

基于图卷积网络的客服对话情感分析^①



孟洁^{1,2}, 李妍^{1,2}, 赵迪^{1,2}, 张倩宜^{1,2}, 刘赫^{1,2}

¹(国网天津市电力公司 信息通信公司, 天津 300010)

²(天津市能源大数据仿真企业重点实验室, 天津 300010)

通信作者: 孟洁, E-mail: rainyymjmj@163.com

摘要: 随着电力业务的发展, 客服环节时刻产生着大量的数据, 然而传统对话数据情感检测方法对于客服质量检测的手段存在着诸多的问题和挑战. 本文根据词语出现的排列和定位构建字图, 对整个语句进行非连续长距离的语义建模; 并针对文档不同组成部分之间的关系, 对语句上下文之间的交互依赖或自我依赖关系进行建模; 最后通过卷积神经网络对所构建的图进行特征提取和邻域节点的特征聚合以得到文本的最终特征表示, 进而实现客服通话过程中的情绪状态检测. 通过实验证明本文提出的模型情感分类性能指标始终高于基线模型, 这表明融合词共现关系、顺序语句上下文编码和交互语句上下文编码结构可以有效提高情感类别检测精度. 该方法为智能化、自动化地检测客服通话过程中的情绪状态提供了更细粒度的分析, 为有效地提高客服服务质量具有重要意义.

关键词: 对话情感分析; 异质网络; 图卷积网络; 注意力机制; 双向门控循环单元

引用格式: 孟洁, 李妍, 赵迪, 张倩宜, 刘赫. 基于图卷积网络的客服对话情感分析. 计算机系统应用, 2022, 31(5): 147-156. <http://www.c-s-a.org.cn/1003-3254/8524.html>

Sentiment Polarity Analysis of Customer Dialogues Based on Graph Convolutional Network

MENG Jie^{1,2}, LI Yan^{1,2}, ZHAO Di^{1,2}, ZHANG Qian-Yi^{1,2}, LIU He^{1,2}

¹(State Grid Tianjin Information & Telecommunication Company, Tianjin 300010, China)

²(Key Laboratory of Energy Big Data Simulation of Tianjin Enterprise, Tianjin 300010, China)

Abstract: With the development of power business, a large amount of data is produced in the link of customer service. However, traditional sentiment analysis methods for dialogues face many problems and challenges in customer service quality detection. In this study, the word graph is constructed according to the arrangement and location of words, and then the discontinuous long-distance semantic modeling of the whole sentence is carried out. Next, according to the relationship among different parts of the document, the self and interaction dependency relationships between sentence contexts are modeled, respectively. Finally, the convolutional neural network (CNN) is applied to the constructed graph for feature extraction and feature aggregation of the neighbor nodes to obtain the final feature representation of the text. In this way, the detection of emotional states is realized in customer dialogues. Experimental results show that the performance of the proposed model is always higher than that of the baseline model, which demonstrates that the fusion of word co-occurrence relationships, as well as sequential context coding and interactive context coding structures, can effectively improve the accuracy of sentiment category detection. This method provides a fine-grained analysis for intelligently and automatically detecting the emotional states in customer dialogues, which is of great significance to effectively improve the quality of customer service.

Key words: dialogue sentiment analysis; heterogeneous network; graph convolutional network (GCN); attention mechanism; bi-directional gated recurrent unit (Bi-GRU)

① 基金项目: 国网科技项目 (KJ20-1-15)

收稿时间: 2021-08-05; 修改时间: 2021-10-11; 采用时间: 2021-10-15; csa 在线出版时间: 2022-04-11

电力行业^[1]是关系到国计民生的基础性行业。电力业务运营过程产生了海量的多种形态的自然语言数据,但是海量的数据会造成信息过载,导致关键信息淹没在大量的其他数据中,增加了工作人员筛选、获取有效信息的难度,导致对已有数据缺乏有效的处理和利用,造成信息资源的浪费。在我们相关的研究中,已实现针对电力业务语音数据设计自然语言识别模型,智能化地提取语音信息中的重要特征并根据语音数据自动生成对应文字内容;进一步地,由于电力行业专业性较强,通过基于自然语言理解技术将非结构化数据结构化,实现电力业务知识库地自动构建。而在本文的研究中,针对对话信息进行有效的分析和理解,改进了传统的繁杂、费时费力且重复性较高的客服电话质检工作,有利于电力系统高效而稳定的运行,将工作人员从繁杂的记录工作中解脱出来,极大地提升数据的利用率及信息服务的自动化、智能化。客服的话务质量关系着用户的满意度,是服务中的一项重要指标。在社交网络平台以及在客服服务等情境中,都会出现大量的对话文本,随着开源会话数据集的增加,会话中的数据挖掘也得到了越来越多研究者的关注^[2,3],目前在客服对话领域数据挖掘工作主要是从对话系统的设计^[4,5]以及在商业领域的客户购物预测^[6]进行的,而在客服对话的情感分析领域的研究也有越来越多的工作,如胡若云等^[7]提出基于双向传播框架,首先从外部语料中获取情感词和评价属性并对其进行扩展,然后使用基于词向量的语料相似度计算方法识别长尾词,最后挖掘出客服对话文本的情感词和评价属性。王建成^[8]通过长短时记忆网络和注意力机制挖掘对话历史中的相关信息,融入句子的语义表示;在解码器中将已解码句子的情感类别的概率分布融入当前句子的情感分类过程中实现客服对话情感分析。Wang等^[9]提出基于主题感知的多任务学习方法,该方法通过捕捉各种主题信息来学习客服对话中主题丰富的话语表达。对话情感识别是识别对话中说话人所说语句的情绪,本质上也可以归纳为文本分类问题。

