

基于胶囊网络的方面级情感分类研究

徐志栋¹, 陈炳阳², 王晓^{3,4}, 张卫山²

(1. 中国人民解放军国防大学国家安全学院, 北京 100091;

2. 中国石油大学(华东)计算机科学与技术学院, 山东 青岛 266580;

3. 青岛智能产业技术研究院, 山东 青岛 266109;

4. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190)

摘要: 由于文本中多种情感极性混合而难以判断, 方面级情感分析成为当前研究的热点。考虑到多面句表达时会在一定程度上对不同目标的多重情感造成特征重叠, 进而影响文本情感分类效果, 提出一种基于胶囊网络的方面级情感分类模型(SCACaps)。模型使用序列卷积分别提取上下文和方面词的特征, 同时引入交互注意力机制, 减少二者对彼此的影响, 并对文本特征表示进行重构后传入胶囊网络。胶囊层间通过引入高层胶囊系数对路由算法进行优化, 整个迭代更新过程的参数全局共享, 以保存较完整的文本特征信息。通过与多个模型进行对比实验发现, SCACaps的分类效果最佳, 同时, 在小样本学习中 SCACaps 也有较好的表现。

关键词: 胶囊网络; 序列卷积; 交互注意力; 方面级; 情感分类

中图分类号: TP301

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202031

Research on capsule network-based for aspect-level sentiment classification

XU Zhidong¹, CHEN Bingyang², WANG Xiao^{3,4}, ZHANG Weishan²

1. School of Security, China People's Liberation Army National Defence University, Beijing 100091, China

2. College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China

3. Qingdao Academy of Intelligent Industries, Qingdao 266109, China

4. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract: Considering the difficulty of judging the mixed multiple sentimental polarities in a text, aspect-level sentiment analysis has become a hot research topic. Multiple sentiments of different targets when expressing multi-faceted in a sentence, it will cause problems such as feature overlap, which will have a negative impact on text sentiment classification. A capsule network-based model for aspect-level sentiment classification (SCACaps) was proposed. Sequential convolution was used to extract the features of context and aspect words separately, and an interactive attention mechanism was introduced to reduce the mutual influence on each other, and then the text feature representation was transmitted into the capsule network after reconstruction. The routing algorithm was optimized by introducing high-level capsule coefficients between the capsule layers, and the global parameters were shared in the entire iterative update process to save relatively complete text feature information. By comparing with multiple models, the SCACaps model has the best classification effect, and the SCACaps model also performs better in small sample learning.

Key words: capsule network, sequential convolution, interactive attention, aspect-level, sentiment classification

收稿日期: 2020-07-30; 修回日期: 2020-09-11

通信作者: 张卫山, zhangws@upc.edu.cn

基金项目: 国家重点研发计划基金资助项目 (No.2018YFE0116700); 山东省自然科学基金资助项目 (No.ZR2019MF049)

Foundation Items: The National Key Research and Development Program of China (No.2018YFE0116700), The Natural Science Foundation of Shandong Province (No.ZR2019MF049)

1 引言

情感分类即挖掘个体用户对产品、服务或者社会舆论等的情感认知^[1], 是自然语言处理领域的一个重要分支。在互联网飞速发展的时代, 人们可以十分便利地通过多种网络平台表达自己的日常情感。在现实生活中, 一句话针对不同方面会存在多种情感极性。例如, 在“这家餐厅菜品不错, 就是服务态度不好”中, “菜品”与“服务态度”这两方面的情感极性是相反的。通过文本方面级情感挖掘可以了解人们的真正意愿, 尤其在发生重大事件、社会转型等关键时期, 对于国家决策具有极其重大的意义。

Liu B^[2]首先提出方面级情感分析, 方面级情感分析属于一种细粒度情感分析, 其针对单句文本中的不同方面产生不同的情感极性。近年来, 方面级情感分类研究主要采用基于深度学习的方法进行特征提取, 许多神经网络模型在情感分析领域取得了一定的效果。长短期记忆(long short-term memory, LSTM)神经网络及其相应变型双向长短期记忆(bi-directional long short-term memory, BiLSTM)神经网络^[3]因具有灵活且简单的特点在方面级情感分析中被广泛使用。Wang Y Q 等人^[4]提出ATAE-LSTM模型, 该模型通过LSTM进行方面嵌入和词语向量表示, 利用方面语义信息搜寻文本中与方面词相关的位置的内容, 通过注意力机制对词语与方面词之间的相互依赖关系进行建模, 增强了方面嵌入的效果。Ma D H 等人^[5]考虑到上下文和方面词在向量表示中会相互影响, 在ATAE-LSTM模型的基础上提出了交互注意力网络(interactive attention network, IAN)模型, 通过注意力矩阵重新对上下文和方面词进行向量表示, 实验效果有所提升。陈虹等人^[6]在IAN模型的基础上提出Bi-IAN模型, 借助BiLSTM模型从两个方向分别提取上下文和目标的语义信息, 获得每个时间上下文和目标的隐藏表示, 取得了较好的实验效果。

