

## 用HEAT进行方面层次的情感分类 (层级注意) 网络

嘉俊成\*, 盛林赵†, 张嘉倪†, 欧文王†, 张馨予\*, 辉宏\*

\* 国防科技大学信息系统与管理学院

长沙, 湖南, 中国

{jiajun.cheng, 惠王} @ nudt.edu.cn, ijunzhang @ hotmail.com

† 香港中文大学计算机科学与工程系

沙田, 新界, 香港

{slzhao, jnzhang, 王} @ cse.cuhk.edu.hk

### 抽象

方面层次的情感分类是一种细致的情感分析任务, 旨在预测文本在不同方面的情感。这个任务的一个关键点是为给定的方面分配适当的情感词汇。最近的工作利用注意力神经网络来分配情感词汇, 并达到最新的表现。然而, 以往的作品只涉及情感信息, 忽略了文本中与情景相关的信息, 这可能会导致情感词与情感词之间在语义上不相关的情况下出现不匹配的情况。为了解决这个问题, 我们提出了一个分层的注意力 (HEAT) 网络方面级情绪分类。HEAT网络包含一个层次关注模块, 包含注意力和注意力。方面注意提取方面相关信息以引导情感注意力更好地分配方面特定的情感文字的话。此外, HEAT网络支持通过引入伯努利注意机制来提取方面术语以及方面级别的情感分类。为了验证所提出的方法, 我们在SemEval的餐厅和笔记本电脑审查数据集上进行了句子层面和审查层面的实验。实验结果表明, 我们的模型更好地分配适当的情感表达式在给定的方面受益于方面的指导。而且, 我们的方法在方面级别的情感分类方面比现有的模型有更好的表现。

### CCS概念

- 信息系统→情感分析; 社交标签;
- 计算方法→神经网络;

### 关键词

情感分类; 方面; 分层注意网络

允许将个人或课堂使用的全部或部分作品的数字或印刷版本免费授予, 前提是复制品不是为了获利或商业利益而制作或发行的, 并且副本在第一页上包含本通知和全部引用。由ACM以外的其他人拥有的组件的版权必须受到尊重。信用抽象是允许的。要复制或重新发布, 在服务器上发布或重新分配到列表, 需要事先的特定许可和/或费用。请求 permissions@acm.org 的权限。

CIKM'17, 2017年11月6 - 10日, 新加坡, 新加坡

© 2017计算机协会。ACM

ISBN

978-1-4503-4918-5/17/11. . .

\$ 15.00

<https://doi.org/10.1145/3132847.3133037>

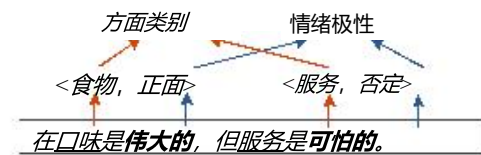


图1: 方面级情感分类的一个例子。下划线是方面术语, 粗体字是情感单词。

### 1 引言

用户生成评论的情感分类是了解个人对产品或服务的重要技术[17]。随着亚马逊, Yelp等在线评论网站和Twitter, 微博等在线社交媒体网站的普及, 评论数量急剧增加, 成为以众包方式调查产品或服务的最重要资源。例如, 通过收集回顾iPhone 7s的推特和分析情绪, 我们可以推断产品是否值得购买, 并知道产品是否会成功并推动苹果的股票。特别是, 产品评论通常包含用户对产品不同方面的评论。因此, 挖掘的评论情绪在方面提供了机会来了解客户的详细反馈。例如, 我们可以从评论中了解到关于餐馆的很多反馈-食物是怎样的? 服务好? 价格可以接受吗?

面向级的情感分类旨在预测对应于某个特定方面的文本的情感极性。图1显示了一个例子, “口味很好, 但服务是恐怖的”。鉴于方面食物的情绪极性是积极的, 而给予方面服务的情感极性是负面的。通过提取用户对特定方面的反馈, 方面级的情感分类可以帮助企业发现产品的具体缺陷, 并进一步改进设计, 并向不同偏好的用户汇报详细信息以供参考。

为给定方面分配适当的情感词是方面层面情感分类的关键。对于图1中的例子, 情感特征应该是“好”的方面食物,

然而, 情绪特征应该是“可怕的”给定方面的服务。但是, 很难对传统情感分类方法[2, 13, 40], 从不同的方面相同的文字提取不同的特征。近来, 关注基于神经网络已经被提出来发现一个给定方面的相关情感词汇, 从而成功地为一个给定的方面分配适当的情感词汇, 并达到方面层面的情感分类的最先进的性能[32]。

但是, 以前的工作只涉及情感信息, 而忽略了文本中与情景相关的信息, 这可能会导致情感词语和方向之间的不匹配, 当一个不相关的情感词在语义上对于给定的方面是有意义的。对于图1中的例子, “好”和“可怕”都是普遍的情感词汇, 可以用

- 所提议的模型可以提取方面术语以及具有伯努利 (S形) 注意力的方面级别情感分类。与标准注意模型相比, 具有伯努利注意力的新模型可以很好地控制多方面词汇和隐含方面表达的实例。
- 实验结果表明, 与最先进的基线相比, 我们的模型可以提高方面级别情感分类的性能。此外, 案例研究表明, 我们的模型可以很好地捕捉文本的方面信息和情感信息。

### 2 相关工作

于方面食品和方面服务。鉴于方面的食物，模型可能兼顾“好”和“可怕”，使人们对情绪的预测感到困惑。解决这个问题的一个简单的方法就是利用方面术语弥补差距。鉴于方面的食物，找到方面术语“口味”要比区分哪个情感词对应于方面更容易得多。在“趣味”这个方面的指导下，我们可以很容易地选择情感词“伟大”，并在这方面决定情感的极性。

