

基于胶囊网络和卷积网络的文本分类对比

胡春涛¹, 夏玲玲^{1*}, 张亮¹, 王超², 韩旭¹

(1. 江苏警官学院 计算机信息与网络安全系, 江苏 南京 210031;

2. 南京市玄武分局, 江苏 南京 210000)

摘要:针对 CNN 应用于文本分类任务中,存在固定大小的卷积核会限制复杂语言结构特征的提取,以及 CNN 在最大值池化时会丢失大量特征信息这两类问题,引入胶囊网络替代 CNN 提取文本特征,根据 Sabour 等人提出的动态路由算法作为胶囊网络内部的传播原理,并将其与 LSTM 连接形成融合神经网络模型,并与文本分类任务中常用的 CNN, LSTM, LSTM 连接 CNN 三个基线模型进行比较。实验结果表明,胶囊网络相较于使用卷积核的 CNN 能够不受特征检测器结构和大小的限制,更加灵活地学习文本整体与部分的内在空间关系;不仅能够检测固定的文本特征,而且能够检测特征的变体及其可能性。LSTM 连接胶囊网络的模型能有效提升文本分类的 F_1 值,达到 0.932,相比其他基线模型拥有最佳性能。

关键词:文本分类;神经网络;胶囊网络;卷积网络;网络舆情

中图分类号:TP391.1

文献标识码:A

文章编号:1673-629X(2020)10-0086-06

doi:10.3969/j.issn.1673-629X.2020.10.016

Comparative Study of News Text Classification Based on Capsule Network and Convolution Network

HU Chun-tao¹, XIA Ling-ling^{1*}, ZHANG Liang¹, WANG Chao², HAN Xu¹

(1. Department of Computer Information and Cyber Security, Jiangsu Police Institute, Nanjing 210031, China;

2. Nanjing Police Station Xuanwu Sub-bureau, Nanjing 210000, China)

Abstract: Concerning the problems that while dealing with text classification tasks, the fixed-volume convolution kernel of CNN will limit the extraction of complex language structure features, and max pooling operation will lose a lot of feature information at the pooling layer, we use an improved fusion model of LSTM connected capsule network which uses the dynamic routing algorithm proposed by Sabour et al. as the internal propagation principle instead of CNN for text features extracting. The experiments are conducted to investigate the performance of improved model by comparing with other three baseline models such as CNN, LSTM and LSTM connected CNN. It is found that the capsule network can learn the inherent spatial relationship between the whole and the part of the context more flexibly than the CNN using the convolution kernel. It can not only detect the fixed text features, and can detect variants of features and their possibilities. The improved model of LSTM connected capsule network can effectively improve the F_1 value to 0.932, which has the best performance compared with other baseline models.

Key words: text classification; neural network; capsule network; CNN; network public opinion

0 引言

近年来,由于神经网络强大的特征自动分析能力,在自然语言处理与图像识别等方面有着极为广泛的应用,其高准确率、高自动化的特性吸引了越来越多的学习者,也推动了人工智能行业的进一步发展。作为自然语言处理的一项重要应用,舆情信息处理技术也因神经网络的推动而发展。类似于开采出的原油需要进

一步的提炼和加工才能获得不同种类、不同用途的产品,舆情信息处理的首要任务就是将文本数据分类,筛选出有价值的信息。随着 TensorFlow, Keras 等深度学习框架的相继推出,初学者能够很容易地上手并搭建起一个基本的神经网络模型,应用于文本分类等一系列任务。然而,模型的通用性给研究者带来便利的同时,却降低了模型适用特定分类任务的准确性,导致现

收稿日期:2019-10-19

修回日期:2020-02-25

基金项目:国家自然科学基金(61802155);国家社会科学基金(13BTQ046);江苏高校哲学社会科学研究重点项目(2018SJDZ1142);江苏省高等学校大学生创新创业训练计划项目(201910329008Z);江苏警官学院科学研究项目(2017SJYZQ01)

作者简介:胡春涛(1998-),男,研究方向为网络攻防、警务大数据和电子数据取证;通信作者:夏玲玲(1988-),女,讲师,博士,研究方向为网络传播动力学、网络爬虫、数据挖掘。

