

基于知识图谱与循环注意力网络的 视角级情感分析

邓立明^{1,2,3}

魏晶晶⁴

吴运兵^{1,2,3}

余小燕^{1,2,3}

廖祥文^{1,2,3}

摘要 现有的视角级情感分析方法难以解决单词在不同语境下“一词多义”问题,因此性能受限. 针对上述问题,文中提出基于知识图谱与循环注意力网络的视角级情感分析方法. 首先,利用动态注意力机制,结合双向长短时记忆网络的文本表示和知识图谱中的同义词信息,获得知识感知状态向量. 再联合位置信息构造记忆内容,并输入多层门限循环单元,计算视角词情感特征,进行视角级文本情感分类. 在3个公开数据集上的实验表明,文中方法分类效果较优.

关键词 视角级情感分析, 知识图谱, 注意力机制, 深度学习

引用格式 邓立明,魏晶晶,吴运兵,余小燕,廖祥文. 基于知识图谱与循环注意力网络的视角级情感分析. 模式识别与人工智能, 2020, 33(6): 479-487.

DOI 10.16451/j.cnki.issn1003-6059.202006001

中图法分类号 TP 391

Aspect Level Sentiment Analysis Based on Knowledge Graph and Recurrent Attention Network

DENG Liming^{1,2,3}, WEI Jingjing⁴, WU Yunbing^{1,2,3}, YU Xiaoyan^{1,2,3}, LIAO Xiangwen^{1,2,3}

ABSTRACT The existing aspect level sentiment analysis methods cannot solve the problem of polysemous word in different contexts. Therefore, a method for aspect level sentiment analysis based on knowledge graph and recurrent attention network is proposed. The text representation of the bidirectional long short-term memory network is integrated with synonym information in knowledge graph using dynamic attention mechanism to obtain the state vector of knowledge perception. To classify aspect level sentiment, the memory content is constructed by combining the location information and inputting the multi-level gated recurrent unit for calculating the sentiment characteristics of aspect terms. The experimental results show that the proposed method achieves better classification results on three open datasets.

收稿日期:2020-03-06;录用日期:2020-05-29

Manuscript received March 6, 2020;

accepted May 29, 2020

国家自然科学基金项目(No. 61976054, 61772135, U1605251)、

国家自然科学基金青年基金项目(No. 41801324)、福建省自然

科学基金面上项目(No. 2017J01755)、模式识别国家重点

实验室开放课题基金项目(No. 201900041)资助

Supported by National Natural Science Foundation of China(No.

61976054, 61772135, U1605251), National Natural Science

Youth Fund Project(No. 41801324), Natural Science Founda-

tion of Fujian Province(No. 2017J01755), Open Project of Na-

tional Laboratory of Pattern Recognition in China(No. 2019000

41)

本文责任编辑 陈恩红

Recommended by Associate Editor CHEN Enhong

1. 福州大学 数学与计算机科学学院 福州 350116

2. 福州大学 福建省网络计算与智能信息处理重点实验室

福州 350116

3. 数字福建金融大数据研究所 福州 350116

4. 福建江夏学院 电子信息科学学院 福州 350108

1. College of Mathematics and Computer Science, Fuzhou Uni-

versity, Fuzhou 350116
2. Fujian Provincial Key Laboratory of Networking Computing
and Intelligent Information Processing, Fuzhou University,
Fuzhou 350116

3. Digital Fujian Institute of Financial Big Data, Fuzhou 350116

4. College of Electronics and Information Science, Fujian Jiangxi
University, Fuzhou 350108

Key Words Aspect Level Sentiment Analysis, Knowledge Graph, Attention Mechanism, Deep Learning

Citation DENG L M, WEI J J, WU Y B, YU X Y, LIAO X W. Aspect Level Sentiment Analysis Based on Knowledge Graph and Recurrent Attention Network. Pattern Recognition and Artificial Intelligence, 2020, 33(6): 479–487.

随着互联网的迅速发展,用户在社交媒体上表达他们对话题的评论呈现爆发式增长,大量信息的流动影响组织的决策过程,对这些评论文本的研究蕴含较大价值^[1]. 视角级文本情感分析^[2]旨在确定每个视角词在文本中表达的情感极性,以便提供比文档级或句子级更全面、深入和细粒度的情感分析,为人们提供便捷自动化的工具,提升互联网信息的利用率.

视角级文本情感分类问题的方法通常提取和学习文本特征,用于构建分类模型,主要包括基于情感词典的方法和基于机器学习的方法. Liu^[3]对含有情感的单词或短语评分,解决因为否定词或转折词而发生情感转移的问题. Moghaddam 等^[4]提取产品的重要信息并评定星级,确定用户的满意程度. 也有研究者采用 SVM 分类器^[5]、*N*-gram^[6]等进行情感极性判断. 上述方法虽然取得一定效果,但需耗费大量的人力、物力和资源设计语义和语法特征,方法性能极大程度依赖这些特征,并且泛化能力较差.

深度学习方法通常从视角词和文本中自动获取低维文本特征表示,并构造神经网络模型以进行视角级文本情感分析,主要包括神经网络模型和注意力机制模型^[7]. Dong 等^[8]提出基于目标依赖的长短时记忆网络 (Target Dependent Long Short Term Memory, TD-LSTM),自适应递归神经网络对情感词向特定视角词自适应传播建模,该过程完全取决于语法依赖树,但可能会出现因为是非标准文本而不起作用的情况. Tang 等^[9]将句子分为视角词的左右两个部分,并使用两个长短时记忆网络 (Long Short-Term Memory, LSTM) 分别模拟视角词与其左上下文和右上下文的相关性. Huang 等^[10]利用参数化的滤波器和门限机制将相位信息融入卷积神经网络中,有效捕捉特定的文本特征信息. Lei 等^[11]提出语义认知网络,模拟人类阅读认知过程,较好地捕捉上下文与视角词之间的关系. 上述方法中每个单词对分类结果是同等重要的,未曾考虑在不同视角情况下单词对情感分类的贡献程度不同的情况.

