

# 基于 ATT-DGRU 的文本方面级别情感分析



尹久 池凯凯 宦若虹

浙江工业大学计算机科学与技术学院 杭州 310023

**摘要** 方面级别情感分类是针对给定文本、分析其在给定方面所表达出的情感极性。现有的主流解决方案中,基于注意力机制的循环神经网络模型忽略了关键词邻近上下文信息的重要性,而结合卷积神经网络(Convolutional Neural Network, CNN)的多层模型不擅长捕捉句子级别的长距离依赖信息。因此,提出了一种基于截断循环神经网络(Disconnected Gated Recurrent Units, DGRU)和注意力机制的方面级别情感分类网络模型(Attention-Disconnected Gated Recurrent Units, ATT-DGRU)。DGRU 网络综合了循环神经网络和 CNN 的优点,既能捕捉文本的长距离依赖语义信息,又可以很好地抽取关键短语的语义信息。注意力机制在推断方面情感极性时捕获每一个单词与给定方面的关联程度,同时生成一个情感权重向量用于可视化。ATT-DGRU 模型在中文酒店评论数据集上进行 ACSA 任务,任务结果表明,其二分类、三分类准确率分别达到 91.53%, 86.61%;在 SemEval2014-Restaurant 数据集进行 ATSA 任务,任务结果表明,其二分类、三分类准确率分别可达 90.06%, 77.21%。

**关键词:** 方面情感分析;深度学习;截断循环神经网络

**中图分类号** TP391

## Aspect-level Sentiment Analysis of Text Based on ATT-DGRU

YIN Jiu, CHI Kai-kai and HUAN Ruo-hong

School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

**Abstract** Aspect-level sentiment classification is to analyze the sentiment polarity in a given aspect for a given text. In the existing mainstream solutions, the attention mechanism-based cyclic neural network model ignores the importance of keyword proximity context information, and the CNN multilayer model is not good at capturing sentence-level long-distance dependency information. This paper proposes an aspect-level emotion classification network model based on disconnected gated recurrent units (DGRU) and attention mechanism, which is abbreviated as ATT-DGRU. The DGRU network used in this model integrates the advantages of circular neural network and CNN. It can not only capture the long-distance dependent semantic information of text, but also extract the semantic information of key phrases. Attention mechanism is used to capture the importance of each word to a specific aspect when deducing the sentiment polarity of a specific aspect, meanwhile generates an emotional weight vector, which can be visualized. Accuracies of two-class and three-class of ATT-DGRU model constructed in this paper can reach 91.11% and 87.76% respectively in ACSA task on Chinese hotel review datasets. Accuracies of two-class and three class of ATT-DGRU model can reach 77.21% and 90.06% respectively in ATSA task on SemEval2014-Restaurant dataset.

**Keywords** Aspect-level sentiment classification, Deep learning, Disconnected gated recurrent unit

## 1 引言

情感分析<sup>[1]</sup>是自然语言处理领域的一项备受关注的研究任务,文本情感分析的主要目的是从原始文本中提取出文本发表者的主观情绪。如今互联网各平台上存在大量用户关于产品和服务的评论信息,这些评论往往以短文本的形式出现,涉及对产品的一个或者多个方面的意见。传统评论挖掘通常粗糙地对一篇文档或者一段文字进行情感分析,只能得到一个情感值的表达,而不能提供高情感值指向的对象的信息,因此,方面级别的观点挖掘<sup>[2-3]</sup>是近年来的研究热点。

方面级别情感分析包括两个子任务:方面词提取任务和基于方面的情感分析<sup>[4]</sup>。基于方面的情感分析(Asspect-Based Sentiment Analysis, ABSA)中的方面(aspect)可以指显式出现在句子中的一个实体(target, aspect term)或者隐式表达的一个描述事物的角度(aspect category)。因此,基于方面的情感分析根据方面词是否显示表达可以被划分为 ATSA (Aspect Term Sentiment Analysis)和 ACSA (Aspect Category Sentiment Analysis)两个子任务<sup>[4]</sup>。方面词可以是一个单词,也可以是一个短语,本文将将其统称为方面词。本文主要研究方面级别的情感分类 ATSA 与 ACSA,即针对句子 S 中所描

收稿日期:2020-05-18 返修日期:2020-07-30 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61872322)

This work was supported by the National Natural Science Foundation of China(61872322).

通信作者:池凯凯(kkchi@zjut.edu.cn)

述的方面 A (Aspect Category 或 Aspect Term), 分析该语句 S 所表达的有关给定 A 方面的情感极性。例如评论:“服务很好,就是房间太旧了都是十几年前的装修了”。该评论文本描述了酒店的“服务”和“房间”两个方面实体项,相应的方面级别情感分类结果分别是“积极”和“消极”。

评论文本往往涉及一个或者多个方面,方面级别的情感分析可以更精确地捕捉产品细粒度的优缺点。因此自动化的文本情感分析成为了必要手段,被广泛应用于医疗、旅游、金融、舆情分析等领域<sup>[4]</sup>。

随着深度学习方法特别是神经语言模型近年来在自然语言处理领域不断取得新的进展,越来越多的方面级别情感分类模型开始采用深度学习模型。早期的模型采用递归神经网络(Recursive Neural Networks, RNN)或其变体长短期记忆网络(Long Short Term Memory Network, LSTM)、门控循环单元(Gated Recurrent Units, GRU),并利用句法结构信息辅助来提高分类准确率,但结果表明此类模型的性能受句法分析错误的影响较大,且无法对口语表达进行有效的语义分析。近期性能表现较好的方面级别情感分类模型大多以循环神经网络或 CNN 及其变体为基础构建而来。