有框架在特定文本分类任务上难以取得良好的效果,尤其是在处理具有个性化需求的分类任务时,只能作为基线模型使用。因此需要模型设计者根据不同分类任务相应地改变神经网络的结构,将不同的神经网络有侧重地进行融合和重组,提高模型对于特定信息的捕捉能力和语义的理解能力,从而使模型在准确率、召回率、 F_1 值、收敛速度、迁移能力等方面相比于普通框架具有显著的提高。

1 研究现状

在文本分类领域研究早期,主要采用向量空间模型、情感词典等方法。朱华宇等人^[1]提出一种基于向量空间模型的中文文本分类方法,通过特征提取、空间降维、层次分类和分类器训练等技术实现分类。姚清耘^[2]提出一种基于向量空间模型的中文文本聚类方法,在搜狗实验室的开源语料上得到了较好的聚类结果。郭叶^[3]提出将句中词语倾向值加权求和以及选用支持向量机两种方法,并在特征选取和规则后处理方面进行了改进,使准确率和召回率有较大提高。周城^[4]提出基于情感词典的分类方法,发现该方法性能上优于基于文本分类的方法。近年涌现出一批将经典深度神经网络应用于文本处理任务的研究。Hochreiter 等人^[5]设计了长短期记忆网络模型(long short-term memory, LSTM),但结构的复杂性导致 LSTM 的计算代价比较高。Cho 等人^[6]提出一种 LSTM 的替代方案——门限循环单元(gated recurrent unit, GRU)。GRU 模型相比 LSTM 模型,结构简单,可以提高模型的训练速度,并且还能解决 RNN 存在的梯度消失问题。除了使用循环神经网络以外,也有一些将卷积神经网络(convolution neural network, CNN)用于文本特征提取的研究。Kim 等人^[7]将预训练好的词向量作为输入,利用 CNN 实现情感分类。Kalchbrenner 等人^[8]提出一种动态卷积神经网络(dynamic convolution neural network, DCNN)模型用于句子特征学习,并取得了较好的效果。但这些都只是使用单层神经网络,特征提取能力、编码能力较弱。因此,需要结构更加复杂的神经网络的组合。陈洁等人^[9]提出使用并行混合神经网络模型,将文本映射成字符级词向量并组成句子矩阵,对句子矩阵分别连接双向 GRU 和 CNN,并分别后接注意力层、池化层,将两个网络的输出向量进行拼接,得到句子的特征表示,最后使用全连接神经网络作为分类器,模型在 AI Challenger2018 用户评论情感数据集上最高准确率达到 0.868 2, F_1 值达到 0.865 6。该方法仅通过不同网络的并行处理获得对文本的不同表示,未能对文本实现更加复杂的抽象和提取。刘月等人^[10]提出利用

CNN 提取短语序列的特征表示,利用 N-LSTM(nested long short-term memory)学习文本的特征表示,引入注意力机制,突出关键短语以优化特征提取,使模型不但能捕捉句子局部特征及上下文语义信息,还能突出输入文本各部分对文本类别的影响程度,在 3 个公开数据集上性能比基线模型有了明显提高。由于在文本上进行 CNN 操作不同于在图像上,细小的信息损失极有可能导致结果偏差。因此 CNN 连接 N-LSTM 的方式有待改进。张冲^[11]设计了 Attention-Based LSTM 模型用于特征提取,得到含有输入序列节点注意力概率分布的语义编码,并将其作为分类器的输入,减少了特征向量提取过程中的信息丢失和信息冗余。在胶囊网络(capsule network)应用于文本分类的研究中,Zhao 等人^[12]在英文语料的基础上,将 Capsule Network 连接卷积网络,受标注成本的限制,作者用单标签数据对模型进行训练,并在多标签数据上进行测试,在训练了 5.8k 的数据后,发现 Capsule Network 的性能远远高于 LSTM、CNN 等单一神经网络。该文仅在英文语料上进行了实验,忽略了中文分词、语法结构对分类结果产生的影响,在文中未能解释胶囊网络性能更优的根本原因。

而胶囊网络的发展也经历了从提出到愈发多样的过程。Sabour 等人^[13]于 2017 年首次提出了胶囊网络,它将神经元的输入由标量变为向量,通过动态路由调整参数加强网络对特征的整合能力。2018 年 Hinton 等人^[14]提出使用期望最大化路由(EM routing)的矩阵胶囊网络,该实验通过图像数据表明胶囊网络还能有效学习物体不同视点的变化。其实,胶囊网络的本质就是聚类算法思想在模型中的实现。Sabour 等人提出的动态路由是基于 K-Means 算法的实现,文献[14]则是将高斯混合模型(GMM)融合到特征整合过程中,通过期望最大化路由对网络参数及进行调整。上述这些使用胶囊网络或基于原始网络改进的模型都在各自的任务上取得了较好的效果,具有广阔的应用前景。

