

# 基于增强多注意力的 LSTM 网络的方面级情感分类

符玉杰<sup>1</sup>, 李旻<sup>1</sup>, 王素格<sup>1,2</sup>, 廖健<sup>1</sup>, 吴卓嘉<sup>1</sup>

(1. 山西大学 计算机与信息技术学院, 山西 太原 030006;

2. 山西大学 计算智能与中文信息处理教育部重点实验室, 山西 太原 030006)

**摘要:** 为了挖掘用户细粒度的情感表达, 方面级情感分析引起了研究者的关注。现有的研究方法通常将方面判别任务与情感分析任务分离, 且在给定方面的基础上进行情感分析。然而现实的评论数据大多没有标注方面标签, 因此, 对于给定的评论文本, 需要同时判断方面类别及其情感极性。本文利用多个分类器同时对评论的方面及其情感进行预测, 使用两层长短期记忆网络对评论的上下文信息和方面信息进行编码, 并利用注意力机制对评论中相关方面的情感信息进行注意, 再对多个注意力模块进行叠加增强建模得到文本的表示, 在此基础上, 预测句子中的方面及其对应的情感。本文在 SemEval 和 BDCI2018 汽车评论的主题及情感识别数据上分别进行了实验, 结果显示本文的方法对方面级的情感分析具有较好的提升效果。

**关键词:** 方面级情感分析; 长短期记忆网络; 多注意力机制

## Multi-attention LSTM Networks Enhanced for Aspect-level Sentiment Classification

Fu Yujie<sup>1</sup>, Li Yang<sup>1</sup>, Wang Suge<sup>1,2</sup>, Liao Jian<sup>1</sup>, Wu Zhuojia<sup>1</sup>

(1. School of Computer & Information Technology, Shanxi University, Taiyuan, Shanxi 030006, China;

2. Key Laboratory of Computation Intelligence and Chinese Information Processing, Shanxi University, Taiyuan, Shanxi 030006, China)

**Abstract:** In order to mine user fine-grained sentiment expression, aspect-level sentiment analysis has attracted the attention of researchers. The existing research methods mostly use pipeline working mode to separate the task of aspect discrimination from the task of sentiment analysis, and conduct sentiment analysis on the basis of given aspect. In addition, most of the review data are mostly unlabeled for aspect. Therefore, for a given review text, it is necessary to judge the aspect and its sentiment polarity at the same time. This paper uses multiple classifier ideas to simultaneously predict the categories of the aspects and sentiment of the review. Two layers of Long Short-Term Memory networks are used to encode the context information and aspect information of the review, the multi-attention mechanism is used to consider the sentiment information of relevant aspects in the review, and multiple attention modules are superposed to obtain the final text representation, which is finally used to predict the aspect and sentiment. Experimental results indicate that the proposed method obtained the better performance on SemEval and BDCI2018 car reviews for aspect-level sentiment analysis.

**Keywords:** Aspect-level sentiment analysis; Long Short-Term Memory Network; Multi-attention mechanism

### 1 引言

方面级情感分析是细粒度情感分析重要的子任务<sup>[1-2]</sup>。对于一个评论和预先给定的方面, 方面级情感分析是对评论句中涉及到给定方面的情感极性进行判别。相比于预测整体的情感极性, 方面级的情感分析可以帮助用户更好的了解产品的细节。对于一条用户评论, 我们希望针对评论所涉及产品的不同方面进行情感分析。

对于方面级情感分析的目的主要是捕捉用户所关注细粒度的信息。传统的文本分类方法对一篇文档或一个评论进行情感分析时, 只能得到篇章的整体单一的情感倾向, 并不能获取代表评论中涉及的具体方面的情感。然而, 一个评论整体的情感并不能代表这个实体所有方

面的情感。通常，用户对于评论的关注会集中在某几个方面，因此，在方面级情感分析中，有两个基本任务：方面类别的情感分析和方面项的情感分析<sup>[3]</sup>。方面类别代表在某个抽象范围或概念层面的描述。方面项则是作为一个特殊的词或短语出现在文本中。相对于方面项的情感分析，方面类别较为抽象和概念化，也具有普遍性。本文主要关注方面类别的情感分析。例如，一部分用户在购买汽车时可能选择“价格”合适和“配置”好的汽车进行选购，另一部分用户偏向于“舒适性”高和“操纵性”好的汽车。因而不同用户对于汽车评价的“好”的标准是不同的。

例 1：森林人的外形，内饰，品牌肯定和奔驰没得比。不过四驱及通过性能应该会比 GLC 强一些。还是想清楚自己需要什么吧。

该例 1 中的方面有：内饰，动力，操控，其情感极性分别为：负，正，正。

现有的研究方法中，大多是在方面标签或者方面项的信息已知的基础上进行情感分析<sup>[11-20]</sup>，然而，真实的评论数据通常缺乏方面的标注，若对其标注需要耗费大量的人力资源。因此，本文针对评论的概念层方面类别判别及情感分析的任务，提出一个基于长短期记忆网的多注意力模型，首先利用 LSTM 捕捉上下文信息，对不同方面类别信息分别进行融合，最后使用注意力机制对各自的方面类别信息进行注意，通过加权得到文本的综合表示。在此基础上，设计了多个注意力模块用于对不同方面类别的注意力信息的加强。通过在 BDCI 和 SemEval 数据集上的实验，验证了本文方法的有效性。