现有的深度学习方法在提取文本特征方面取得了一定的效果, 但仍然存在如下问题: 一方面, 借助循环神经网络处理方面级情感的模型在一定程度上仍存在梯度消失、运行速度较慢等问题, 而基于卷积神经网络(convolutional neural network, CNN)的模型在处理文本时序任务时也存在一定的困难; 另一方面, 传统的注意力机制仅考虑了句子

中的一部分上下文信息, 忽略了句子所传达的整体意义, 在面对复杂句子时可能会引入噪声, 进而影响对情感极性的预测。如在反讽句“也许这款笔记本销量的提高并不是因为它的性能吧”中, 传统注意力机制会因“提高”预测此句为积极情感。为此, 本文提出了基于胶囊网络的方面级情感分类模型(capsule network-based model for aspect-level sentiment classification, SCACaps), 在胶囊层间传递参数的过程中, 模型通过高层胶囊系数调整每个低层胶囊选择高层胶囊的概率, 整个过程中全局共享参数。此外, 通过序列卷积分别对上下文和方面词进行建模, 借助交互注意力机制对它们的向量表示进行重构。在SemEval2014任务4数据集上验证模型的有效性, 实验结果表明, SCACaps具有较好的分类效果。本文的主要贡献如下。

(1) 使用胶囊网络处理方面级情感分类任务, 并引入高层胶囊系数来调整每个低层胶囊对高层胶囊的贡献, 在一定程度上解决了单句文本中的特征重叠问题。

(2) 通过序列卷积提取上下文和方面词的隐含特征, 同时引入交互注意力机制对上下文和方面词的向量表示进行重构, 以减少二者的相互影响。

2 相关研究

2.1 情感分析研究

传统神经网络常用LSTM来解决文本时序性问题, 但LSTM无法记忆完整的文本信息且计算较慢。CNN在图像处理方面具有出色的局部特征提取能力, 因此相关研究人员逐渐将CNN应用到文本处理任务中, 同时考虑文本序列的特点, 对CNN进行变形, 使其能够在特征提取时将文本中的所有词汇都考虑在内。Kim Y^[7]首次提出Text-CNN(text-convolutional neural network)模型, 针对句子级别进行情感分类, 并取得了不错的分类效果。Xue W 等人^[8]借助门控单元从文本中抽取与方面相关的信息, 结合CNN提出了GCAE(gated convolutional network with aspect embedding)模型, 取得了更高的正确率。韩建胜等人^[9]提出一种基于双向时间卷积网络(bi-directional temporal convolutional network, Bi-TCN)的情感分析模型, 该模型使用单向多层空洞因果卷积结构分别对文本进行前向和后向特征提取, 将两个方向的序列特征

融合后进行情感分类。高巍等人^[10]提出一种基于序列卷积神经网络的情感分类模型,将因果卷积和扩张卷积结合,用卷积层代替全连接层,以改善网络特征选取的局限性。Fan F F 等人^[11]提出 MGAN 模型,该模型分别使用了细粒度和粗粒度的注意力,进一步描述了方面和上下文单词之间的交互,取得了不错的效果。曾锋等人^[12]通过双层注意力捕获不同单词和不同句子的重要性,利用双向循环神经网络获取更多的序列上下文依赖信息,进而进行情感极性预测。曾义夫等人^[13]提出基于双记忆注意力的方面级别情感分类模型,借助循环神经网络的序列学习能力得到语句编码,并通过构造相应的注意力机制从语句编码中提取关于给定方面词的情感表达。

2.2 胶囊网络研究

胶囊网络^[14-15]被提出后,逐渐在文本情感分类任务中被广泛使用,与传统神经元以标量值表示特征不同,其网络节点通过向量值表示文本中词汇的多属性特征,借助动态路由机制代替池化层结构进行参数传递。Chen Z 等人^[16]提出一个转换胶囊网络 (transfer capsule network, TransCap) 模型,该模型将句子级语义表示封装为来自方面级和文档级数据的语义胶囊,从而将非方面级情感转换为方面级情感进行分类。Xiao L Q 等人^[17]提出一种基于胶囊的多任务学习体系结构,利用胶囊特征聚类的优点,对网络中每个任务的特征进行聚类,减少了任务之间的干扰。胶囊网络可以抓取部分-部分、部分-整体的位置信息^[18],从而获取文本的局部特征和整体情感极性之间的关系。杨云龙等人^[19]借助门控循环单元 (gated recurrent unit, GRU) 捕捉文本上下文的全局特征,在初始胶囊层获取文本整体属性的向量化特征信息,通过主胶囊层进行特征组合,提出一种融合 GRU 和胶囊特征的情感分析模型 G-Caps (gated recurrent unit-capsules)。王敏蕊等人^[20]针对序列生成模型在多标签分类任务中易产生累积误差等问题,使用 BiLSTM + Attention 进行语义编码,利用动态路由结构进行解码,通过共享全局参数来减少误差累积,构建了基于动态路由的序列生成模型 (sequence generation model based on dynamic routing, DR-SGM)。胶囊网络在文本处理任务中表现出较强的特征提取能力,越来越多的研究人员逐渐将其应用到跨领域文本处理任务中。林悦等人^[21]提出利用目标领域的标记信息来改善模型