当前胶囊网络多用于英文语料的分类,鉴于中文和英文在语法、语义结构等方面存在差异,文中在中文语料上对模型的结构进行相应改进。在总结上述方法的优缺点后,构建了 LSTM 连接胶囊网络的文本分类模型,其中胶囊网络基于 K-Means 动态路由算法试验,并与多个基线模型进行比较,通过准确率、召回率、 F_1 值等指标来评价模型性能优劣。

2 数据集与模型介绍

2.1 数据集

文中爬取了“今日头条”公开的新闻数据集,原始

数据集总共分为 15 大类,包括民生、文化、娱乐、体育、财经、房产、汽车、教育、科技、军事、旅游、国际、证券、农业、电竞。由于平台自身数据不均衡,证券类新闻数量明显少于其他类别,因此将证券类划归入财经类,使得数据分布较为均匀,表 1 为各类别的样本数据统计情况。

表 1 各类别的样本数据统计情况

| 分类类别 | 分类类别(英文) | 样本数量 |
|------|--------------------|-------|
| 民生 | news_story | 2 117 |
| 文化 | news_culture | 2 259 |
| 娱乐 | news_entertainment | 6 045 |
| 体育 | news_sports | 6 195 |
| 财经 | news_finance | 4 641 |
| 房产 | news_house | 2 103 |
| 汽车 | news_car | 5 940 |
| 教育 | news_edu | 4 510 |
| 科技 | news_tech | 5 740 |
| 军事 | news_military | 2 783 |
| 旅游 | news_travel | 3 099 |
| 国际 | news_world | 3 902 |
| 证券 | stock | 58 |
| 农业 | news_agriculture | 3 492 |
| 电竞 | news_game | 5 305 |

2.2 分类模型的构造

模型第一层去除停用词后将词语编码成词向量。文中使用的词向量模型 word2vec 基于“Zhihu_QA 知乎问答”语料训练,将词语转化为 256 维的向量。

第二层在词向量层上添加双向 LSTM 对文本进行编码。经统计,文本平均长度为 276,通过计算将神经元个数设置为 300 可以覆盖 99% 的文本长度。为避免过拟合,在两层 LSTM 中间添加隐藏层,将 20% 的权重归为零。

第三层在双向 LSTM 后接胶囊网络,设置输入胶

囊的个数为 64。

第四层在胶囊网络之后连接全连接网络层,用 softmax 函数输出分类结果,模型结构如图 1 所示。

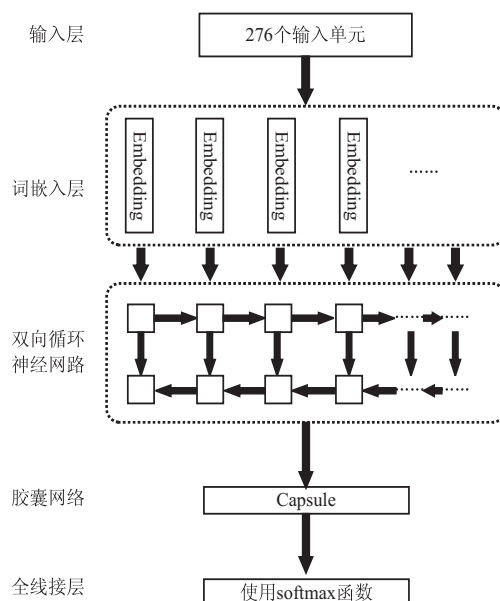


图 1 模型结构

为了便于比较,文中所有模型均使用相同数据集,并设计了 CNN 单模型,双向 LSTM 单模型,和胡西祥^[15]提出的双向 LSTM 连接 CNN 的融合模型,在同等条件下使用控制变量的方法对多组结果进行比较。

第一组将 CNN 单模型、双向 LSTM 单模型与 LSTM 连接 CNN 的模型进行比较,其中 CNN 均使用一维卷积,卷积核个数设置为 64,通过多次实验发现卷积核大小设置为 3 * 256 时模型取得最高 F₁ 值,后接全连接网络使用 softmax 函数作为分类器,借以证明融合神经网络模型分类效果是否优于单模型。