传统的情感分析(又称观点挖掘)是从文本中分析人们的观点和情感,是自然语言处理中一个活跃的研究领域^[10]。现有的基于文本的情感分类研究,作为情感分析的一个重要组成部分,可以分为基于句法分析、基于词典分析与基于机器学习分析3大类。基于句法分析的情感分析主要采用句法知识对文本进行分析,

从而确定文本情感类别;基于词典的情感分析方法通常将文本看成词的集合,通过统计文本中的关键词再使用相关算法对文本情感进行评分,从而判断情感类别;基于机器学习的情感分类方法主要基于Pang等^[11]的工作,将文本情感分类看作文本分类问题,利用从情感词选取的大量嘈杂的标签文本作为训练集直接构建分类器(朴素贝叶斯、支持向量机等)。近年来,随着深度学习技术的发展,人们提出了各种神经网络模型,并在文本情感分类方面取得了显著的效果。

Kim^[12]首次将神经网络用于情感分类,该方法是卷积神经网络(convolutional neural networks, CNN)的直接应用,具有预先训练好的词嵌入功能。Kalchbrenner等^[13]提出了一个动态CNN方法,用于句子语义建模,来处理长度不同的输入句子,并在句子上生成一个特征图。这种模型能明确地捕捉短期和长期的关系;Xu等^[14]提出了一个缓存的长短期记忆网络(long-short term memory, LSTM)模型来捕获长文本中的整体语义信息;Zhou等^[15]设计了由两个基于注意力的LSTM模型用于跨语言情感分类;宋曙光等^[16]提出基于注意力机制的特定目标情感分析方法;Wang等^[17]提出利用CNN-LSTM组合模型预测文本情感极性。Poria等^[18]使用双向门控循环单元(bi-directional gated recurrent unit, Bi-GRU)提取对话句子表征,指出对话中的信息主要依赖于话语中的序列上下文信息;王伟等^[19]将Bi-GRU模型与注意力机制相结合,并成功提成了情感分类效果。孟仕林等^[20]提出利用情感词的情感信息构造出情感向量,与语言模型生成的向量相结合来表示文本,并使用双向LSTM模型作为分类器进行情感分析。邱宁佳等^[21]提出融合语法规则构建双通道中文情感模型,首先设计语法规则对文本进行预处理,然后使用CNN和BiLSTM模型提取局部情感特征并挖掘到文本时间跨度更大时的语义依赖关系。Shen等^[22]和Wang等^[23]利用双向注意网络捕捉单轮问答对话中的语义匹配信息。以上对话情感分析研究主要集中在话语间语境联系的建模上,而没有考虑对话的整体特征,也没有考虑特定说话人的特征。Majumder等^[24]提出采用注意力机制,将每个目标话语的全部或部分对话信息集中起来。然而,这种聚合机制没有考虑说话人的话语之间信息以及其他话语与目标话语的相对位置关系。

从理论上分析,RNN、LSTM、GRU等网络模型,可以传播长期的上下文信息,然而在实际应用中,这些

方法过度依赖于对话上下文和序列信息,容易忽略对话中的上下文情感动态建模,并且由于缺乏标准数据集对文本的有效标记,难以实现对文本的意图、个性以及主题等方面的建模。

近年来,图卷积神经网络(graph convolutional networks, GCN)受到越来越多的关注.在最近对相关自然语言处理任务的研究中,Yao等^[25]提出在整个语料库上构建一个异构图来进行文本级分类.Wu等^[26]反复消除GCN层之间的非线性,并将函数折叠成线性变换,以降低GCN的复杂性.Zhang等^[27]将句法依赖树引入到GCN中对句法结构进行编码,并将新的GCN结构用于句子和体层的情感分析.陈佳伟等^[28]提出基于自注意力门控图卷积网络的特定目标情感分析模型,并将预训练BERT应用到该任务中实现情感极性的分析。

由于图神经网络在文本分类^[25,29]、链接预测^[30]、文档摘要^[31]和关系分类^[32]等方面取得了成功,受以上研究的启发,本文将图网络应用于客服对话的情感识别问题。

首先,在对话过程中,词语是构成语句的主要成分,词语在决定句子的情感极性方面起着最重要的作用,词语的出现、排列和定位以及文档不同组成部分之间的关系对于理解文档是必要的和有价值的.另外,在双方对话的交互语句进行建模时,说话人信息是建模说话人之间依赖的必要条件,即模型能够理解说话人是如何影响其他说话人的情感状态;同样地,对说话人自身话语的建模有助于模型理解说话人个体的情感状态变化.此外,在对话过程中,上下文话语的相对位置决定了过去话语会对未来话语产生影响。

基于以上因素,本文提出了一种新的基于异构图卷积网络的对话情感识别模型(conversation sentiment analysis based on graph convolution network, CS-GCN).在该模型中,首先,根据词共现关系构建词语关系图,对整个语句进行非连续长距离的语义建模;其次使用Bi-GRU模型来捕捉连续顺序语句之间的上下文信息,并根据对话双方语句信息建立顺序语句的自我上下文依赖关系和交互语句上下文依赖关系,并通过邻居语句特征聚合进而得到每个语句的特征.最后通过以上图模型的建立,将词结构嵌入表示、语句上下文之间的自我顺序依赖和交互依赖关系得到的语句嵌入表示聚合得到文本信息的最终嵌入表示,实现对话情感状

态检测。

1 基于图卷积网络的对话情感识别

1.1 模型设计

在大数据时代,文本是存储数据和元数据最普遍的形式之一,数据表示是数据挖掘中特征提取阶段的关键步骤.因此,如何构建一个合适的文本表示模型,能够更好地描述文本数据的固有特征,仍然是一个挑战.本文从词共现关系、语句的自依赖关系和互依赖关系两方面构建分别构建词图模型和有向语句图模型,并采用图卷积网络提取文本特征,进一步得到文本最终的嵌入表示用于识别对话情感状态.该模型主要由4部分组成:字图编码结构、顺序上下文语句编码结构、交互上下文语句编码结构和情感分类器.所提出框架的总体架构如图1所示。

