

文章编号: 2096-1618(2020)05-0524-07

基于胶囊特征聚合的评价词和评价对象抽取

康 刚, 吴四九, 方 睿

(成都信息工程大学计算机学院, 四川 成都 610225)

摘要:为提高传统门控制单元(GRU)的特征抽取能力,受独立循环神经网络(independently recurrent neural network, IndRNN)的启发,提出独立门控制单元(independently gate recurrent unit, IndGRU),实验结果表明 IndGRU 在3个数据集上的特征提取能力优于传统 GRU 和长短期记忆网络(LSTM),证明了 IndGRU 的有效性。针对传统评价词和评价对象抽取方法不能很好地利用抽象特征之间的关联关系问题,提出一种基于胶囊特征聚合的评价词和评价对象抽取模型,模型使用多个双向 IndGRU 并行提取文本上下文信息,构造胶囊特征,使用动态路由算法利用特征间关系实现胶囊特征的聚合,最后使用条件随机场(CRF)完成序列标注。模型在3个基准数据集上取得了比目前的先进方法较好或相当的效果,证明了模型的有效性。

关 键 词:胶囊网络;动态路由算法;评价词;评价对象;独立门控制单元;序列标注

中图分类号:TP391.1

文献标志码:A

doi:10.16836/j.cnki.jcuit.2020.05.008

0 引言

情感分析就是对文本数据中的情感信息的分析和挖掘,而在评价文本中包含了两种对情感分析至关重要的词:评价词和评价对象。在商品评论中,评价词就是表达评论者对产品或产品某一方面的看法或情感倾向的词,评价对象就是描述产品或产品某方面特征的词^[1]。评价词能直接影响对评论者的情感倾向的判断,评价对象词包含评论者的评论所针对的具体细粒度方面信息,有时需要将评价词和评价对象所反映的评价实体的特有特征属性相结合,才能决定评价文本的情感倾向。评价词和评价对象的抽取是情感分析中的重要任务。

评价词和评价对象间具有较强的关联性,利用这种关联关系可以促进评价词和评价对象的抽取效果^[2-3]。GRU 是一种结构简单、参数较少、易于扩展的循环神经网络,使用 GRU 去获取文本中的上下文关联信息。Li 等^[4]提出独立循环神经网络,使用 Hadamard 积替换 RNN 隐藏层计算过程中的矩阵运算,将 RNN 层内神经元解耦,提高网络的可解释性,IndRNN 比传统 LSTM 具备更好的长期记忆能力。受文献^[4]的启发,提出 IndGRU 结构,提高模型对长距离的关联信息的识别效果。

胶囊网络是一种利用胶囊特征的神经网络,包含两个重要组件:胶囊和动态路由算法。胶囊是一组神经元的集合,用来表示一类特征,胶囊的输入和输出都是矢量,与传统标量特征相比,胶囊矢量有更加丰富的特征表达^[5]。动态路由算法是实现胶囊间信息传递

的方法,能够获取局部和整体间的特征关系^[6],实现胶囊特征的聚合。文献^[7]提出胶囊式 LSTM,将胶囊网络思想应用到时间序列数据的特征抽取中,通过动态路由算法处理特征之间的语义关联信息。受文献^[7]的启发,提出一种结合动态路由算法、双向独立门控制单元(BiIndGRU)和 CRF 的评价词和评价对象联合抽取模型,是一个使用胶囊特征聚合机制和 CRF 相结合的模型(capsule feature aggregation conditional random fields, CFACRF)。

1 相关工作

评价词和评价对象的抽取方法主要有基于规则的方法、基于统计机器学习的方法。基于规则的方法主要是直接利用评价文本中的文本特性,手动定义抽取规则,再利用模式匹配等方法实现词的获取。文献^[8]利用句子的依存关系树(dependency tree)和单词词性获取评价词和评价对象。文献^[9]通过设计算法对手工定义规则的筛选,优化抽取规则,从而提高词抽取的准确性。基于规则的方法需要一定的专家知识的支撑,且容易受到定义规则的完备性和正确性的影响。