注意力机制模型主要计算上下文单词对文本的贡献程度,有效解决神经网络对每个单词都是同等考虑的问题. Wang 等^[12]提出带有视角词的基于注意力的长短时记忆网络 (Attention-Based LSTM with

Aspect Embedding, ATAE-LSTM),因视角词不同而重点关注不同的上下文单词. Tang 等^[13]提出端到端的记忆网络模型 (Memory Network, MemNet),通过多注意力机制捕捉外部记忆中每个上下文单词的重要性,但并未结合不同的注意力结果. Ma 等^[14]设计交互注意力网络 (Interactive Attention Networks, IAN),通过两个注意力网络对视角词和上下文进行交互建模. Chen 等^[15]提出基于记忆块的循环注意力网络 (Recurrent Attention Network on Memory, RAM),利用双向长短时记忆网络 (Bi-directional LSTM, BiLSTM) 并结合位置信息构造记忆块,然后计算多层注意力的结果,最后利用门限控制单元进行非线性的结合. Wang 等^[16]提出利用线性链式条件随机场层捕捉视角词与观点词之间的依赖关系,较好地处理复杂的句子. 上述方法仅考虑文本内容,未考虑在不同语境下一个单词可能表现出不同的语义信息,导致方法的预测性能下降.

针对上述问题,本文提出基于知识图谱与循环注意力网络 (Knowledge Graph and Recurrent Attention Network, KGRAN) 的视角级情感分析方法. 首先通过深层双向长短时记忆网络 (Deep BiLSTM, DBiLSTM) 学习文本中词与词之间的上下文关系,设计动态注意力机制融合知识图谱的过程:结合当前单词向量表示与知识图谱中的同义词向量表示,构建记忆内容,并引入哨兵向量避免外部知识误导方法,获取符合当前语境的语义信息,进一步提升文本分类的预测性能. 本文使用 SemEval 2014 Task4 的公开数据集 Laptop、Restaurant 和 Dong 等^[8]采集的 Twitter 数据集进行实验,验证本文方法在视角级文本情感分析问题上的有效性.

1 基于知识图谱与循环注意力网络的视角级情感分析方法

1.1 问题形式化定义

视角级文本情感分类任务旨在学习一个分类器,将给定一个文本及文本中的一个视角词输入分类器中,判断该文本关于给定视角的情感类别标签,

这实质上是一个三分类问题。

本文的问题形式化定义如下:假设给定一个文本

$$S = \{w_1, w_2, \dots, w_i, w_{i+1}, \dots, w_{i+l}, \dots, w_N\},$$

其中文本中的视角词

$$w_a = \{w_i, w_{i+1}, \dots, w_{i+l}\},$$

类别集合 $y = \{y_1, y_2, y_3\}$. 视角级文本情感分类任务的目标是学习一个分类模型,将文本 S 和视角词 w_a 映射至对应的类别标签 y_i ,即

$$f(S, w_a) \rightarrow y_i,$$

其中,文本 S 包含 N 个单词,视角词 w_a 包含 $l + 1$ 个单词,类别集合 y_1, y_2, y_3 分别表示情感极性为正面、中立和负面。

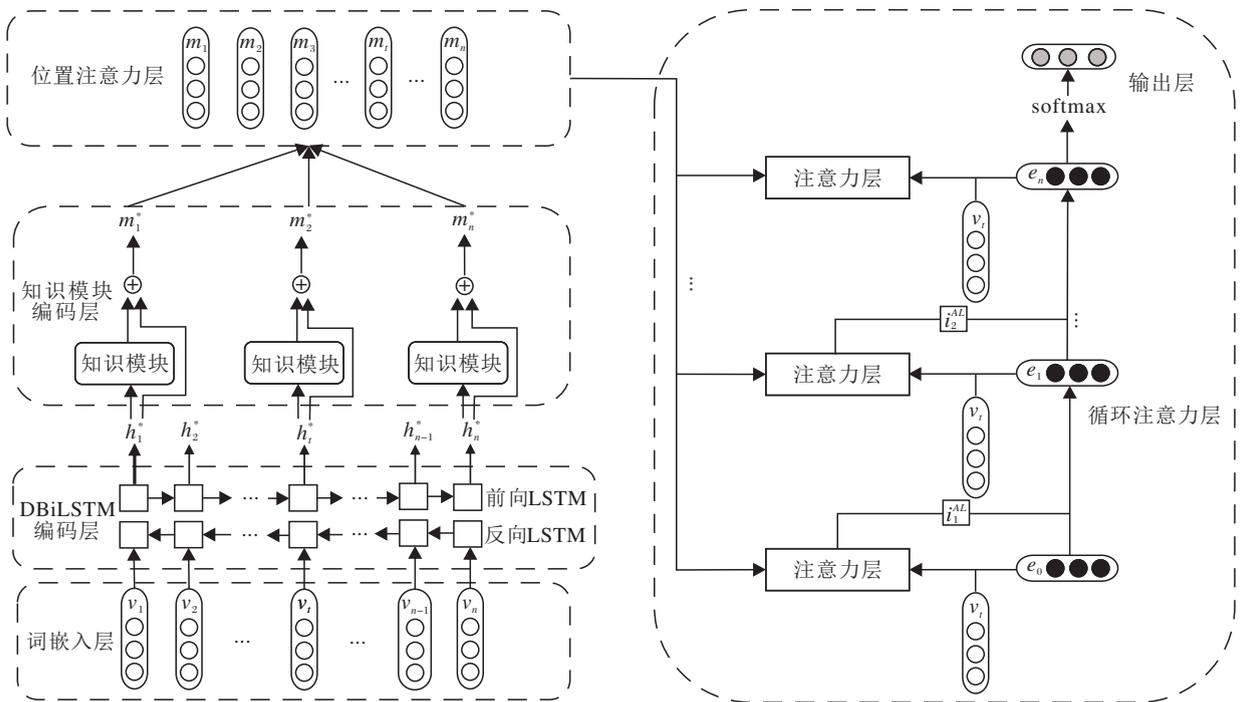


图1 基于知识图谱与循环注意力网络的视角级情感分析方法框图

Fig. 1 Framework of aspect level sentiment analysis based on knowledge graph and recurrent attention network