## 2 相关工作

在文本情感分析任务中，方面级情感分类通常被视为一种分类问题<sup>[4]</sup>。早期的方法大多利用语言解析器、情感词汇等外部资源设计有效的特征，然后利用传统的统计学习算法进行预测。文献[5-8]基于统计的方法，主要利用人工设计一些特征，例如，特定情感的词嵌入、情感词汇、n-gram 和依存信息等。这些方法往往需要繁重的特征工程和大量的语言资源，在建模时未能利用方面与上下文之间较为重要的语义关联信息。

伴随着深度学习模型的发展，越来越多的神经模型被提出并用于自动学习句子级的方面表示。文献[11-20]使用长短期记忆网络<sup>[9]</sup>(Long Short-Term Memory, LSTM)从词嵌入向量中提取情感信息，再利用注意力机制<sup>[10]</sup>对特定方面信息进行关注，得到相应的文本表示，实现方面的情感分析。Tang 等人<sup>[11]</sup>提出了 TD-LSTM 模型，基本思路是使用两个 LSTM 模型对方面项之前和之后的上下文分别建模，将方面项作为 LSTM 的最后一个输入单元，并充分利用方面项的语义信息，从而更好地判断其情感倾向。TD-LSTM 虽然考虑了方面项的作用，但是没有充分利用方面项和它相应的上下文之间的关联信息。Wang 等人<sup>[12]</sup>提出了基于注意力机制的长短期记忆网络的 ATAE-LSTM 模型，当涉及不同的方面时，这个模型能够注意到句子的不同部分，该模型在输入层和隐藏层分别加入方面信息，通过注意力机制对句子中重要程度的单词给予重要的权重，并加权求和得到句子表示，从而实现对情感极性的较好的分析。Cheng 等人<sup>[13]</sup>提出了一个 HEAT 模型，包括方面注意力和情感注意力。方面注意力提取与方面相关的信息，引导情感注意力更好地定位到文本与方面相关的情感词。另外，引入伯努利注意力机制，使其层次模型较好的处理多个方面词的实例。由于单注意力机制存在对信息的捕捉不够全面的问题，研究者们开始使用多层注意力机制抽取较为抽象的特征。Tang 等人<sup>[14]</sup>使用深度记忆网络在方面级上进行情感分类，模型由多层参数共享计算层构成。每一层都是基于上下文和位置注意力的模型，可以使每个上下文单词学习到对应的权重，并且利用这些信息进行文本表示。Chen 等人<sup>[15]</sup>提出了 RAM 模型，将双向 LSTM 产生的时间步骤状态作为记忆单元，并按距离方面项的相对位置加权，使同一句子中的不同方面项有自身特有的记忆。模型对位置加权的记忆单元进行了多重注意，将注意力结果使用 GRU 编码，编码结果用来预测输出。Zhu 等人<sup>[16]</sup>提出了一种具有辅助记忆的深层记忆网络。该模型构建了

两个记忆模块，其中，情绪记忆模块被用来捕获重要的上下文词进行情绪分类，方面记忆模块来隐式地相互转换方面和方面项，并将它们提供给情绪记忆模块。两者相互作用来同时学习方面和情感特征。

在方面项的情感分析任务中，很多研究者认为，距离方面项越近的词应赋予较高的权重。因此，很多研究者将位置信息引入方面级情感分析。Gu 等人<sup>[17]</sup>提出了 PBAN 模型，作者认为当一个方面项出现在一个句子中，与其相邻的词比其他长距离的词更应引起注意。该模型不仅关注方面项的位置信息，而且通过双向注意力机制相互建模方面项与上下文之间的关系。He 等人<sup>[18]</sup>提出一个基于句法的注意力模型，该模型通过句法依存关系对句子进行分析，考虑句中每个词与方面项的句法依存距离，同时方面项被映射到一个方面矩阵重新表示，以此来参与注意力权重的计算。

相较于传统模型中将方面项和上下文分开建模，Ma 等人<sup>[19]</sup>利用注意力机制将方面项和上下文联系起来。作者认为在细粒度的情感分类任务中，方面项与上下文应该具有交互性，即上下文应该是特定于方面项的，方面项也应该是特定于上下文的。因此，利用注意力机制实现二者交互统一建模。Huang 等人<sup>[20]</sup>介绍了一个 AOA 神经网络。以一种联合的方式对方面项和句子建模，它通过 AOA 模块对方面项和句子进行交互来共同学习方面和句子的表示，并自动关注句子中的重要部分。实验表明其方法优于以前基于 LSTM 的框架结构。