受上述直觉的启发，我们提出了一个层次关注（HEAT）网络，通过提取方面表达式来改善方面层次的情感分类。HEAT网络通过分层关注模块共同学习从文本中模拟方面信息和方面特定的信息。层次关注模块首先在给定方面的指导下对正式信息进行处理，然后在给定层面的方向和提取的文本方面信息下注意情感信息。采取图1中的句子作为一个例子，给定方面的食物，等级关注模块首先在食物的指导下注意“口味”一词，然后用方便食物的信息和“口味”一词找到“伟大”的单词。在这样的过程中，层次关注机制在方面术语的引导下更好地定位了文本的特定方面的情感表达，从而提高了方面层面情感分类的性能。具体而言，我们引入位置遮罩层来表示方面项和情感表达式的位置信息，从而提高了HEAT网络中情感表达式的注意力计算。此外，HEAT网络支持将方面术语与方面级别一起提取引入了伯努利的注意力机制[11]，最后，我们评估我们在句子层面的四个基准数据集和在审查层面的三个数据集的方法。实验结果表明，HEAT模型比方法级别情感分类的最新方法表现更好。

总之，我们的工作主要贡献如下：

- 我们提出了HEAT网络来进行方面层面的情感分类。该模型捕捉文本的方面信息，以帮助捕获句子的特定方面的情感信息，从而提高方面情感分类的准确性。

98

在本节中，我们首先简要介绍一下方面级别的分类。然后，我们展示了关注网络的相关研究，这是我们方法的关键技术。

## 2.1 方面层面的情感分类

方面层面的情感分类是一个细粒度的情感分类任务，需要同时考虑文本的方面信息和情感信息。研究方面级情感分类可分为三个不同的线路。

第一个分别提取一个句子的方面和情感，然后把它们联系起来。一个句子的方面是提取语言模式[25,36]，有监督的序列标签[3,38]或分类算法[35]，一个句子的情感通常被归类为一般的情感分类方法，如基于规则的方法[2]，基于特征的分类器[13,34]或神经网络[27,29,40]。然而，这些方法只给句子分配了一种情感极性，因此对于句子在不同方面表达不同意见的情况，不能产生正确的结果，如图1中的例子。

二是依赖于目标的情感分类，旨在根据句子中提到的给定目标词来推断句子的情感极性。依赖目标的情感分类通常被定义为文本分类问题，通过添加一些目标特定的特征[9,12]或设计一些特定的神经网络结构[21,30]来考虑目标词。然而，依赖于目标的情感分类不能处理隐含的方面表达，而且由于不将目标词分为方面类别，所以也是有限的。

三是近期趋势层面的情感分类，利用神经网络预测句子的情感极性。特别是，王等人。[32]提出了一种基于注意机制的LSTM来预测一个句子的情感极性对于给定的纵横类别和达到状态-the-艺术表现。面向方面的情感分类的注意模型，既包括隐含的，也包括显性的方面表达，并将情感自动分成方面类别，克服了上述两条线的缺点。然而，[32]中的模型直接关注了这个问题具有关注层的特定于方面的情感信息，当不相关的情感词在语义上对于目标方面有意义时，这可能导致情感词和方面的不匹配。

会议1C: 情绪

CIKM'17, 2017年11月6 - 10日, 新加坡

因此，我们提出了一个层次关注网络来解决这个问题，通过提取方面术语和使用方面术语来帮助捕获方面特定的情感信息。

## 2.2 注意网络

注意网络是一个精心设计的神经网络，可以在查询的指导下有选择地关注特定的输入或特征。注意网络已被用于各种自然语言处理（NLP）任务，如机器翻译[1,18,19]，问题回答[6,37,39]，文本总结[20,26]和情感分析[30-32]。特别是，具有多层关注的记忆网络主宰了许多NLP任务，如问答[14,28]。另外，Lin等人[16]采用自注意模型的句子和句子找到的重要组成部分。Kim等人[11]提出了一个结构化的关注网络来选择文本的潜在结构信息，并在机器翻译，问答和语言推理方面取得了良好的效果。大多数关注网络是以端到端的方式进行训练的[1,28,41]，其中输入的注意力是在没有监督的情况下在最终任务的指导下推断的。然而，有监督的注意力可以达到更好的表现，并学习更清晰的线索[19,33]。因此，是否对注意力进行监督是一个权衡这取决于任务和数据集。本文试图利用注意机制对方面信息和情感信息进行建模，从而提高方面层面情感分类的性能。对于学习注意力权重的方式，我们只对情景注意力进行监督，并从最后的情感分类任务中推断情感注意力。

## 3问题陈述

在方面层面的情感分类文献中，这个层面是一个抽象的，层次的概念，具有几个不同的定义[17,22,24,30]。在本文中，我们遵循[17]给出的最流行的定义，这也在[24,32,35]中使用。下面我们给出方面和方面术语的定义来帮助理解HEAT模型。

**定义3.1（方面）** 产品的一个方面（也称为方面类别）是产品的类似部分或属性的类别。

以笔记本电脑评论为例，屏幕和电池是笔记本电脑的一部分。价格和可用性是笔记本电脑的属性。他们都被视为笔记本电脑的方面。

**定义3.2（方面术语）**。一个方面术语，也被称为显式方面表达式，是一个词或短语，出现在文本中明确指出一个方面类别。

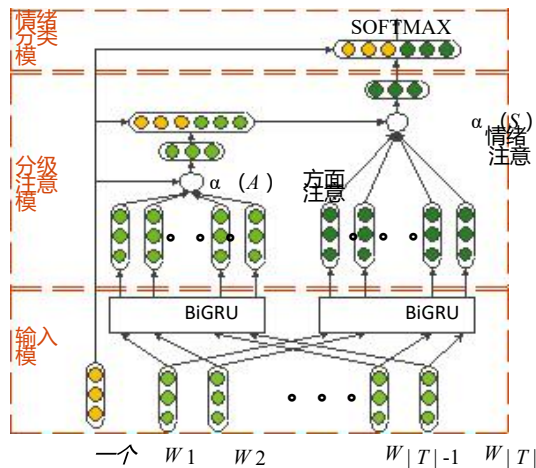


图2：用于方面级别情感分类的HEAT网络的图示。

方面级情感分类的目标是预测对应于某个特定方面的句子或段落的情感极性。通过以上定义，方面层面的情感分类可以自动地将意见归纳为类别，并且包含显式和隐式方面表达式。

## 4 型号

本节介绍我们的HEAT网络架构。首先，我们概述了HEAT网络中的模块。然后，我们显示每个模块的细节，并介绍了培训目标函数。最后，我们展示了如何用伯努利注意机制来提取方面术语。在下面的描述中，我们用粗体大写字母表示带有粗体小写字母和矩阵的向量。