第二组将含有胶囊网络的模型与双向 LSTM 连接 CNN 进行比较,借以证明胶囊网络^[16]在文本特征提取能力方面是否优于卷积神经网络。

仔细分析一中的 Capsule 层,得到如图 2 所示的胶囊网络内部结构。

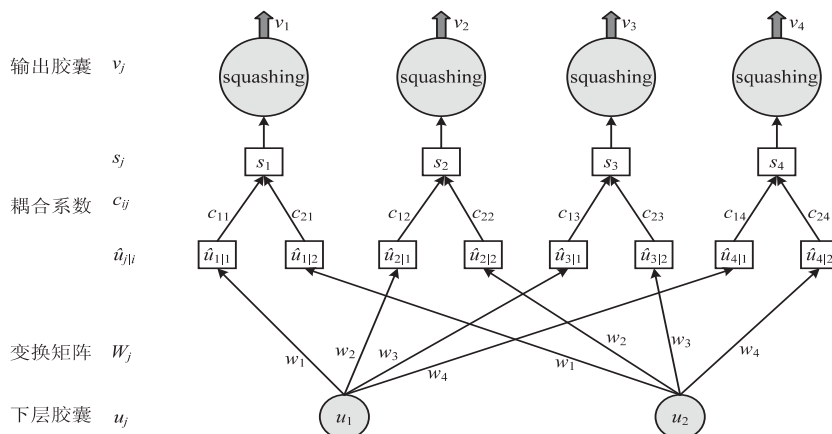


图 2 权值共享的胶囊网络结构示意图

其中包含多层的胶囊层,每一层都被分成许多组向量神经元(胶囊),因此每个神经元具有长度和方向的概念。在文本中,长度用来衡量某个词存在的概率,方向代表该词的各种属性(如位置、语意特征、语境含义等)。图 2 中每个底层胶囊对应 4 个变换矩阵,即对底层神经元进行四个维度信息的提取。由于文本长度通常是不固定的,使用权值共享的胶囊网络,即对于固定的上层胶囊 S_j ,它与所有底层胶囊连接的变换矩阵是共用的,如图 2 中 u_1 和 u_2 共享同一个 W_1 。

低级别的向量神经元需要决定如何将其经过不同维度提取的输出向量发送到更高一级的向量神经元,该过程由耦合系数 c 来决定,耦合系数 c 本质上是一个权重,定义了低层神经元到高层神经元的概率分布,采用“动态路由”算法^[13]进行更新。下面通过对该算法的介绍将更有利于说明胶囊网络相比 CNN 在传播原理上的差异。

首先,下层的胶囊(向量) u 经过变换矩阵 W 做点积,得到 \hat{u} ,如式(1)^[13]所示:

$$\hat{u}_{ji} = W_{ij} \cdot u_i \quad (1)$$

在确定 \hat{u}_{ji} 后,需要进行第二个阶段的权值分配以计算输出节点, S_j 由式(2)^[13]得到:

$$S_j = \sum_i c_{ij} \cdot \hat{u}_{ji} \quad (2)$$

其中, c 为耦合系数,一层胶囊网络所有的耦合系数和为 1,即图 2 中 $c_{11} + c_{21} + c_{12} + c_{22} + c_{13} + c_{23} + c_{14} + c_{24} = 1$ 。该耦合系数由动态路由(dynamic routing softmax)算法迭代更新,且 softmax 函数中的 b_{ij} 初始化为 0,耦合系数 c_{ij} 计算方式如式(3)^[13]所示:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (3)$$

b_{ij} 通过测量当前层的输出 v_j 和前面层级的 \hat{u}_{ji} 的向量一致性,由式(4)^[13]多次迭代得出:

$$b_{ij} \leftarrow b_{ij} + \hat{u}_{ji} \cdot v_j \quad (4)$$

即计算点积更新 b ,进而优化 c 。这种更新的做法可解释为:点积运算接收两个向量,并输出一个标量。对于给定长度但方向不同的两个向量而言,点积有下列几种情况:正值、零、负值。当 \hat{u} 和 v 的相乘结果为正时,代表两个向量指向的方向相似, b 更新结果变大,那么耦合系数就高,说明该 \hat{u} 和 v 十分匹配。相反,若是 \hat{u} 和 v 相乘结果为负, b 更新结果变小,那么耦合系数就小,说明不匹配。通过迭代确定 c ,也就确定了一条路线,路线的尽头就是所求的最优化参数。