基于统计机器学习的方法,一般是将词抽取问题转化为序列标注问题。传统统计机器学习方法需要利用特征工程构造特征,常见特征有词性、位置特征或语法特征,再利用统计机器学习获取特特征和标签间的映射关系。常使用 CRF^[10]、隐马尔科夫模型^[11]等实现序列标注。深度学习是机器学习的一种,深度学习可以自动学习文本特征,避免复杂的特征工程。文献^[12]将长短时记忆神经网络模型应用在评价词和评价对象的抽取任务中。结合文本的语法规则可以提高

深度学习方法的效果,一种是利用文本语法得到文本特征再输入深度学习模型中,文献[3]使用递归神经网络学习文本的语法依赖树信息,得到评价词和评价对象间双向传播特征,再利用 CRF 实现序列标注。文献[13]利用语法规则构造关系嵌入(relation embedding)和子树嵌入(subtree embedding)两种特征,再利用卷积神经网络(convolutional neural network,CNN)来结合这两种特征。另外一种是利用语法规则构造用于深度学习模型训练的伪标签数据(pseudo labeled data),文献[1]通过对预定义的语法规则的匹配,构造模型预训练数据,从而提高模型的词抽取效果。文献[14]使用语法规则获取文本中的词,并设计过滤算法优化抽取的结果,构造伪标签数据,最后使用伪标签数据来训练深度 GRU 网络。

传统基于统计机器学习的方法主要使用的是标量特征,文中提出的 CFACRF 模型使用了胶囊特征,胶囊中使用矢量表达特征,矢量具有长度和方向,能表达的特征信息更加丰富,模型使用动态路由算法利用胶囊特征间的空间关系,实现胶囊特征的聚合,实现对评价词和评价对象抽取效果的提升。

2 基于胶囊特征聚合的评价词和评价对象抽取模型

评价词和评价对象抽取问题是一个序列标注问题,主要关注词在句子中的位置,文中采用 BIO(begin,inside,outside)标签方案^[15]。定义 5 种标签 {B-T,I-T,B-O,I-O,O},分别表示 {评价对象的开始,评价对象的内部,评价词的开始,评价词的内部,其他词},一个简短示例如表 1 所示,“love”是评价词,标签是 B-O,operating 是评价对象的开始,标签是 B-T,system 是评价对象中的词,标签是 I-T。

表 1 标注示例

句子	I	love	the	operating	system
标签	O	B-O	O	B-T	I-T

基于胶囊特征聚合的评价词和评价对象抽取模型结构如图 1 所示,模型主要包含词嵌入层、并行双向 IndGRU 层、动态路由层和 CRF 层 4 部分。

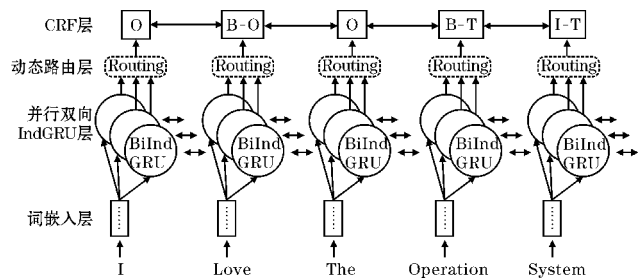


图 1 CFACRF 模型结构

2.1 词嵌入层

使用 word2vector 方法预训练长度为 100 的词矢量。词嵌入层接受文本句子作为输入,利用预训练的词矢量,把输入文本转换成矢量矩阵。

2.2 并行双向独立 GRU 层

GRU 是一种循环神经网络,通过内部的门控单元,控制时间步间的信息流通,防止长期记忆中的梯度消失和梯度爆炸问题。GRU 有两个门控制单元:更新门和重置门,两个门的值都是通过上一时刻的隐藏层状态和当前时刻的输入计算得到的。本文主要使用 Hadamard 乘积替换原始 GRU 单元中隐藏状态的部分矩阵计算,最终得到的 IndGRU 单元中隐藏层输出计算过程如下:

$$r_t = \text{sigmoid}(W_r x_t + U_r \odot h_{t-1} + b_r) \quad (1)$$

$$z_t = \text{sigmoid}(W_z x_t + U_z \odot h_{t-1} + b_z) \quad (2)$$

$$\tilde{h}_t = \tanh(W_h x_t + r_t \odot U_i \odot h_{t-1} + b_h) \quad (3)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