### 4.1 HEAT网络架构

图2演示了HEAT网络架构，它包含三个模块：输入模块，分层关注模块和情感分类模块。

输入模块。输入模块将文本和目标方面编码为分布向量表示,并提取文本中每个单词的上下文特征。

分层注意模块。分层关注模块分为两个关注层,即关注方面和情感关注,抓取方面信息和方面特定情感信息。方面的注意力旨在在目标方面的指导下关注方面信息,即方面术语。情感关注旨在在目标方面和提取的方面信息的指导下捕捉文本的情感特征。

情绪分类模块。情感分类模块以情感特征和给定方面预测文本在目标方面的情感极性。



更接近方面通常比另一方更重要[30]。因此，我们使用位置遮罩层来表示方面项和情感表达式的位置信息。由于我们没有方面项的确切位置，我们使用方面的注意力权重向量 $\alpha^{(a)}$ 来计算位置掩码。具体来说，我们定义一个位置矩阵 $M \in \mathbb{R}^{|T| \times |T|}$ 代表文本中每个单词的接近度，

$$m_{ij} = 1 - \frac{|i - j|}{|T|}, \quad (12)$$

where  $|T|$  是文本的长度， $i, j \in \{1, 2, \dots, |T|\}$ 。然后根据方面的关注权重计算出位置蒙版 $m^{(l)}$ ：

$$m^{(l)} = M \alpha^{(a)}. \quad (13)$$

现在接近方面项的单词在 $m^{(l)}$ 中将具有较大的值，因为方面项在 $\alpha^{(a)}$ 中将具有较大的值。最后，情绪关注分数计算如下，

$$i_i = \frac{1}{|T|} \exp \left( \frac{v(S)}{v(S)} \right) \quad (15)$$

目标方面文本的情感特征是文本中所有词语的上下文特征与权重向量 $\alpha^{(s)}$ 的加权和，

$$\frac{1}{|T|} \sum_{j=1}^{|T|} \alpha_j^{(s)} v_j$$

4.4情感分类模块

情感分类模块在S2EMD模块之后预测目标方面的文本情感 $\hat{y}$ ，其中 $\alpha^{(a)}$ 和 $\alpha^{(s)}$ 是权重不同的情感极性为不同的方面，我们级联方面向量 $w$ 和一个情感特征 $v(S)$ 作为神经输入到预测文本的情感极性上的目标方面，

$$i_j = \frac{1}{|T|} \exp(x_j)$$

4.5 目标函数

为了确保情景注意和情感注意分别提取情景信息和情感信息，我们在情景注意和最终情感分类上都使用了监督。对于方面的关注，训练集提供了目标方面每个文本的方面术语。我们使用一个方面术语的面具矢量 $m \in \mathbb{R}^{|T|}$ 以标记方面项，其中，当词 $w_i$ 是目标方面的方面项时， $m_i = 1$ ，否则 $m_i = 0$ 。那么方面的关注的损失是

$$L_{\text{aspect}} = - \frac{1}{|T|} \sum_{i=1}^{|T|} m_i \log \alpha^{(a)}(w_i) \quad (18)$$

$$L_{\text{aspect}} = - \frac{1}{|T|} \sum_{i=1}^{|T|} m_i \log \alpha^{(a)}(w_i)$$

对于情感分类， $y$ 是标记的情感分布， $\hat{y}$ 是预测的情感分布。我们使用 $y$ 和 $\hat{y}$ 之间的交叉熵损失作为分类的损失，

$$L_{\text{sent}} = - \sum_{k=1}^d \left( y_k \log(\hat{y}_k) + (1 - y_k) \log(1 - \hat{y}_k) \right) \quad (19)$$

最后，目标函数是

$$L = \lambda L_{\text{aspect}} + (1 - \lambda) L_{\text{sent}} + \lambda \| \theta \|^2, \quad (20)$$

其中 $\lambda \in [0, 1]$ 是用于平衡方面关注损耗和情感分类损失超参数， $\lambda$ 是L2-正则化项，并且 $\theta$ 是所述参数集。

4.6 Bernoulli注意的方面术语提取

方面条款是方面级情感分析的重要信息。例如，在餐厅评论中，食物是一个方面。“味道”，“汉堡包”和“披萨”是食物的方面术语。距离

从整个情感到各个方面，企业和顾客都需要更多的细节，比如哪种食物（比如披萨或鸡肉）是好的。因此，提取方面术语和方面级别的情感分类是有意义的。然而，标准的注意机制不能表示明确的方面术语，特别是当文本具有多个方面术语或一个方面没有方面术语时，因为这些单词的权重是