1.2 字图编码

1.2.1 词共现图构建

词语是构成语句的主要成分,词语在决定句子的情感极性方面起着最重要的作用,词语的出现、排列和定位以及文档不同组成部分之间的关系对于理解文档是必要的和有价值的.对话情感识别的第一个关键问题是对对话语句词语进行建模并提取特征,尽管有各种基于图形的文档表示模型,但词共现图是表示社交媒体内容中一个词和另一个词之间关系的有效方法.在文献[33]的启发下,本文构建窗口大小为3的词图.图2给出了一个示例句子的图,当两个节点在一个最大长度为3的窗口内时,则认为两节点之间具有连边。

1.2.2 字图编码结构

在本文提出的方法中,为每个数据样本构造了一个异构图 $G_W = (V, E_W)$,其中, V 和 E_W 分别是字节点集合和字-字之间的边集合, $v_i \in V, i \in [1, 2, \dots, M], M$ 为语句中包含的字数,使用邻接矩阵 A 表示字图中的节点关系.然后,将该图输入到GCN网络中,根据图中的关系,利用卷积运算对局部信息进行有效编码.对于 l 层GCN节点表示定义为 Z^l_W ,其中, $l = 1, 2, \dots, L$.第 l 层的第 i 节点表示被定义为 $z^l_{W_i}$.然后,每个节点的表示根据式(1)中的卷积运算来更新:

$$z^l_i = \text{ReLU} \left(\sum_{j=1}^M c^j A_{ij} (W^l z^l_{j-1} + b^l) \right) \quad (1)$$

其中, $c^i = 1/D_{ii}$ 为归一化常数, D 表示节点的度矩阵, D_{ii} 表示节点 i 在图中的度, W^l 和 b^l 为可训练的权重和偏差. 最后一层的节点表示为 $z_W^L = \{z_{W_1}^L, z_{W_2}^L, \dots, z_{W_M}^L\}$,

GCN 各层之间传播的信息表示为:

$$Z_W^l = \text{ReLU}(\tilde{A}Z^{(l-1)}W^l + b^l) \quad (2)$$

其中, $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$.

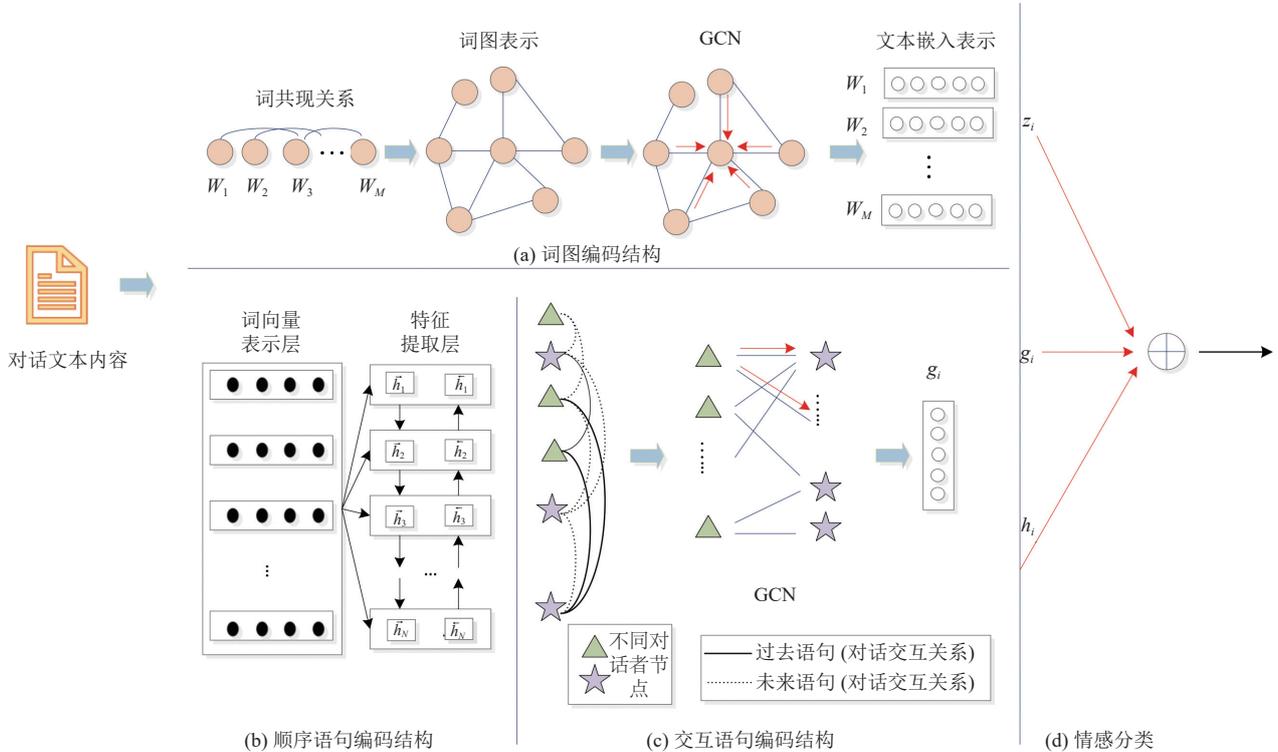


图1 基于图卷积网络的客服对话情感分析框架



图2 窗口为3的示例词图

1.3 语句编码

1.3.1 构建对话语句有向图

对话情感识别另一个关键的是对语句上下文进行建模, 语句上下文包括说话者自身语句的上下文以及双方对话形成的上下文. 受 Majumder 等^[24] 的启发, 针对每一句对话本文都通过两种上下文关系进行建模. 但考虑到在对上下文语句进行建模时, 情感动态会随之改变, 即对话双方未来语句的情感状态不仅会受到自身过去语句的影响, 还会受到对方过去语句的影响, 这种对话语句自身的依赖性和语句交互的依赖性分别

称为顺序语句自身依赖和交互语句相互依赖关系. 在本文提出的方法中, 将这两种截然不同但相关的上下文信息 (顺序语句编码和交互语句编码) 结合起来, 实现增强的上下文表示, 从而更好地理解会话系统中的情绪状态. 如图3所示为语句自依赖和交互语句依赖关系.