其中 U_r, U_z, U_i 都是一维矢量,“ \odot ”表示矢量中对应元素相乘的 Hadamard 乘积。普通的 IndGRU 只保存当前时间步以前的信息,文本中单词间的关系是相互的,所以需要同时获取当前时间步以后的信息,需要使用双向 IndGRU。如图 2 所示,BiIndGRU 便是用两个单独的 IndGRU 网络结构分别获取前面时间步和后面时间步的信息, \vec{h}_t 为向前的 IndGRU 单元隐藏状态输出, \overleftarrow{h}_t 为向后的 IndGRU 隐藏层输出,使用连接向量 $h_t = [\vec{h}_t; \overleftarrow{h}_t]$ 为 BiIndGRU 在 t 时刻隐藏状态输出。

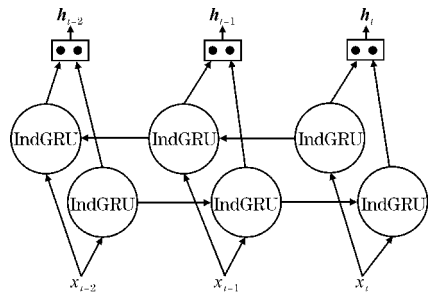


图 2 BiIndGRU 结构

并行双向 IndGRU 层使用多个 BiIndGRU 在同一时刻并行的提取特征,并添加了 Dropout 操作来提高模型泛化效果,通过 IndGRU 构造胶囊特征的计算公式如下:

$$h_t^i = \text{BiIndGRU}_i(x_t) \quad (5)$$

其中, x_t 表示 t 时刻的输入, BiIndGRU_i 表示第 i 个双向 IndGRU 的数据操作函数, h_t^i 表示第 i 个双向 IndGRU 在

t 时刻的隐藏层特征,表示初始的特征胶囊。

2.3 动态路由层

动态路由算法是胶囊网络中相邻两层胶囊间信息传递的算法,动态路由算法可以实现胶囊特征的聚合。算法过程如图3所示,其中, \mathbf{h}_t^i 表示 t 时刻得到的第 i 个初始胶囊, \mathbf{v}_t^j 表示聚合后的第 j 个输出胶囊, $\hat{\mathbf{u}}_{j|i}$ 表示初始胶囊 \mathbf{h}_t^i 对输出胶囊 \mathbf{v}_t^j 生成的预测胶囊。算法主要包括两个阶段:预测矢量 $\hat{\mathbf{u}}_{j|i}$ 的计算阶段和最终输出胶囊 \mathbf{v}_t^j 计算阶段。

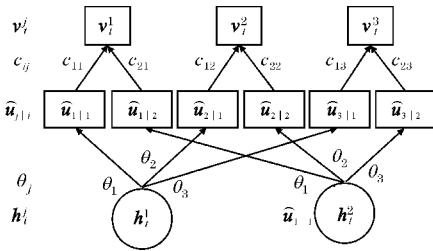


图3 动态路由算法

在动态路由算法中初始胶囊和输出胶囊间采用全连接形式链接,预测矢量 $\hat{\mathbf{u}}_{j|i}$ 通过一个全连接层得到,计算方式如下:

$$\hat{\mathbf{u}}_{j|i} = f(\mathbf{h}_t^i, \theta_j) \quad (6)$$

其中的 θ_j 是全连接层的参数。设 c_{ij} 为初始胶囊 \mathbf{h}_t^i 和输出胶囊 \mathbf{v}_t^j 间的耦合系数 (coupling coefficient), 决定初始胶囊传递多少信息到输出胶囊中, 则输出胶囊 \mathbf{v}_t^j 就是所有初始胶囊对其的预测矢量的线性加权和, 计算方式如下:

$$\mathbf{v}_t^j = \sum_i c_{ij} \hat{\mathbf{u}}_{j|i} \quad (7)$$

设 b_{ij} 为初始胶囊 \mathbf{h}_t^i 连接到输出胶囊 \mathbf{v}_t^j 的先验概率, 初始时值为0, 则耦合系数 c_{ij} 是通过 b_{ij} 得到, 为过滤掉文本中背景信息的干扰, 采用文献[16]提出的 Leaky-Softmax 函数计算胶囊间的耦合系数, 计算方式如下:

$$c_{ij} = \frac{\exp(b_{ij})}{1 + \sum_{k=1}^m \exp(b_{ik})} \quad (8)$$

其中 m 为初始胶囊的个数。在胶囊网络中, 胶囊矢量表示某类特征, 其长度可以表示某类特征存在的概率, 设 $\hat{\mathbf{v}}_t^j$ 表示输出胶囊 \mathbf{v}_t^j 代表的某类特征的特征矢量, $\hat{\mathbf{v}}_t^j$ 通过 Squashing 函数得到, Squashing 函数是文献[6]提出的非线性映射函数, 确保短矢量长度缩减到0附近, 长矢量缩减到略小于1的长度, 计算方式如下:

$$\hat{\mathbf{v}}_t^j = \frac{\|\mathbf{v}_t^j\|^2}{1 + \|\mathbf{v}_t^j\|^2} \frac{\mathbf{v}_t^j}{\|\mathbf{v}_t^j\|} \quad (9)$$

在胶囊网络中, 胶囊矢量的方向代表某类特征实

例, 为使预测胶囊矢量传递更多信息到与其方向相近的输出胶囊中, 所以通过预测胶囊矢量和最终输出胶囊矢量的点积来更新 b_{ij} , 计算方式如下:

$$b_{ij} = b_{ij} + \hat{\mathbf{u}}_{j|i} \cdot \hat{\mathbf{v}}_t^j \quad (10)$$

动态路由算法中, 需要通过公式(7)~(10)所示步骤迭代式地动态更新 c_{ij} 的值, 强化方向相近的胶囊间的连接强度, 让预测胶囊信息更多地传递到方向相近的输出胶囊矢量中。

动态路由层在 t 时刻的最终输出 \mathbf{v}_t 是通过输出胶囊的拼接得到的, 计算方式如下:

$$\mathbf{v}_t = [\mathbf{v}_t^1; \mathbf{v}_t^2; \dots; \mathbf{v}_t^j] \quad (11)$$

使用的动态路由算法和原始的动态路由算法的最大区别是胶囊矢量的输入输出矢量未经过 Squashing 函数的压缩, CFACRF 是复合模型, 动态路由算法只用于特征的聚合部分, Squashing 可能会带来数据信息的损失, 所以 Squashing 函数只用于胶囊间耦合系数的计算。

利用不同路由策略来实现了4个CFACRF模型。

2.3.1 CFACRF

模型使用上面介绍的原始路由策略, 胶囊间耦合系数的计算如式(8)所示。

2.3.2 CFACRF-Mod

模型使用文献[16]提出的胶囊间连接强度的修正策略。由于胶囊网络中, 胶囊矢量长度可以代表特征存在的概率, 因此, 可以用初始胶囊矢量的长度修正胶囊间的耦合系数, 计算方式如下:

$$\mathbf{u}_i = \text{squashing}(\mathbf{h}_t^i) \quad (12)$$

$$a_i = \|\mathbf{u}_i\| \quad (13)$$

$$c_{ij} = a_i \cdot \text{leaky-softmax}(\mathbf{b}_{ij}) \quad (14)$$

其中 \mathbf{u}_i 表示初始胶囊 \mathbf{h}_t^i 经过 Squashing 函数得到的特征矢量, Squashing 函数如式(9)所示, a_i 表示矢量 \mathbf{u}_i 的长度, Leaky-Softmax 函数如式(8)所示。

