Abstract

方面层次情感分类的目的是识别特定语境句子中某些方面所表达的情感。本文提出了一种面向方面情感分类的attention-over-attention（AOA）神经网络。我们的方法以一种联合的方式对方面和句子进行建模，并具体捕获方面和上下文句子之间的交互。通过AOA模块，我们的模型可以共同学习方面和语义的表示，并自动关注句子中的重要部分。我们在笔记本电脑和餐厅数据集上的实验表明，我们的方法优于以前基于LSTM的体系结构。

1 Introduction

与文档级情感分类任务[4,15]不同，方面级情感分类是一个更细粒度的分类任务。它旨在识别语境句子中某一特定方面的情感极性（如积极、消极、中性）。例如，在一句“食物很棒，但服务很糟糕”中，“食物”和“服务”两个方面的情绪极性分别是正的和负的。

方面情感分类克服了在一个句子中出现多个方面时文档级情感分类的局限性。在我们前面的例子中，有两个方面，整个句子的总体情绪混合着正负极性。如果我们忽略方面信息，很难确定指定目标的极性。这种误差普遍存在于一般情感分类任务中。在最近的一项研究中，江等人。人工评估了一个Twitter情感分类器，结果显示40%的情感分类错误是因为没有考虑目标[6]。

针对方面级情感分类提出了多种方法。典型的方法是通过监督训练建立机器学习分类器。在这些基于机器学习的方法中，主要有两种不同的类型。一种是基于手工创建的特征构建分类器[6,26]。另一种是基于神经网络，使用端到端的训练，没有任何先验知识[11,25,28]。由于神经网络具有无需特征工程就可以从数据中学习表示的能力，因此在这项研究中得到了广泛的应用。

由于神经网络的优点，我们提出了一种基于长短期记忆（LSTM）神经网络的方面级情感分类问题。以往基于LSTMs的方法主要集中在文本建模上[23,28]，而我们的方法使用LSTMs同时建模方面和文本。此外，LSTMs生成的目标表示和文本表示通过attention-over-attention（AOA）模块相互作用[2]。AOA不仅自动生成方面到文本的相互关注，而且自动生成文本到方面的相互关注。这是受到这样一种观察的启发，即一个句子中只有很少的词对某一方面的情绪有贡献。很多时候，这些带有感情色彩的词语与这些方面高度相关。在前面的例子中，在“开胃菜没问题，但服务很慢”这句话中有“开胃菜”和“服务”两个方面。根据我们的语言经验，我们知道“慢”这个否定词更倾向于描述“服务”，而不是“开胃菜”。同样，对于一个方面短语，我们也需要关注最重要的部分。这就是为什么我们选择AOA来处理方面和句子中最重要的部分。与之前的方法相比，我们的模型在2014年SemEval的笔记本电脑和餐厅数据集上表现更好[17]

2 Related work

Sentiment Classification

情感分类旨在检测文本的情感极性。对于这一研究问题，有多种方法被提出[12]。大多数现有的作品使用机器学习算法来分类文本的监督方式。朴素贝叶斯和支持向量机（SVM）等算法在该问题中得到了广泛的应用[10,15,27]。大多数方法要么依赖于n-gram特征，要么依赖于手工设计的特征。为此目的构建了多个情感词典（14，18，22]。

近年来，神经网络在情感分类方面取得了显著的进展。基于神经网络的方法自动学习特征表示，不需要密集的特征工程。研究人员提出了多种神经网络结构。经典方法包括卷积神经网络[7]、递归神经网络[9,24]、递归神经网络[19,29]。这些方法在情绪分析方面取得了很好的效果。

Aspect Level Sentiment Classification

方面级情感分类是情感分类的一个分支，其目的是识别句子中某一特定体的情感极性。一些早期的工作设计了基于规则的方面级情感分类模型，如[3,13]。Nasukawa等人首先对句子进行依赖性分析，然后使用预定义的规则来确定对方面的看法[13]。Jiang等人基于句子的语法结构，通过创建多个目标相关特征来改进目标相关情感分类[6]。这些与目标相关的特征与其他内容特征一起进一步输入到支持向量机分类器中。