适应能力的思想,并设计了额外的胶囊网络层促进目标领域的适应,构建了一个基于胶囊网络的跨领域情感分类模型。Zhang B W 等人^[22]通过胶囊网络对内在空间部分-整体关系进行编码,构成领域不变知识,从而将源领域和目标领域联系起来,此外,为了增强综合句子表示学习,模型还在胶囊网络中融入了语义规则,实现了跨领域文本的情感分析。在情感分析任务中,胶囊网络使用路由机制来生成高级特征,根据不同词汇被考虑的程度分配不同的耦合系数,通过聚类解决特征重叠的问题,但目前关于其被用于方面级情感分析的研究较少。

3 模型构建

SCACaps 的模型结构如图 1 所示。首先在词嵌入层对上下文和方面分别进行向量表示,然后采用序列卷积提取各自的特征,接着引入交互注意力机制分别求出上下文和方面注意力矩阵,并重构它们的向量表示,最后将提取到的特征向量加入胶囊网络,借助动态路由算法实现文本情感分类。图 1 中, $(w_1^c, w_2^c, \dots, w_n^c)$ 为上下文的向量表示, $(h_1^c, h_2^c, \dots, h_n^c)$ 为经过卷积序列提取的上下文隐含特征, $(w_1^a, \dots, w_m^a)$ 和 $(h_1^a, \dots, h_m^a)$ 分别为方面词的向量表示和隐含特征。

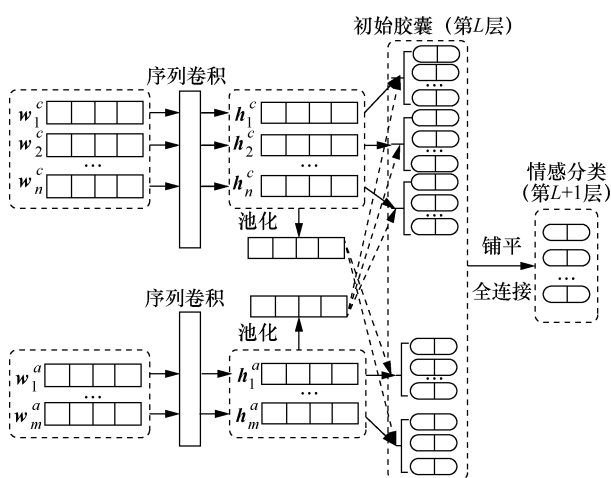


图 1 SCACaps 的模型结构

3.1 词嵌入层

词嵌入指借助词典将词映射为向量并进行分布式表示的过程。由于预训练的词向量模型能够从大量未标记文本中获取词汇的句法和语音信息,因此实验中采用经过预训练的词向量 GloVe 进行初始

化, 分别得到上下文和方面词的向量表示 w_n^c 和 w_m^a , 同时考虑词汇在上下文的位置信息 k , 从而得到词向量矩阵 $M \in R_{d \times |v|}^k$, 其中 d 是词向量的维数, $|v|$ 是词汇表的大小。

借助词汇表将每条文本数据转化为整数序列, 根据每个词汇在词向量矩阵中对应的列, 将输入文本中的每个词表征为一个 300 维的向量。按照词汇在文本中的顺序进行排列拼接, 可以将输入文本数据表示为矩阵形式, 进而实现对文本特征的向量化表示。

3.2 序列卷积

虽然 LSTM 考虑了文本序列的特性, 但仍无法获取文本的完整信息。考虑到 CNN 在文本处理任务中的良好表现以及其能够保留文本序列特征, 本文提出的模型引入了序列卷积, 借助因果空洞卷积进行批标准化和残差连接, 再将空洞卷积的输出链入卷积层进行特征提取, 分别得到上下文和方面词的隐含特征 h_n^c 和 h_m^a 。序列卷积的结构如图 2 所示。其中, x 代表输入的向量, u 代表提取到的特征。

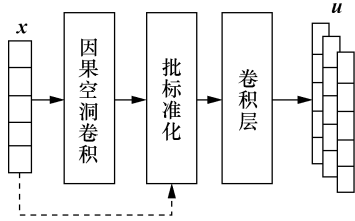


图2 序列卷积的结构

因果空洞卷积对过去的输入 (即 0 到 $t-1$ 时刻的文本序列) 进行计算, 对 t 时刻的文本特征进行预测。若想因果卷积具有长期记忆, 就需要为其提供较大的感受野, 因此模型引入因果空洞卷积以保证一定的序列长度以及计算复杂度, 即向卷积核中添加值为零的权重来增加观察序列长度, 从而保证较为完整的特征, 因果空洞卷积如图 3 所示。