1.2.1 输入文本词嵌入层

本文方法输入包含两部分:1) 句子文本序列

$$S = \{w_1, w_2, \dots, w_i, w_{i+1}, \dots, w_{i+l}, \dots, w_N\};$$

2) 句子中出现的视角词序列

$$w_a = \{w_i, w_{i+1}, \dots, w_{i+l}\}.$$

其中,视角词 w_a 为句子 S 的子序列,视角词可以为一个单词,也可以为多个单词.每个单词 $w_i \in \mathbf{R}^{1|V|}$, $|V|$ 为词汇大小.假设 $L \in \mathbf{R}^{d \times |V|}$ 表示利用全局词向量(Global Vector, GloVe)生成的预训练词向量查找表,其中 d 为词向量维度。

1.2 方法组成部分

基于知识图谱与循环注意力网络的视角级文本情感分类方法框图如图1所示.本文方法由6个部分组成.1)输入文本词嵌入层:将文本进行词向量化.2)基于深层双向长短时记忆网络的词编码层:捕获文本的双向语义依赖关系.3)融合知识模块的编码层:对知识图谱中的同义词信息与文本信息进行建模.4)融合位置注意力信息的文本表示:对单词的位置信息进行建模,计算每个单词对文本情感分类的贡献程度.5)基于循环注意力网络的视角级情感特征表示:计算每个记忆内容注意力得分,通过门限控制单元提取文本特征.6)输出层:输出文本的情感分类结果。

因此,本文将每个单词 w_i 通过词嵌入矩阵 L 映射成与之对应的词向量表示 $v_i \in \mathbf{R}^{d \times 1}$,输入的句子文本序列 S 映射的词嵌入表示为

$$V_s = \{v_1, v_2, \dots, v_i, v_{i+l}, \dots, v_N\},$$

视角词序列 w_a 的词嵌入表示为

$$V_a = \{v_i, v_{i+1}, \dots, v_{i+l}\}.$$

如果视角词是一个单词,直接取词嵌入后的向量作为此视角词序列的表示,若视角词为多个单词构成,取多个单词的词嵌入向量平均值作为此视角词序列的表示。

1.2.2 基于深层双向长短时记忆网络的词编层

词编层采用 DBiLSTM 获取文本隐藏层表示。假设在每个时刻 t , 前向 LSTM 在输出 l 层的隐藏层状态向量 \mathbf{h}_t^l 的同时在隐藏层中维持一个记忆单元 \mathbf{c}_t^l , 其中初始隐藏层向量 $\mathbf{h}_0^l = \mathbf{v}_l$, 更新过程如下:

$$\begin{aligned} \mathbf{i}_t &= \sigma(\mathbf{W}_i \mathbf{h}_t^{l-1} + \mathbf{U}_i \mathbf{h}_{t-1}^l), \\ \mathbf{f}_t &= \sigma(\mathbf{W}_f \mathbf{h}_t^{l-1} + \mathbf{U}_f \mathbf{h}_{t-1}^l), \\ \mathbf{o}_t &= \sigma(\mathbf{W}_o \mathbf{h}_t^{l-1} + \mathbf{U}_o \mathbf{h}_{t-1}^l), \\ \hat{\mathbf{c}}_t &= \tanh(\mathbf{W}_c \mathbf{h}_t^{l-1} + \mathbf{U}_c \mathbf{h}_{t-1}^l), \\ \mathbf{c}_t^l &= \mathbf{f}_t \odot \mathbf{c}_{t-1}^l + \mathbf{i}_t \odot \hat{\mathbf{c}}_t, \\ \mathbf{h}_t^l &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t^l). \end{aligned}$$

其中: σ 表示 logistic sigmoid 函数; \tanh 表示 Hyperbolic tangent 函数; \odot 表示点乘操作符; \mathbf{i}_t 表示 t 时刻的输入门、 \mathbf{f}_t 表示 t 时刻的遗忘门、 \mathbf{o}_t 表示 t 时刻的输出门、 \mathbf{c}_t^l 表示 t 时刻的记忆单元的激活向量, 分别控制是否根据当前输入这些向量与隐藏层状态向量 \mathbf{h}_t^l 具有相同的维度; $\mathbf{W}_i, \mathbf{W}_f, \mathbf{W}_o, \mathbf{W}_c$ 表示前向 LSTM 的输入门、遗忘门、输出门、记忆单元在 l 层关于输入向量的模型参数; $\mathbf{U}_i, \mathbf{U}_f, \mathbf{U}_o, \mathbf{U}_c$ 表示前向 LSTM 的输入门、遗忘门、输出门、记忆单元在 l 层关于隐藏层向量的模型参数。本文使用 Karpathy 等^[17] 提出的 2 层 BiLSTM 堆叠而成, 产生的初始记忆块如下:

$$\mathbf{H}^* = \{\mathbf{h}_1^*, \mathbf{h}_2^*, \dots, \mathbf{h}_i^*, \dots, \mathbf{h}_n^*\},$$

其中

$$\mathbf{h}_i^* = (\vec{\mathbf{h}}_i^l, \overleftarrow{\mathbf{h}}_i^l) \in \mathbf{R}^{\vec{d}_l + \overleftarrow{d}_l}.$$

1.2.3 融合知识模块的编码层

融合知识模块的编码层如图 2 所示, 输入为单词向量 \mathbf{v}_i 经过 DBiLSTM 进行编码后的隐藏向量 \mathbf{h}_i^* 及从 WordNet 知识库中获取与 \mathbf{v}_i 相关的一组候选同义词 Syn_{v_i} , 并使用对应关系空间中的翻译方法 (Translation in the Corresponding Relation Space, TransR)^[18] 等知识图谱嵌入方法学习候选词的表示, 每个候选同义词 $k \in Syn_{v_i}$ 都经过预训练词向量查找表找到与之对应的向量 $\mathbf{t}_k \in \mathbf{R}^{d \times 1}$.