相对。

在HEAT网络中，方面关注模块可以强调每个单词在表示目标方面的重要性。 $\alpha^{(a)}$ 中权重较高的词作为方面词的相对概率，这促使我们将情景词与情感分类任务一起提取出来。此外，我们使用Bernoulli (sigmoid) 注意力[11]来计算 $\alpha^{(a)}$ 来提取方面项而不是softmax注意力（例如方程(8)），

$$\alpha^{(a)}(w_i) = \frac{\exp(x_i)}{\sum_{j=1}^{|T|} \exp(x_j)} \quad (21)$$

在这样的过程中，注意力方面成为一个序列标注模型，其中 $\alpha^{(a)}$ 中 $w_i$ 被视为词的瓦特我是一个方面的术语。我们只是假设权重高于0.5的词作为方面术语。使用伯努利注意机制来计算权重， $\alpha^{(a)}$ 可以为文本中的所有条款为零。根据等式(13)，位置掩码 $m^{(l)}$ 也将为零，这导致等式(14)中的情绪注意的失效。为了避免这个问题，我们在位置掩码中添加一个小数 $\epsilon$ ，

$$\alpha^{(s)}(w_i) = \frac{\exp \left( \frac{m^{(l)}(i) + \epsilon}{|T|} \right)}{\sum_{j=1}^{|T|} \exp \left( \frac{m^{(l)}(j) + \epsilon}{|T|} \right)}$$

引入伯努利注意机制后，注意力损失函数为 $\alpha^{(a)}$ 与 $m$ 之间的交叉熵损失，

表1: 具有评论数量 (R) , 句子 (S) 和“观点”, 情感元组 (T) 的数据集。

数据集	训练			测试			总		
	#R	#S	#T	#R	#S	#T	#R	#S	#T
REST14	-	3401	3,713	-	800	1025	-	4201	4,738
REST15	254	1,315	1476	96	685	730	350	2000	2206
REST16	350	2000	2,107	90	676	703	440	2676	2810

值得注意的是，我们使用不同的损失函数来标准方面的关注和伯努利方面的关注。对于标准方面的关注，损失函数（方程 (18)）只涉及标注方面项的权重。因为权重的总和是1，所以方面项的权重越高意味着其他词的权重越低。然而，对于伯努利方面的注意，词的权重没有直接关系，所以我们必须评估所有词的损失（方程 (23)）。处理隐式方面表达式时，不同的损失函数具有另一个优点。当一个文本隐含地向某个方面表达情感时，没有方面术语。情感关注层可能成为情绪关注层的噪声。然而，在训练过程中，标准方面注意力的损失将归零。然后，层次关注模型变成了具有两个关注层的分类模型。对于伯努利方面的注意，注意力损失将所有方面的注意力权重归结为零。情感关注层仅使用方面嵌入和单词的上下文信息。综上所述，两种不同关注类型所设计的损失函数能很好地评估方面关注的损失，避免了模型在处理隐式方面表达时的不足。

5实验

5.1 数据集

我们用评论数据集评估我们的模型在句子级别和评论级别。具体而言，审查级别情感分类旨在预测某一方面的元评论的情绪。句子层次情感分类的目的在于通过结束语在特定方面预测从评论中划定的每个句子的情感。数据集取自SemEval 2014任务4 [24]，SemEval挑战赛2015任务12 [23]和SemEval挑战赛2016任务5 [22]。每个数据集的详细信息显示在表1中。“REST”意思是餐馆域，“LAPT”意思是笔记本电脑域。REST15和REST16已经标记了“方面，方面术语，情感极性”元组，这对我们的模型是足够的。最初的REST14数据集分别标注每个句子的方面项，情感极性和观点，情感极性。为了将我们的模型与最先进的 [32]进行比较，我们手动将每个方面术语映射到相应的方面类别。原始的LAPT15数据集只标注每个句子的“方面”，“情感极性”元组。为了评估我们的电子产品评论模型，与餐厅评论有很大的不同，我们为每个句子手动标注每个方面的方面术语。保持一致

与[32]一样，我们将产品的属性和组件作为方面，忽略了REST15，REST16和LAPT15中组件的方面。由于REST14不具有句子的审阅信息，我们使用其他三个数据集来评估我们的模型在审查水平。

5.2 实验设置

如4.2节所述，我们训练词汇嵌入在两个领域特定的语料库中，而不是标注的数据集。继[31]，我们使用Yelp挑战数据集1的餐厅域，它包含了不同的餐厅4.1M的评论，我们使用Amazon的电子数据集[7]笔记本电脑领域，其中包含超过150万条评论。对于两个语料库，我们用word2vec工具来训练100维的词嵌入。2

为使实验可重复，我们解释参数设置如下。我们用Tensorflow 3和TensorLayer来实现我们的模型。4除文字嵌入外的所有参数均以均匀分布U (-0.1,0.1) 初始化。方面嵌入的维度和隐藏层的大小为32。不同模型和数据集的方面注意力损失的权重是不同的，并且用5倍交叉验证进行调整。L2正则化权重λ 设定为0.001。训练采用随机梯度下降，动量为0.9，批量为30。我们在训练过程中使用了学习速率衰减和提前停止。具体而言，对于每个数据集，将10%的训练数据作为验证集提出。初始学习率是0.001。当验证集上的分类损失增加时，我们将学习率衰减一半，当学习率低于0.00001时，停止训练。为了缓解不同初始化参数的影响，我们在每个数据集上对不同的初始参数运行每个模型5次，并将平均值作为模型的结果报告给数据集。

5.3 比较的候选模型

我们比较以下型号的性能。AT-LSTM：基于注意的 LSTM (AT-LSTM) [32]使用LSTM 提取每个单词的上下文特征，并在给定方面的指导下参与情感表达。在具体的方面，这取决于方面嵌入矢量和所述上下文特征的情绪信息被提取。

1 [https://www.yelp.com/dataset\\_challenge](https://www.yelp.com/dataset_challenge)  
2 <https://radimrehurek.com/gensim/models/word2vec.html>  
3 <https://www.tensorflow.org>  
4 <http://tensorlayer.org>