2.3.3 CFACRF-Rev

模型采用文献[17]提出的反转动态路由 (reversed dynamic routing) 策略。普通路由策略中, c_{ij} 是决定初始胶囊分别传递多少信息到各个输出胶囊中。而在反转路由策略, c_{ij} 确定输出胶囊分别接受多少初始胶囊的信息。因此普通路由策略和反转动态路由策略主要区别在于由 b_{ij} 得到 c_{ij} 的计算方式。在反转动态路由策略中, c_{ij} 的计算方式如下:

$$c_{ij} = \frac{\exp(b_{ij})}{1 + \sum_{k=1}^m \exp(b_{kj})} \quad (15)$$

其中 m 为初始胶囊的个数。

2.3.4 CFACRF-Mod-Rev

模型中既使用反转路由策略, 也使用修正路由策

略,则 c_{ij} 的计算过程如式 (12) ~ (14) 所示,其中式 (14) 的 Leaky-Softmax 函数计算方式如式 (15) 所示。

2.4 CRF 层

CRF 是由 Lafferty 等^[18] 提出的概率模型,常用于序列标注问题。CRF 能够利用序列元素的标签间的关系,得到全局最优的标签序列。CFACRF 模型使用一个全连接层将动态路由层的输出输入到 CRF 中,实现对文本序列的最终标注。

3 实验

3.1 数据集与评测标准

使用 3 个数据集:SemEval-2014 Restaurants (SE14-L)、SemEval-2014 Laptops (SE14-R)、SemEval-2015 Restaurants (SE15-R),3 个数据集来源于国际语义评测比赛任务原始数据集只有评价对象标签,经过 Wang 等^[3,19] 的标注,增加评价词的标签数据集详细信息如表 2 所示。

表 2 数据集信息				
数据集	数据集划分	样本数	评价对象词数	评价词词数
SE14-L	训练集	3048	2373	2520
	测试集	800	654	678
SE14-R	训练集	3044	3699	3528
	测试集	800	1134	1021
SE15-R	训练集	1315	1279	1216
	测试集	685	597	517

实验使用 3 个指标,分别是准确率 P 、召回率 R 、 $F1$,在测试集上的计算方式如下:

$$P = \frac{\text{模型抽取出的正确词数}}{\text{模型抽取出的全部词数}} \tag{16}$$

$$R = \frac{\text{模型抽取出的正确词数}}{\text{测试集中的全部词数}} \tag{17}$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{18}$$

3.2 模型参数

标签方案中,定义了 5 种标签,模型中的输出胶囊个数设为 5,每个输出胶囊表示一种标签对应的特征。不同动态路由迭代次数下 CFACRF 模型在 SE14-R 上的评价词和评价对象联合抽取 $F1$ 值对比如图 4 所示。

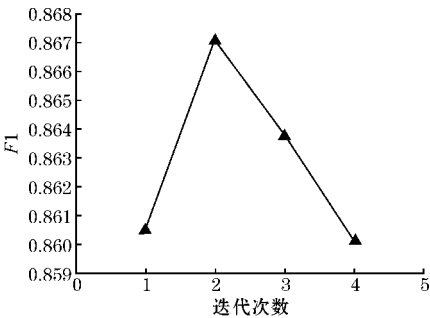


图 4 不同迭代次数的 $F1$ 值对比

从图 4 可知,当迭代次数为 2 时,模型的 $F1$ 值最高。不同 BiIndGRU 个数下 CFACRF 模型在 SE14R 上的 $F1$ 值对比如图 5 所示。

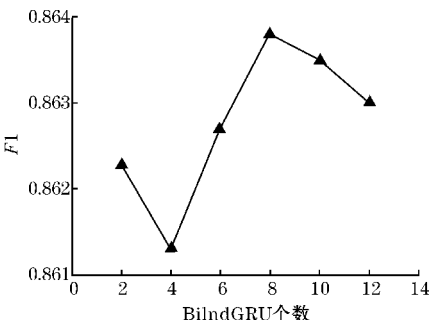


图 5 不同 BiIndGRU 个数的 $F1$ 值对比

从图 5 可知,当模型使用 8 个 BiIndGRU 时,模型取得最高的 $F1$ 值。模型最终的参数数据如表 3 所示。

表 3 模型参数	
参数	取值
IndGRU 隐藏层特征的数量	100
BiIndGRU 个数	8
动态路由迭代次数	2
dropout 率	0.5
学习率	0.001
批处理大小	25
输出胶囊个数	5
输出胶囊维度	12

