组中进行共享.通过实验,在分组  $D=32$  时,文本分类准确率达到最高的 87.21%,可见胶囊的分组对准确率有很大贡献.而当分组达到 64 时,由于参数的增加,网络训练较为困难,且过多的特征信息会导致网络过拟合,降低分类准确率(由 87.21%降低为 85.07%),因此我们将分组数量选为 32 组进行训练.

表 2 在 20 news 文本数据集上的胶囊分组结果  
Tab. 2 Capsule grouping results on 20 news text dataset

Numbers	Parameters/M	Accuracy/%
$D=1$	0.69	79.01
$D=2$	0.76	82.58
$D=4$	0.91	84.66
$D=8$	1.20	85.93
$D=16$	1.79	86.85
$D=32$	2.97	87.21
$D=64$	5.33	85.07

4.3 实验结果与分析

本文与其他方法在模型参数数量,训练时间和准确率上进行了对比实验,实验结果如表 3 和表 4 所示.

表 3 参数数量、训练时间对比结果  
Tab. 3 Comparison of the number of parameters and training time

Model	Parameters/M	Training time/s	Accuracy/%
Capsule-A <sup>[4]</sup>	27.2	141	79.88
Capsule-B <sup>[4]</sup>	81.6	345	79.72
CapsNet-dynamic-routing <sup>[6]</sup>	13.7	79	86.26
CapsNet-static-routing <sup>[6]</sup>	9.4	77	86.34
文献 <sup>[21]</sup>	3.3	45	83.94
CapsNet-GSR	2.97	42	86.89

表 3 为本方法与其他方法在 20 news 数据集上的参数数量和训练时间的对比结果.批量大小均设置为 8,运行时间为训练中每个 epoch 的运行时间.从表 3 中可以看出,我们的模型比其他胶囊网络模型有着更少的参数数量,同时每个 epoch 所消耗的时间也是最少的,在 20 news 数据集的分类准

确率更达到了最高的 86.89%。Capsule-A<sup>[4]</sup> 采用了动态路由策略,与文献[2]的网络结构类似,且采用共享权重矩阵与非共享权重矩阵相结合,增加了网络的参数。CapsNet-GSR 与 Capsule-A 相比,参数明显减少了很多,因此训练时间也大大缩短。而 CapsNet-static-routing<sup>[6]</sup> 虽然采用的是静态路由,

但与 CapsNet-GSR 采用“胶囊压缩+静态路由”的组合策略相比,模型参数仍然更多。文献[21]中的模型通过组合编码胶囊构造词嵌入来减少词嵌入的参数,并采用 K 均值路由算法代替动态路由算法,因此在参数数量和训练时间上与本文模型相差不大,但在准确率上比本文模型要低 2.95%。

表 4 模型准确率对比结果  
Tab. 4 Comparison results of model accuracy

Model	AG's /%	MR(2005) /%	TREC/%	MPQA/%	20 news/%
CNN-rand <sup>[16]</sup>	92.2	76.1	91.2	83.4	—
CNN-static <sup>[16]</sup>	91.4	81.0	92.8	89.6	—
CNN-non-static <sup>[16]</sup>	92.3	81.5	93.6	89.5	86.6
LSTM <sup>[19]</sup>	86.1	75.9	86.8	—	—
BiLSTM <sup>[19]</sup>	88.2	79.3	89.6	—	—
Capsule-A <sup>[4]</sup>	92.1	81.3	91.8	88.28	80.39
Capsule-B <sup>[4]</sup>	92.6	82.3	92.8	88.38	80.03
CapsNet-dynamic-routing <sup>[6]</sup>	—	81.00	94.60	90.01	86.68
CapsNet-static-routing <sup>[6]</sup>	—	80.98	94.80	90.10	86.74
Capsule-LSTM <sup>[7]</sup>	91.8	81.7	93.2	—	—
CapsNet-GSR	92.63	82.19	94.97	90.06	87.21

表 4 显示了本方法与其他方法在 5 个文本数据集上的准确率实验结果。从表 4 可知,本文方法在 5 个数据集中有 3 个获得了最高的准确率。

CapsNet-GSR 与传统的 CNN 模型 CNN-non-static<sup>[16]</sup> 相比,在 5 个数据集上分别提高了 0.33%、0.69%、1.37%、0.56% 和 0.61%。与传统的 LSTM 模型 BiLSTM<sup>[19]</sup> 相比,在前 3 个数据集上分别提高了 4.43%、2.89% 和 5.37%。容易发现,本文模型在 TREC 数据集上提高较明显。究其原因有以下两点:(1) 胶囊网络对小样本数据集较为友好;(2) CapsNet-GSR 利用胶囊建立起局部-整体对应关系,并在静态路由过程中找到哪个局部属于哪个整体对象,从而保留了文档和类别的实例化参数,提高了文本的分类准确率。

与胶囊网络模型 Capsule-B 相比,本文模型在 AG's news、TREC、MPQA 和 20 news 数据集上分别有 0.03%、2.17%、1.68%、7.18% 的提高。实验表明,CapsNet-GSR 利用分组,共享局部转换矩阵,减少了参数,经过胶囊的整形转化,局部特征被进一步得到提取与保留;通过胶囊压缩合并了相似胶囊,剔除噪声胶囊,提高了胶囊的质量;利用静态路由,进一步降低参数数量,防止过拟合,降低计算复杂度,提高了分类准确率。