表2: 方面级别情感分类的准确性 (句子级别)。

模型	REST14		REST15		REST16		LAPT15	
	宾。苏氨酸。	宾。苏氨酸。	宾。苏氨酸。	宾。苏氨酸。	宾。苏氨酸。	宾。苏氨酸。	宾。苏氨酸。	宾。苏氨酸。
AT-LSTM	89.6	83.1	81.0	77.2	87.6	83.0	86.3	82.1
ATAE-LSTM	89.9	84.0	80.9	77.4	87.2	82.7	85.8	82.3
AT-BiGRU	90.4	84.3	82.8	79.2	90.4	86.7	87.0	84.3
HEAT-GRU	89.6	84.3	81.2	79.1	89.7	85.8	87.4	84.5
HEATB-GRU	89.4	84.0	81.8	79.6	89.2	85.4	87.3	84.2
HEAT-BiGRU	91.3	85.1	83.0	80.1	90.8	87.1	87.9	84.9
HEATB-BiGRU	91.1	84.9	83.4	80.5	91.1	87.5	88.0	85.1

表3: 方面级别情绪分类的准确性 (审查级别)。

模型	REST15		REST16		LAPT15	
	宾。苏氨酸。	宾。苏氨酸。	宾。苏氨酸。	宾。苏氨酸。	宾。苏氨酸。	宾。苏氨酸。
AT-LSTM	78.3	75.0	83.2	82.1	79.6	74.6
ATAE-LSTM	78.8	76.4	83.6	81.8	79.3	74.2
AT-BiGRU	79.7	75.8	83.5	80.5	83.6	80.2
HEAT-GRU	83.9	81.0	86.5	82.3	83.8	80.4
HEATB-GRU	84.4	81.5	86.6	82.9	84.3	80.4
HEAT-BiGRU	84.2	81.7	86.3	83.0	84.9	81.7
HEATB-BiGRU	84.8	82.1	85.6	82.9	85.5	82.4

负面)和“Thr.”表示三级预测 (正面，负面和中性)。最好的分数是粗体。我们获得以下意见。1) 我们提出的HEAT网络比最先进的注意模型获得更好的结果。另外，所有最好的成绩都是由HEAT-BiGRU和HEATB-BiGRU完成的，其中大部分都是由HEATB-BiGRU实现的。2) 所有双向模型比相应的单向模型获得更高的精度。双向模型可以同时掌握每个单词的前后两个语境信息，因此比单向模型更有效。3) 大多数分层关注网络比相应的基线模型获得更好的结果。分层模型可以提取方面的信息，这有助于捕捉句子的特定方面的情感信息。此外，标准方面的注意力和伯努利方面的注意力得到相似的预测精度。

值得注意的是，我们在REST14上实现AT-LSTM和ATAE-LSTM的结果比[32]要低。AT-LSTM和ATAE-LSTM的准确度分别为87.9和87.8，三级分类为80.8和81.0。这可能是由于我

ATAE-LSTM： 基于注意力的方面嵌入 LSTM (ATAE-LSTM) [32]是通过连接方面嵌入到LSTM层中的词嵌入来改进AT-LSTM。在这样的过程中，LSTM层可以利用上下文特征提取方面的信息，从而可以捕获更多的重要信息来响应这个方面。ATAE-LSTM已经在REST14上实现了最先进的性能。

AT-BiGRU： 基于注意力的 BiGRU (AT-BiGRU) 将AT-LSTM中的LSTM层替换为BiGRU，其中包含与我们的模型类似的情感关注模块。我们比较它来验证添加方面的注意模块的改进。

HEAT-BiGRU： HEAT-BiGRU是我们在方面关注模块中提出的标准注意的HEAT网络。该模型捕获文本的方面信息，并使用方面信息来捕获方面特定的情感信息。

HEATB-BiGRU： HEATB-BiGRU是我们在Bernoulli注意模块中提出的HEAT网络。该模型使用伯努利注意力来标注方面关注层中的方面术语。

HEAT-GRU， HEATB-GRU： HEAT-GRU和HEATB-GRU分别是HEAT-BiGRU和HEATB-BiGRU的单向版本，使用GRU提取上下文信息作为分层关注模块的输入。我们与它们比较来验证经常性能网络的选择。

5.4 评估

句子级别。表2显示了句子级别评估的实验结果，其中“Bin”表示二元分类（正，

们的嵌入质量较低造成的。王等人 [32]训练单词嵌入在一个外部语料库上，有四千四百亿条文字，比我们的要大得多。由于我们也对REST14进行了评估，我们在表2中报告了[32]的结果：AT-LSTM和ATAE-LSTM的二进制分类分别为89.6和89.9，三级分别为83.1和84.0 分类。

审查级别。表3给出评价水平评价结果。类似于句子级别的结果，我们的分层关注神经网络比基线模型获得显著更好的性能。但是，对于句子级别的结果还是有一些不同的观察结果 首先，BiGRU没有像句子级别评估那样带来显著的改进。对于REST16数据的二进制分类任务，双向模型甚至比相应的单向模型稍差。这种现象可以解释为：一个方面的审视情绪主要取决于当地的情境特征。然而，审查级文本通常包含几个句子，其中双向模型增加了包含来自与给定方面无关的句子的情绪信息的风险。第二，层次关注的提高比句子层次评价更为重要。评论文本更可能包含这样的情况，即对于给定的方面，几个情感词在语义上是有意义的。这种情况比我们提出的更容易与先前的注意模式不匹配

会议1C: 情绪

CIKM'17, 2017年11月6 - 10日, 新加坡

方面	标准注意				伯努利的关注			
(1) 环境:	0.997				0.998			
(2) 装饰:	需要升级, 但食物是惊人的。				装饰 需要升级, 但食物是惊人的。			
(3) 餐饮:	装饰需要升级, 但是	0.999	餐饮 是惊人的。		装饰需要升级, 但是	0.997	餐饮 是惊人的。	
(4) 一般:	0.107	0.228	0.664		0.999	0.998	0.999	
(1) 环境:	我喜欢 披萨, 特别是	蘑菇	披萨。		我喜欢 披萨, 特别是	蘑菇	披萨。	
(2) 装饰:	0.271	0.115	0.365	0.111				
(3) 餐饮:	一世 将 是 背部对于 当然				我会回来的	当然。		