3.3 实验环境

试验在 Windows 操作系统的计算机上开展,使用 Pycharm 开发平台,编程语言使用 python3.6,利用 tensorflow-gpu1.12 实现深度学习模型,计算机处理器为 Intel core i7-8750,GPU 为 GTX1060。

3.4 实验结果

为对比分析 IndGRU 的有效性,使用 CRF 结合单个双向门控制单元 (BiGRU) 组建模型 BiGRUCRF,使用 CRF 结合单个双向长短时记忆 (BiLSTM) 组建模型 BiLSTMCRF,使用 CRF 结合单个 BiIndGRU 组建复合模型 BiIndGRUCRF。3 个模型的结构如图 6 所示,3 个模型都使用词嵌入层实现文本到词矢量的转换,都使用一个全连接层将获取到的特征数据输入 CRF 中,3 个模型的

区别便是使用不同的循环神经网络提取特征。

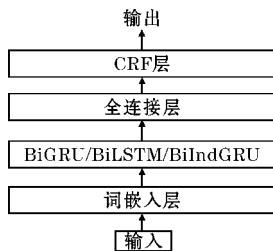


图 6 BiGRUCRF、BiLSTMCRF、BiIndGRUCRF 结构

在 3 个数据集上开展实验,将评价词和评价对象看作同一类抽取目标词,计算评价词和评价对象联合抽取总的指标值,最后的实验结果如表 4 所示。从表 4 可以看出,与 BiGRUCRF 和 BiLSTMCRF 两个模型相比,BiIndGRUCRF 模型在 3 个数据集上的 $F1$ 值分别提高了2.55%、1.36%、0.8%,说明 IndGRU 在 3 个数据集上的特征获取能力比传统 GRU 和 LSTM 都好,IndGRU 结构是有效的。

表 4 模型联合抽取评价指标对比 %

模型	SE14-L			SE14-R			SE15-R		
	P	R	$F1$	P	R	$F1$	P	R	$F1$
BiGRUCRF	78.53	76.05	77.27	82.7	84.97	83.82	67.13	69.69	68.39
BiLSTMCRF	77.75	75.83	76.78	84.7	83.99	84.34	68.85	70.2	69.51
BiIndGRUCRF	78.75	80.82	79.82	85.71	85.69	85.7	69.68	70.97	70.31

为分析模型的评价词和评价对象抽取效果,文中选择 7 个公开模型分别在 3 个数据集上对比评价词和评价对象的 $F1$ 值。7 个公开模型介绍如下:

(1)DP 是 Qiu 等^[2]提出的基于手工规则的评价词和评价对象抽取方法,利用评价词和评价对象的语法和结构关联关系,通过手工定义 8 条规则,完成词的抽取。

(2)RNCRF 是 Wang 等^[3]提出的结合句法规则和神经网络的组合模型,先获取文本的句法规则树,并作为递归神经网络的输入,最后利用 CRF 模型完成单词的序列标注。

(3)CMLA 是 Wang 等^[19]提出的一个基于注意力机制的深度神经网络模型,是一个多层注意力网络,模型利用注意力机制获取评价词和评价对象间的直接或间接关联关系,帮助评价词和评价对象的识别。

(4)NCRF-AE 是 Zhang 等^[20]提出的使用 CRF 的神经网络自编码模型。模型包含两部分,一个由深度神经网络增强的 CRF 编码器和一种试图重建输入的解码器,模型采用半监督学习方式完成序列结构预测任务。

(5)HAST 是 Li 等^[21]提出的基于截断历史注意力 (truncated history-attention) 和选择性转移网络 (selective transformation network) 的评价对象抽取方法。

(6)DE-CNN 是 Xu 等^[22]提出的使用通用的词嵌入和特定域的词嵌入的双嵌入 (double embeddings) 作为卷积神经网络输入的评价对象抽取方法。