本文使用带有哨兵向量的注意力机制, 可动态决定是否引入外部知识, 并且区分哪一项外部知识有效, 从而更好地衡量同义词与上下文信息的关系。哨兵向量

$$\mathbf{s}_t = \sigma(\mathbf{W}_b \mathbf{h}_{t-1}^* + \mathbf{U}_b \mathbf{x}_t).$$

其中: $\mathbf{W}_b, \mathbf{U}_b$ 为要学习的权重参数矩阵; \mathbf{h}_{t-1}^* 为上一个隐藏状态的输出向量, 尽可能保存当前输入之前的文本表示; \mathbf{x}_t 为当前输入的文本向量。

同义词向量 \mathbf{t}_k 和哨兵向量 \mathbf{s}_t 的注意权重:

$$\begin{aligned} S(\mathbf{t}_k, \mathbf{h}_i^*) &= (\mathbf{t}_{kb})^T \tanh(\mathbf{W}_t \mathbf{t}_k + \mathbf{W}_{ht} \mathbf{h}_i^* + \mathbf{b}_t), \\ S(\mathbf{s}_t, \mathbf{h}_i^*) &= (\mathbf{s}_{tb})^T \tanh(\mathbf{W}_s \mathbf{s}_t + \mathbf{W}_{hs} \mathbf{h}_i^* + \mathbf{b}_s), \\ \alpha_{t_k} &= \frac{\exp(S(\mathbf{t}_k, \mathbf{h}_i^*))}{\sum_{k=1}^{num} S(\mathbf{t}_k, \mathbf{h}_i^*) + S(\mathbf{s}_t, \mathbf{h}_i^*)}, \\ \beta_t &= \frac{\exp(S(\mathbf{s}_t, \mathbf{h}_i^*))}{\sum_{k=1}^{num} S(\mathbf{t}_k, \mathbf{h}_i^*) + S(\mathbf{s}_t, \mathbf{h}_i^*)}. \end{aligned}$$

其中: $\mathbf{W}_t, \mathbf{W}_{ht}, \mathbf{W}_s, \mathbf{W}_{hs}$ 为方法需要学习的权重参数矩阵; $\mathbf{t}_{kb}, \mathbf{s}_{tb}, \mathbf{b}_t, \mathbf{b}_s$ 为方法需要学习的权重参数向量; $S(\cdot)$ 用于计算同义词 \mathbf{t}_k 和哨兵向量 \mathbf{s}_t 在 \mathbf{h}_i^* 状态下的重要性得分; α_{t_k}, β_t 分别表示同义词 \mathbf{t}_k 和哨兵向量 \mathbf{s}_t 与当前状态 \mathbf{h}_i^* 的相关性。

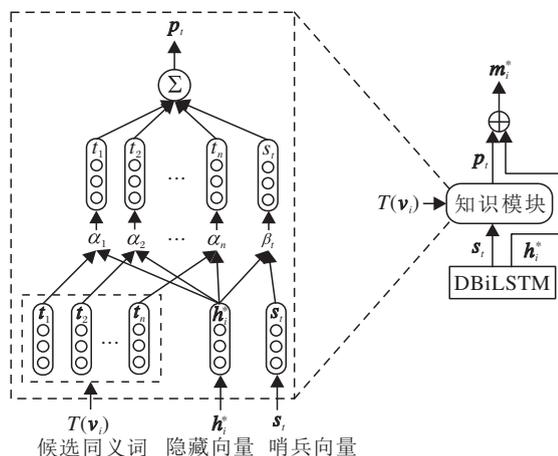


图 2 知识图谱嵌入结构图

Fig. 2 Embedded structure of knowledge graph

同时, 为了将知识图谱中的同义词信息融入编码层, 通过 α_{t_k}, β_t 计算知识状态向量 \mathbf{p}_t , 并将知识状态向量 \mathbf{p}_t 与经过 DBiLSTM 获得的隐藏层向量 \mathbf{h}_i^* 结合, 得到隐藏层向量 \mathbf{m}_i^* , 称为知识感知状态向量:

$$\begin{aligned} \mathbf{m}_i^* &= \mathbf{p}_t + \mathbf{h}_i^*, \\ \mathbf{p}_t &= \sum_{k \in Syn_{v_i}} \alpha_{t_k} \mathbf{t}_k + \beta_t \mathbf{s}_t, \\ \sum_{k \in Syn_{v_i}} \alpha_{t_k} + \beta_t &= 1. \end{aligned}$$

如果 $Syn_{v_i} = \emptyset$, 则令 $\mathbf{p}_t = \mathbf{0}$, 即知识感知状态向量与隐藏层向量一致。其中, 知识状态向量 \mathbf{p}_t 与隐藏层向量 \mathbf{h}_i^* 具有相同维度。

使用知识感知状态向量 \mathbf{m}_i^* 作为文本表示的记忆块:

$$\mathbf{M}^* = \{\mathbf{m}_1^*, \mathbf{m}_2^*, \dots, \mathbf{m}_i^*, \dots, \mathbf{m}_n^*\}.$$

1.2.4 融合位置注意力信息的文本表示

一个上下文单词靠近视角词会比远离视角词的上下文单词更重要,对视角级情感分类的作用更大.因此,本文构造上下文词语与视角词之间的位置信息关系,用于衡量每个单词对文本情感分类的贡献程度.在第*i*个单词的位置权重如下:

$$w_i = 1 - \frac{|i - t|}{t_{\max}}, u_i = \frac{|i - t|}{t_{\max}},$$

其中,*t*为视角词的位置, t_{\max} 为输入文本的单词总数, u_i 为上下文单词与视角词的相对位置距离.