随后，介绍了各种基于神经网络的方法来解决这一方面的情感分类问题。典型的方法是基于LSTM神经网络。TD-LSTM通过开发两个LSTM网络来模拟方面目标的左右上下文来解决这个问题[23]。该方法利用这两个LSTMs的最后隐藏状态来预测情绪。为了更好地抓住句子中的重要部分，王等人使用一个方面词嵌入来生成一个注意力向量，集中在句子的不同部分[28]。按照这些思路，Ma等人使用两个LSTM网络分别对句子和方面进行建模[11]。它们进一步利用句子生成的隐藏状态，通过池操作计算对方面目标的注意，反之亦然。因此，他们的IAN模型可以兼顾句子和目标的重要部分。他们的方法和我们的相似。然而，该方法忽略了句子与目标词对之间的交互作用，实验表明我们的方法优于它们的模型。

3 Method

Problem Definition

在这个方面层次的情感分类问题中，我们给出了一个句子s=[w 1 ,w 2 ,...,w i ,..,w j ,...,w n ]和一个方面目标t = [w i ,w i+1 ,...,w i+m−1 ]。方面目标可以是单个单词或长短语。目的是对句子中的方面目标的情感极性进行分类。

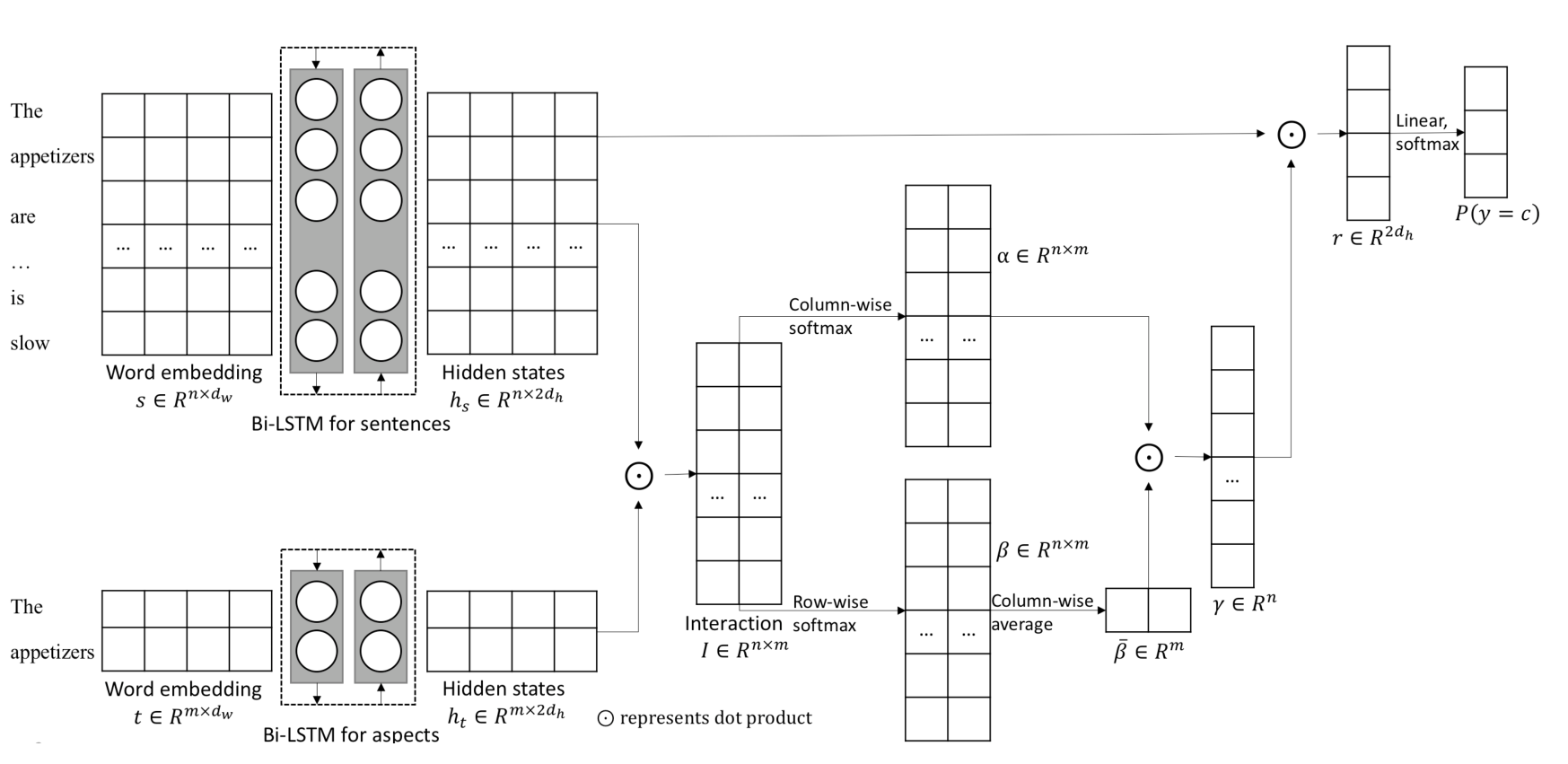


图1。我们的方面层次情感分类模型的总体架构。

我们的神经模型的总体架构如图1所示。它主要由四个部分组成：单词嵌入、双向长短时记忆（Bi LSTM）、Attention-over-Attention模块和最终预测。

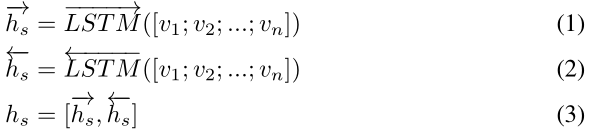
Word Embedding

给定一个长度为n的句子s=[w 1，w 2，…，w i，…，w j，…，wn]和一个长度为m的目标t=[wi，wi+1，…，wi+m-1]，我们首先将每个单词映射到一个低维实值向量，称为单词嵌入[1]。对于每个单词w i，我们可以从M v×d w得到一个向量vi∈rdw，其中v是词汇的大小，dw是嵌入的维数。在嵌入查找操作之后，我们得到了两组词向量[v 1；v 2；…；v n]∈rn×d w和[v i；vi+1；…；vi+m-1]∈rm×dw，分别用于句子和方面短语。