(7)RINANTE 是 Dai 等^[1]提出的一种结合 BiLSTM 和 CRF 的模型,模型使用利用语法规则构造的辅助训练数据对模型进行预训练。

模型在 3 个数据集上评价词和评价对象分别的 $F1$ 值对比如表 5 所示,其中“-”表示没有数据。

表 5 评价词和评价对象的 $F1$ 值对比

模型	SE14-L		SE14-R		SE15-R	
	评价词	评价对象	评价词	评价对象	评价词	评价对象
DP	55.29	19.19	65.94	38.72	46.31	27.32
RNCRF	77.03	75.28	83.93	82.23	63.75	65.39
CMLA	79.16	73.63	84.67	82.46	70.50	68.22
NCRF-AE	75.44	74.32	85.23	83.28	70.16	65.33
HAST	—	79.52	—	85.61	—	69.77
DE-CNN	—	81.59	—	85.20	—	68.28
RINANTE	81.96	81.37	86.34	86.76	72.09	71.89
CFACRF	82.1	79.46	88.1	85.45	71.75	70.64
CFACRF-Mod	79.22	81.84	87.5	86.15	72.11	69.31
CFACRF-Rev	82.09	79.91	86.96	85.9	71.48	72.08
CFACRF-Mod-Rev	81.7	79.75	86.83	85.73	70.44	71.19

由表5可知,CFACRF模型对SE14-R数据集中的评价词抽取效果提升最多, $F1$ 值比RINANTE模型提高了1.76%。但4个基于胶囊特征聚合的模型在SE14-R中的评价对象抽取效果表现最差,可能是由于文中模型是评价词和评价对象的联合抽取模型,评价词和评价对象的抽取效果可能会存在此消彼长的情况,导致模型在SE14-R中评价词的 $F1$ 值较高,而评价对象的抽取效果较差。使用4种不同路由策略的模型中,结合反转路由策略和修正路由策略的CFACRF-Mod-Rev模型在数据集上的表现最差,CFACRF和CFACRF-Mod表现效果较好。CFACRF模型在SE14-L上的评价词提升了0.14%。CFACRF-Mod模型在SE14-L的评价对象的抽取 $F1$ 值提高了0.25%。在SE15-R上,评价词和评价对象的 $F1$ 值分别提高了0.02%、0.19%。结果证明提出的模型通过动态路由算法利用特征间关系聚合胶囊特征,获取了更有效的特征信息,基于胶囊特征聚合的模型是有效的,并具有先进性。

4 结束语

通过对传统GRU单元内部计算的简单改进,提出IndGRU结构,并通过实验证明IndGRU在3个数据集上的特征抽取效果优于传统GRU和LSTM。提出基于胶囊特征聚合的序列标注模型,研究了将动态路由算法用于评价词和评价对象抽取问题中,模型使用多个BiIndGRU并行获取上下文特征信息,利用动态路由算法利用特征间关系实现特征的聚类,并通过实验证明了模型的有效性。在3个英文数据集上开展实验,下一步可以将模型应用到中文数据集,或其他自然语言处理任务中。可以进一步优化词矢量的预训练过程,如采用ELMo或BERT等。未来还可以寻找更有效的动态路由算法。

参考文献:

- [1] Dai H, Song Y. Neural aspect and opinion term extraction with mined rules as weak supervision [C]. Proceedings of the 57th Conference of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2019: 5268–5277.
- [2] Qiu G, Liu B, Bu J, et al. Opinion word expansion and target extraction through double propagation [J]. Computational Linguistics, 2011, 37(1): 9–27.
- [3] Wang W, Pan S J, Dahlmeier D, et al. Recursive neural conditional random fields for aspect-based sentiment analysis [C]. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Stroudsburg, PA: ACL, 2016: 616–626.
- [4] Li S, Li W, Cook C, et al. Independently recurrent neural network (indrnn): building a longer and deeper rnn [C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2018: 5457–5466.
- [5] Hinton G E, Krizhevsky A, Wang S D. Transforming auto-encoders [C]. 21st International Conference on Artificial Neural Networks. Berlin, Germany: Springer, 2011: 44–51.
- [6] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules [C]. Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017. California, USA: NIPS Proceeding, 2017: 3856–3866.
- [7] 郑毅. 时间序列数据的胶囊式LSTM特征提取算法研究 [D]. 武汉: 华中师范大学, 2018.
- [8] Xu L, Liu K, Lai S, et al. Walk and learn: a two-stage approach for opinion words and opinion targets co-extraction [C]. Proceedings of the 22nd International Conference on World Wide Web. Rio de Janeiro, Brazil: International World Wide Web Conferences Steering Committee, 2013: 95–96.
- [9] Liu Qian, Gao Zhiqing, Liu Bing, et al. Automated rule selection for aspect extraction in opinion mining [C]. Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2015: 1291–1297.
- [10] Niklas Jakob, Iryna Gurevych. 2010. Extracting opinion targets in a single and cross-domain setting with conditional random fields [C]. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2010: 1035–1045.
- [11] Jin W, Ho H H, Srihari R K. A novel lexicalized HMM-based learning framework for web opinion mining [C]. Proceedings of the 26th annual international conference on machine learning. New York: ACM, 2009: 465–472.
- [12] Liu P, Joty S, Meng H. Fine-grained opinion mining with recurrent neural networks and word embeddings [C]. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2015: 1433–1443.