另外，还有研究者将卷积神经网络(CNN)用于方面级情感分析任务中。Li 等人<sup>[21]</sup>使用 CNN 替代基于 RNN 的注意力机制去提取重要的分类特征。他们设计了一个特征变换组件来将实体信息引入到单词的语义表示当中。这个特征变换过程可能会使上下文信息丢失，所以文中提出了两种“上下文保留”机制，将上下文信息和变换之后的特征结合起来，再通过 CNN 对特征进行表示。Xue 等人<sup>[3]</sup>提出了一个基于 CNN 和门控机制的模型，该模型利用门控机制，根据给定的方面信息有效地控制情感信息的传递，两个卷积窗口分别对方面和情感信息进行抽取，得到的表示用于生成最终的特征向量。

以上的研究方法都是在方面标签或者方面项的信息给定的基础上进行情感分析，同时，面向方面项的情感分析研究较多，而方面类别的情感分析工作较少。本文针对方面类别的情感分析任务，使用基于 LSTM 改进注意力机制，通过增强的多注意力模块对重要的信息进行注意，从而提升方面类情感的性能。

### 3 基于 LSTM 的多注意力的网络模型(LSTM-Matt)

为了有效的对上下文与方面词间的情感关联程度进行建模，我们采用 LSTM+attention 作为模型的基本框架，基于多个分类器的思想，提出一个基于长短期记忆网的多注意力模型(LSTM-Matt)。模型结构如图 1 所示。

该模型利用 LSTM 生成有上下文信息嵌入的表示，同时融合方面信息，经过多个注意力模块分别对句子进行注意，在全连接层进行融合，最后将融合后的特征向量输入 softmax 分类器中，以实现每个方面的情感极性判断。模型整体分为三个部分：

#### (1) 词嵌入

给出一个句子  $S=\{w_1, w_2, \dots, w_n\}$ ，我们将每个词映射到一个连续的向量空间  $e_w \in R^{v \times d}$ ， $v$  为词表大小， $d$  为词嵌入的维度。词向量与方面表示都对应一个词嵌入表示。方面集合表示为  $ASP=\{a_1, a_2, \dots, a_m\}$ ，本文使用随机初始化的方法对方面向量进行初始表示。

#### (2) 第一层 LSTM

LSTM 已经被证明是一种将上下文信息融合到词嵌入表示中的有效方式<sup>[23]</sup>，通过第一层 Bi-LSTM 得到的隐藏层状态作为上下文嵌入的序列表示。我们将前向 LSTM 单元在  $w_i$  上的隐藏状态表示为  $\overrightarrow{LSTM}(w_i)$  和后向 LSTM 的隐藏状态表示  $\overleftarrow{LSTM}(w_i)$  拼接作为单词表示，得

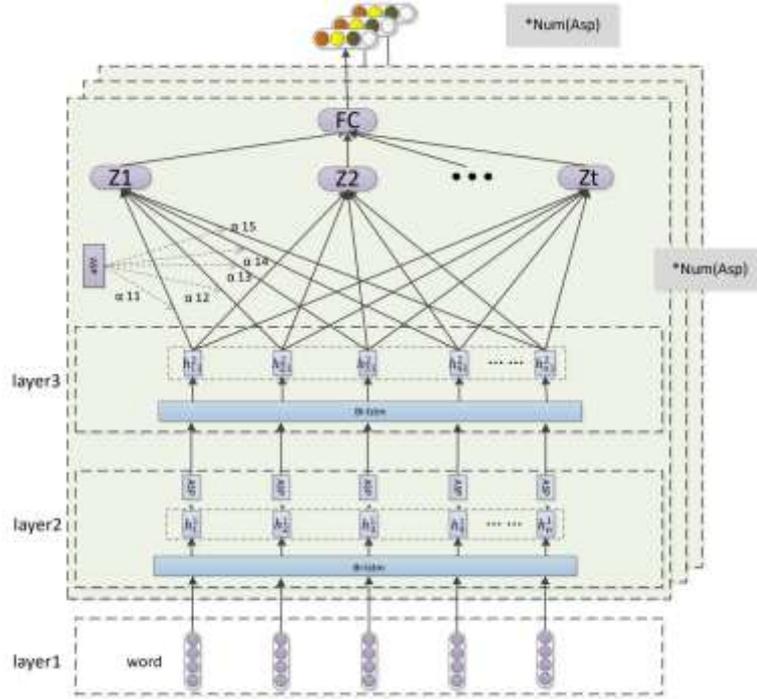


图 1 LSTM-Matt 模型结构

到上下文信息嵌入的单词表示  $h_i^{(1)} \in \mathbb{R}^{2d}$ , 如下公式 (1) 所示。

$$h_i^{(1)} = [LSTM(w_i); LSTM(w_i)] \quad i \in [1, n] \quad (1)$$

### (3) 第二层 LSTM

在方面级情感分析中, 由于每个评论文本都可能涉及多个方面的情感, 而不同的方面, 可能有不同的极性。因此, 在对句子的某一方面的感情极性进行分类时, 方面信息显得尤为重要。为了更好地利用方面信息, 同时考虑多个方面, 得到特定方面的上下文表示, 最终判定该方面的感情类别标签 (将文本未涉及的方面的判定为 Null 标签)。我们将第一层 LSTM 的输出  $h_i^{(1)}$  分别与每个方面嵌入向量  $a_j$  进行拼接。通过对 Bi-LSTM 上下文和方面之间的内部关联进行建模, 形成方面信息与上下文信息融合嵌入的表示  $h_{ij}^{(2)} \in \mathbb{R}^{3d}$ , 如下公式 (2) 所示。