循环神经网络是一类以序列数据为输入、在序列的演进方向进行递归且所有节点(循环单元)按链式连接的递归神经网络<sup>[5]</sup>,其擅长对整个句子进行建模,捕捉长距离依赖信息。然而,研究表明,RNN 对整个句子的建模有时会成为一种负担,使模型忽略关键的短语信息。CNN 模型则正好相反,更擅长抽取局部的位置不变特征,不擅长捕捉长距离依赖信息。而方面级别情感分析不仅需要重点提取方面词向量邻近语义信息,而且更需要提取丢失文本的长距离语义信息。

为了在 RNN 的基础上结合 CNN 的优点,文献[6]设计了基于截断门控循环单元的网络来进行情感分类。DGRU 在 GRU 的基础上做出了改进。GRU<sup>[7]</sup>是 RNN 的一种特殊类型,可以通过重置门和更新门来捕捉序列间长距离依赖关系,其计算过程相比 RNN 的另一种类型 LSTM<sup>[8]</sup>更高效。DGRU 模型是通过限制 DGRU 层信息的流动来将位置不变性引入 GRU 模型中。DGRU 模型也可以被认为是一种特殊的 CNN 模型,其本质是将 CNN 模型中的卷积核替换成了 DGRU,使得 DGRU 模型既能捕捉长距离依赖信息,又可以很好地抽取关键短语信息。

综上,如果 DGRU 能在捕捉信息长距离依赖过程中保留位置不变性信息,那么理论上其可以准确地实现对复杂语句的方面级别情感分类。本文提出了一个新颖的基于 DGRU 和注意力机制的模型 ATT-DGRU,该模型同时遵循 Encoder-Decoder 框架。ATT-DGRU 首先将句子视为单词序列,以 DGRU 作为单词序列学习工具,能够在隐状态层获得句法层面的有效表达。然后,利用注意力机制对由 DGRU 模块得到的文本向量语义编码信息进行解码,提取出关于给定方面词的情感权重向量。其中,情感权重向量由每一个单词被分配的不同注意力权重系数组成,注意力权重系数用于区分文本中信息的重要程度。

本文的主要贡献如下:

(1)提出了一种新的方面级别情感分类模型 ATT-DGRU。该模型首次将 DGRU 应用于方面级别情感任务中,并

表现出较主流的 LSTM,GRU 网络更优秀的方面情感关键信息的提取效果。

(2)在酒店评论数据集上进行实验,数据集根据方面词是否出现在样本文本中,即方面词为实体方面项或方面角度而被分为显性数据集和隐性数据集两类。通过对比这两类数据集的实验结果发现,方面级别情感分析的准确率的提高不仅依赖于分类模型的性能,同时还取决于方面情感分析的另一子任务,即方面词提取任务中方面主题词的归纳。

## 2 相关工作

方面情感分类是情感分析领域的一个细粒度任务,其目标是识别句子中针对给定方面的情感极性<sup>[4]</sup>。早期的方面情感分析任务主要是通过机器学习方法结合语言学特征来进行探索,这是断定给定方面实体情感的典型方法。文献[9]构造了一个二进制支持向量机(Support Vector Machine, SVM)来对目标无关的特征和扩展信息进行分类。文献[10]将一组修剪特征作为整体分类器的输入,在方面项提取和情感分析方面表现出了有效性。文献[11]依赖 SentiWordNet,并根据特定条件在未标记的评论中创建了特定领域的情感词典。尽管这些方法已经取得了较好的分类结果,但是它们很少考虑目标实体和相应上下文之间的交互。此外,此类方法很大程度上依赖于人的先验或领域知识,且学习到的是浅层的特征,因此难以识别复杂的活动。

神经语言模型和表示学习近年来在自然语言处理领域不断取得新的进展,并更多地使用各种网络模型对文本语言进行编码,尽可能地对文本语义进行复现,在解码器模块对编码的表示进行解读,进一步获取文本语义中更有用的信息,生成最终的语义表达形式。编码器与解码器对同一文本所生成的表示不能相差太大,因此可以通过调整编码器与解码器的参数使得重构误差较小。

### 2.1 神经语言模型

近年来,方面情感分类领域中热门的神语言模型大多以循环神经网络或 CNN 及其变体为基础进行构建。

CNN 最初被应用于计算机视觉任务中,而后因为其能识别局部的位置不变特征,所以也被应用于自然语言领域中并取得了出色的结果。文献[12]首次将一维 CNN 应用于词性标注、命名实体标注、语义角色标注任务中。为了进一步捕捉文本词语间的联系,文献[13]提出了一种新的动态  $k$ -max 池的 CNN 模型。文献[14]建立了一个基于多层次关注的 CNN 结构 ATT-CNN,该结构在嵌入层和共享层分别实现了自我关注和集中关注。以上文献都证实了基于 CNN 的网络在自然语言任务中的有效性。

RNN 非常适合用于处理和预测序列数据,因此许多以 RNN 变体为基础的模型(如许多基于 LSTM,GRU 的模型)被广泛用于解决目标实体与其上下文之间缺乏交互的问题。文献[15]中的模型使用 LSTM 网络对文本序列进行语义建模,通过 LSTM 建模后的句子表达能体现文本上下文的语义联系。文献[16]运用双向 LSTM 建模目标实体与上下文之间的复杂关系,并提出了目标相关 LSTM 和目标连接 LSTM。为解决以上模型由信息远距离传输导致的性能不佳问题,文献[17]应用门控循环单元结构来控制情绪信息流并

提高了分类精度。文献[18]构建了一个基于双重 LSTM 的位置感知注意力模型 AE-DLSTM,该模型的双重 LSTM 结构可以捕获方面词正反两个方向的语义信息,并通过捕获位置信息来为上下文词生成不同的权重。文献[19]所构建的交互注意力网络(Interactive Attention Networks, IAN)模型也用到了两个 LSTM 结构,其中一个用于获取方面目标项的注意力权重,另一个用于获取上下文注意力权重,对两者进行串联分析后得到分类结果。