胶囊网络使用 Squash 作为激活函数,输出 v 的计算方式如式(5)^[13]所示:

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

该激活函数前一部分是输入向量 s_j 的缩放尺度,后一部分是 s_j 的单位向量,既保留了输入向量的方向,又将输入向量的模压缩到 $[0, 1)$ 之间,能有效避免梯度爆炸问题的出现。在整个网络中,耦合系数 c_{ij} 通过动态路由进行更新,而整个网络其他参数包括 W_{ij} 在内都需要根据损失函数进行更新。在损失函数方面,文中使用多分类对数损失函数(categorical crossentropy),优化器使用 Adam。

表 2 将胶囊网络和卷积网络的传播过程进行对比,进一步分析其差异。

表 2 胶囊网络和卷积网络传播对比

| | 向量神经元 | 标量神经元 |
|-------|---|----------------------------|
| 输入 | u | x |
| 转换 | $\hat{u}_{ji} = W_{ij} \cdot u_i$ | - |
| 加权求和 | $S_j = \sum_i c_{ij} \cdot \hat{u}_{ji}$ | $a_j = \sum_i w_i x_i + b$ |
| 非线性激活 | $v_j = \frac{\ s_j\ ^2}{1 + \ s_j\ ^2} \frac{s_j}{\ s_j\ }$ | $h_i = g(a_j)$ |
| 输出 | v_j | h_i |

动态路由算法的运行逻辑可以做如下解释:

初始化让所有 b_{ij} 初始化为 0,目的是为了保证耦合系数 c_{ij} 的均匀分布,同时说明低层的向量神经元要传输到哪个高层神经元是不确定的。

动态路由就是找到每个低层神经元最有可能传输到哪个高层神经元,具体到实例就是含有“华为”和“小米”的句子最有可能属于科技、文化或是其他类别。动态路由在每一次循环中都做了归一、预测、加总、压缩和更新五个步骤。

起初由于 b_{ij} 全为 0,那么:

$$c_{11} = c_{12} = c_{21} = c_{22} = \dots = \frac{1}{8} \quad (6)$$

通过变换矩阵进行预测,由“华为”的语义向量点乘“华为”对应科技类别的变换矩阵,得到由“华为”的语义推出整句语义类别的高层向量 \hat{u}_{111} :

$$\hat{u}_{111} = W_{11} \cdot u_1 \quad (7)$$

同理,由“小米”的语义向量点乘“小米”对应科技类别的变换矩阵,得到由“小米”的语义推出整句语义类别的高层向量 \hat{u}_{112} :

$$\hat{u}_{112} = W_{21} \cdot u_2 \quad (8)$$

由“华为”的语义向量点乘“华为”对应文化类别的变换矩阵,得到由“华为”的语义推出整句语义类别的高层向量 \hat{u}_{211} :

$$\hat{u}_{211} = W_{12} \cdot u_1 \quad (9)$$

由“小米”的语义向量点乘“小米”对应文化类别的变换矩阵,得到由“小米”的语义推出整句语义类别的高层向量 \hat{u}_{212} :

$$\hat{u}_{212} = W_{22} \cdot u_2 \quad (10)$$

通过耦合系数 c_{ij} 进行加权求和:

$$S_1 = c_{11} \cdot \hat{u}_{111} + c_{21} \cdot \hat{u}_{112} \quad (11)$$

$$S_2 = c_{12} \cdot \hat{u}_{211} + c_{22} \cdot \hat{u}_{212} \quad (12)$$

加权求和可以理解成“华为”和“小米”等多个低层神经元共同作用得到融合多个语义的高层语义类别 s_1, s_2 。

压缩,即单位化语义类别向量:

$$v_1 = \frac{\|s_1\|^2}{1 + \|s_1\|^2} \frac{s_1}{\|s_1\|} \quad (13)$$

$$v_2 = \frac{\|s_2\|^2}{1 + \|s_2\|^2} \frac{s_2}{\|s_2\|} \quad (14)$$

通过预测结果更新参数 b_{ij} ,就是从“华为”和“小米”推出科技或文化类别的可能性。

$$b_{11} \leftarrow b_{11} + \hat{u}_{111} \cdot v_1 \quad (15)$$

$$b_{21} \leftarrow b_{21} + \hat{u}_{112} \cdot v_1 \quad (16)$$

其中, \hat{u}_{111} 为低层神经元对词语属于哪一类别的单独预测, v_1 为多个神经元融合的“共识预测”,当两者预测结果越相似(即向量方向较为一致),点积就越大, b_{11} 和 b_{21} 就变大,即由“华为”和“小米”推出科技类的可能性就大。该点积的过程如图 3 所示。