$$h_{ij}^{(2)} = [LSTM([h_i^{(1)}; a_j]); LSTM([h_i^{(1)}; a_j])] \quad i \in [1, n], j \in [1, m] \quad (2)$$

### (4) 注意力层

标准的 Bi-LSTM 模型并不能针对某个方面识别出文本对于此方面感情类别有贡献的部分。本文引入了注意机制用于捕捉句子针对特定方面的关键信息。注意层的输入是第二层 LSTM 的输出  $h_{ij}^{(2)}$ , 注意力机制是根据方面嵌入  $a_j$  和  $h_{ij}^{(2)}$  计算每个状态的权重。 $\alpha_j$  为每个  $h_{ij}^{(2)}$  的权重, 是在推断方面  $a_j$  的感情极性时, 状态  $h_{ij}^{(2)}$  被正确注意到的概率。 $\alpha_j$  通过一个注意力模型来计算得到, 其核心在于通过一个得分函数将隐层状态表示  $h_{ij}$  和注意力向量表示  $a_j$  进行联系, 再利用 softmax 函数将其归一化得到概率分布。通用的注意力模型结构如下公式

(3) - (4) 所示:

$$\alpha_i = \frac{\exp(\text{score}(h_{ij}, a_j))}{\sum_{i=1}^n \exp(\text{score}(h_{ij}, a_j))} \quad (3)$$

$$z_s = \sum_{i=1}^n \alpha_i h_{ij} \quad (4)$$

其中, 得分函数  $\text{score}()$  主流的计算方法<sup>[24]</sup>有以下三种:

$$\text{score}(h_{ij}, a_j) = \begin{cases} h_{ij} a_j & \text{dot} \\ h_{ij} W a_j & \text{general} \\ W[h_{ij}; a_j] & \text{concat} \end{cases} \quad (5)$$

其中,  $W$  为参数矩阵。

此外, 在线性变换中, 元素间的差是拼接的一种特殊情况<sup>[27]</sup>。单词嵌入研究表明, 向量的线性偏移量可以捕获两个单词之间的关系<sup>[22]</sup>。因此, 我们也使用差注意力来捕捉方面与文本的关系。得分函数如公式 (6) 所示。

$$\text{score}(h_{ij}, a_j) = V^T \tanh(W(h_{ij} - a_j) + b) \quad (6)$$

其中  $W$  为参数矩阵,  $V$  为参数向量,  $b$  为偏置

#### (5) 多注意力机制

多注意力机制在阅读理解的任务中显示出较好的效果<sup>[25]</sup>。多注意允许模型在每一次传递过程中关注输入的不同部分<sup>[15]</sup>。Vaswani 等人<sup>[26]</sup>也验证了不同的学习注意力表示  $z_s$  进行多次线性投影对情感分类作用。我们采用多注意力机制使模型关注来自不同表示子空间的信息, 对多个注意力模块并行地执行, 将得到的注意力表示进行拼接并再次投影, 最终得到文本表示如公式 (7) 示:

$$\text{Multi} = [z_1; z_2; \dots; z_m] W^c \quad (7)$$

其中,  $z_t$  为第  $t$  个注意力块得到的文本表示,  $W$  为参数矩阵。

#### (6) 输出层

将得到的文本表示输入全连接层进行分类, 将概率最高的标签设置为最终结果, 最小化交叉熵损失函数对模型进行优化训练, 优化过程如下公式 (8) - (11) 所示:

$$L^* = -\sum_{j=1}^n \sum_{i=1}^x \hat{y}_{a_{ij}} \log(y_{a_{ij}}) \quad (8)$$

$$\beta_{ij} = \exp \frac{\min(\text{num}(a_{ij}))}{\text{num}(a_{ij})} \quad (9)$$

$$\beta_{ij}^* = \frac{\beta_{ij}}{\sum_i \beta_{ij}} \quad (10)$$

$$L = -\sum_{j=1}^n \sum_{i=1}^x \beta_{ij}^* \hat{y}_{a_{ij}} \log(y_{a_{ij}}) \quad (11)$$

公式 (8) 中  $\hat{y}_{a_{ij}}$  为样本  $j$  在方面  $a_i$  的真实标签,  $y_{a_i}$  为预测标签,  $L$  为交叉熵损失。由

于数据中存在不平衡现象，需要对损失进行平衡。公式（9）中， $num(a_{ij})$ 为方面  $a_i$  下第  $j$  种标签的个数。公式（11）为最终的目标函数。