图3: 学习方面的关注分数的例子。单词顶部的数字是单词的侧面注意分数。低于0.1的分数没有标记。粗体字是标签方面的术语。颜色深度表示单词的重要程度。

分层神经模型。因此，我们提出的层次神经模型比层次分析法在层次评估方面取得了更为显著的改善，这也证明了层次关注机制的有效性。

通过比较复习水平和句子水平的结果，发现REST16和LAPT15在复习水平上的准确率明显低于句子水平。这是因为评论经常提到更多的方面而不是一个句子，在预测目标方面的情感极性时带来更多的噪音信息。然而，REST15的审查水平的准确性高于句子结果。我们检查REST15的测试集，发现它有更多的句子超出了我们模型的控制范围（见5.6节）比其他数据集。这也是为什么REST15的句子级别结果明显比REST14和REST16的结果差。然而，审查级别情感分类可以缓解这个问题，因为审查通常包含更多关于某个方面的信息而不是一个句子。例如，“我们吃完饭之前吃得很好”的句子表达了对方面食物的负面情绪，但是我们的模型没有预测到这一点。这句话的回顾是“一切都很顺利，直到我们吃完饭。我看了一眼鸡，我感到震惊。这是一个非常未煮熟的菠菜和土豆泥的床上，“它包含更多的方面的食物信息，模型可以正确预测情绪，因此，REST15的审查水平分类结果比句级结果，并且与REST16 的评论级分类结果相当。

5.5 可视化的注意

看点关注。图3显示了标准关注和伯努利注意力的表现，以找到四个例子中的方面术语。例（1）和（2）对于不同的方面是相同的句子。这两种注意机制都可以明显地找到方面的术语示例（3）是针对目标方面具有多个方面术语的句子。我们可以发现，两个注意机制的得分明显高于其他词语。然而，由于分数的总和被限制为1，所以标准注意的第一“匹萨”和“蘑菇”的分数是相当小的。例子（4）是没有任何方面术语的句子。所有伯努利注意机制的得分低于0.1，正确表明没有

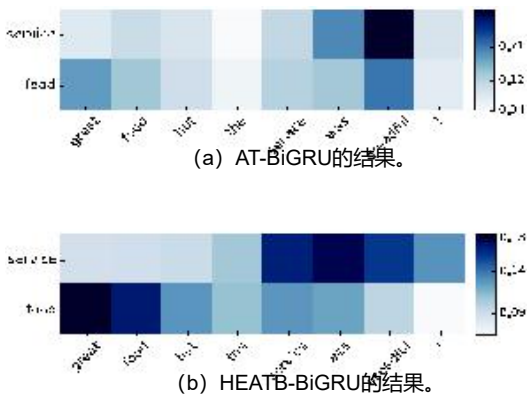


图4: HEATB-BiGRU和AT-BiGRU对两个不同情绪句子的情感关注结果。颜色深度表示情感关注矢量 $\alpha(s)$ 中的权重。左边的字显示目标方面。这两个子图显示了我们的层级关注机制的影响。

方面术语被标记。但是，标准的注意机制被迫留意一些词语。从最后两句话，我们可以发现，如果存在一个以上的方面词或者没有方面词时，标准方面的注意力会被混淆，而伯努利方面的注意力很好的解决了这个问题。

情感关注。我们比较了我们模型的情感关注结果和不同方面的句子的基线，看看等级关注的效果。我们选



择 HEATB-BiGRU 和 AT-BiGRU 作为例子，因为 HEATB-BiGRU在我们提出的模型中大多达到最好，AT-BiGRU是所有基线中最强的竞争者。图4显示了 HEATB-BiGRU 和 AT-BiGRU对于两个方面句子的情感关注结果。比较图4 (a) 和图4 (b)，我们观察到HEAT-BiGRU表现比AT-BiGRU好多在一个句子中包含多个在不止一个方面具有语义意义的情感词。在

(1)	价钱 是 大, 希望 它 没有窗户8。	<价格, 正面>
(2)	价格很好, 希望 它做了 不 有 视窗 8。	<OS, 否定>
(3)	它需要一个 长 时间 至 加载 任何 页。	<性能, 否定>
(4)	互联网是快速和可靠的, 电池 生活持续 一个长 时间。	<电池, 正面>
(5)	有 至 返回 该 电脑。	<一般, 否定>
(6)	一世将 无疑 走 背部 再次。	<一般, 正面>
(7)	怎么样 是 这个 餐厅 仍然 打开?	<一般, 否定>

图5: 情感关注和分类结果的例子 前5分的词是有色的，颜色深度表示情感关注词的重量。

换句话说，AT-BiGRU更容易无关情绪字不匹配比一个方面HEAT-BiGRU当不相关的情绪字对于给定的纵横语义上有意义的。特别地，我们发现AT-BiGRU在图4 (a) 中困惑于定位方面食物的情感词。在图4 (b) 中，HEATB-BiGRU很好地解决了这个问题，“服务是可怕的”这个表达得分高于其他的表达方式服务和表达“大食”实现给定方面的食物。相比前得分AT-BiGRU，我们HEATB-BiGRU参与方面术语“食物”，并通过层次关注机制利用它来正确地找到相应的情感词“great”。

此外，从图4中我们发现，情感关注并不是严格的关注情感词汇。当给予方面服务时，“服务是可怕的！”的得分相当高。我们认为这是因为有些情感词语在不同的语境中表达了不同的情感，即使是在同一方面。例如，“快”一词在“电池快速充电”的句子中表达了积极的情绪，而在“力量快速”的句子中则表达了负面的情绪。因此，关注整个情感表达为情感分类提供了更多的上下文信息。