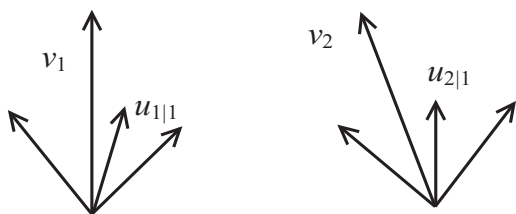


图 3 计算 v_j 和 \hat{u}_{ji} 的向量一致性更新 c_{ij}

最后用 softmax 函数更新概率分布,由“华为”推出科技类的概率:

$$c_{11} = \frac{e^{b_{11}}}{e^{b_{11}} + e^{b_{12}}} \quad (17)$$

$$c_{21} = \frac{e^{b_{21}}}{e^{b_{21}} + e^{b_{22}}} \quad (18)$$

最终 $b_{11} > b_{12}$, 即 $c_{11} > c_{12}$, $b_{21} > b_{22}$, 即 $c_{21} > c_{22}$, 故两个词路由到科技类的概率变大。图 2 中演示了 4 种变换矩阵,即向量神经元有四种不同的方向,可以从 4 个维度对信息进行提取,如语义、位置、上下文语境、词语类别等,原理和上述相同。

重复以上预测,加权,压缩,更新的步骤,直到每个耦合系数 c_{ij} 趋于稳定。

图 4 展示了胶囊网络内部向量神经元经过一个维度矩阵变换的传播过程。

其中,对所有的低层神经元 u_1, u_2, u_3, \dots ,使用同一个共享矩阵 W ,得到高层语义向量 $\hat{u}_{j1}, \hat{u}_{j2}, \hat{u}_{j3}, \dots$,经过耦合系数 $c_{1j}, c_{2j}, c_{3j}, \dots$,加权求和得到其

中一个维度的高层向量 v_j ,其他维度依此类推。

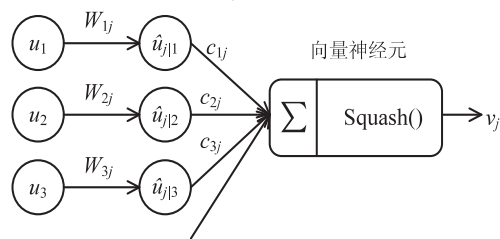


图 4 胶囊网络内部传播过程示意图

3 实验结果分析和比对

不同模型的最优分类结果如表 3 所示。

表 3 不同模型最优分类结果

| 对比模型 | F1 值 | 精确度 | 召回率 |
|---------------------------------|---------|---------|---------|
| CNN(模型 1) | 0.754 1 | 0.754 8 | 0.753 6 |
| BiLSTM(模型 2) | 0.773 1 | 0.768 0 | 0.778 3 |
| BiLSTM+CNN(模型 3) | 0.919 9 | 0.904 2 | 0.938 4 |
| BiLSTM + Capsule Network (模型 4) | 0.932 0 | 0.931 2 | 0.932 8 |

对比模型 1 和模型 2 发现,单模型 BiLSTM 分类性能优于 CNN。对比模型 2 和模型 3 发现融合模型的性能要高于单模型, LSTM 适用于对文本信息进行编码, CNN 相比而言更适用于对特征进行提取。对比模型 3 和模型 4 发现连接胶囊网络的模型在精确率和召回率两方面均优于连接 CNN 的模型,其原因主要有以下几点:

(1) CNN 使用特征检测器,从一维的向量序列中提取局部模式,并使用最大值池化减少参数数量。但作为一个空间不敏感模型, CNN 在处理文本中更长更复杂的结构和语句顺序时(例如倒装,前置等复杂结构),就会由于卷积核长度的限制而难以提取到有关特征。

(2) CNN 的池化操作会丢失一些空间信息,尤其是对高层信息的池化。在对图像的操作中,池化能在保证图像重要特征不缺失的情况下减少参数,使得全连接层能以较少的参数,较高的准确率完成分类任务。但在自然语言处理中,语义并不像图像特征,可能一个词语顺序的颠倒就会导致整句语义的变化。因此池化容易忽略蕴藏在文本中的细节特征,造成最后分类结果的误差,池化操作如图 5 所示。