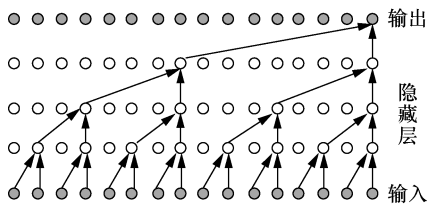


图3 因果空洞卷积

在图 3 中, 卷积核的大小为 2, 卷积感受野依次扩大了 1 倍、2 倍、4 倍、8 倍, 可观察到的序列

范围呈指数增长, 当卷积感受野扩大为原来的 8 倍时, 卷积可以观察到 16 个输入数据。

然后将输出接入卷积层进行特征提取, 将低层特征组成低级别胶囊。卷积核参数 $W \in R^{n \times 300}$, 其中 300 为词向量长度, n 为卷积核的大小。该卷积核提取的特征如下:

$$c_i = f(w \times x_{i:i+h-1} + b) \quad (1)$$

其中, w 为权重, b 为偏置项, f 为 ReLU 激活函数。

以 n 为窗口大小进行窗口滑动遍历, 可得到 $l-n+1$ 个特征, 其中 l 为输入文本矩阵的长度, 将特征进行组合可得到一个特征图 $C = [C_1, C_2, \dots, C_{l-n+1}]$ 。为了提取文本中不同的特征, 使用 m 个卷积核得到 m 个特征图。将 m 个特征图拼接在一起可得到低层胶囊的输入矩阵 Z 。 Z 是经过卷积神经网络得到的特征图矩阵, 其由 m 个列向量组成, 每个列向量由不同卷积核提取得到, 代表对同一文本的多角度特征, 计算过程如下:

$$Z = [C_1, C_2, \dots, C_m] = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{l-n+1,1} & c_{l-n+1,2} & \dots & c_{l-n+1,m} \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{l-n+1} \end{bmatrix} \quad (2)$$

3.3 交互注意力机制

上下文和方面词会影响彼此的向量表示, 因此引入交互注意力机制以减少上下文和方面词的相互影响, 即将 Q (query) 和一组 K-V (key-value) 对映射为标记单词的权重, 进而提高情感分类效果。对上下文隐含特征 h_n^c 和方面词隐含特征 h_m^a 进行求均值操作, 然后通过可训练参数将它们转换为统一维度 d_k , 得到 Q 矩阵, 计算式如下:

$$q_c = \frac{\sum_n h_n^c}{N} \times W_c^q \quad (3)$$

$$q_a = \frac{\sum_m h_m^a}{M} \times W_a^q \quad (4)$$

同样需要对上下文和方面词的 K 矩阵进行向量表示, 维度也转换为 d_k , 计算式如下:

$$k_n^c = h_n^c \times W_c^k \quad (5)$$

$$k_m^a = h_m^a \times W_a^k \quad (6)$$

其中, W_c^q 、 W_a^q 、 W_c^k 、 $W_a^k \in R^{2d_i \times d_k}$ 。

通过上下文特征矩阵和方面词均值可求得上下文对方面词的注意力矩阵 a_n^c ，类似地，可求得方面词对上下文的注意力矩阵 a_m^a ，计算式如下：

$$S_n^c(q_a, k_n^c) = \frac{k_n^c \times q_a^\top}{\sqrt{d_k}} \quad (7)$$

$$S_m^a(q_c, k_m^a) = \frac{k_m^a \times q_c^\top}{\sqrt{d_k}} \quad (8)$$

$$a_n^c = \frac{\exp(S_n^c(q_a, k_n^c))}{\sum_{n=1}^N \exp(S_n^c(q_a, k_n^c))} \quad (9)$$

$$a_m^a = \frac{\exp(S_m^a(q_c, k_m^a))}{\sum_{m=1}^M \exp(S_m^a(q_c, k_m^a))} \quad (10)$$

其中，除以 $\sqrt{d_k}$ 是为了避免 softmax 函数的梯度消失，减少矩阵维度对注意力得分的影响。

将上下文和方面词乘以各自的注意力矩阵以重构它们的特征表示，从而将上下文和方面词合并成初始胶囊，每层神经胶囊的 u （即经过序列卷积提取到的特征）都反映了文本语义、位置等属性特征。

3.4 胶囊网络

胶囊网络可以获得文本的局部特征和整体情感极性之间的关系，本文提出的模型将胶囊网络分为低胶囊层、高胶囊层、分类层3层。各胶囊层间共享全局参数以减少误差的累积，同时采用优化的动态路由算法迭代更新参数。此外，在胶囊网络的最后一层引入 Orphan 类别，以减少停用词等文本背景信息对文本分类效果的负面影响。胶囊网络整体架构如图4所示，其中初始胶囊为低层胶囊的第一层。