Bi-LSTM

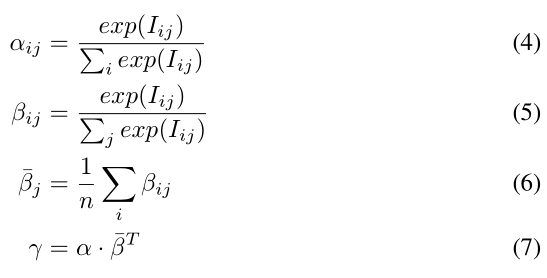
在得到词向量后，将这两组字向量分别送入两个双向LSTM网络。我们使用这两个Bi LSTM网络来学习句子和目标中单词的隐藏语义。每一个双LSTM是通过叠加两个LSTM网络得到的。使用LSTM的优点是它可以避免梯度消失或爆炸问题，并且能够很好地学习长期依赖性。

在输入s=[v 1；v 2；…；v n]和前向LSTM网络的情况下，我们生成了一个隐藏态序列hs1∈rn×dh，其中dh是隐藏态的维数。我们通过将s输入另一个反向LSTM来生成另一个状态序列hs2。在Bi LSTM网络中，通过将hs1与hs2级联产生最终输出隐藏态hs∈R n×2dh。我们用同样的方法计算方面目标t的隐藏语义状态ht。



Attention-over-Attention

给出了文本的隐藏语义表示和Bi LSTMs生成的方面目标，通过AOA模块计算了文本的注意权重。这是受问答中使用AOA的启发[2]。在给定目标表示ht∈R m×2dh和句子表示hs∈R n×2dh的情况下，我们首先计算一个成对的交互矩阵I=hs·ht，其中每个条目的值表示句子和目标之间的词对的相关性。通过列式和行式两种方法，我们得到了目标对句子的注意α和句子对目标的注意β。在对β进行列平均之后，我们得到目标级注意β∈R m，这表示方面目标中的重要部分。最后一个句子级注意γ∈R n由每个目标对句子注意α的加权和计算，由等式（7）给出。通过显式地考虑每个方面词的贡献，我们可以学习句子中每个词的重要权重。



Final Classification

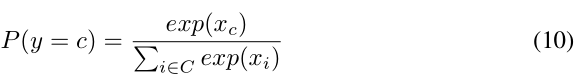
最后一个句子表示是利用AOA模块中的句子注意对句子隐藏语义状态的加权和。



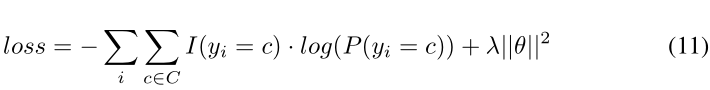
我们把这个句子表示作为最终的分类特征，并将其输入到一个线性层中，将r投射到目标C类的空间中。