## 4 实验

为了验证算法的性能，我们分别验证多种注意力的差别及多注意力模块的有效性，通过设置对比实验。模型的最佳参数组合均通过实验确定。

### 4.1 数据集

SemEval 2014 task4<sup>1</sup> 评测中提出了 4 个子任务，本文的研究工作针对于 task4 中的 subtask3 和 subtask4，即方面类别判别和方面类别情感分析。选择 Restaurant 2014 数据作为实验英文数据集。文献[3]通过合并 2014-2016 年三年的餐厅评论，获得一个较大的数据集，称为 Restaurant-Large，我们同样在此数据集上对我们的方法进行分析。

数据集的统计信息如表 1 所示：

表 1：情感数据集统计信息

数据集	方面数	情感	最大句长	Positive		Negative		Neutral	
				Train	Test	Train	Test	Train	Test
Restaurant 2014	5	3	69	2179	657	839	222	500	94
Restaurant-Large	8	3	69	2710	1505	1189	680	757	241

中文数据集实验选用 CCF BDCI 2018-汽车行业用户观点主题及情感识别数据集，此数据集为用户在汽车论坛中对汽车相关内容的评价。训练数据集中评论个数为 10655，句子的最大长度为 128 个汉字。评价中方面总共有 10 类，包括：动力、价格、内饰、配置、安全性、外观、操控、油耗、空间、舒适性。其标签类别为-2、-1、0、1，分别表示不涉及此方面、负向、中立、正向。同一条评论中可能会包含多个方面。由于赛道系统关闭，给出的测试集未进行标注，因此，我们使用对训练集进行 5 折交叉验证，实验结果为 5 折的均值。

### 4.2 评价标准

本实验按照“方面+情感分析”识别数量和结果（是否正确）进行判断，需要识别文本中可能包含的多个“方面”。实验结果采用五折交叉验证，精确率 P、召回率 R 以及 F1 值见公式(12)-(14)所示。

$$P = \frac{T_p}{T_p + F_p} \quad (12)$$

$$R = \frac{T_p}{T_p + F_N} \quad (13)$$

$$F1 = \frac{2 * P * R}{P + R} \quad (14)$$

$T_p$ : 判断正确的数量：对“方面+情感”的判断结果完全正确则计入  $T_p$ ;

$F_p$ : 判断错误或多判的数量：识别出的“方面+情感值”数量多于测试样本中实际包含的数量，超出的数量计入  $F_p$ ；如果对“方面”或“情感值”的判断结果错误则计入  $F_p$ ；

$F_N$ : 漏判的数量：识别出的“方面+情感值”数量少于测试样本中实际包含的数量，或未对某个测试样本数据给出结果，缺少的数量计入  $F_N$ ；

### 4.3 实验设计及结果分析

<sup>1</sup><http://alt.qcri.org/semEval2014/task4/>

### 4.3.1 超参数设计

为了实现第3节中模型的测试，模型中的超参数设置见表2所示：

表2：超参数设置

超参数	值
embedding size	300
LSTM hidden dim	300
batch size	32
learn rate	0.001
optimizer	Adam
dropout	0.5
random_state	10

### 4.3.2 基于不同注意力机制的方面判别对比实验

本实验为了比较点积 LSTM-Dot-Att、双线性 LSTM-General-Att、拼接 LSTM-Concat-Att 以及差 LSTM-Minus-Att 四种注意力方式在方面判别性能，比较实验结果见表3所示。

表3：基于不同注意力机制的方面判别比较实验结果

models	P	R	F
<b>LSTM-Dot-Att</b>	88.946	87.178	88.053
<b>LSTM-General-Att</b>	88.241	87.743	87.991
<b>LSTM- Concat -Att</b>	88.612	87.830	88.220
<b>LSTM-Minus-Att</b>	<b>88.714</b>	<b>87.846</b>	<b>88.278</b>