如果视角词是由多个单词组成的,计算上下文单词与视角词的距离*i*-*t*可以判断上下文单词在视角词的左侧还是右侧.若在左侧,*t*表示视角词最左边单词的位置;若在右侧,*t*表示视角词最右边单词的位置;从而计算上下文单词与视角词的距离.结合位置注意力机制计算的权重 w_i 与融合知识图谱的感知状态向量 m_i^* ,生成位置权重记忆块:

$$M = \{m_1, m_2, \dots, m_i, \dots, m_n\},$$

$$m_i = (w_i m_i^*, u_i) \in \mathbf{R}^{\vec{d}_i + \overleftarrow{d}_i + 1}.$$

1.2.5 基于循环注意力网络的视角级情感特征表示

本文采用多层注意力机制,从融合知识模块和位置注意力的记忆内容中提取信息,在获得融合知识模块和位置注意力的记忆内容 M 后,计算每个记忆块 m_j 的注意力得分:

$$g_j^i(m_j, e_{i-1}, v_i) = W_i^{AL}(m_j, e_{i-1}, v_i) + b_i^{AL},$$

$$\alpha_j^i = \frac{\exp(g_j^i)}{\sum_k \exp(g_k^i)}, i_i^{AL} = \sum_{j=1}^N \alpha_j^i m_j,$$

其中,设 e_0 为0向量, W_i^{AL} 为模型需要学习的权重参数矩阵, b_i^{AL} 为模型需要学习的权重参数向量.

同时利用门限循环单元(Gated Recurrent Unit, GRU),在每个记忆块的注意力得分计算后更新每层的 e_i ,将上一轮的计算结果 e_{i-1} 与本轮的表示 i_i^{AL} 非线性地结合,引导方法加入一些有用信息,更好地保留当前注意力机制之前的结果,有利于模型的情感预测.更新 e_i 的过程如下:

$$r = \sigma(W_r i_i^{AL} + U_r e_{i-1}),$$

$$z = \sigma(W_z i_i^{AL} + U_z e_{i-1}),$$

$$\tilde{e}_i = \tanh(W_x i_i^{AL} + W_g(r \odot e_{i-1})),$$

$$e_i = (1 - z) \odot e_{i-1} + z \odot \tilde{e}_i.$$

其中: $W_r \in \mathbf{R}^{H \times (\vec{d}_i + \overleftarrow{d}_i + 1)}$, $W_z \in \mathbf{R}^{H \times (\vec{d}_i + \overleftarrow{d}_i + 1)}$, $W_x \in \mathbf{R}^{H \times (\vec{d}_i + \overleftarrow{d}_i + 1)}$; $U_r \in \mathbf{R}^{H \times H}$, $U_z \in \mathbf{R}^{H \times H}$, $W_g \in \mathbf{R}^{H \times H}$ 为方

法需要学习的权重参数矩阵; H 为GRU隐藏的神经单元数.

1.2.6 输出层

本文将最后一次的输出向量 e_n 作为特征,并将其输入到softmax层中进行视角级文本情感分类预测.该模型通过最小化交叉熵误差进行模型的训练,损失函数如下:

$$loss =$$

$$\sum_{(s, v_a) \in T} \sum_{c \in y} P_c^g(s, v_a) \cdot \ln(P_c(s, v_a)) + \lambda \|\theta\|^2,$$

其中, T 表示训练集中的所有句子, y 表示情感类别的标签集合, (s, v_a) 表示句子-视角组成的训练数据对, P_c^g 表示文本的真实情感类别标签, P_c 表示预测的情感类别标签值,参数集合 θ 在模型训练过程中进行调整.在实验过程中,使用自适应低阶矩估计优化器(Adaptive Estimates of Lower-Order Moments, Adam)更新规则作为优化方法更新参数,同时采用 L_2 正则化和随机失活(Dropout)减小模型过拟合的影响.

2 实验及结果分析

2.1 实验数据集及实验设置

本文在3个公开标准数据集上进行实验:饭店(Restaurant)、笔记本(Laptop)数据集采用SemEval 2014 Task4^[2]公开的数据集;推特(Twitter)领域数据集采用Dong等^[8]采集的数据集.这3个数据集都含有3个情感极性标签:正面(Positive)、负面(Negative)、中立(Neutral).具体情况如表1所示.

表1 实验数据集

Table 1 Experimental datasets

名称		正面	负面	中立	总计
Laptop	训练集	994	870	464	2328
	测试集	341	128	169	638
Restaurant	训练集	2164	807	637	3608
	测试集	728	196	196	1120
Twitter	训练集	1561	1560	3127	6248
	测试集	173	173	346	692

本文使用WordNet,利用Bordes等^[19]提供的预处理数据,训练知识图谱嵌入方法,其中包含151 442对三元组,40 943个语法集,18个关系.

实验采用GloVe词向量查询表进行词嵌入初始化,在训练过程中不再更新.词嵌入维度和隐藏状态维度均设置为300,学习率设置为0.005, L_2 正则项

的权重系数 $\lambda = 0.001$, dropout 系数设置为 0.5.

本文使用 Tensorflow 实现 KGRAN 及方法对比, 使用相同输入、词向量查询表、词嵌入维度、隐藏状态维度和优化器等.