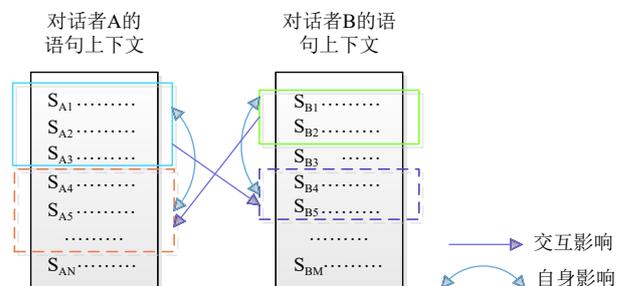


图3 语句自依赖和交互语句依赖关系图

1.3.2 顺序语句编码

对话的本质是一种序列信息, RNN 模型被广泛用

于序列信息建模, GRU 是一种与 LSTM 相似的递归神经网络, 它也被用来解决反向传播中的长期记忆和梯度问题. GRU 的输入输出结构类似于 RNN, 其内在思想类似于 LSTM. 与 LSTM 相比, GRU 具有更少的“选通”和更少的参数, 但可以实现与 LSTM 相同的性能. 因此考虑到计算开销, 本文选择 Bi-GRU 来捕捉上下文信息. 首先在编码语句时不区分说话人, 对全部对话语句信息进行特征表示. 以前向传播 GRU 为例, 每个 GRU 单元包含更新门和重置门, 给定状态向量 $\mathbf{h}^t \in \mathbb{R}^F$, 当前隐藏状态可计算如下:

$$\begin{cases} \mathbf{z}^t = \sigma(\mathbf{W}_z \mathbf{h}^t + \mathbf{U}_z \mathbf{h}^{t-1}) \\ \mathbf{r}^t = \sigma(\mathbf{W}_r \mathbf{h}^t + \mathbf{U}_r \mathbf{h}^{t-1}) \\ \tilde{\mathbf{h}}^t = \tanh(\mathbf{W}_s \mathbf{h}^t + \mathbf{U}_s (\mathbf{r}^t \odot \mathbf{h}^{t-1})) \\ \mathbf{h}^t = (1 - \mathbf{z}^t) \odot \mathbf{h}^{t-1} + \mathbf{z}^t \odot \tilde{\mathbf{h}}^t \end{cases} \quad (3)$$

其中, $\mathbf{z}^t \in \mathbb{R}^F$ 更新门; $\mathbf{r}^t \in \mathbb{R}^F$ 是重置门; $\mathbf{W}_z, \mathbf{W}_r, \mathbf{W}_s \in \mathbb{R}^{F \times D}$, $\mathbf{U}_z, \mathbf{U}_r, \mathbf{U}_s \in \mathbb{R}^{D \times F}$ 是可训练参数; $\sigma(\cdot)$ 是激活函数; \odot 是哈达玛乘积. Bi-GRU 分别在时间维以前向和后向依次处理输入序列, 并将每个时间步 RNN 的输出拼接成为最终的输出层, 进而得到的网络输出为:

$$\mathbf{h}_i = \left[\vec{\mathbf{h}}^i \oplus \overleftarrow{\mathbf{h}}^i \right] \quad (4)$$

这样每个时间步的输出节点, 都包含了输入序列中当前时刻完整的过去和未来的上下文信息.

1.3.3 交互语句上下文编码

定义上下文语句有向图 $G = (V, E, R)$, 每条语句表示一个顶点, V 为所有顶点集合, $v_i \in V, i \in [1, 2, \dots, K]$, K 为语句总数, 由于 $|V| + |R| \geq 2$, 因此该图为异质图. 在异质图 G 中, 每个顶点 v_i 的初始向量都由顺序语句编码得到的特征矩阵 \mathbf{h}_i 进行初始化. E 为所有边关系的集合, $\mathbf{a}_{ij} \in E$ 表示句子对 (v_i, v_j) 之间的连边关系. 在对话文本中, 由于语句表达有先后顺序, 因此在本文中把语句 v_i 前后的语句分别称为过去语句和未来语句, 并通过设置两种大小的窗口来构建语句之间的连边关系, 即过去语句的窗口大小为 P , 未来语句的窗口大小为 Q , 语句 v_i 与过去语句 $v_{i-1}, v_{i-2}, \dots, v_{i-P}$ 和未来语句 $v_{i+1}, v_{i+2}, \dots, v_{i+Q}$ 分别组成语句对. 其次, 与语句 v_i 构成连边的关系有两种, 一种是说话者本人的语句, 另一种是对方的语句, 因此设置 $r \in R$, 表示不同种类的连边关系. 而针对每一组语句对, 其不同语句针对目标语句的重要性并

不完全相同, 因此在特定连边关系 r 情况下, 为了计算语句 v_i 与不同语句之间的连边权重, 本文采用基于相似度的注意力算法计算, 权重计算如下:

$$a_{ij}^r = \frac{\exp(\mathbf{v}_i^T, \mathbf{v}_j)}{\sum_{i=1}^K \exp(\mathbf{v}_i^T, \mathbf{v}_j)}, j = i - P, \dots, i + Q \quad (5)$$

顶点 v_i 通过聚合局部邻域信息 (由过去和未来的上下文窗口大小指定的邻域语句), 利用边信息利用节点对 (v_i, v_j) 之间边的权重 a_{ij} 以及节点的隐藏状态 \mathbf{h}_i^t 和 \mathbf{h}_j^t 信息得到更新信息 \mathbf{m}_i^{t+1} :

$$\mathbf{m}_i^{t+1} = \sum_{r \in R} \sum_{j \in N(i)} f(\mathbf{h}_i^t, \mathbf{h}_j^t, a_{ij}^r) \quad (6)$$