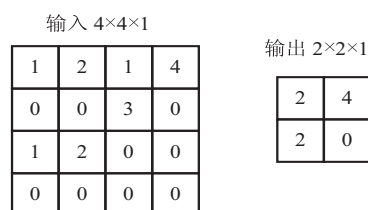


图 5 池化操作示意图

池化操作类似于两层网络之间的信使,它告诉下一层在输入图片的左上角最大值是2,右上角最大值是4,但不知道这个2和4是具体从哪里来的。就好比在池化文本信息时,只提取了特征性最强的信息,但忽略了文本蕴含的语句结构信息。

(3)胶囊网络通过变换矩阵和耦合系数去自动学习哪些是文本重要信息,重要的信息其输出的概率(耦合系数)更大,次要信息输出的概率较小。因此能够实现对信息的有效鉴别和提取。另外,胶囊网络不受特征检测器结构和大小的限制,能够学习文本整体与部分的内在空间关系,自动归纳出新的感受域,从而让网络能灵活地学习文本内在复杂的关系结构。正如Hinton^[14]所说,胶囊网络不仅用来捕捉特定变体的特征,更是用来捕捉特征及其变体的可能性。因此训练胶囊网络的目的不仅在于检测特征,还在于学习变体,相比于CNN更适合对文本信息的特征提取。

4 结束语

在中文新闻分类数据集标注准确和分布均匀的基础上,针对卷积神经网络不能有效提取复杂语意结构和池化操作容易丢失大量特征信息等问题,使用了胶囊网络代替CNN并提出了一种LSTM连接Capsule Network的多网络融合模型。通过对比实验有效证明了胶囊网络在文本特征提取能力上相较于CNN的优势。在后续的工作中,将尝试使用BERT(bidirectional encoder representation from transformers)模型对中文文本的每个字给出一个动态向量表示,并结合胶囊网络进一步改善模型的分类效果。

参考文献:

- [1] 朱华宇,孙正兴,张福炎. 一个基于向量空间模型的中文文本自动分类系统[J]. 计算机工程,2001,27(2):15-17.
- [2] 姚清耘. 基于向量空间模型的中文文本聚类方法的研究[D]. 上海:上海交通大学,2008.
- [3] 郭叶. 中文句子情感倾向分析[D]. 北京:北京邮电大学,2010.
- [4] 周城. 面向中文Web评论的情感分析技术研究[D]. 长沙:国防科学技术大学,2011.
- [5] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation,1997,9(8):1735-1780.
- [6] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]//Proceedings of the 2014 conference on empirical methods in natural language. Doha:CoRR Press,2014:1724-1734.
- [7] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 conference on empirical methods in natural language processing. New York:ACM Press,2014:1746-1751.
- [8] KALCHBRENNER N, GREFFENSTETTE E, BLUNSON P. A convolutional neural network for modelling sentences[C]//Proceedings of the 2014 annual meeting of the association for computational linguistics. Baltimore:ACL,2014:655-665.
- [9] 陈洁,邵志清,张欢欢,等. 基于并行混合神经网络模型的短文本情感分析[J]. 计算机应用,2019,39(8):2192-2197.
- [10] 刘月,翟东海,任庆宁. 基于注意力CNLSTM模型的新闻文本分类[J]. 计算机工程,2019,45(7):303-308.
- [11] 张冲. 基于Attention-Based LSTM模型的文本分类技术的研究[D]. 南京:南京大学,2016.
- [12] ZHAO Wei, YE Jianbo, YANG Min, et al. Investigating capsule networks with dynamic routing for text classification[C]//Proceeding of the 2018 conference on empirical methods in natural language processing. New York:ACM Press,2018:3110-3119.
- [13] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[C]//Proc of annual conference on neural information processing systems. Cambridge,MA:MIT Press,2017:3859-3869.
- [14] SABOUR S, FROSST N, HINTON G. Matrix capsules with EM routing[C]//6th international conference on learning representations. [s.l.]:ICLR,2018:1-15.
- [15] 胡西祥. 基于深度学习的微博评论情感倾向性分析[D]. 哈尔滨:哈尔滨工业大学,2019.
- [16] 范文豪,吴晓富,张索非. 动态路由胶囊网络的可视化研究[J]. 计算机技术与发展,2019,29(8):71-75.