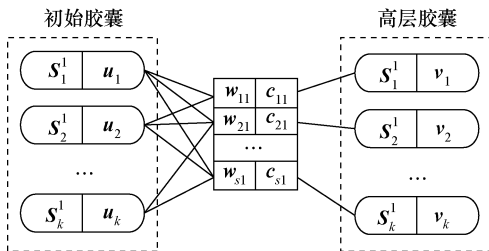


图4 胶囊网络整体架构

不同于传统神经网络通过加权求和得到的每个神经元都是一个标量值，胶囊网络中的每个胶囊都是可以对文本多个特征进行表示的向量，其

连接方式为全连接。由于低层胶囊传入的每一个胶囊神经元都是向量，因此对应的权值也为向量，同样利用反向传播来更新。在胶囊网络中，低层胶囊始终将其输出传递给对此表示“同意”的高层胶囊。

第 L 层初始胶囊的输入 u_i 是融合初阶信息后的向量，经过迭代更新后将参数传到 $L+1$ 层，得到输出 v_j 。与全连接神经网络相似，低层胶囊同样进行线性加权求和操作，计算代表低层胶囊的权重 c_{ij} ，并预测下一层胶囊。胶囊网络的输入 s_j 由式(11)得到：

$$s_j = \sum_i c_{ij} \hat{u}_{j|i}, \quad \hat{u}_{j|i} = w_{ij} u_i \quad (11)$$

其中， w_{ij} 是每个映射胶囊向量要乘的权重矩阵。动态路由算法的计算过程可以被看作上一层每一个胶囊神经元以不同强弱的连接输出到后一层的某一个神经元。动态路由算法如图5所示。

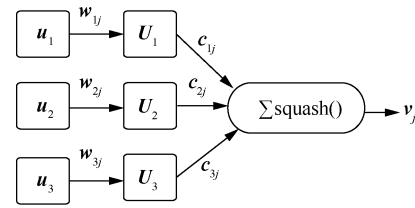


图5 动态路由算法

低层胶囊的所有权重 c_{ij} 通过 softmax 函数得到，以保证权重均为非负数且总和为1，从而说明 c_{ij} 的概率性质。计算式为：

$$c_{ij} = \text{softmax}(b_{ij}) \cdot \lambda_j \quad (12)$$

其中， b_{ij} 是一个临时变量，其初始值为0，它的值会在迭代过程中被更新并被保存在 c_{ij} 中； λ_j 为 v_j 的模，是被引入高层胶囊的权重系数，被用来修正下次迭代时低层胶囊与高层胶囊的连接强度，以此对动态路由算法进行优化。在胶囊网络中，使用非线性压缩函数 squash 作为激活函数：

$$v_j = \text{squash}(s_j) = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (13)$$

$$\lambda_j = \|v_j\| \quad (14)$$

squash 函数保留了向量原有的方向，同时将长度压缩到区间(0,1)，并生成传给所有高层胶囊的输出向量 v_j 。然后进行权重更新，计算式如下：

$$\mathbf{b}_{ij} = \mathbf{b}_{ij} + \hat{\mathbf{u}}_{ji} \cdot \mathbf{v}_j \quad (15)$$

通过点积运算检验胶囊网络的输入与输出的相似性,通过迭代更新 $\mathbf{b}_{ij}$ 来确定 $\mathbf{c}_{ij}$,即确定了一条路线,这条路线上的胶囊神经元的模都特别大,低层胶囊被高层胶囊选择的概率也很大,路线的尽头是那个被正确预测的胶囊。最终得到父层胶囊向量 $\mathbf{V}=[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s]$,其中 s 为输入文本情感极性的数量, $\mathbf{v}_i \in R^{n_s}$, n_s 为高层胶囊的维数。

然后通过全连接层进行降维,并将有效的特征进行整合。二维胶囊矩阵 $\mathbf{V}$ 通过铺平操作得到一维向量 $\mathbf{T}$,然后借助全连接层将其映射为一个长度为 l 的向量,其长度表示情感极性的数量。最后通过 softmax 函数计算出每个情感极性出现的概率:

$$\hat{y} = \text{softmax}(\mathbf{W}_l \times \mathbf{T} + \mathbf{b}) \quad (16)$$

其中, $\mathbf{W}_l$ 为权重, $\mathbf{b}$ 为偏置项,已知数据集标签为 y ,采用交叉熵作为损失函数,计算式如下:

$$\text{Loss}(\theta) = -\frac{1}{l} \sum_{i=1}^l y_i \log \hat{y} + \lambda \|\theta\|^2 \quad (17)$$