其中, $f(\cdot)$ 为连接操作, $N(i)$ 为节点 v_i 在关系 r 下邻居节点的集合. 在得到邻居聚合信息后, 进一步利用节点本身隐藏状态信息 \mathbf{h}_i^t 和更新信息 \mathbf{m}_i^{t+1} 得到节点的状态信息 \mathbf{g}_i^{t+1} :

$$\mathbf{g}_i^{t+1} = \sigma(\mathbf{h}_i^t + \mathbf{m}_i^{t+1}) \quad (7)$$

其中, $\sigma(\cdot)$ 是激活函数, 式 (6) 和式 (7) 有效地聚合语句上下文有向图中每个语句的邻域信息.

1.4 情感分类与模型优化

为了实现对话情感状态的分析, 我们设置了 3 种聚合函数 $f(\cdot)$ (求和聚合、最大池化聚合、串联聚合) 将上述 3 种编码器得到的顺序语句编码特征、交互上下文语句特征以及字图编码特征聚合得到对话文本的嵌入表示 e_i , 3 种聚合函数分别为:

$$e_i = f(z_i, h_i, g_i) = z_i + h_i + g_i \quad (8)$$

$$e_i = f(z_i, h_i, g_i) = \text{element-wise-max}(z_i, h_i, g_i) \quad (9)$$

$$e_i = f(z_i, h_i, g_i) \quad (10)$$

在实验中我们进一步测试了不同聚合函数的性能, 在将以上 3 种嵌入表示聚合后, 进一步基于注意机制来获得最终的句子表示:

$$\begin{cases} \beta_i = \text{Softmax}(\mathbf{e}_i^T \mathbf{W}_\beta [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N]) \\ \tilde{\mathbf{e}}_i = \beta_i [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N]^T \end{cases} \quad (11)$$

最后, 使用全连接网络对话语进行分类:

$$\begin{cases} l_i = \text{ReLU}(\mathbf{W}_l \tilde{\mathbf{e}}_i + \mathbf{b}_l) \\ P_i = \text{Softmax}(\mathbf{W}_l l_i + \mathbf{b}) \\ \hat{y}_i = \arg \max(P_i[k]) \end{cases} \quad (12)$$

使用分类交叉熵和L2正则化作为训练期间损失的度量:

$$L = -\frac{1}{\sum_{i=1}^N s(i)} \sum_{i=1}^N \sum_{j=1}^{s(i)} \log p_{(i,j)}[\hat{y}_{i,j}] + \lambda \|\theta\|_2 \quad (13)$$

其中, N 是样本(对话)的数量, $s(i)$ 是样本 i 中的话语数量, P_{ij} 是对话 i 中语句 j 的情感标签的概率分布; y_{ij} 是对话 i 中语句 j 的期望标签, λ 是正则化权重, θ 是所有可训练参数的集合. 在实验中使用基于随机梯度下降的Adam优化器来训练. 为避免网络模型训练过程中出现过拟合现象, Dropout设置为0.5.

2 实验

在实验中, 首先将所提出的CS-GCN方法与基线和性能较好的情感分析方法进行了比较, 验证了不同上下文窗口对情感分析准确率以及F1值的影响, 并且验证了不同聚合函数的性能, 最后设置了消融实验来验证采用词共现关系, 顺序语句编码器和交互语句上下文编码器进行对话情感分析的有效性. 实验环境为: Intel Xeon 2.30 GHz, RAM 256 GB 和 Nvidia Quadro P620 GPU; 操作系统和软件平台是 Ubuntu 16.04、TensorFlow 1.14 和 Python 3.5.

2.1 数据集与评价指标

IEMOCAP: 南加州大学的SAIL实验室收集的成人对话数据. 该数据集包括151段对话, 共7433句. 该数据集标注了6类情绪: 中立, 快乐, 悲伤, 愤怒, 沮丧, 兴奋, 其中, 非中立情绪占比77%. 为简化实验, 实验中对6种情绪状态只进行积极、消极以及中立3种情感分类.

AVEC2012: SEMAINE数据库收集的多模态对话数据, 由4个固定形象的机器人與人进行对话, 曾用于AVEC2012挑战赛使用的数据. 该数据有95段对话, 共5798句. 标注了4个情感维度: Valence, Arousal, Expectancy, Power. Valence表示情感积极的程度, Arousal表示兴奋的程度, Expectancy表示与预期相符的程度, Power表示情感影响力. 其中, Valence、Arousal和Expectancy为[-1, 1]范围内的连续值, Power为大于等于0的连续值.

为了评价不同模型的情感分析准确率, 采用准确率和F1值作为评价指标

2.2 情感分析对比

为了全面评估CS-GCN, 将本文提出的模型与以下方法进行比较, 选择串联操作作为聚合函数, 并且在参数设置过程中, 尽量保持4种文本模型在参数设置上的一致性, 由于模型结构的不同而无法达到一致性时, 采用最优的参数设置. IEMOCAP和AVEC2012中没有提供预定义的训练集和验证集分割, 因此使用10%的训练对话作为验证集.

CNN^[12]: 基于卷积神经网络的基线模型, 该模型是上下文无关的, 不使用上下文语境中的信息. 该模型中, 词向量维度为100, 学习率为0.001, 滤波窗口选择3、4和5; Dropout为0.5, 批大小为64.