由表3可以看出，差注意机制在捕获方面重要信息上具有较好的性能。

为了展示4种注意力方式在方面判别上注意力权重的不同，我们展示了样例的可视化结果，如图2所示。

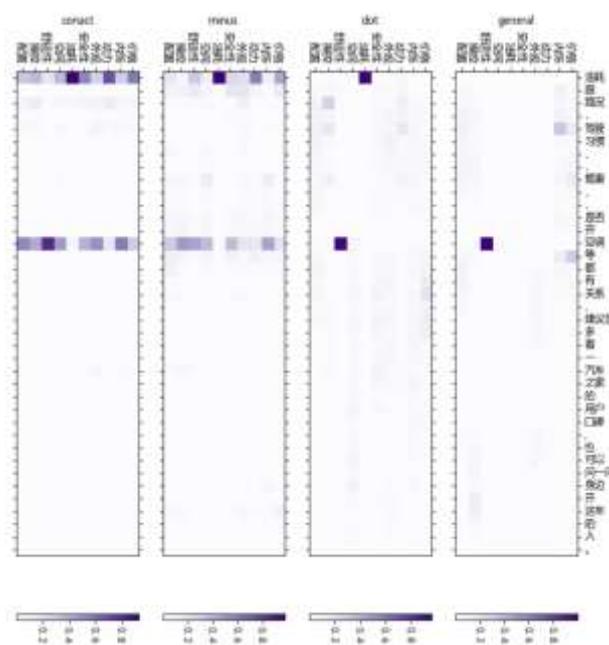


图2 4种不同的注意力方式的“油耗”方面注意力权重热力图示例

图 2 中，样例中涉及到的方面应为“油耗”，但是 Dot、General、Concat 三种注意力都错误的注意到了“空调”，对于舒适性给出了错误的中性标签，LSTM-Dot-Att 和 LSTM-General-Att 对于通用的方面相关的修饰词识别容易出现错误，而 Minus 对于方面的识别较为准确。

#### 4.3.3 基于不同注意力机制的方面判别及情感分析对比实验

为了比较点积 LSTM-Dot-Att、双线性 LSTM-General-Att、拼接 LSTM-Concat-Att 以及差 LSTM-Minus-Att 四种注意力方式在方面判别与情感分析方面的性能，对它们进行了实验，其结果见表 4 所示。

表 4 不同注意力的方面判别及情感分析实验结果

models	P	R	F
<b>LSTM-Dot-Att</b>	63.987	83.847	72.583
<b>LSTM-General-Att</b>	64.381	83.111	72.557
<b>LSTM-Concat-Att</b>	64.401	83.987	72.902
<b>LSTM-Minus-Att</b>	<b>64.849</b>	<b>84.085</b>	<b>73.224</b>

由表 4 可知，对于方面级情感分析，差注意力同样可以取得较好的注意文本中与方面相关的词，从而提升了方面情感的识别效果。我们同样给出了四种注意力方式在方面情感分析上注意力权重的可视化样例，如图 3 所示。

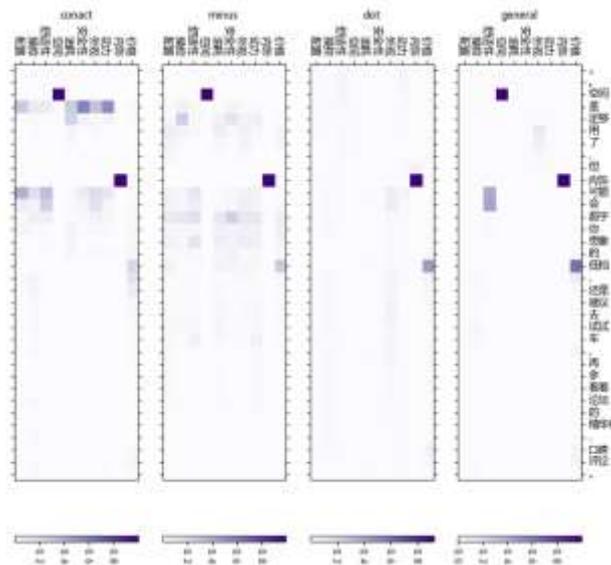


图 3 4 种不同的注意力方式的“内饰”及“空间”方面情感注意力权重热力图示例

如图 3 所示，样例中提及到两个方面“内饰”及“空间”。四种注意力的方法都对方面进行了正确的判别，但是在情感极性的判定中出现了错误。点乘注意力得到的情感标签为 0,0。双线性注意力得到的情感标签为 0,1。拼接注意力得到的情感标签为 0,1。仅仅涉及到方面的预测，并没有对情感进行正确的判断。差注意力得到的标签为-1,1，对方面及情感都进行了正确预测。

#### 4.3.4 多注意力模块效果实验

多注意力模块的数量也是影响模型性能的一个主要因素，我们设置了 1 到 4 个注意模块来测试模型的性能，实验结果如表 5 所示。其中，LSTM-CNN-Att\*n, n=1, 2, 3, 4 分别表示使用 1,2,3,4 个注意力模块。

表 5 多注意力模块实验结果

LSTM-CNN-Att*n 模型	P	R	F
LSTM_Minus_Att*1	64.849	84.085	73.224
LSTM_Minus_Att*2	<b>64.628</b>	<b>85.010</b>	<b>73.431</b>
LSTM_Minus_Att*3	65.446	83.571	73.406
LSTM_Minus_Att*4	64.701	84.704	73.363

由表 5 可知,模型使用 2 或 3 个注意力模块实验的性能最好。使用多个注意力模块会有一些的效果提升,这说明在复杂的情况下,多注意力模块可以捕捉更多情感特征。另一方面,性能并不是随着关注次数的增加而单调增加。因为随着模型 s 复杂度的增加,模型的训练难度增加,泛化能力降低。