类递归神经网络<sup>[5]</sup>擅长对整个句子进行建模,捕捉长距离依赖信息,而 CNN 模型更擅长抽取局部的位置不变特征,但不擅长捕捉长距离依赖信息。类递归神经网络往往容易忽视关键的短语信息,因此无法很好地完成方面级别情感分类任务。不少研究者尝试设计一种能融合 CNN 与 RNN 的优点的网络结构。文献[20]提出一个基于 CNN 和 Gating Mechanisms 的方面级别情感分类模型 GCAE,该模型使用新颖的 Tanh-ReLU 门单元选取文本中给定方面的情感特征,并针对 ACSA 任务和 ATSA 任务设计了不同结构的卷积模块。文献[21]用深度 LSTM 代替了卷积过滤器在查询分类领域取得了很好的成果。文献[6]在文献[19]的基础上对模型做了进一步改进:1)提出使用更大的窗口尺寸以更好地捕捉长距离依赖信息;2)使用 Max-pooling 以更好地包含位置不变性信息;3)使用 GRU 替代 LSTM 作为循环单元。以上改进使 DGRU 模型在 7 个数据集上均取得了较好的分类结果。以上文献通过对 CNN 与 RNN 的融合进行改进,使 DRNN 类模型能更好地编码方面词与文本的语义信息,但是单纯基于 DGRU 的模型还是无法很好地捕捉到复杂语句中相对距离较远的情感极性词与方面词之间的潜在关联。后续的研究表明,通过引入注意力机制有助于解决该问题。

## 2.2 注意力机制

一条复杂语句中可能包含多个方面词,句中的每个词和方面词之间的联系能传达出关于该方面词的情感极性信息。通过引入注意力机制,能够从复杂语句中捕捉到与方面词关联较大的单词,降低对次要信息的关注度,甚至无视无关的信息。例如,文献[22]提出了一种基于注意力的 LSTM 模型 AB-LSTM,该模型根据每个单词对特定目标实体的情感重要性对每个单词的特征进行加权。文献[14]建立了一个基于多层次关注的 CNN 结构 ATT-CNN,借助注意力机制在嵌入层和共享层分别实现了自我关注和集中关注。文献[23]提出的 RAM 模型采用多注意力机制与 RNN 非线性相结合,在处理多并发时可增强模型的表达能力。

本文所设计的注意力模型首先将 DGRU 生成的语义隐状态构成外部记忆,然后将其与注意力机制所学习到的注意力分布向量进一步连接,最后生成最终能体现对给定方面词关注的完整语义向量。

## 3 构建模型

本文提出的 ATT-DGRU 模型采用了编码器-解码器网络结构。其中,基于 DGRU 的编码器用于读入文本序列并建模生成包含语义信息抽象表达的隐状态向量;解码器对该向量进行解读,输出关于给定方面词的情感语义向量,并将其输

入 Softmax 分类器,得到最终方面级别情感分类结果。最终模型的呈现将采用由简到繁的方式,本文将分成 3 步对其进行介绍。

首先,给出一个基于 GRU 的简单网络模型 GRU-MLP,用于说明模型的输入和输出,该模型事实上是一个句子级别的情感分类模型,在本文中作为基准模型,用于辅助评估 ATT-DGRU 模型的性能。

然后,构建了一个简单的基于 DGRU 的模型,该结构保留了 GRU 循环单元,并参考了 CNN 的过滤器结构,其每一个 DGRU 子单元只与邻近的  $k$  个单词有关。DGRU 作为 ATT-DGRU 的语义编码部分,直接对 DGRU 合成的文本语义编码进行分类,将得到的结果与 ATT-DGRU 进行比较,可以观察引入注意力机制的有效性。

### 3.1 GRU-MLP

RNN 是传统前馈神经网络的扩展。然而,标准的 RNN 存在梯度消失或爆炸问题。为了解决这些问题,LSTM<sup>[8]</sup>被开发出来并具有优异的性能。GRU<sup>[7]</sup>则是 LSTM 的一个变体,在 LSTM 的基础上将遗忘门与输入门合成一个更新门,同时没有单独的存储单元,因为其在保持相同效果的同时结构更加简单,所以近年来被广泛用于处理文本语义。GRU 结构如图 1 所示,通过门控单元可以控制信息的流动,一定程度上解决了梯度消失问题。GRU 通过重置门  $r_t$  和更新门  $z_t$  来控制信息流。 $t$  时间步时隐藏状态  $h_t = \text{GRU}(h_{t-1}, x_t)$ ,GRU 具体数据流的计算如下:

$$r_t = \sigma(W_r \times [h_{t-1}, x_t]) \quad (1)$$

$$z_t = \sigma(W_z \times [h_{t-1}, x_t]) \quad (2)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \times [r_t * h_{t-1}, x_t]) \quad (3)$$

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

其中, $W_r, W_z, W_{\tilde{h}} \in \mathbf{R}^{d \times d}$  分别为重置门  $r_t$ 、更新门  $z_t$ 、候选集  $\tilde{h}_t$  的权重矩阵, $x_t$  为当前时间步的输入, $h_{t-1}$  为  $t-1$  时间步的隐状态, $\sigma$  表示 Sigmoid 激活函数,\* 表示 Hadamard 乘积。