其中Wl和bl分别是权重矩阵和偏差。在线性层之后，我们使用softmax层计算情感极性为c∈C的句子s朝向a方面的概率为：



一个方面目标的最终预测情绪极性就是概率最高的标签。我们用L2正则化训练我们的模型以最小化交叉熵损失



其中I（·）是一个指示函数。λ是L2正则化参数，θ是LSTM网络和线性层中的一组权矩阵。我们进一步应用dropout以避免过度拟合，即随机丢弃LSTM单元的部分输入。

我们使用带有Adam[8]更新规则的小批量随机梯度下降来最小化模型中关于权重矩阵和偏差项的损失函数。

4 Experiments

Dataset

我们在SemEval 2014任务4[26]中针对笔记本电脑和餐厅的两个领域特定数据集上进行了实验。有经验的注释者标记了句子的体词及其极性。表1给出了情绪极性类别的分布。

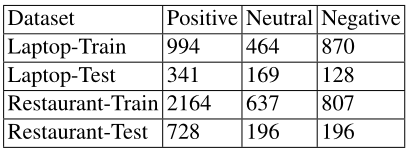


表1 SemEval 2014任务4数据集的情绪极性类别分布。表中的数字表示句子方面对的数目。

Hyperparameters Setting

在实验中，我们首先随机选取20%的训练数据作为验证集来优化超参数。所有权重矩阵从均匀分布U（-10-4，10-4）随机初始化，所有偏差项设置为零。L2正则化系数设置为10-4，dropout保持率设置为0.2[20]。单词嵌入由300维Glove向量初始化[16]，并在训练期间固定。对于词汇量不足的单词，我们从均匀分布U（-0.01,0.01）随机初始化它们。LSTM隐藏状态的维度设置为150。Adam优化器的初始学习率为0.01。如果每三个阶段的训练损失都没有减少，学习率就会降低一半。批量设置为25。

Model Comparisons

我们分别在这两个SemEval数据集上训练和评估我们的模型。我们用准确度来衡量性能。为了进一步验证模型的性能，我们将其与几种基线方法进行了比较。我们列出如下：

Majority是一种基本的基线方法，它将训练集中最大的情感极性分配给测试集中的每个样本。

LSTM使用一个LSTM网络对句子进行建模，最后一个隐藏状态作为句子的表示形式进行最终分类。

TD-LSTM使用两个LSTM网络来模拟围绕方面术语的前后上下文。这两个LSTM网络的最后隐藏状态被连接起来以预测情感极性[23]。

AT-LSTM首先通过LSTM模型对句子进行建模。然后将LSTM中的隐藏状态与方面项嵌入相结合，生成注意向量。最后的句子表示是隐藏状态的加权和[28]。

ATAE-LSTM通过将方面嵌入到每个词向量中来进一步扩展AT-LSTM[28]。

IAN使用两个LSTM网络分别对句子和方面项进行建模。它利用句子中隐藏的状态为目标生成一个注意向量，反之亦然，并基于这两个注意向量输出一个句子表示和一个目标表示进行分类[11]。

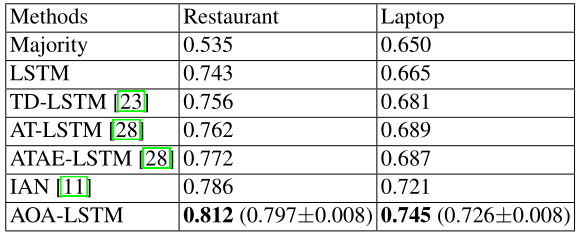


表2 比较结果。对于我们的方法，我们运行10次，并显示“最佳（平均值±标准偏差）”。基线的性能是从他们的原始论文中引用的。

在我们的实现中，我们发现性能随着不同的随机初始化而波动，这是训练神经网络的一个众所周知的问题[21]。因此，我们运行了10次我们的训练算法，并报告了平均精度以及我们在表2中得到的最佳精度。所有的基线方法在他们的论文中只报告了一个最佳数字。平均而言，我们的算法优于这些基线方法，我们的最佳训练模型在很大程度上优于它们。

Case Study

在表3中，我们列出了测试集中的五个例子。为了分析哪一个词对方面情绪极性的贡献最大，我们在表3中可视化了最后一个句子的注意向量γ。颜色的深度表示一个单词在句子中的重要性，颜色越深越重要。在前两个例子中，开胃菜和服务在“开胃菜没问题，但服务很慢”这句话中有两个方面，我们可以观察到，当句子中有两个方面时，我们的模型可以自动指向每个方面的正确情感指示词。同样的事情也发生在第三和第四个例子中。在最后一个例子中，方面是一个短语“启动时间”，从句子内容“启动时间非常快，从35秒到1分钟”，这个模型可以学习到“时间”是方面中最重要的一个词，这有助于进一步发现表示“非常快”的情感。