#### 4.3.5 与其他模型的比较实验

为了验证我们方法的有效性,我们设计了以下对比实验,分别在 SemEval 和 BDCI 数据集上进行验证。

##### (1) 对比模型介绍

A, B 分别为 BDCI 评测<sup>2</sup>的第一名在测试集 A, B 榜上的测试结果,其使用流水线的工作方式,先判断其方面,再判断方面涉及的情感。由于赛道关闭,现将其成绩作为参考。

TDLSTM 为文献[8]提出的简单长短期记忆网络;AE-LSTM, AT-LSTM 为文献[12]提出的一种模型变体;MemNN 为文献[14]提出的一种多层记忆网络。

ATAE<sup>[12]</sup>: 一个基于 LSTM 的注意力模型,在输入时融合方面编码的信息,在 LSTM 输出的隐藏状态后添加了注意力机制。

HEAT<sup>[13]</sup>: 通过捕捉方面信息来帮助捕捉句子特定方面的情感信息,从而提升方面级情感分析的准确率。

GCAE<sup>[3]</sup>: 门控的卷积神经网络,在卷积过程中加入方面信息,用来针对不同的方面进行情感分类。

LSTM-CNN-Att: 本文模型的一个变体,在第二层 LSTM 使用了 CNN 代替,利用 CNN 可以抽取局部特征,为注意力做准备。卷积窗口设置为 1 时取得最佳的实验结果。

LSTM-CNN-Att\*n: 为 LSTM-CNN-Att 拼接多个注意力模块。

上述方法中,对于相关论文在对应数据集上有实验结果的,我们进行列举,并用“\*”标记。对于 ATAE, HEAT, GCAE 3 个比较先进的模型中不存在实验结果的数据集,我们进行了复现。

##### (2) 给定方面类别的情感分析实验

此实验目标为根据给定的评论和方面类别标签,预测该评论在此方面的情感极性,评价标准为准确率,实验结果如表 6 所示。

表 6 给定方面的情感分析对比实验

models	Restaurant 2014	Restaurant-Large	CCF BDCI 2018
TDLSTM <sup>[8]</sup>	82.6*	-	-
AE-LSTM <sup>[12]</sup>	82.5*	-	-
AT-LSTM <sup>[12]</sup>	83.1*	-	-
MemNN <sup>[14]</sup>	84.28*	-	-
ATAE <sup>[12]</sup>	84.0*	83.91 ± 0.49*	73.52
HEAT <sup>[13]</sup>	<b>85.1*</b>	85.12 ± 0.42	73.56
GCAE <sup>[3]</sup>	84.6	85.92 ± 0.27*	73.52
<b>Our-model</b>	84.89	<b>86.48 ± 0.34</b>	<b>73.81</b>

<sup>2</sup><https://www.datafountain.cn/competitions/310/ranking?sch=1356>

由表 6 的实验结果可以看出, Restaurant 2014 的数据集较小, 本文的方法优于大多已有模型, 略低于 HEAT<sup>[13]</sup>方法。在 Restaurant-Large 数据集上, 本文的方法有一定的提升。在 CCF BDCI 2018 数据集上, 我们对 3 个参考文献的方法进行了复现实验, 结果表明我们的模型也有不错的表现。

### (3) 方面判别和情感分析实验

此实验的目标为给定一个评论文本, 我们对其所属方面以及方面所对应的情感极性进行预测, 评价标准为 4.2 节定义的 F1 值。已有的方法是我们复现的结果。对比实验结果见表 7。

表 7 在三种数据集上的方面判别和情感分析比较实验

models	Restaurant 2014	Restaurant-Large	CCF BDCI 2018
<b>A</b>	-	-	70.491
<b>B</b>	-	-	70.849
ATAE <sup>[12]</sup>	80.41 ± 0.46	76.06 ± 0.40	71.975
HEAT <sup>[13]</sup>	<b>81.08 ± 0.62</b>	78.03 ± 0.23	72.323
GCAE <sup>[3]</sup>	79.97 ± 0.48	78.02 ± 0.26	71.366
LSTM-CNN-Att	80.32 ± 0.40	77.57 ± 0.18	72.708
LSTM-CNN-Att*2	80.43 ± 0.19	77.72 ± 0.18	73.275
LSTM-CNN-Att*3	80.33 ± 0.37	77.75 ± 0.16	73.200
<b>Our-model</b>	80.72 ± 0.19	<b>78.38 ± 0.17</b>	<b>73.431</b>

由表 7 实验结果得知, 在 Restaurant 2014 数据集上, 由于数据集较小, 模型训练不充分, 多种模型表现的不太稳定, 导致多种模型在此数据集上的结果较为接近。在 Restaurant-Large 与 CCF BDCI 2018 数据集上, 随着数据集的增大, 我们的模型越来越稳定, 性能逐渐变好。

LSTM+attention 的结构在数据集上取得了较好的效果, 而 CNN 的结构相比之下表现略低。

LSTM-CNN-Att 上层 CNN 使用的是词窗口卷积, 相当于对固定窗口内的局部信息做了一次矩阵变换, 但 CNN 无法对较长的上下文信息进行建模, 当评论中涉及不同方面的不同情感倾向时, 由于无法对于窗口外的信息进行有效的捕捉, 导致情感判断错误。