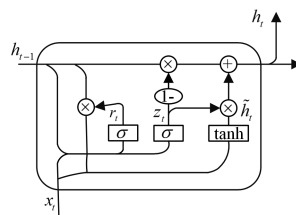


图 1 GRU 循环单元的内部结构

Fig. 1 Internal structure of GRU unit

本文所设计的基于 GRU 网络的多层感知机模型 GRU-MLP 的结构如图 2 所示。

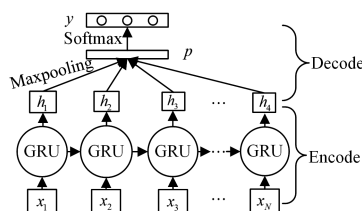


图 2 GRU-MLP 的结构

Fig. 2 Structure of GRU-MLP model

输入模型矩阵  $\mathbf{X} = \{\mathbf{X} \in \mathbf{R}^{2d \times N} | (x_1, x_2, x_3, \dots, x_i, \dots, x_{N-1}, x_N)\}$ , 其中,  $N$  为文本序列的长度,  $d$  为词向量长度,  $x_i = \begin{bmatrix} a \\ w_i \end{bmatrix}$ ,  $w_i$  为文本序列的词向量,  $a$  为该样本所给定方面词向量, 两向量的长度均为  $d$ . 经过 GRU 网络计算编码得到的隐状态  $\mathbf{H} = \{\mathbf{H} \in \mathbf{R}^{d \times N} | (h_1, h_2, h_3, \dots, h_i, \dots, h_{N-1}, h_N)\}$  包含每个输入单词的语义特征, 以及与给定方面词语的语义联系, 可将  $\mathbf{H}$  视为文本句子级别的表达。

GRU-MLP 解码器部分首先对  $\mathbf{H}$  进行最大池化采样操作, 将池化窗口设置为  $N \times 1$ , 设置滑动步长为 1, 生成一个长度为  $N$  的情感表达向量  $\mathbf{p}$ . 最大池化是一种捕捉重要信息的采样方式, 相比平均池化可以更好地保持输入特征的位置不变性<sup>[6]</sup>.

接下来对  $\mathbf{p}$  进行 ReLU 线性变换后送入 Softmax 分类器, 得到关于该语句的最终情感分类结果, 即:

$$y = \text{Softmax}(W_p p + b_p) \quad (5)$$

其中,  $\mathbf{y} \in \mathbf{R}^{|\mathcal{c}| \times 1}$  是概率分布向量, 当模型进行两分类时,  $|\mathcal{c}| = 2$ , 当模型进行三分类时,  $|\mathcal{c}| = 3$ .

### 3.2 DGRU

GRU 网络虽然擅长对整个句子进行建模, 捕捉长距离依赖信息, 但是往往容易忽视关键的短语信息, 因此无法很好地完成方面级别情感分类任务。为了减少建模整个句子的负担, 文献[6]首次提出的截断的门控循环单元网络模型模仿 CNN 的卷积过滤器结构限制了 GRU 信息流的距离, 通过分离 GRU 的信息流, 将位置不变性引入 GRU 的变体中。DG-GRU<sup>[6]</sup>的具体结构如图 3 所示:

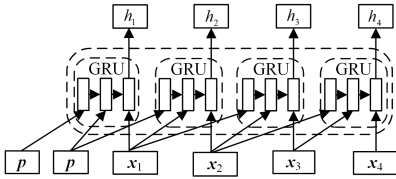


图 3 DGRU 的网络结构

Fig. 3 Structure of DGRU

与其他 RNN 变体一样, DGRU 将输入序列输入 GRU 单元中, 并在每一步生成一个输出向量。与普通 RNN 变体的一个重要区别是, DGRU 每一步的状态只与前面的  $k-1$  个单词有关, 而与其他单词无关。因此, 每一步的输出  $h_i$  可表示为:

$$h_i = \text{GRU}(x_i, x_{i-1}, x_{i-2}, \dots, x_{i-k+1}) \quad (6)$$

当  $t$  为 1 时,  $h_1$  由  $k-1$  个  $\mathbf{p}$  和  $x_1$  计算而来,  $h_1$  可表示为:

$$h_1 = \text{GRU}(\mathbf{p}, \dots, \mathbf{p}, x_1) \quad (7)$$

其中,  $\mathbf{p}$  是长为  $2d$  的零向量,  $k$  是需要设置的一个窗口大小的超参数。因为每一步的输出只取决于前  $k-1$  个单词和当前单词, 所以输出可以被看作是一个包含  $k$  个单词的短语的表示。具有相同  $k$  个单词且顺序一样的短语无论在文本的哪个位置都具有相同的表示。

### 3.3 ATT-DGRU

为引入对局部关键短语信息的关注, DGRU 网络牺牲了一定的对长距离依赖信息的捕捉能力。而在方面级别文本情

感分类中, 影响给定方面词情感的关键信息多来源于方面词周边。为解决这个问题, 本文进一步引入基于 DGRU 的注意力网络模型。DGRU 网络用于方面级别文本情感分类模型, 有利于捕捉给定方面词所依赖的语义信息, 而注意力机制的引入可以进一步将分类器的注意力引向与给定方面词有关的短语位置。最终, 本文构建了基于 DGRU 和注意力机制的方面级别情感分类网络模型 ATT-DGRU, 其结构如图 4 所示。

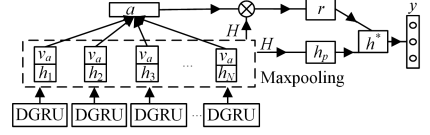


图 4 ATT-DGRU 网络模型的结构

Fig. 4 Structure of ATT-DGRU model

用 DGRU 替代 GRU-MLP 中的 GRU 作为编码器, 将编码器模块输出的隐状态表示向量输入到带有修正线性单元 (Tanh) 激活的多层感知机 (Multilayer Perceptron, MLP) 中, 并将 MLP 的输出发送到一个 Softmax 函数中, 以得到一个注意力权重向量, 该向量的每一个值代表对应位置单词的情感权重。具体计算过程如下, 定义  $\mathbf{a} \in \mathbf{R}^{N \times 1}$  为注意力情感权重向量, 以及一个文本情感权重表达隐状态,  $\mathbf{r} \in \mathbf{R}^{d \times 1}$ ,  $\mathbf{r}$  可由以下表达式得到:

$$\mathbf{M} = \tanh \left( \begin{bmatrix} \mathbf{W}_h \mathbf{H} \\ \mathbf{W}_v \mathbf{v}_a \otimes \mathbf{e}_N \end{bmatrix} \right) \quad (8)$$

$$\mathbf{a} = \text{Softmax}(\mathbf{w}^T \mathbf{M}) \quad (9)$$

$$\mathbf{r} = \mathbf{H} \mathbf{a}^T \quad (10)$$

其中,  $\mathbf{M} \in \mathbf{R}^{2d \times N}$ ,  $\mathbf{W}_v \in \mathbf{R}^{d \times d}$ ,  $\mathbf{W}_h \in \mathbf{R}^{d \times d}$ ,  $\mathbf{w} \in \mathbf{R}^{2d \times 1}$  都为权重矩阵或权重向量。  $\mathbf{v}_a \otimes \mathbf{e}_N = [\mathbf{v}_a; \mathbf{v}_a; \dots; \mathbf{v}_a]$ , 即方面词向量  $\mathbf{v}_a$  重复出现  $N$  次,  $\mathbf{e}_N$  表示长为  $N$  的单位向量。