5.6 质量分析

如上所示，我们的模型获得了方面级别情感分类的最新性能。在这一小节中，我们展示了我们的模型的优点，并通过一些典型的例子来分析错误所在。

实例探究。在图5中，我们列举了来自测试集的情感分类结果的一些例子。示例（1）-（5）是笔记本电脑评论，（6）-（7）是餐厅的评论。在例（1）和（2）中，句子“价格很高，希望没有Windows 8.”提到价格和操作系统（OS）两个方面。结果表明，我们的模型可以区分不同方面的情感极性。在例（3）和（4）中，两个句子具有不同的情感极性的“长时间”相同的情感词语。“长时间”是电池寿命的一个优势，而加载internet网页的缺点。我们的模型在两个句子中都使用了“长时间”这个术语，也准确地区分了不同的情感极性。结果表明，我们的模型可以很好地捕捉上下文信息，区分不同上下文的不同词义。例（5）和（6）是两个表达情感而没有任何感情的句子

字。由于这两个句子比较简单，在评论中经常出现一些类似的句子，我们的模型可以正确地学习正则化，预测它们的情感极性。例（7）是一个具有修辞性问题的句子，由于情绪被问题语调颠倒，难以预测情绪。我们的模型给出了“如何”的最高得分，并正确预测了情感的极性，表明我们的模型具有很好的捕捉文本语义的能力。

错误分析。我们对模型进行错误分析，发现一些类似的情况，我们的模型经常出错。对于情感分类，三种句子经常混淆我们的模型。首先是比较情绪的句子。一个例子是“坦率地说，这里的中国菜是我能在家做得更好的东西”。该模型表示“更好”一词，但不能区分该评价中的目标是否为正面。其次是一些晦涩的情绪表达。一个例子是“节省你的钱和时间，去别的地方”。这句话是一家餐厅的评论，并向餐厅表达负面情绪。类似的情况经常出现在对餐厅方面服务和笔记本电脑方面支持的评论中。这些评论表达了没有情感的意见，而是隐含的表达，如故事，建议或计划。我们的模型可以处理一些简单而频繁出现的模型，例如图（5）和（6）但是，因为它们很少出现，需要推理的含义，我们不能处理一些复杂的。第三个是有条件的句子，例如“如果鼠标垫正常工作，将是一个非常好的笔记本电脑”。句子对方面鼠标表示负面情绪。这种评论经常出现在笔记本评论中以表达用户的意愿。我们的模型可以关注“正常工作”，但不能捕获条件指标“if”，从而得到不正确的预测。对于方面的关注，有两个主要的问题。首先是未知的话。有些文字，特别是餐厅评论中的一些食物的名字，很少使用，并没有出现在前列车语料库。我们没有嵌入它们，也不能在注意层面注意到它们。第二个是“句子中的”中国风味的印度食品“这个复杂的方面表达方式，在句子“不是一个很花哨的地方，但很好的中国风格的印度食品”和“瓶装啤酒的选择”中，“瓶装啤酒的选择并不多”。该模式只能标注“印度食品”和“啤酒”，未能获得资格。

6结论和未来的工作

在本文中，我们提出了用于方面级别情感分类的HEAT网络。该模型的关键思想是通过学习句子的方面信息来帮助捕捉句子

第三十三届国际机器学习会议论文集. 1378-1387.  
[14] 赖思维, 刘康, 何世柱, 赵君. 如何生成一个好词嵌入. IEEE智能系统31, 6 (2016), 5-14.  
[16] 林周汉, 冯敏伟, 西塞罗·诺盖拉·桑托斯, 莫瑜, 冰翔, 鲍文周和Yoshua Bengio. 结构化的自我关注句嵌

的情感,从而提高方面层面情感分类的性能。此外,我们提高了方面关注的注意力机制,以及方面层面的情感分类。情绪级和审查级数据集的实验结果表明,我们的模型在基线上获得了更好的性能。

虽然我们提出的模型改善了方面层面的情感分类,并且比最先进的方法有更好的表现,但是我们的模型却有一些特殊的情况,比如比较情绪(在5.6节中分析过)。第5.6节所示的特殊情况在方面层面的情感分类方面仍然没有解决。将来,我们将尝试设计特定的神经结构来处理这些特殊情况。

## 致谢

本文所描述的工作得到了国家自然科学基金(No. 71331008)和中国香港特别行政区研究资助委员会(普通研究基金中国科大第14208815号)的部分支持,

## 参考

- [1] Dzmitry Bahdanau, Kyunghyun Cho 和 Yoshua Bengio. 神经机器翻译通过联合学习来对齐和翻译。计算机科学(2014年)。
- [2] Danushka Bollegala, 大卫威尔和约翰 Carroll. 2013年。使用情感敏感词库的跨域情感分类。Knowledge and Data Engineering IEEE Transactions on 25.8 (2013) 1719-1731.
- [3] 郑家俊, 张新, 裴利, 张胜, 丁朝云, 王辉. 探索中国公众人物民意调查微博文本的情感分析。Applied Intelligence (2016) 1-14.
- [4] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk 和 Yoshua Bengio. 学习使用 rnn 编码器-解码器进行统计机器翻译的短语表示。arXiv 预印本 arXiv: 1406.1078 (2014)。
- [5] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho 和 Yoshua Bengio. 门控递归神经网络在序列建模上的经验评估。arXiv 预印本 arXiv: 1412.3555 (2014)。
- [6] David Golub 和 Xiaodong He. 2016年。与注意力回答的字符级问题。arXiv 预印本 arXiv: 1604.00727 (2016)。
- [7] 毁了他和朱利安·麦考利。起起伏伏: 通过一流的协同过滤, 对流行趋势的视觉演变进行建模。在第25届国际万维网会议论文集中。国际万维网会议指导委员会, 507-517.
- [8] 塞普·霍赫里特和Jürgen 施米德休伯。长期的短期记忆。神经计算 9.8 (1997) 1735-1780.
- [9] 龙江, 莫宇, 周明, 刘小华, 赵铁军. 目标依赖的 Twitter 情绪分类。在计算语言学协会第49届年会: 人类语言技术第1卷计算语言学协会 151-160.
- [10] 拉法尔 Jozefowicz, 沃伊切赫萨伦巴, 和伊利亚 Sutskever. 经常性网络架构的实证研究。在第32届国际机器学习会议 (ICML) 的论文集中。2342至2350年。
- [11] 尹金, 卡尔丹顿, Luong Hoang 和 Alexander M. Rush. 结构化的关注网络。arXiv 预印本 arXiv: 1702.00887 (2017)。
- [12] 斯韦特兰娜 Kiritchenko, 小旦诺科林樱桃, 和赛义夫穆罕默德。2014年NRC-加拿大-2014: 检测方面和情绪顾客评论。在 SemEval 2014. 437-442.
- [13] Svetlana Kiritchenko, 朱晓丹, Saif M. Mohammad. 情绪分析简短的非正式文本。人工智能研究杂志 50 (2014) 723-762.
- [14] Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus 和 Richard Socher. 动态记忆网络用于自然语言处理。在