LSTM-CNN-Att\*n 的 3 个模型结果表明, 使用了多个注意力块后, 模型可以通过多注意力机制可对不同子空间的特征进行捕捉, 从而加强对方面及情感判别的精度, 较 LSTM-CNN-Att 有了明显的提升, 再次证明了多注意块的有效性。

## 5 结束语

本文通过构建一个基于多注意机制的 LSTM-Matt 模型, 对方面级情感分析问题进行了研究。该文提出的模型同时考虑上下文信息和方面信息, 并结合差注意力与多个注意力机制, 在 SemEval2014 与 CCF BDCI 2018-汽车行业用户观点主题及情感识别数据集上的方面情感分析任务中, 验证了本文方法的有效性。在未来的工作里, 基于上述实验中表现出的不足, 我们将考虑对方面项的情感分析问题进行研究, 引入方面的位置信息。进一步改进多种注意力的融合方式, 另外, 对于上下文与方面的交互机制研究也是我们下一步的重点工作。

## 参考文献

- [1] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis[J]. Foundations and Trends in Information Retrieval, 1-2(2), 2008: 1–135.
- [2] Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis[J]. Mining Text Data, 2012: 415-463.

- [3] Wei Xue and Tao Li. Aspect based sentiment analysis with gated convolutional networks[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 2514-2523.
- [4] Ruidan He, Wee Sun Lee, Hwee Tou Ng, et al. Effective attention modeling for aspect-level sentiment classification[C]//Proceedings of the 27th International Conference on Computational Linguistics, 2018: 1121–1131.
- [5] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 2011: 151–160.
- [6] Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews[C]//Proceedings of International Workshop on Semantic Evaluation, 2014: 437–442.
- [7] Joachim Wagner, Piyush Arora, Santiago Cortes, et al. DCU: Aspect-based polarity classification for SemEval task 4[C]//Proceedings of International Workshop on Semantic Evaluation, 2014: 223–229.
- [8] Duy-Tin Vo and Yue Zhang. Target-dependent twitter sentiment classification with rich automatic features[C]// Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015: 1347– 1353.
- [9] Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735–1780.
- [10] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate[C]//Proceedings of the International Conference on Learning Representations. 2015: CoRR abs–1409.0473.
- [11] Duyu Tang, Bing Qin, et al. Effective lstms for target-dependent sentiment Classification[C]//Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers, 2016: 3298–3307.
- [12] Yequan Wang and Minlie Huang, et al. Attention-based lstm for aspect-level sentiment classification[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: 606–615.
- [13] Jiajun Cheng, Shenglin Zhao, Jiani Zhang, et al. Aspect-level sentiment classification with HEAT (hierarchical attention) network[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017: 97-106.
- [14] Duyu Tang, Bing Qin, Ting Liu. Aspect level sentiment classification with deep memory network[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: 214–224.
- [15] Peng Chen, Zhongqian Sun, Lidong Bing, 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.
- [16] Peisong Zhu, Tiejun Qian. Enhanced aspect level sentiment classification with auxiliary memory[C]// Proceedings of the 27th International Conference on Computational Linguistics, 2018: 1077–1087.
- [17] Shuqin Gu, Lipeng Zhang, Yuexian Hou. A position-aware bidirectional attention network for aspect-level sentiment analysis[C]//Proceedings of the 27th International Conference on Computational Linguistics, 2018: 774–784.
- [18] Ruidan He, Wee Sun Lee, Hwee Tou Ng, et al. Effective attention modeling for aspect-level sentiment classification[C]//Proceedings of the 27th International Conference on Computational Linguistics, 2018: 1121–1131.
- [19] Dehong Ma, Sujian Li, Xiaodong Zhang, et al. Interactive attention networks for aspect-level sentiment classification[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017. pages 4068-4074.
- [20] Binxuan Huang, Yanglan Ou, Kathleen M. Carley. Aspect level sentiment classification with

- attention-over-attention neural networks[C]//Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, 2018: 197-206.
- [21] Xin Li, Lidong Bing, Wai Lam, et al. Transformation networks for target-oriented sentiment classification[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 946-956.
- [22] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems, 2013: 3111-3119.
- [23] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. Recurrent convolutional neural networks for text classification[C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015: 2267–2273.
- [24] Minh-Thang Luong, Hieu Pham, Christopher D. Manning. Effective approaches to attention-based neural machine translation[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 1412–1421.
- [25] Chuanqi Tan, Furu Wei, Wenhui Wang, et al. Multiway attention networks for modeling sentence pairs[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 2018: 4411-4417.
- [26] Vaswani, Ashish and Shazeer, Noam and Parmar, et al. Attention is all you need[C]//Proceedings of the 31th International Conference on Neural Information Processing Systems, 2017: 5998-6008.
- [27] Lili Mou, Rui Men, Ge Li, et al. Natural language inference by tree-based convolution and heuristic matching[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 130–136.