其中, θ 为模型参数, λ 为正则化系数。

4 实验结果及分析

4.1 实验数据

在情感分析实验中,采用 SemEval2014 任务 4 的酒店评论 (Restaurant 数据集) 和笔记本评论 (Laptop 数据集) 作为实验数据集,数据集中包含 3 种情感极性:积极情感、中性情感、消极情感,分别用 1、0、-1 表示,数据集详细信息见表 1。

4.2 模型参数

在固定其他参数的情况下,依次对某一可变参数进行调整,以取得在最好分类效果时的模型参数。对未登录的单词使用均匀分布 $U(-0.1, 0.1)$ 进行采样初始化,使用 0.001 学习率的 Adam 优化算法

来训练模型,进行 5 次重复实验,最终取平均值作为实验结果。在序列卷积中采用 ReLU 函数作为激活函数,并设置 dropout 为 0.2,从而将输出的结果传入胶囊网络,序列卷积层参数见表 2。

表 2 序列卷积层参数

参数	参数值
词向量维度	300
因果空洞卷积层数	5
卷积核大小	5
空洞因子	$2n$
残差模块数	4
激活函数	ReLU
优化函数	Adam
dropout	0.2

在胶囊层使用动态路由算法更新参数时,初始胶囊的个数根据句子长度确定。实验发现,当采用 $a=0.5$ 的 squash 函数时,模型性能比原始 squash 函数提升近 2%,因此采用 $\frac{\|\mathbf{s}_j\|^2 + \varepsilon}{0.5 + \|\mathbf{s}_j\|^2}$ 作为激活函数,胶囊网络参数见表 3。

表 3 胶囊网络参数

胶囊网络	可调参数	参数值
初始胶囊	胶囊维度	8
	层数	1
	参数共享	是
高层胶囊	胶囊维度	16
	胶囊个数	2
	迭代次数	3
	层数	1
	参数共享	是
	激活函数	$\frac{\ \mathbf{s}_j\ ^2 + \varepsilon}{0.5 + \ \mathbf{s}_j\ ^2}$

表 1

数据集详细信息

数据集	积极情感		中性情感		消极情感	
	训练集/个	测试集/个	训练集/个	测试集/个	训练集/个	测试集/个
Restaurant	2 164	728	637	196	807	196
Laptop	994	341	464	169	870	128

4.3 实验对比模型

为了对本文提出的模型进行分类效果进行客观的评估, 通过与以下模型进行对比分析, 研究 SCACaps 的情感分类效果。

(1) BiLSTM^[3]: 分别从前向和后向提取文本序列特征, 将二者整合并作为最终的分类结果, 然后输出。

(2) Seq-CNN^[10]: 将因果卷积和扩张卷积结合, 最后一层用卷积层代替全连接层构建序列卷积网络, 以改善网络特征选取的局限性。

(3) ATAE-LSTM^[4]: 将方面嵌入和单词表示结合起来, 通过 LSTM 提取隐藏特征, 然后引入注意力机制预测情感极性。

(4) IAN^[5]: 用 LSTM 分别提取上下文和方面词的隐藏特征, 通过交互注意力机制减少上下文和方面词向量表示的相互影响, 实现情感分类。

(5) MGAN^[11]: 模型分别利用细粒度和粗粒度的注意力机制, 进一步描述方面词和上下文之间的交互。

(6) Bi-IAN^[6]: 用 BiLSTM 分别提取上下文和方面词的隐藏特征, 引入交互注意力机制预测情感极性。

4.4 实验结果与分析

在实验过程中, 对比模型的模型参数尽量与第 4.2 节设定的参数保持一致, 使用 Early Stopping 策略训练模型, 以防止发生过拟合现象, 通过调整多组超参数来选择最佳分类效果。以 accuracy 和 F1 分数为评价指标, 实验结果见表 4、表 5。

从表 4 和表 5 可知, Seq-CNN 和 BiLSTM 的分类效果较为接近, 但在实验中, Seq-CNN 的运行速度比 BiLSTM 快一些。ATAE-LSTM 通过方面嵌入

表 4 多种模型实验结果 (accuracy)

模型	Restaurant	Laptop
BiLSTM	77.13%	68.76%
Seq-CNN	77.19%	68.78%
ATAE-LSTM	78.48%	69.87%
IAN	78.61%	72.18%
MGAN	81.24%	75.38%
Bi-IAN	79.38%	72.24%
SCACaps	81.64%	76.26%

表 5 多种模型实验结果 (F1 分数)

模型	Restaurant	Laptop
BiLSTM	63.74%	62.42%
Seq-CNN	63.79%	62.48%
ATAE-LSTM	63.38%	62.37%
IAN	68.13%	65.82%
MGAN	71.87%	72.38%
Bi-IAN	69.94%	66.74%
SCACaps	72.34%	72.43%