- 入。arXiv 预印本 arXiv: 1703.03130 (2017)。
- [13] 刘兵. 情感分析和意见挖掘。人类语言技术综合讲座 5.1 (2012) 1-167.
- [18] 胡志明升陈德良, 总管 Hieu 范, 和 Christopher D. 曼宁. 注意神经机器翻译的有效方法。计算机科学 (2015)。
- [19] 海涛米, 王治国和安倍晋三. 监督神经机器翻译的注意力。arXiv 预印本 arXiv: 1608.00112 (2016)。
- [20] P. Nema, M. Khapra, A. Laha 和 B. Ravindran. 基于查询的抽象概括的多样性驱动型注意模型。ArXiv 电子版打印 arXiv: 1704.08300 (2017年4月)。
- [21] Thien Hai Nguyen 和 Kiyooki Shirai. PhraseRNN: 用于基于方面的情感分析的短语递归神经网络。在 EMNLP. 2509至2514年。
- [22] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, 赵艳艳, 秦兵, Orpheus De Clercq. SemEval-2016 任务5: 基于方面的情感分析。在国际语义评估研讨会上。
- [23] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, 和 Ion Androutsopoulos. 2015年 SemEval - 2015年任务12: 观点基于情感分析。(2015年)。
- [24] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos 和 Suresh Manandhar. 2014年 SemEval 2014 任务4: 观点基于情感分析。(2014), 27-35 国际语义评估研讨会论文集。
- [25] 光秋, 刘兵, 贾家补, 陈春. 意见词扩大和目标提取通过双重繁殖。计算语言学 37.1 (2011) 9-27.
- [26] 亚历山大 M. 拉什, 萨米特普拉, 和 Jason 韦斯顿. 抽象句总结的神经注意模型。计算机科学 (2015)。
- [27] 理查德 Socher, 亚历克斯 Perelygin, 让武, 杰森, 克里斯托弗·d 人为宁, 安德鲁·吴, 和克里斯托弗·波茨。在情绪树库上的递归深度模型用于语义组合性。在 EMNLP, 卷。1642年。引用 1642年。
- [28] Sainbayar 苏赫巴托尔, 亚瑟 Szlam, 杰森韦斯顿, 和 Rob 弗格森。2015年端端的端到端存储网络。计算机科学 (2015)。
- [29] 杜 Tang, 秦兵, 刘挺. 用门控递归神经网络进行情感分类的文档建模。在2015年会议的经验方法在自然语言处理。1422年至1432年。
- [30] 杜 Tang, 秦兵, 刘挺. 深度记忆网络的方面情感分类。arXiv 预印本 arXiv: 1605.08900 (2016)。
- [31] 王文雅, 新加坡潘霖, 达尔·阿尔, 肖小葵. 融合多层次注意力共同提取方面和意见条款。第31届 AAAI 人工智能会议 (AAAI-17)。AAAI 出版社。
- [32] 王治全, 黄敏烈, 朱小燕, 李钊. 2016年基于注意力 LSTM 的观点级情感分类。在自然语言处理实证方法会议上。606-615.
- [33] 杰森韦斯顿萨米特普拉和安托博尔德。内存网络。arXiv 预印本 arXiv: 1410.3916 (2014)。
- [34] 吴海兵, 顾晓东. 2014年, 减少过度加权的监督期限加权的情感分析。在 COLING. 1322年至1330年。
- [35] 吴海兵, 顾一伟, 孙上帝, 顾晓东. 2016年方面为基础的与卷积神经网络 Opin-离子综述。2016年国际神经网络联合会议 (IJCNN)。IEEE. 3157-3163.
- [36] 吴元斌, 张琦, 黄宣景, 吴立德. 2009 意见挖掘的短语依赖关系解析。在2009年会议的经验方法在自然语言处理会议: 第3卷, 第3卷计算语言学协会 1533年至1541年。
- [34] 熊来明, 钟伟雄和理查德·索瑟。用于问答的动态 Coattention 网络。arXiv 预印本 arXiv: 1611.01604 (2016)。
- [38] 杨碧珊和克莱尔·卡迪. (1) 中的细粒度意见提取的联合推理。1640-1649.
- [39] 尹义鹏, 莫雨, 冰翔, 鲍文周, 辛里奇. 用关注卷积神经网络回答简单问题。arXiv 预印本 arXiv: 1606.03391 (2016)。
- [40] 余剑飞, 姜静. 学习句子嵌入与跨域情感分类的辅助任务。计算语言学协会。
- [41] 张加妮, 史行健, 欧文·金, 杨延炎. 用于知识追踪的动态键值记忆网络。在第26届国际万维网会议论文集中。国际万维网会议指导委员会, 765-774.