本文应用 Maxpooling 将文本语义表达矩阵转化为文本语义表达向量, 并提取最重要的信息和位置不变的特征。最终给定方面的文本情感表达由文本语义表达隐状态  $h_p$  与文本情感权重表达状态  $\mathbf{r}$  相结合, 其表达式如下:

$$h_p = \text{Maxpooling}(\mathbf{H}) \quad (11)$$

$$h^* = \tanh(\mathbf{W}_r \mathbf{r} + \mathbf{W}_x h_p) \quad (12)$$

其中,  $h^* \in \mathbf{R}^{d \times 1}$ ,  $\mathbf{W}_r, \mathbf{W}_x$  是训练过程中需要学习的权重参数,  $h^*$  即视为给定方面的文本情感表达。

将  $h^*$  送入 Softmax 分类器, 得到情感分类的概率分布向量  $\mathbf{y} \in \mathbf{R}^{|\mathcal{c}| \times 1}$ , 如式 (13) 所示:

$$y = \text{Softmax}(\mathbf{W}_s h^* + b_s) \quad (13)$$

其中,  $\mathbf{W}_s, b_s$  分别为 Softmax 分类器的参数矩阵和偏置项。

## 4 实验分析

### 4.1 数据集制作与预处理

为验证基于 DGRU 和注意力机制的方面级别情感分类模型 ATT-DGRU 的有效性, 在来自网络的酒店评论数据集上进行实验。数据集包含 3425 条中文酒店评论, 经过预处理后, 仅保留评论字数在 25~130 之间的评论, 共 3072 条。数据集主要包含网络用户对酒店卫生、服务、位置等方面的评价。

参考 SemEval2014 测评任务中的数据集, 本文定义样本

由文本、方面以及情感这 3 个要素组成,这里的情感是指文本关于给定方面的情感。本文对 3072 条评论进行人工标注,其中,标注结果包含积极、消极、中立 3 个极性。标注过程中产生两个样本集,其中显性样本集用于 ACSA 分类任务,隐性样本集用于 ATSA 任务。

评论文本显性样本示例 1:{文本:设施空调不好用,电视没设置好信号,淋浴水不方便,服务效率低;方面:设施;情感:消极};显性样本示例 2:{文本:设施空调不好用,电视没设置好信号,淋浴水不方便,服务效率低;方面:服务;情感:消极};隐性样本示例 3:{文本:桌子什么的都没有擦,有很明显的污渍,浴巾也是破的。楼下有个全家还算方便,离火车站挺近;方面:卫生;情感:消极}。

一条文本评论通常包含多个方面的情感,通过人工标注方面情感之后,样本数量最终有 7664 条,样本类型具体分布以及训练集、测试集划分如表 1 所列。

表 1 数据集的统计信息  
Table 1 Statistics of data set

	数据集			总计
	正样本	负样本	中性样本	
显性训练集	2702	589	317	3608
显性测试集	500	400	101	1001
隐性训练集	1678	414	230	2322
隐性测试集	300	300	133	733
总计	5180	1703	781	7664

隐性样本数据集、显性样本数据集各方面样本数量的分布分别如图 5、图 6 所示。

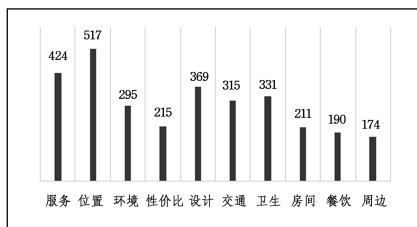


图 5 隐性样本集方面项分布

Fig. 5 Aspect-term distribution of hidden sample set

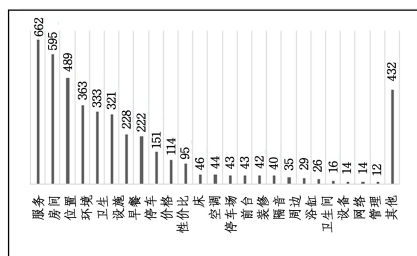


图 6 显性样本集方面项分布

Fig. 6 Aspect-term distribution of dominant sample set

隐性样本集中方面词一共有 10 个,为给定方面集合,符合 ACSA 分类任务标准。

显性样本集中方面词一共有 265 个,均为提取自评论文本的单词。显性样本集中出现次数少于 10 的其他项方面词有 240 个,累计 432 次。81.0% 的显性样本中方面词仅占总方面词数量的 4.1%。

综上,隐性数据集方面词分布更均匀,方面词的主题性更强。

## 4.2 任务描述及参数设定

本文所构建的分类器模型旨在解决方面级别情感分类任务问题。给定一条文本和方面项,任务是得到该文本关于给定方面的情感极性。本文所构建模型的目的是通过文本和方面两项输入得到情感极性。例如输入“价格太贵,服务也跟不上”与“价格”,得到“消极”这一结果。

本文采用斯坦福大学公开发布的 Glove 词向量作为词典,选用单词向量维度为 100 的 100d-6B 数据包。Glove 词典中向量间的表达差异可由欧氏距离来判定,且两个意思相近的词向量表达也相似,可在一定程度上减少人工标注过程中不同标注人员总结方面词时产生的差异。