- [13] Zhao Y, Qin B, Liu T. Encoding syntactic representations with a neural network for sentiment collocation extraction[J]. Science China Information Sciences, 2017, 60(11): 3–14.
- [14] Wu C, Wu F, Wu S, et al. A hybrid unsupervised method for aspect term and opinion target extraction[J]. Knowledge-Based Systems, 2018, 148: 66–73.
- [15] Ratnoff L, Roth D. Design challenges and misconceptions in named entity recognition[C]. Proceedings of the Thirteenth Conference on Computational Natural Language Learning. Stroudsburg, PA: ACL, 2009: 147–155.
- [16] 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. Stroudsburg, PA: Association for Computational Linguistics, 2018: 3110–3119.
- [17] Gong J, Qiu X, Wang S, et al. Information aggregation via dynamic routing for sequence encoding[C]. Proceedings of the 27th International Conference on Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2018: 2742–2752.
- [18] John Lafferty, Andrew McCallum, Fernando Pereira. Conditional random fields: probabilistic models for segmenting and labeling sequence data[C]. Proceedings of the 8th International Conference of Machine Learning. San Mateo, CA: Morgan Kaufmann, 2001: 282–289.
- [19] Wang W, Pan S J, Dahlmeier D, et al. Coupled multi-layer attentions for co-extraction of aspect and opinion terms[C]. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2017: 3316–3322.
- [20] Zhang X, Jiang Y, Peng H, et al. Semi-supervised structured prediction with neural CRF autoencoder[C]. Proceedings of the 2017 conference on empirical methods in natural language processing (EMNLP 2017). Stroudsburg, PA: Association for Computational Linguistics, 2017: 1701–1711.
- [21] Li X, Bing L, Li P, et al. Aspect term extraction with history attention and selective transformation[C]. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018). San Mateo, CA: Morgan Kaufmann, 2018: 4194–4200.
- [22] Xu H, Liu B, Shu L, et al. Double embeddings and cnn-based sequence labeling for aspect extraction[C]. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2018: 592–598.

Extraction of Opinion Terms and Opinion Targets based on Capsule Feature Aggregation

KANG Gang, WU Sijiu, FANG Rui

(College of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: In order to improve the feature extraction ability of Gate Recurrent Unit, inspired by the work of Independently Recurrent Neural Network(IndRNN), Independently Gate Recurrent Unit(IndGRU) is proposed. The experimental results show that the ability of feature extraction of IndGRU is superior to the traditional GRU and LSTM on three benchmark datasets, which prove the IndGRU is effective. Due to that the traditional model for extraction of opinion targets and opinion terms can not make good use of correlation between abstract features, an opinion targets and opinion terms extraction model based on capsule feature aggregation is proposed. This model extracts different context information features by using multiple bidirectional IndGRU in parallel, capsule feature is constructed based on these extracted features, and relationship between features is used by the dynamic routing algorithm to aggregate the capsule features. Finally, the aggregated features were used in Conditional Random Field to complete the sequence labeling. The model has achieved performance better than or comparable with the current advanced methods on three benchmark data sets, which proves the effectiveness of this model.

Keywords: capsule network; dynamic routing algorithm; opinion targets; opinion terms; IndGRU; sequence labeling