对方面词进行向量表示, 增强上下文和方面词间的交互, 分类效果有所提升。IAN、MGAN 与 Bi-IAN 引入交互注意力机制, 对上下文和方面词的特征向量进行重构, 从而使分类效果得到进一步提升, 其中 MGAN 分别以细粒度和粗粒度的方式进行情感分析, 显著提升了分类效果。

SCACaps 在所有模型中表现最好, 主要原因是胶囊网络可以抓取文本的局部特征和整体特征的信息, 从而获得方面词和上下文的关系, 进而更好地进行特征聚类及处理特征重叠的问题。

4.5 小样本数据集实验与分析

因为在真实场景训练模型时, 可能没有足够的已标注语料集, 所以笔者还在小样本数据集中进行了实验, 以研究模型的有效性。在实验中, 笔者分别从 Restaurant 数据集中随机选择 300 条数据作为训练语料, 进行 5 次实验, 并求其 accuracy 和 F1 分数的平均值, 实验结果见表 6。

表 6 小样本数据集实验结果

模型	accuracy	F1 分数
BiLSTM	63.76%	—
Seq-CNN	64.03%	—
ATAE-LSTM	65.42%	—
IAN	65.57%	—
MGAN	71.84%	65.27%
Bi-IAN	66.32%	62.69%
SCACaps	73.28%	67.14%

从小样本数据集的实验结果可以发现, 所有模型分类效果都有所下降。其中, 尽管 ATAE-LSTM、IAN、Bi-IAN 考虑了上下文和方面词对彼此的影响, 但与 BiLSTM、Seq-CNN 相比, 性能提升不大,

这说明交互注意力机制的引入对性能提升影响不大。通过实验可以得到, 在小样本中方面词与上下文对彼此的影响较小, 重构其向量表示没有太大意义。同时发现, 从不同的层次、粒度进行情感极性预测的 MGAN 表现突出, 其 accuracy 和 F1 分数皆较高, 因为其对文本进行了多层次情感粒度特征的抽取。SCACaps 仍表现最佳, 因为胶囊网络使用路由机制来生成高级特征, 利用其优越的空间信息提取能力, 可以提取出更丰富的文本信息。

5 结束语

针对方面级情感分类任务中存在特征表达重叠等问题, 同时考虑到胶囊网络出色的特征提取能力, 本文提出一种基于胶囊网络的方面级情感分类模型。模型通过序列卷积来较快地提取上下文和方面词的特征, 同时通过交互注意力机制减弱二者之间的相互影响, 在胶囊层间传递参数的过程中引入高层胶囊权重系数对算法进行优化, 一定程度上解决了同一文本中存在针对不同方面的多重情感的难题。通过对比实验分析发现, 本文提出的模型在文本分类任务中效果最佳。此外, 通过在小样本和学习任务中进行实验发现, 本文提出的模型也有不错的表现。在接下来的工作中, 可以将多层粒度分析引入胶囊网络进行方面级情感预测, 探讨模型的分类能力。

参考文献:

- [1] 赵传君, 王素格, 李德玉. 跨领域文本情感分类研究进展[J]. 软件学报, 2020, 31(6): 1723-1746.
ZHAO C J, WANG S G, LI D Y. Research progress on cross-domain text sentiment classification[J]. Journal of Software, 2020, 31(6): 1723-1746.
- [2] LIU B. Sentiment analysis and opinion mining[J]. Synthesis lectures on Human Language Technologies, 2012, 5(1): 1-167.
- [3] AUGENSTEIN I, ROCKTÄSCHEL T, VLACHOS A, et al. Stance detection with bidirectional conditional encoding[C]// The 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL Press, 2016: 876-885.
- [4] WANG Y Q, HUANG M L, ZHU X Y, et al. Attention-based LSTM for aspect-level sentiment classification[C]// The 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL Press, 2016: 606-615.
- [5] MA D H, LI S J, ZHANG X D, et al. Interactive attention networks for aspect-level sentiment classification[C]// The 26th International Joint Conference on Artificial Intelligence. CA: AAAI, 2017: 4068-4074.
- [6] 陈虹, 杨燕, 杜圣东. 用户评论方面级情感分析研究[J]. 计算机科学与探索, 2020: 已录用.
CHEN H, YANG Y, DU S D. Research on aspect-level sentiment analysis of user reviews[J]. Journal of Frontiers of Computer Science and Technology, 2020: accepted.
- [7] KIM Y. Convolutional neural networks for sentence classification[J]. arXiv preprint, 2014, arXiv: 1408.5882.
- [8] XUE W, LI T. Aspect based sentiment analysis with gated convolutional networks[C]// The 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL Press, 2018: 2514-2523.
- [9] 韩建胜, 陈杰, 陈鹏, 等. 基于双向时间深度卷积网络的中文文本情感分类[J]. 计算机应用与软件, 2019, 36(12): 225-231.
HAN J S, CHEN J, CHEN P, et al. Chinese text sentiment classification based on bidirectional temporal deep convolutional network[J]. Computer Applications and Software, 2019, 36(12): 225-231.
- [10] 高巍, 孙盼盼, 李大舟, 等. 基于序列卷积网络的 Twitter 情感分类[J]. 计算机工程与设计, 2020, 41(5): 1314-1319.
GAO W, SUN P P, LI D Z, et al. Twitter sentiment classification based on sequential convolution network[J]. Computer Engineering and Design, 2020, 41(5): 1314-1319.
- [11] FAN F F, FENG Y S, ZHAO D Y. Multi-grained attention network for aspect-level sentiment classification[C]// The 2018 Conference on Empirical Methods in Natural Language Processing. [S.l.:s.n.], 2018.
- [12] 曾锋, 曾碧卿, 韩旭丽, 等. 基于双层注意力循环神经网络的方面级情感分析[J]. 中文信息学报, 2019, 33(6): 108-115.
ZENG F, ZENG B Q, HAN X L, et al. Double attention neural network for aspect-based sentiment analysis[J]. Journal of Chinese Information Processing, 2019, 33(6): 108-115.
- [13] 曾义夫, 蓝天, 吴祖峰, 等. 基于双记忆注意力的方面级别情感分类模型[J]. 计算机学报, 2019, 42(8): 1845-1857.
ZENG Y F, LAN T, WU Z F, et al. Bi-memory based attention model for aspect level sentiment classification[J]. Chinese Journal of Computers, 2019, 42(8): 1845-1857.
- [14] HINTON G E, KRIZHEVSKY A, WANG S D. Transforming auto-encoders[C]// International Conference on Artificial Neural Networks. Heidelberg: Springer, 2012: 44-51.
- [15] HINTON G E, SABOUR S, FROSST N. Matrix capsules with EM routing[C]// International Conference on Learning Representations. [S.l.:s.n.], 2018.
- [16] CHEN Z, QIAN T Y. Transfer capsule network for aspect level sentiment classification[C]// The 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL Press, 2019: 547-556.
- [17] XIAO L Q, ZHANG H L, CHEN W Q, et al. Mcapsnet: capsule network for text with multi-task learning[C]// The 2018 Conference on Empirical Methods in Natural Language Processing. [S.l.:s.n.], 2018.
- [18] ZHAO W, YE J B, YANG M, et al. Investigating capsule networks with dynamic routing for text classification[C]// The 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL Press, 2018: 3110-3119.
- [19] 杨云龙, 孙建强, 宋国超. 基于 GRU 和胶囊特征融合的文本情感分析[J]. 计算机应用, 2020: 已录用.
YANG Y L, SUN J Q, SONG G C. Text sentiment analysis based on

feature fusion of gated recurrent unit and capsule[J]. Journal of Computer Applications, 2020: accepted.

- [20] 王敏蕊, 高曙, 袁自勇, 等. 基于动态路由序列生成模型的多标签文本分类方法[J]. 计算机应用, 2020: 已录用.

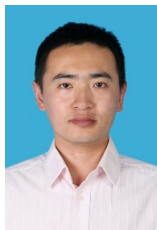
WANG M R, GAO S, YUAN Z Y, et al. Sequence to sequence model with dynamic routing for multilabel text classification[J]. Journal of Computer Applications, 2020: accepted.

- [21] 林悦, 钱铁云. 基于胶囊网络的跨领域情感分类方法[J]. 南京信息工程大学学报, 2019, 11(3): 286-294.

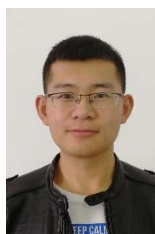
LIN Y, QIAN T Y. Cross-domain sentiment classification by capsule network[J]. Journal of Nanjing University of Information Science & Technology, 2019, 11(3): 286-294.

- [22] ZHANG B W, XU X F, YANG M, et al. Cross-domain sentiment classification by capsule network with semantic rules[J]. IEEE Access, 2018, 6: 58284-58294.

[作者简介]



徐志栋 (1981-), 男, 博士, 中国人民解放军国防大学国家安全学院研究员, 主要研究方向为自然语言处理、社会计算。



陈炳阳 (1993-), 男, 中国石油大学 (华东) 计算机科学与技术学院博士生, 主要研究方向为数据挖掘、自然语言处理、情感识别。



王晓 (1987-), 女, 博士, 青岛智能产业技术研究院常务副院长, 中国科学院自动化研究所复杂系统管理与控制国家重点实验室副研究员, 主要研究方向为社会计算、智能交通与车联网、人工智能等。



张卫山 (1970-), 男, 博士, 中国石油大学 (华东) 计算机科学与技术学院教授、博士生导师, 黄岛区“智能大数据处理创新人才团队”负责人, 中国石油大学“石油大数据处理”团队负责人, 主要研究方向为大数据智能处理、人工智能。