2.2 评价指标

本文采用正确率 (Accuracy, Acc)^[8] 和宏平均下的 F1 值 (Macro-F1, Mac-F1) 作为实验评价指标衡量总体实验效果, 具体计算公式如下:

$$Acc = \frac{total_true}{total_gold},$$

$$Mac-F1 = \frac{1}{class_num} \sum_{l=1}^{class_num} F1_l,$$

$$F1_l = \frac{2 \cdot P_l \cdot R_l}{P_l + R_l},$$

$$P_l = \frac{n_true_l}{n_predict_l}, R_l = \frac{n_true_l}{n_gold_l},$$

其中, n_true_l 为方法准确预测类别为 l 的数量, $n_predict_l$ 为预测类别为 l 的数量, n_gold_l 为实际属于类别 l 的数量, P_l 为精确率, R_l 为召回率.

2.3 对比方法

选取的对比方法如下.

1) SVM^[4]. 利用 SVM 提取一系列特征, 在 Sem Eval2014 Task4 上取得良好效果.

2) LSTM^[9]. 使用 LSTM 计算句子的特征表示, LSTM 是一个循环模型, 把最后一个单词对应的隐藏层状态输出向量作为句子特征.

3) TD-LSTM^[8]. 利用 2 个 LSTM 网络对视角词的左侧和右侧上下文进行建模, 再连接建模后的两个上下文表示, 用于预测视角词的情感极性.

4) ATAE-LSTM^[12]. 在每个单词后面附加视角词, 增强视角词对文本情感极性预测的影响, 并使用带有注意力机制的 LSTM 网络预测视角词的情感极性.

5) IAN^[14]. 设计注意力机制, 可以交互学习上下文表示和视角词表示.

6) MemNet^[13]. 一个多层的共享参数计算层, 每层都是基于上下文信息和位置注意力的模型, 可为每个上下文单词学习一个权重, 并利用这些信息计算文本表示, 用于视角词情感极性预测.

7) RAM^[15]. 通过 DBiLSTM 改进 MemNet 构造的记忆块, 并利用门限控制单元非线性结合多个注意力输出句子的表示.

8) 基于内容注意力的视角情感分类模型 (Content Attention Based Aspect Based Sentiment Classification Model, Cabasc)^[7]. 利用句子级注意力机制

从全局捕获关于给定视角的重要信息, 并通过上下文注意力机制考虑单词及其相关性顺序, 解决记忆模型的短视问题.

9) 基于语法导向的混合注意力网络 (Syntax-Directed Hybrid Attention Network, SHAN)^[20]. 利用全局注意力机制捕获关于视角的粗略信息, 利用语法导向的局部注意力机制从语法上观察接近视角的单词, 利用信息门综合局部和全局注意力结果的信息, 自适应地生成一个噪声更小、面向情感的文本表示.

2.4 有效性对比

为了验证本文方法的有效性, 各方法在 3 个数据集上的 ACC 和 Mac-F1 值对比如表 2 所示. 表中 SVM、IAN、RAM、SHAN 为从相关论文中获取的实验结果, LSTM、TD-LSTM、ATAE-LSTM、MemNet、Cabasc 为根据相关论文复现的结果.

表 2 各方法在 3 个数据集上的实验结果对比

Table 2 Experimental results comparison of different algorithms on 3 datasets

方法	Laptop		Restaurant		Twitter	
	Acc	Mac-F1	Acc	Mac-F1	Acc	Mac-F1
SVM	70.49	-	80.16	-	-	-
LSTM	66.77	61.78	74.29	62.58	66.33	63.37
TD-LSTM	67.71	60.25	75.36	64.48	69.51	67.13
ATAE-LSTM	68.50	61.52	77.32	64.99	68.93	67.22
IAN	72.10	63.10	78.60	67.40	68.80	66.60
MemNet	71.79	67.07	79.96	69.09	70.09	66.81
RAM	74.49	71.35	80.23	70.80	69.36	67.30
Cabasc	75.07	70.13	80.54	70.76	71.53	69.79
SHAN	74.64	-	81.02	-	-	-
KGRAN	75.54	71.67	82.05	72.55	72.25	71.02

由表 2 可看出, 本文方法取得最优结果. KGRAN 的性能优于 Cabasc、SHAN, 因为 KGRAN 中考虑到不同单词在不同语境中可能会有不同的语义信息, 通过动态注意力机制衡量知识图谱中的同义词信息对当前信息的影响, 更好地刻画单词在当前语境下的语义信息, 为情感分类提供更细粒度的特征, 促使性能得到进一步提升.

相比神经网络的 LSTM 和 TD-LSTM, SVM 取得更优结果, 在 Laptop、Restaurant 数据集上正确率分别达到 70.49%、80.16%. 这主要是因为 LSTM 欠缺对视角词信息的考虑, TD-LSTM 因为无法捕获视角词和上下文的交互, 并且获取特征主要集中在靠近

视角词的部分,可能会忽略长距离的情感特征而导致性能较差,因此特征工程有助于性能提升,但是需要耗费大量资源。

在基于注意力机制的神经网络中,相比 MemNet, RAM 在 Laptop、Restaurant 数据集上的正确率分别提升 2.7% 和 0.27%。这主要是因为 MemNet 仅使用一层注意力机制计算,而 RAM 是将不同注意力计算层的结果非线性结合用于计算,因此取得优于 MemNet 的结果。相比 RAM, Cabasc 在 3 个数据集上的正确率分别提升 0.58%、0.31%、2.17%, 主要是因为 Cabasc 设计句子级注意力机制,从全局角度捕获特定视角词的重要信息,通过上下文注意力机制衡量单词与视角词之间的顺序和关系,为每个视角词定制记忆内容,因此 Cabasc 性能优于 RAM。SHAN 充分利用基于句法的局部信息和全部信息实现注意力权重的动态调整,在 Restaurant 数据集上取得优于 Cabasc 的性能。