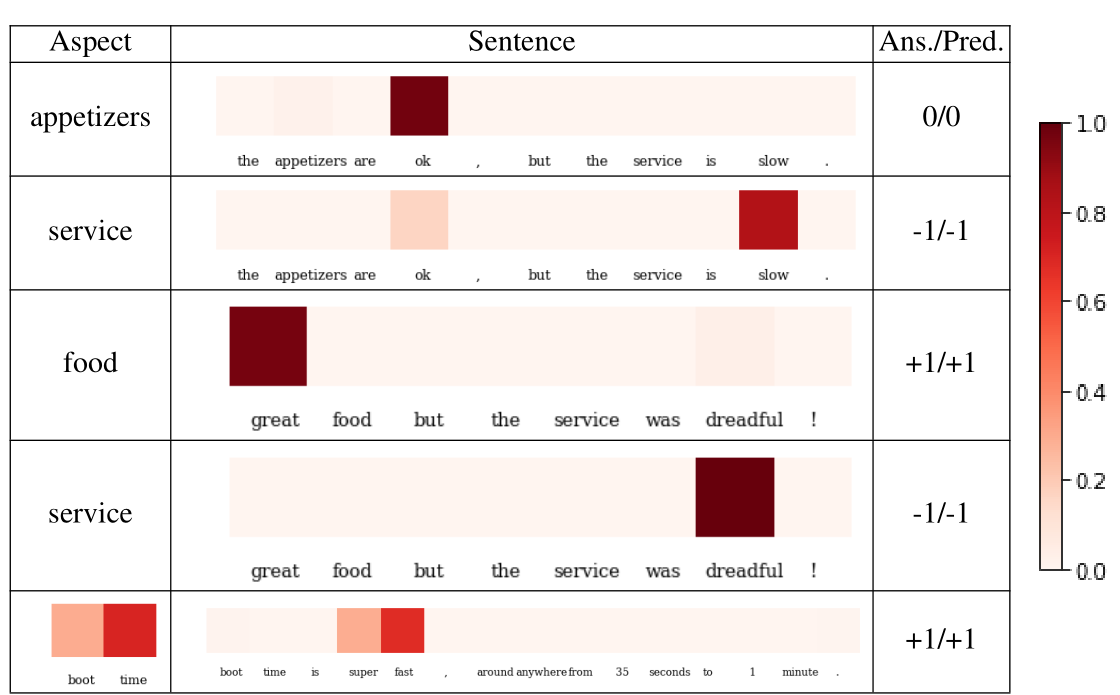


表3 句子最后注意权重的例子。颜色深度表示注意向量γ中权重的重要程度。

Error Analysis

第一类主要错误来自于以前作品中也出现过的非构图情感表达[25]。例如，在“我们的两道菜大约花了2个半小时才上完”这句话中，对“上完”这一方面没有直接的感情表达。第二类错误是由句子中使用的习语引起的。例子包括“这项服务是正点的——你还期望丽思酒店提供什么服务？”其中“服务”是方面词。在这种情况下，我们的模型无法理解习语“正点”所表达的情感。第三个因素是复杂的情感表达，如“我从未吃过糟糕的饭菜（或糟糕的服务）@pigalle”。我们的模型仍然不理解这种复杂的表达的含义，即使它能处理简单的否定句，如“菜到了，满口青椒，人绝对吃不下”中的“绝对吃不下”。

5 Conclusion

本文提出了一种面向方面情感分类的神经网络模型。我们的模型利用Attention-over-Attention模块来学习方面和句子中的重要部分，从而生成句子的最终表示。在SemEval 2014数据集上的实验表明，与那些基线方法相比，我们的模型具有更好的性能。我们的案例研究也表明，我们的模型能够有效地学习句子和目标语中的重要部分。

在我们的错误分析中，有些情况下我们的模型不能有效地处理。一是复杂的情感表达。一种可能的解决方案是将句子的语法结构合并到分类模型中。另一种错误来自不常见的习语。在未来的工作中，我们将探讨如何将先前的语言知识整合到这样的神经网络模型中。