Memnet^[34]: 每一个话语都被反馈到网络中, 与先前话语相对应的记忆以多跳方式不断更新, 最后利用记忆网络的输出进行情感分类. 该模型中, 学习率是0.01, 每25个epoch学习率自动减半, 直到100个epoch. 在梯度下降中, 没有采用动量和权重衰减. 矩阵采用Gaussian distribution随机初始化, 均值为0, 方差为0.1. 批大小为32. 采用L2正则化.

c-LSTM^[35]: 通过使用双向LSTM网络捕获上下文内容进行建模, 然后进行情感分类, 但是上下文LSTM模型与说话人之间的交互无关, 即不对语句间交互依赖关系建模. 该模型词向量维度为100, Bi-LSTM隐藏层有64个神经元, 学习率为0.001, Dropout为0.5, 同时设置双向LSTM的层数为128层.

CMN^[36]: CMN使用两个不同的GRU模型为两个说话人从对话历史中建模话语语境. 最后, 将当前话语作为查询反馈给两个不同的记忆网络, 得到两个说话人的话语表示. 对于CMN模型, 卷积层的滤波器大小为3、4和5, 每个滤波器具有50个特征映射, 在这些特征图上使用最大池, 池窗口大小为2, 最后使用100个神经元的完全连接层.

DialogueRNN^[24]: 基于说话者、过去话语的上下文和过去话语的情绪, 应用了3个GRU来对这3方面进行建模, 该模型是目前在对话情感分析领域取得最好性能的模型, 因此在本文中选用原文的参数进行实验.

在实验中将所有的实验都独立运行5次, 并取平均值作为最终实验结果, 如表1和表2所示.

在IEMOCAP数据集上, CS-GCN的准确率和F1值分别为68.34%和71.26%, 相比于DialogueRNN模型的准确率和F1值分别提高了5%和6%左右. 在

AVEC2012数据集上, CS-GCN在所有4个情感维度上的准确率和F1值也同样都优于其他模型. 与CNN, Memnet, c-LSTM和CMN 4个基线模型相比, CS-GCN和DialogueRNN都对交互语句级别的上下文进行建模, 而其他基线模型都没有对交互语句之间的依赖关系进行编码, 因此其情感分类的准确率优于基线模型. 其次, CS-GCN和DialogueRNN都对交互语句级别的上下文进行建模, 但CS-GCN同时采用了词共现关系对语句进行编码, 其次CS-GCN通过卷积神经网络对语句的邻域信息进行聚合进而得到该语句的特征表示, 因此相比于DialogueRNN模型, CS-GCN模型获得了更好的情感分类性能.

表1 IEMOCAP数据集在不同模型上的情感分类准确率对比(%)

模型	积极		中立		消极	
	准确率	F1	准确率	F1	准确率	F1
CNN	43.53	54.62	34.32	39.25	49.23	50.28
Memnet	55.33	60.28	58.21	52.74	53.43	62.57
c-LSTM	56.24	56.12	54.22	51.43	58.67	60.21
CMN	59.37	58.56	52.47	52.38	61.14	59.37
DialogueRNN	62.83	64.52	58.34	59.12	63.22	65.24
CS-GCN	68.34	71.26	62.36	65.41	68.81	70.24

表2 AVEC2012数据集在不同模型上的情感分类准确率对比(%)

模型	Valence		Arousal		Expectancy		Power	
	准确率	F1	准确率	F1	准确率	F1	准确率	F1
CNN	53.21	50.33	49.24	50.16	47.28	50.13	49.35	51.32
Memnet	52.53	53.86	53.16	54.31	48.21	50.34	53.25	56.86
c-LSTM	54.24	56.48	52.58	55.66	53.85	58.56	55.12	57.13
CMN	58.75	59.76	60.63	58.78	61.12	66.31	59.81	60.34
DialogueRNN	64.36	63.54	63.38	65.43	68.43	70.25	66.45	63.26
CS-GCN	70.12	71.35	69.34	70.26	73.21	72.33	69.83	70.77

2.3 上下文窗口设置

在实验中分别验证了不同的上下文窗口对情感分类准确率的影响, 分别设置 $P=10, Q=10$; $P=8, Q=8$; $P=4, Q=4$; $P=0, Q=0$ 四组实验进行对比, 在两个数据集上得到的准确率和F1值如图4所示.

由图4可知, 在两类数据集上, 随着上下文窗口的减小, CS-GCN模型的情感分类性能也逐步下降, 当上下文窗口大小为 $P=0, Q=0$, CS-GCN模型只对词共现结构和顺序上下文进行编码, 不再对交互语句之间的依赖关系进行编码, 因此情感分类性能较差.

2.4 聚合函数测试

为了验证不同聚合函数对情感分类准确率的影响,

我们测试了3种聚合函数在两个数据集上的性能, 实验准确率如表3所示.

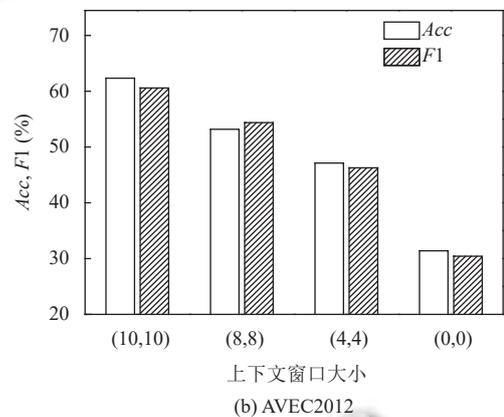
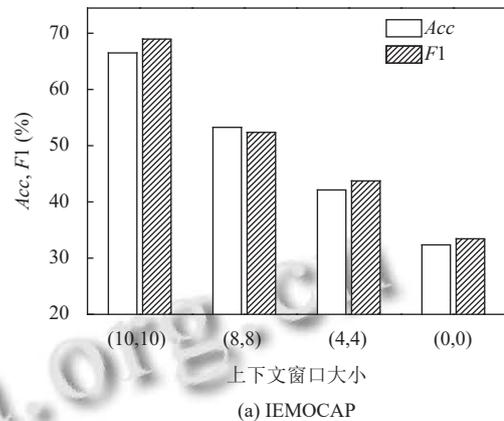


图4 不同上下文窗口大小情感分类性能对比

表3 不同聚合函数测试准确度(%)

数据集	聚合函数类型		
	求和	最大池化	串联
IEMOCAP	56.24	53.31	69.35
AVEC2012	58.72	56.83	70.23

显然, 串联融合达到了最佳准确度, 而其他两个聚合函数的性能相对较差. 这是因为串联操作在嵌入聚合期间保留比求和和最大池更多的信息, 有助于学习更好的节点信息表示.