2.5 不同知识图谱嵌入方式的影响

本节分析不同的知识图谱嵌入方式对方法性能提升的影响,具体实验结果如表 3 所示。由表可看出,使用 TransR 对方法性能提升最多,在 Laptop、Restaurant 和 Twitter 这 3 个数据集中正确率分别达到 75.54%、82.05%、72.25%。这主要是因为 TransR 对于不同关系关注实体的不同属性,相似实体在实体空间中彼此靠近,而在特定的不同方面对应的关系空间彼此远离。而翻译嵌入方法 (Translating Embedding, TransE)^[19] 只能在一对一的简单关系模型中和大规模稀疏知识库上取得良好的性能和可扩展性。基于超平面翻译方法 (Translating on Hyperplanes, TransH)^[21] 虽然解决 TransE 不能处理复杂关系的缺点,但由于实体向量被投影到关系的语义空间中,导致它们具有相同维度,限制模型的表达能力。因此,TransR 能将同义词更准确地映射到低维空间中,取得最优结果。

表 3 各嵌入方式对本文方法性能的影响

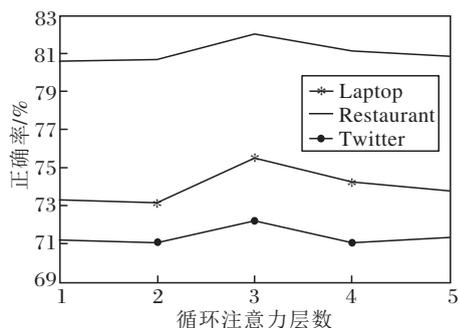
Table 3 Effect of different knowledge graph embedding on performance of the proposed method

嵌入方式	Laptop		Restaurant		Twitter	
	Acc	Mac-F1	Acc	Mac-F1	Acc	Mac-F1
TransH	73.82	68.75	80.63	71.06	69.65	67.74
TransE	74.14	68.90	81.25	71.87	70.23	68.90
TransR	75.54	71.67	82.05	72.55	72.25	71.02

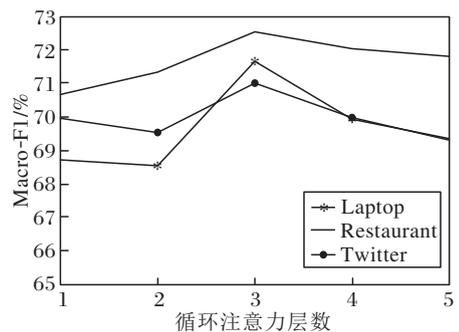
2.6 注意力计算层数的影响

本节分析不同的注意力计算层数对性能的影响,

具体结果如图 3 所示。由图可看出,本文方法随循环注意力层数增加先递增后递减,在层数为 3 的时候 Acc 和 Mac-F1 值达到最佳,使用单层注意力计算层数的性能总是不如使用多层注意力计算层数的性能。这说明在复杂情况下,单层的注意力计算层数可能无法有效捕捉文本中的情感特征关键信息。同时,方法性能并非随着注意力计算层的增加而增加。由图可看出,当注意力计算层数为 3 时,本文方法在 3 个数据集上的 Acc、Mac-F1 值取得最优值,但是超过 3 层之后方法性能有所下降。这主要是由于随着层数的增加,复杂度增加使方法变得更难以训练,泛化能力下降。



(a) 正确率
(a) Accuracy



(b) Mac-F1

图 3 注意力计算层数对本文方法性能的影响

Fig. 3 Effect of multiple computation layers on performance of the proposed method

2.7 注意力机制的影响

为了更好地说明注意力机制对文本表示的影响,选择如下样例进行注意力可视化。

例:Great food but the service was dreadful!

译:食物不错,但服务太差了!

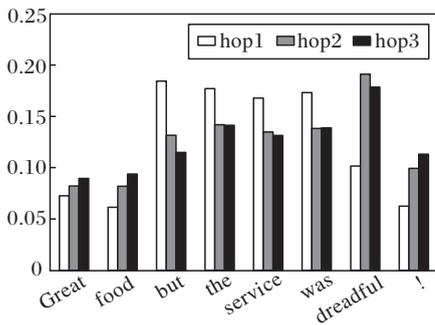
视角词:food 极性:Positive

视角词:service 极性:Negative

例子中含有 2 个不同的视角词,在不同视角词下具有不同的情感极性,图4(a)为视角词“food”下各个单词的注意力权重可视化效果图,可看出经过 3 层注意力层计算之后,单词“Great”注意力权重最高,并正确预测文本在视角词“food”下的情感极性为积极。(b)为视角词“service”下各个单词的注意力权重可视化效果图,此时单词“dreadful”注意力权重最高,正确预测文本在视角词“service”下的情感极性为消极。由此说明注意力机制可有效刻画不同视角下文本单词对情感分类的重要性,较好地衡量单词对文本情感分类的贡献度。



(a) food



(b) service

图 4 两个视角词的注意力权重

Fig. 4 Attention weight of 2 aspect words

3 结束语

本文提出基于知识图谱与循环注意力网络的视角级情感分析方法。首先通过深层双向长短时记忆网络获取输入文本的上下文编码信息,利用注意力机制将知识图谱融入神经网络模型中,解决单词在不同句子中具有不同含义的问题。再利用单词与视角词的相对位置计算每个单词对视角词的位置注意

力得分,并将该结果与融合知识图谱的记忆块结合,构造融合知识图谱的位置注意力记忆块。最后,利用门限控制单元将前一轮计算结果与该轮注意力得分进行非线性结合,将最终一轮结果作为文本的特征输入分类器中,得到预测结果。实验表明,本文方法在 3 个数据集上的正确率和宏平均 F1 值都具有一定提升。今后将考虑引入对抗训练,增加方法的泛化能力和鲁棒性。