ATT-DGRU 模型的整个超参数设置如表 2 所列。实验过程中  $k$  的试验范围为 {4, 6, 8, 10, 12},  $k$  取不同值时在显性数据集与隐性数据集上的准确率如图 7 所示。由图可知,当 DGRU 窗口长度为 8 时,即生成文本语义表达矩阵  $H$  时,每个词仅与邻近的 8 个单词相关,模型效果最好。在数据集上进行测试,模型的其他超参数已调整到最佳值。

表 2 参数设置

Table 2 Setting of parameters

参数名称	参数符号	参数值
词向量维度	$d$	100
各隐层维度	$d$	100
文本最大长度	$N$	130
DGRU 窗口长度	$k$	8
学习率	$Learning\ Rate$	0.001
L2 范数	$l2\_reg$	0.001
Dropout 概率	$keep\_prob1$	0.8

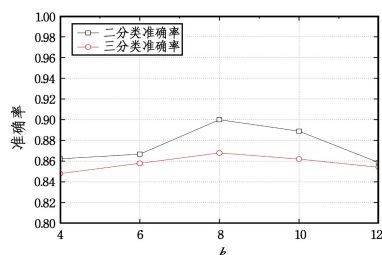


图 7 DGRU 窗口长度  $k$  对准确率的影响

Fig. 7 Influence of window length  $k$  on the accuracy of DGRU

模型训练过程中本文使用交叉熵损失函数以及 Adam 优化器。Adam 算法利用梯度的一阶矩估计和二阶矩估计来动态调整每个参数的学习率。交叉熵能够衡量同一个随机变量中的两个不同概率分布的差异程度,这在模型训练过程中被表示为真实概率分布与预测概率分布之间的差异。

## 4.3 对比模型

为评估 ATT-DGRU 模型的性能,本文在酒店评论数据集上与 LSTM, ATAE-LSTM<sup>[20]</sup>, GRU-MLP, DGRU<sup>[6]</sup> 进行对比,在 SemEval2014-Restaurant (aspect-term) 数据集上与 LSTM, ATAE-LSTM<sup>[20]</sup>, IAN<sup>[23]</sup>, AE-DLSTM<sup>[18]</sup> 进行对比。

(1) LSTM。LSTM 擅长捕捉长距离信息依赖,被广泛用于生成句子上下文语义表达。但标准的 LSTM 捕捉不到句子中与给定方面有关的特征,因此即使要获取不同方面的情感,也只会得到一个关于句子的唯一情感极性。

(2) ATAE-LSTM。文献[20]首先将方面词向量拼接

句子每个单词的向量上,然后再利用注意力机制计算每个单词的权重,从而生成加权句子表示。该模型引入的注意力机制可获取每个单词的情感权重。

(3)IAN。文献[23]所构建的交互注意力网络模型 IAN 用到了两个 LSTM 结构,一个用于获取方面目标词的注意力权重,另一个用于获取上下文注意力权重,对两者串联进行分析后得到分类结果。

(4)AE-DLSTM。文献[18]构建了一个基于双重 LSTM 的位置感知注意力模型 AE-DLSTM,该模型的双重 LSTM 结构可以捕获方面词正反两个方向的语义信息,并通过捕获位置信息为上下文词生成不同的权重,得到分类结果。

(5)GRU-MLP。GRU 为 LSTM 的变体,GRU 处理文本时生成的句子上下文语义表达效果与 LSTM 类似,对计算量进行了优化。GRU-MLP 在标准 GRU 的基础上还使用了 MaxPolling-MLP 来捕捉与方面相关的情感关键信息。

(6)DGRU。文献[6]试图让 RNN 也能像 CNN 一样捕捉到单词的位置和位置不变性特征。使用非连续的  $k$  个 GRU 循环单元对邻近的  $k$  个单词进行编码,相比普通的 GRU 变体,DGRU 更擅长捕捉局部关键信息。因此,DGRU 较 GRU 和 LSTM 更适合处理方面情感任务模型中的文本语义信息。

#### 4.4 算法准确率比较分析

表 3 列出了 ATT-DGRU 模型与相关对比模型在酒店文本隐性数据集、显性数据集上方面级别情感分类的准确率结果。对于 ATAE-LSTM 模型,本文使用原作者发布的代码进行了实验重现。

表 3 不同模型在酒店数据集上的分类结果

Table 3 Accuracy of different models on hotel datasets  
(单位:%)

模型名称	隐性数据集		显性数据集	
	三分类	二分类	三分类	二分类
LSTM	80.70	87.71	79.53	86.46
ATAE-LSTM	85.62	89.95	83.43	<b>89.21</b>
GRU-MLP	83.20	89.53	82.81	88.73
DGRU	84.21	88.59	81.47	87.25
ATT-DGRU	<b>86.61</b>	<b>91.53</b>	<b>85.83</b>	88.89

为测试所提模型对中性情感的分类性能,本文设计了三分类实验组。将数据集中情感极性为中性的样本剔除,将得到的二分类数据集用于训练,将二分类分类器与三分类分类器结果进行比对。

由表 3 可知,每一行表示相应模型在不同实验条件下得到的预测准确率,每一列得分最高的一项用粗体标出。总体来说,本文提出的 ATT-DGRU 模型优于其他相关模型。

通过分析实验数据还可以得到以下认知。本文进行实验的模型绝大部分在隐性数据集上的分类准确率要高于在显性数据集上的分类准确率。因此,当数据集所涉及方面词多为总结性的方面词时,可以进一步提高分类的准确率。具有总结性是指隐性数据集中“点心”“甜品”会被标记为“餐饮”这一类具有概括性的方面,显性数据集中更多的是出现次数少、表达方面较冷门的词语(如“阁楼”“遮光”“镜子”),因此显性数据集的词在训练过程中得不到充分的学习和理解。这一发