2.5 消融实验

为了验证采用词共现关系, 顺序语句编码器和交互语句上下文编码器进行对话情感分析的有效性, 在两个数据集上对几种模型的情感预测性能进行了实验比较. 首先, 将只有顺序语句编码器模型视为对话语句情感分析的基线模型, 记为CS-GCN_N. 此外, 为了验证词共现关系和交互语句上下文编码对情感分析的有效性, 在基线模型的基础上加入了词共现编码结构记为

CS-GCN_W. 为了验证交互语句上下文编码的适用性, 在基线模型的基础上增加了交互语句上下文编码结构, 记 CS-GCN_S. 最后, 将词共现关系编码结构和交互语句上下文编码结构嵌入到基线模型之上, 进一步验证组合的优越性. 模型类型如表 4 所示, 实验结果如图 5 所示.

表 4 不同模型类型

模型	顺序语句上下文编码	词图编码	交互语句上下文编码
CS-GCN _N	√	×	×
CS-GCN _W	√	√	×
CS-GCN _S	√	×	√
CS-GCN	√	√	√

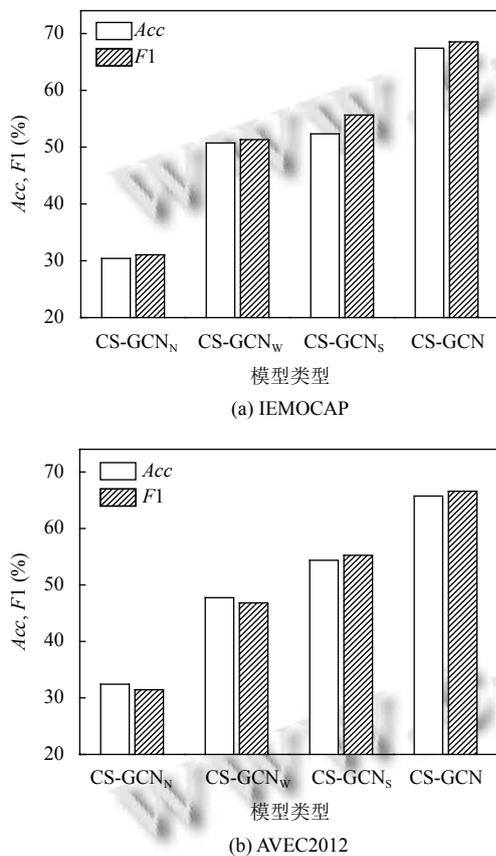


图 5 不同模型类型情感分类性能对比

从图 5 可以看出, 随着词共现关系编码器, 交互语句上下文编码器结构的加入, 情感分类的准确率和 $F1$ 得到了逐步提高. 在数据集 IEMOCAP 上, 与 CS-GCN_N 模型相比, CS-GCN_W 模型的准确率和 $F1$ 值分别提高了 10.38% 和 10.05% 左右. 与 CS-GCN_N 模型相比, CS-GCN_S 模型的准确率和 $F1$ 值分别提高了 13.66% 和 10.75%. 通过词共现关系编码器, 交互语句

上下文编码器结构的加入, 证明了 CS-GCN_W 和 CS-GCN_S 模型能够提高情感分类的性能, 具有普遍适用性. 其次, 在 CS-GCN 模型中, 实现了 3 类编码结构的嵌入和融合用于情感分类, 与 CS-GCN_N 模型相比, CS-GCN 模型的准确率和 $F1$ 值分别提高了 40.98% 和 38.06%. 最后, 将 CS-GCN 模型与 CS-GCN_W 和 CS-GCN_S 模型相比较, 准确率和 $F1$ 值同样得到了显著提高. 综上所述, 这些对比实验充分验证了本文提出的 CS-GCN 模型通过引入词共现关系编码器, 交互语句上下文编码器结构, 显著提升了对话情感分析的性能.

3 结论

受异质图卷积网络的启发, 本文提出了一种基于图卷积网络的客服对话情感分析方法. 提出的 CS-GCN 模型与传统的基于 RNN 模型的对话数据情感检测方法不同, CS-GCN 模型将词共现关系, 语句的顺序上下文特征以及不同说话人语句之间的交互依赖关系进行融合编码, 通过图卷积网络实现对文本特征提取与聚合得到最终的文本表示, 进一步实现对对话情感的状态识别. 在实验中, 将 CS-GCN 模型与基础 CNN, Memnet, c-LSTM, CMN 和 DialogueRNN 模型在两个数据集进行情感分类准确率的对比, 验证了 CS-GCN 模型可有效提高情感识别准确度. 其次, 设计了相关的消融实验验证了融合词共现编码, 顺序语句上下文编码以及交互语句上下文编码 3 种结构对于对话情感识别准确率提高的有效性. 在未来的工作中可将多模态信息整合到 CS-GCN 网络中, 进一步提高对说话人之间的情感转换检测, 为客服的质检工作提供及时、客观、有效的分析.