参 考 文 献

- [1] ELGAMAL M. Sentiment Analysis Methodology of Twitter Data with an Application on HAJJ Season. *International Journal of Engineering Research and Science*, 2016, 2(1): 82–87.
- [2] PONTIKI M, GALANIS D, PAYLOPOULOS J, *et al.* SemEval-2014 Task 4: Aspect Based Sentiment Analysis // *Proc of the 8th International Workshop on Semantic Evaluation*. Stroudsburg, USA: ACL, 2014: 27–35.
- [3] LIU B. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 2012, 5(1): 1–167.
- [4] MOGHADDAM S, ESTER M. Opinion Digger: An Unsupervised Opinion Miner from Unstructured Product Reviews // *Proc of the 19th ACM International Conference on Information and Knowledge Management*. New York, USA: ACM, 2010: 1825–1828.
- [5] KIRITCHENKO S, ZHU X D, CHERRY C, *et al.* NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews // *Proc of the 8th International Workshop on Semantic Evaluation*. Stroudsburg, USA: ACL, 2014: 437–442.
- [6] WAGNER J, ARORA P, CORTES S, *et al.* DCU: Aspect-Based Polarity Classification for SemEval Task 4 // *Proc of the 8th International Workshop on Semantic Evaluation*. Stroudsburg, USA: ACL, 2014. 223–229.
- [7] LIU Q, ZHANG H B, ZENG Y F, *et al.* Content Attention Model for Aspect Based Sentiment Analysis // *Proc of the World Wide Web Conference*. Washington, USA: IEEE, 2018: 1023–1032.
- [8] DONG L, WEI F R, TAN C Q, *et al.* Adaptive Recursive Neural Network for Target-Dependent Twitter Sentiment Classification // *Proc of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)*. Stroudsburg, USA: ACL, 2014: 49–54.
- [9] TANG D Y, QIN B, FENG X C, *et al.* Effective LSTMs for Target-Dependent Sentiment Classification // *Proc of the 26th International Conference on Computational Linguistics (Technical Papers)*. Stroudsburg, USA: ACL, 3298–3307.
- [10] HUANG B X, CARLEY K M. Parameterized Convolutional Neural Networks for Aspect Level Sentiment Classification // *Proc of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, USA: ACL, 2018: 1091–1096.
- [11] LEI Z Y, YANG Y J, YANG M, *et al.* A Human-Like Semantic Cognition Network for Aspect-Level Sentiment Classification // *Proc of the 33rd AAAI Conference on Artificial Intelligence*. Palo

Alto, USA; AAAI Press, 2019: 6650–6657.

- [12] WANG Y Q, HUANG M L, ZHU X Y, *et al.* Attention-Based LSTM for Aspect-Level Sentiment Classification // Proc of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA; ACL, 2016: 606–615.
- [13] TANG D Y, QIN B, LIU T. Aspect Level Sentiment Classification with Deep Memory Network // Proc of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA; ACL, 2016: 214–224.
- [14] MA D H, LI S J, ZHANG X D, *et al.* Interactive Attention Networks for Aspect-Level Sentiment Classification // Proc of the 26th International Joint Conference on Artificial Intelligence. New York, USA; ACM, 2017: 4068–4074.
- [15] CHEN P, SUN Z Q, BING L D, *et al.* Recurrent Attention Network on Memory for Aspect Sentiment Analysis // Proc of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA; ACL, 2017: 452–461.
- [16] WANG B L, LU W. Learning Latent Opinions for Aspect-Level Sentiment Classification // Proc of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto, USA; AAAI Press, 2018: 5537–5544.
- [17] KARPATHY A, JOHNSON J, LI F F. Visualizing and Understanding Recurrent Networks[C/OL]. [2020-02-26]. <https://arxiv.org/pdf/1506.02078.pdf>.
- [18] LIN Y K, LIU Z Y, SUN M S, *et al.* Learning Entity and Relation Embeddings for Knowledge Graph Completion // Proc of the 29th AAAI Conference on Artificial Intelligence. Palo Alto, USA; AAAI Press, 2015: 2181–2187.
- [19] BORDES A, USUNIER N, GARCIA-DURAN A, *et al.* Translating Embeddings for Modeling Multi-relational Data // BURGESS C J C, BOTTOU L, WELLING M, *et al.*, eds. Advances in Neural Information Processing Systems 26. Cambridge, USA; The MIT Press, 2013: 2787–2795.
- [20] WANG X Y, XU G L, ZHANG J Y, *et al.* Syntax-Directed Hybrid Attention Network for Aspect-Level Sentiment Analysis. IEEE Access, 2019, 7: 5014–5025.
- [21] WANG Z, ZHANG J W, FENG J L, *et al.* Knowledge Graph Embedding by Translating on Hyperplanes // Proc of the 28th AAAI Conference on Artificial Intelligence. Palo Alto, USA; AAAI Press, 2014: 1112–1119.

作者简介



邓立明, 硕士研究生, 主要研究方向为观点挖掘、情感分析、自然语言处理。E-mail: n180320048@fzu.edu.cn.

(**DENG Liming**, master student. His research interests include opinion mining, sentiment analysis and natural language processing.)



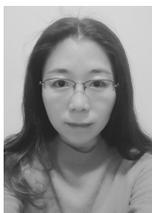
魏晶晶, 博士, 讲师, 主要研究方向为观点挖掘。E-mail: weijj@fjxxu.edu.cn.

(**WEI Jingjing**, Ph. D., lecturer. Her research interests include opinion mining.)



吴运兵, 硕士, 副教授, 主要研究方向为机器学习、数据挖掘、知识表示。E-mail: wyb5820@fzu.edu.cn.

(**WU Yunbing**, master, associate professor. His research interests include machine learning, data mining and knowledge representation.)



余小燕, 硕士, 讲师, 主要研究方向为文本信息处理。E-mail: 35490823@qq.com.

(**YU Xiaoyan**, master, lecturer. Her research interests include Chinese information processing.)



廖祥文(通讯作者), 博士, 教授, 主要研究方向为观点挖掘、情感分析、自然语言处理。E-mail: liaoxw@fzu.edu.cn.

(**LIAO Xiangwen** (Corresponding author), Ph. D., associate professor. His research interests include opinion mining, sentiment analysis and natural language processing.)