现可以指导我们在方面分析情感任务的方面词提取这一子任务中对方面词进行主题提取归纳。

对比都使用了注意力机制但语义编码模块不同的 ATAE-LSTM 与 ATT-DGRU 模型可以发现,DGRU 比 LSTM 更擅长捕捉方面级别情感,这证实了 DGRU 在捕捉长距离依赖信息过程中保留位置不变性信息的这一特征,该特征在方面情感分类任务中表现出了有效性。

对比本文所构建的由简到繁的 GRU-MLP, DGRU, ATT-DGRU 这 3 个模型分类效果可知,DGRU 相比 GRU 更能在方面情感分类任务中捕捉到有效的语义信息,提升分类效果;ATT-DGRU 在 DGRU 的基础上引入注意力机制,弥补了 DGRU 只对邻近  $k$  个单词进行编码的缺陷,进一步捕捉到了复杂语句中相对距离较远的情感极性词与方面词之间的潜在关联。实验结果表明,DGRU 编码器合成的文本语义表达结合给定方面情感注意力权重系数  $a$ ,所合成最终的文本情感表达  $h^*$  所包含的信息相比无注意力机制解码的文本语义表达更有利于提高模型的理解能力。

为进一步证实本文所构建模型的有效性,我们将所构建的 ATT-DRNN 在数据集 SemEval2014 上的 Restaurants 进行了 ATSA 实验,其三分类准确率为 77.21%,二分类准确率为 90.06%。此外,我们还将其与相关工作中在此数据集上进行过 ATSA 实验的 LSTM<sup>[15]</sup>,ATAE-LSTM<sup>[20]</sup>,IAN<sup>[23]</sup>和 AE-DLSTM<sup>[18]</sup>进行比较,准确率对比如表 4 所列。对比模型准确率数据参考文献[18]。本文所构建的模型的准确率高于大部分对比模型。其中,准确率最高的 AE-DLSTM 模型使用了两个结合注意力机制的 LSTM 子网络进行并行计算,再进行归并计算得出最后分类结果。但 AE-DLSTM 计算过程所需的运行内存大约是本文所构建模型 ATT-DGRU 的两倍,若考虑该成本因素,则 ATT-DGRU 具有一定的优势。

表 4 不同模型在 SemEval2014-Res(AT)数据集上的分类结果

Table 4 Accuracy of different models on SemEval2014-Res(AT) dataset  
(单位:%)

模型名称	三分类	二分类
LSTM	74.30	89.25
ATAE-LSTM	75.18	89.43
IAN	76.87	89.85
AE-DLSTM	79.57	91.69
ATT-DGRU	77.21	90.06

#### 4.5 模型有效性分析

$P$ ,  $R$ ,  $F$  分别代表精确率、召回率和  $F_1$  值。精确率是被正确判别的样本数与被判别的样本总数之比。召回率是针对原样本而言的,它表示被正确判别的样本数与应当被判别的样本数之比。然而这两个指标往往互相矛盾,为此采用  $F$ -测量值作为综合评估标准。具体计算公式如下:

$$P = TP / (TP + FP) \quad (14)$$

$$R = TP / (TP + FN) \quad (15)$$

$$F = \frac{2 * P * R}{P + R} \quad (16)$$

其中, $TP$  为被检索到的正样本实际也是正样本的数量; $FP$  为被检索到的正样本但实际上是负样本的数量; $FN$  为未被检

索到的正样本但实际是正样本的数量;  $TN$  为未被检索到的正样本实际是负样本的数量。正样本是被正确判别的样本,负样本是被错误判别的样本。

由表 5 可知,ATT-DGRU 可以有效地理解积极方面情感与消极方面情感,但其中性情感的  $F$  值较两者偏低。可能因为:1)酒店文本数据集中的中性样本较少,训练样本少导致得不到准确的理解;2)中性情感的表达较为复杂且分散在句子的不同位置,容易被擅长捕捉局部信息的 DGRU 忽略。

表 5 ATT-DGRU 在数据集上的各情感极性的精准度

情感极性	指标	隐性数据集		显性数据集	
		三分类	二分类	三分类	二分类
Positive	$P$	94.53	94.18	92.07	94.85
	$R$	82.74	86.45	82.74	87.24
	$F$	88.24	90.14	87.16	90.89
Negative	$P$	80.01	86.16	81.26	83.97
	$R$	85.21	88.01	83.03	84.27
	$F$	82.53	87.52	82.14	84.12
Neutral	$P$	77.32		73.51	
	$R$	50.82	无	49.15	无
	$F$		61.32		58.91

#### 4.6 ATT-DGRU 模型注意力机制的可视化

最后,以两个可视化案例来说明文本提出的注意力机制的有效性,随机抽取两例测试集样本作为测试对象。ATT-DGRU 模型测试文本时,可抽取该文本的情感注意力权重分布  $a$  用于可视化。情感注意力权重分布  $a$  可以直观地观察到每个词与该文本的给定方面情感极性的关联程度的高低。每一个词的权重系数即为该词对给定方面情感极性的影响概率值,所有权重值相加为 1。

本文所抽取的样本为{文本:酒店卫生不太好,不主动提供洗漱用品,环境一般;方面:环境;情感:中立}、{文本:服务态度很好,办理入住及退房也快,交通方便;方面:交通;情感:积极}。图 8 给出了注意力权重单词上的分布情况。其中,颜色的深度代表权重系数的大小,颜色越深代表权重系数越大。在示例样本一中,权重最大的为“一般”,其他与方面词相关程度低的单词注意力分布相对较低且均匀。在示例样本二中,除方面词“交通”外,“方便”权重最高,“快”虽然可以形容“交通”,但因为在此语境中其修饰的是“退房”,所以“方便”的权重更高,其他非相关单词权重均很低。以上可视化分析结果表明,ATT-DGRU 模型所构造的注意力机制能准确地关注与方面情感相关的关键词汇。