参考文献

- 1 王林信, 杨鹏, 江元, 等. 智能电网大数据隐私保护技术研究. 电力信息与通信技术, 2019, 17(12): 24-30.
- 2 Kratzwald B, Ilić S, Kraus M, et al. Deep learning for affective computing: Text-based emotion recognition in decision support. arXiv: 1803.06397, 2018.
- 3 Colnerič N, Demšar J. Emotion recognition on twitter: Comparative study and training a unison model. IEEE Transactions on Affective Computing, 2020, 11(3): 433-446. [doi: 10.1109/TAFFC.2018.2807817]
- 4 雷鸣. 电商客服领域对话系统的设计与实现 [硕士学位论文]. 北京: 北京交通大学, 2020.

- 5 吕诗宁, 张毅, 胡若云, 等. 融合神经网络与电力领域知识的智能客服对话系统研究. 浙江电力, 2020, 39(8): 76–82.
- 6 杨昕. 基于客服会话的客户购物意愿预测模型研究与应用 [硕士学位论文]. 南昌: 南昌大学, 2020.
- 7 胡若云, 孙钢, 丁麒, 等. 基于双向传播框架的客服对话文本挖掘算法. 沈阳工业大学学报, 2021, 43(2): 188–192. [doi: 10.7688/j.issn.1000-1646.2021.02.12]
- 8 王建成. 面向对话文本的情感分类方法研究 [硕士学位论文]. 苏州: 苏州大学, 2020.
- 9 Wang JC, Wang JJ, Sun CL, *et al.* Sentiment classification in customer service dialogue with topic-aware multi-task learning. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(5): 9177–9184. [doi: 10.1609/aaai.v34i05.6454]
- 10 Brown PF, deSouza PV, Mercer RL, *et al.* Class-based N -gram models of natural language. Computational Linguistics, 1992, 18(4): 467–479.
- 11 Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, Volume 10. Philadelphia: ACM, 2002. 79–86.
- 12 Kim Y. Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: ACL, 2014. 1746–1751.
- 13 Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore: ACL, 2014. 655–665.
- 14 Xu JC, Chen DL, Qiu XP, *et al.* Cached long short-term memory neural networks for document-level sentiment classification. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: ACL, 2016. 1660–1669.
- 15 Zhou XJ, Wan XJ, Xiao JG. Attention-based LSTM network for cross-lingual sentiment classification. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: ACL, 2016. 247–256.
- 16 宋曙光, 徐迎晓. 基于多注意力网络的特定目标情感分析. 计算机系统应用, 2020, 29(6): 163–168. [doi: 10.15888/j.cnki.csa.007427]
- 17 Wang J, Yu LC, Lai KR, *et al.* Dimensional sentiment analysis using a regional CNN-LSTM model. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: ACL, 2016. 225–230.
- 18 Poria S, Majumder N, Mihalcea R, *et al.* Emotion recognition in conversation: Research challenges, datasets, and recent advances. IEEE Access, 2019, 7: 100943–100953. [doi: 10.1109/ACCESS.2019.2929050]
- 19 王伟, 孙玉霞, 齐庆杰, 等. 基于 BiGRU-attention 神经网络的文本情感分类模型. 计算机应用研究, 2019, 36(12): 3558–3564.
- 20 孟仕林, 赵蕴龙, 关东海, 等. 融合情感与语义信息的情感分析方法. 计算机应用, 2019, 39(7): 1931–1935. [doi: 10.11772/j.issn.1001-9081.2018112375]
- 21 邱宁佳, 王晓霞, 王鹏, 等. 融合语法规则的双通道中文情感模型分析. 计算机应用, 2021, 41(2): 318–323.
- 22 Shen CL, Sun CL, Wang JJ, *et al.* Sentiment classification towards question-answering with hierarchical matching network. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: ACL, 2018. 3654–3663.
- 23 Wang JJ, Sun CL, Li SS, *et al.* Aspect sentiment classification towards question-answering with reinforced bidirectional attention network. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 3548–3557.
- 24 Majumder N, Poria S, Hazarika D, *et al.* DialogueRNN: An attentive RNN for emotion detection in conversations. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 6818–6825.
- 25 Yao L, Mao CS, Luo Y. Graph convolutional networks for text classification. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 7370–7377.
- 26 Wu F, Zhang TY, de Souza AH Jr, *et al.* Simplifying graph convolutional networks. Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019. 6861–6871.
- 27 Zhang C, Li QC, Song DW. Aspect-based sentiment classification with aspect-specific graph convolutional networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: ACL, 2019. 4568–4578.
- 28 陈佳伟, 韩芳, 王直杰. 基于自注意力门控图卷积网络的特定目标情感分析. 计算机应用, 2020, 40(8): 2202–2206.
- 29 Zuo EG, Zhao H, Chen B, *et al.* Context-specific heterogeneous graph convolutional network for implicit sentiment analysis. IEEE Access, 2020, 8: 37967–37975. [doi: 10.1109/ACCESS.2020.2975244]

- 30 Zhang CX, Song DJ, Huang C, *et al.* Heterogeneous graph neural network. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 793–803.
- 31 Wang DQ, Liu PF, Zheng YN, *et al.* Heterogeneous graph neural networks for extractive document summarization. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 6209–6219.
- 32 Xie YX, Xu H, Li JE, *et al.* Heterogeneous graph neural networks for noisy few-shot relation classification. Knowledge-Based Systems, 2020, 194: 105548. [doi: [10.1016/j.knosys.2020.105548](https://doi.org/10.1016/j.knosys.2020.105548)]
- 33 Violos J, Tserpes K, Psomakelis E, *et al.* Sentiment analysis using word-graphs. Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics. Nîmes: ACM, 2016. 1–9.
- 34 Sukhbaatar S, Szlam A, Weston J, *et al.* End-to-end memory networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 2440–2448.
- 35 Poria S, Cambria E, Hazarika D, *et al.* Context-dependent sentiment analysis in user-generated videos. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver: Association for Computational Linguistics, 2017. 873–883.
- 36 Hazarika D, Poria S, Zadeh A, *et al.* Conversational memory network for emotion recognition in dyadic dialogue videos. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans: Association for Computational Linguistics, 2018. 2122–2132.