与胶囊网络模型 CapsNet-static-routing<sup>[6]</sup> 相比,CapsNet-GSR 在 MR(2005)、TREC 和 20 news 数据集上分别提高了 1.21%、0.17% 和 0.47%。由此可见,采用胶囊分组与胶囊压缩操作对提高分类性能有促进作用。通过利用 CapsNet-static-routing 和 CapsNet-GSR 与其他模型比较,可见静态路由策略对提高文本分类准确率有较好效果,而且结合胶囊分组能使胶囊提取的特征更加利于预测分类。胶囊的压缩可能会丢失部分细节信息,但压缩的仅仅是相似胶囊或噪声胶囊,对分类结果影响较小。

5 结 论

本文基于胶囊网络提出了 CapsNet-GSR 文本分类模型。该模型采用胶囊分组策略提取文本局部信息,提高了胶囊相关性,减少了模型参数,利于胶囊的预测分类;再利用胶囊压缩操作合并相似胶囊,去除噪声胶囊;最后通过静态路由机制降低参数数量,减少模型训练时间。实验证明,CapsNet-GSR 在参数数量和训练时间上均有大幅降低,在分类准确率上也有相应地提升。下一步研究中,我们将考虑在多标签分类领域中使用该模型来检验其分类性能。

参考文献:

[1] 杨棣威. BiGRU-CapsNet 文本分类模型研究[D]. 厦门: 厦门大学, 2019.

[2] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]//Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS). California: Mit Press, 2017.

[3] Ding X P, Wang N N, Gao X B, *et al.* Group feed-back capsule network [J]. IEEE T Image Process, 2020, 29: 6789.

[4] Zhao W, Ye J, Yang M, *et al.* Investigating capsule networks with dynamic routing for text classification [C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: ACL, 2018.

[5] 杨巨成, 韩书杰, 毛磊, 等. 胶囊网络模型综述[J]. 山东大学学报: 工学版, 2019, 49: 1.

[6] Kim J, Jang S, Choi S, *et al.* Text classification using capsules [J]. Neurocomputing, 2020, 376: 214.

[7] Yang M, Zhao W, Chen L, *et al.* Investigating the transferring capability of capsule networks for text classification [J]. Neur Net, 2019, 118: 247.

[8] 何明. 一种基于改进信息增益特征选择的最大熵模型文本分类方法[J]. 西南师范大学学报: 自然科学版, 2019, 44: 113.

[9] 刘云, 黄荣乘. 基于指数分布族的类特定文本分类算法[J]. 重庆邮电大学学报: 自然科学版, 2019, 31: 694.

[10] 王盼盼. 基于 BiGRU-CapsNet 模型的文本分类研究[D]. 西安: 陕西师范大学, 2019.

[11] Hayes P J, Weinstein S P. CONSTRUE/TIS: a system for content-based indexing of a database of news stories [C]//Proceedings of the Conference on Innovative Applications of Artificial Intelligence. Washington: AAAI, 1990.

[12] Maron M E. Automatic indexing: an experimental inquiry [J]. J Acm, 1961, 8: 404.

[13] Cover T, Hart P. Nearest neighbor pattern classification [J]. IEEE T Inform Theory, 1967, 13: 21.

[14] Quinlan R J. Induction of decision trees [J]. Mach Learn, 1986, 1: 81.

[15] Joachims T. Text categorization with Support Vector Machines: Learning with many relevant features [C]//Proceedings of the Conference on Machine Learning. Berlin: Springer, 1998.

[16] Kim Y. Convolutional neural networks for sentence classification [EB/OL]. [2020-12-15]. <https://arxiv.org/pdf/1408.5882.pdf>.

[17] 高云龙, 吴川, 朱明. 基于改进卷积神经网络的短文本分类模型[J]. 吉林大学学报: 理学版, 2020, 58: 923.

[18] Tomás M, Martin K, Lukás B, *et al.* Recurrent neural network based language model [C]//Proceedings of the Eleventh Annual Conference of the International Speech Communication Association. Makuhari: ACM, 2010.

[19] Cho K, van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation [EB/OL]. [2020-12-15]. <https://arxiv.org/pdf/1406.1078.pdf>.

[20] Zhao W, Peng H Y, Steffen E, *et al.* Towards scalable and reliable capsule networks for challenging NLP applications [C]//Proceedings of the 57th Annual Meeting of the Association-for-Computational-Linguistics (ACL). Florence: ACL, 2019.

[21] Ren H, Lu H. Compositional coding capsule network with K-means routing for text classification [EB/OL]. [2020-12-17]. <https://arxiv.org/pdf/1810.09177.pdf>.

[22] Mikolov T, Sutskever I, Chen K, *et al.* Distributed representations of words and phrases and their compositionality [C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. California: ACM, 2013.

引用本文格式:

中 文: 朱海景, 余谅, 盛钟松, 等. 基于静态路由分组胶囊网络的文本分类模型[J]. 四川大学学报: 自然科学版, 2021, 58: 062001.

英 文: Zhu H J, Yu L, Sheng Z S, *et al.* Text classification model based on static routing and grouped capsule network [J]. J Sichuan Univ: Nat Sci Ed, 2021, 58: 062001.