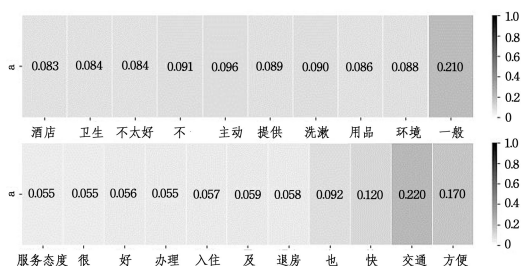


图 8 注意力分布的可视化结果

Fig. 8 Visualization results of attention distribution

情感分类网络模型 ATT-DGRU,该模型的构建通过使用编码-解码框架以及注意力机制有效提高了对复杂语句的方面级别情感分类准确性。ATT-DGRU 首先在 DGRU 编码层对输入的单词文本与方面词进行编码,得到句子级别文本语义表达,然后,为进一步精确得到方面词有关的情感信息,使用注意力机制对文本语义表达提取情感权重系数,最后结合文本语义表达与情感权重表达得到给定方面情感表达。ATT-DGRU 在酒店评论数据集上进行实验,实验结果表明,其不论是在二分类还是在三分类任务上都有较相关工作更高的准确率。

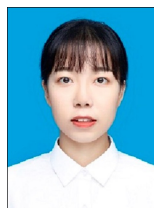
值得重视的是,这是 DGRU 首次被应用于方面级别分类任务中,结合注意力模型的 DGRU 具有针对与方面项有关联的局部词汇的优秀的捕捉能力,这一优势有利于满足情感来源的可视化需求,这也是本文所构建 ATT-DGRU 的独特优势。

#### 参考文献

- [1] PANG B, LEE L, VAITHYANATHAN S, et al. Thumbs up? Sentiment Classification using Machine Learning Techniques [C]// The ACL-02 Conference on Empirical Methods in Natural Language Processing. 2002:79-86.
- [2] ZHANG L, LIU B. Sentiment Analysis and Opinion Mining [M]. Morgan: Claypool Publishers, 2017: 1-50.
- [3] PANG B, LEE L. Opinion Mining and Sentiment Analysis [J]. Foundations and Trends in Information Retrieval, 2008, 2(1): 1-135.
- [4] PONTIKI M, GALANIS D, PAVLOPOULOS J, et al. SemEval-2014 Task 4: Aspect Based Sentiment Analysis [C]// International Conference on Computational Linguistics. 2014: 27-35.
- [5] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning [M]. Massachusetts: The MIT Press, 2016: 24-31.
- [6] WANG B. Disconnected Recurrent Neural Networks for Text Categorization [C]// the 56th Annual Meeting of the Association for Computational Linguistics. 2018: 2311-2320.
- [7] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation [J]. arXiv:1406.1078, 2014.
- [8] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to Forget: Continual Prediction with LSTM [J]. Neural Computation, 2000, 12(10): 2451-2471.
- [9] ZHANG Z, ZOU Y, GAN C. Textual sentiment analysis via three different attention convolutional neural networks and cross-modality consistent regression [J]. Neurocomputing, 2018, 275: 1407-1415.
- [10] AKHTAR M S, GUPTA D, EKBAL A, et al. Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis [J]. Knowledge Based Systems, 2017, 125(1): 116-135.
- [11] HAN H, ZHANG J, YANG J, et al. Generate domain-specific sentiment lexicon for review sentiment analysis [J]. Multimedia Tools & Applications, 2018, 77(16): 21265-21280.

结束语 本文构建了一个基于截断 GRU 的注意力方面

- [12] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural Language Processing(Almost) from Scratch[J]. *Journal of Machine Learning Research*, 2011, 12(1):2493-2537.
- [13] KALCHBRENNER N, GREFFENSTETTE E, BLUNSOM P, et al. A Convolutional Neural Network for Modelling Sentences [C]//*Meeting of the Association for Computational Linguistics*. 2014;655-665.
- [14] WANG L, CAO Z, DE MELO G, et al. Relation Classification via Multi-Level Attention CNNs[C]//*the 54th Annual Meeting of the Association for Computational Linguistics*. 2016; 1298-1307.
- [15] TANG D, QIN B, LIU T, et al. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification [C] // *The 2015 Conference on Empirical Methods in Natural Language Processing*. 2015;1422-1432.
- [16] TANG D, QIN B, FENG X, et al. Effective LSTMs for Target-Dependent Sentiment Classification[J]. *arXiv: 1512. 01100v2*, 2016.
- [17] ZHANG M, ZHANG Y, VO D T, et al. Gated neural networks for targeted sentiment analysis [C] // *Thirtieth AAAI Conference on Artificial Intelligence*, 2016;3087-3093.
- [18] SHUANG K, REN X, YANG Q, et al. AELA-DLSTMs: Attention-enabled and location-aware double LSTMs for aspect-level sentiment classification[J]. *Neurocomputing*, 2019, 334:25-34.
- [19] MA D, LI S, ZHANG X, et al. Interactive attention networks for aspect-level sentiment classification [J]. *arXiv: 1709. 00893*, 2017.
- [20] XUE W, LI T. Aspect based sentiment analysis with gated convolutional networks[J]. *arXiv: 1805. 07043*, 2018.
- [21] SHI Y, YAO K, TIAN L, et al. Deep LSTM based Feature Mapping for Query Classification[C]// *The 2016 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies*. 2016;1501-1511.
- [22] YANG M, TU W, WANG J, et al. Attention Based LSTM for Target Dependent Sentiment Classification [C] // *Thirty-First AAAI Conference on Artificial Intelligence*. 2017;5013-5014.
- [23] CHEN P, SUN Z, BING L, et al. Recurrent attention network on memory for aspect sentiment analysis[C]// *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2017;452-461.



**YIN Jiu**, born in 1996, postgraduate. Her main research interest include machine learning and so on.



**CHI Kai-kai**, born in 1980, Ph. D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include wireless networks and machine learning.