Abstract

方面层次情感分类是情感分析中的一项精细化任务。由于它提供了更加完整和深入的结果，方面层次的情绪分析近年来受到了广泛的关注。本文揭示了句子的情感极性不仅由内容决定，而且与所关注的方面密切相关。例如，“开胃菜没问题，但服务很慢。”，对于口味，极性是正的，而对于服务，极性是负的。因此，探讨一个方面和一个句子的内容之间的联系是值得的。为此，我们提出了一种基于注意的长短期记忆网络，用于体层情感分类。当以不同方面作为输入时，注意机制可以集中在句子的不同部分。我们在SemEval 2014数据集上进行了实验，结果表明，我们的模型在方面级别的情感分类上达到了最先进的性能。

Introduction

情绪分析（Nasukawa and Yi，2003）也被称为意见挖掘（Liu，2012），是近年来备受关注的NLP关键任务。方面层次情感分析是一个细粒度的任务，可以提供完整和深入的结果。本文对体层情感分类进行了研究，发现句子的情感极性对内容和体层都有很大的依赖性。比如“员工不是很友善，但品味涵盖一切”的情感极性，如果是食物的方面，则是积极的，而在考虑服务的方面，则是消极的。当考虑不同方面时，极性可能是相反的。

神经网络在机器翻译（Lample et al.，2016）、释义识别（Yin et al.，2015）、问答（Golub and He，2016）和文本摘要（Rush et al.，2015）等多种NLP任务中取得了最新的性能。然而，神经网络模型在处理微观层次的情绪分类方面还处于初级阶段。在一些研究中，目标相关情绪分类可以从考虑目标信息中获益，例如在目标相关LSTM（TD-LSTM）和目标连接LSTM（TC-LSTM）中（Tang等人，2015a）。然而，这些模型只考虑目标而不考虑方面信息，这被证明是方面级分类的关键。

注意力已经成为获得优异结果的有效机制，如图像识别（Mnih等人，2014年）、机器翻译（Bahdanau等人，2014年）、蕴含推理（Rocktaschel等人，2015年）和句子总和化（Rush等人，2015年）所示。更重要的是，神经注意力可以提高阅读理解能力（Hermann等人，2015）。在这篇文章中，我们提出了一个注意机制，以强制模式关注句子的重要部分，以回应特定的方面。我们设计了一个方面到句子注意机制，可以集中在句子的关键部分给予方面。

我们探讨了在方面层情感分类中方面和情感极性的潜在相关性。为了捕获特定方面的重要信息，我们设计了一个基于注意的LSTM。我们在包含餐厅和笔记本电脑数据的基准数据集（Pontiki等人，2014）上评估我们的方法。

我们工作的主要贡献概括如下：

•我们提出基于注意的长-短期记忆用于方面层面的情感分类。当涉及到不同方面时，这些模型能够处理句子的不同部分。结果表明，注意机制是有效的。

•由于方面在这项任务中起着关键作用，我们提出了两种在注意过程中考虑方面信息的方法：一种是将方面向量连接到句子隐藏表示中以计算注意权重，另一种是将方面向量附加到输入词向量中。

•实验结果表明，与多个基线相比，我们的方法可以提高性能，并且进一步的例子证明了注意机制对于方面级别的情感分类是有效的。

本文的其余部分结构如下：第2节讨论了相关的工作，第3节详细描述了我们基于注意力的建议，第4节进行了大量的实验来证明我们的建议的有效性，第5节总结了这项工作和未来的方向。

Related Work

在这一部分中，我们将简要回顾方面级情绪分类和神经网络用于情绪分类的相关工作。

Sentiment Classification at Aspect-level

在文献中，方面级情感分类通常被认为是一个分类问题。如前所述，方面级情感分类是一个细粒度的分类任务。目前大多数的方法试图检测整个句子的极性，而不管所提到的实体或方面如何。解决这些问题的传统方法是手动设计一组特性。随着丰富的情感词典（拉奥和Ravigand，2009；Perez Rosas等人，2012；Kaji和喜连川，2007），词汇为基础的特点建立情感分析（穆罕默德等人，2013）。这些研究主要集中在构建情感分类器，这些特征包括单词和情感词典，使用SVM（Mullen和Celever，2004）。然而，结果在很大程度上取决于特征的质量。此外，特征工程是劳动密集型的。

Sentiment Classification with NeuralNetworks

由于提出了一种简单有效的学习分布式表示的方法（Mikolov等人，2013），神经网络在很大程度上促进了情绪分析。经典模型包括递归神经网络（Socher et al.，2011；Dong et al.，2014；Qian et al.，2015）、递归神经张量网络（Socher et al.，2013）、递归神经网络（Mikolov et al.，2010；Tang et al.，2015b）、LSTM（Hochreiter and Schmidhuber，1997）和树LSTM（Tai et al., 2015年）目前已应用于情绪分析。利用句子的句法结构，基于树的LSTMs已经被证明对许多NLP任务非常有效。然而，这种方法可能会遇到语法分析错误，这在缺乏资源的语言中很常见。

LSTM在各种NLP任务中取得了巨大的成功。TD-LSTM和TC-LSTM（Tang等人，2015a）考虑了目标信息，在目标相关情绪分类方面取得了最新的性能。TC-LSTM通过平均目标短语包含的词的向量来获得目标向量。然而，简单地平均目标短语的单词嵌入量不足以表示目标短语的语义，从而导致次优性能。

尽管这些方法是有效的，但是在细粒度层面上区分不同的情绪极性仍然是一个挑战。因此，我们有动机设计一个强大的神经网络，可以充分利用方面信息进行情感分类。

**Attention-based LSTM with Aspect Embedding**

Long Short-term Memory (LSTM)

递归神经网络（RNN）是传统前馈神经网络的一种扩展。然而，标准RNN存在梯度消失或爆炸问题。为了克服这些问题，开发了长-短期记忆网络（LSTM），并取得了优异的性能（Hochreiter和Schmidhuber，1997）。在LSTM体系结构中，有三个门和一个单元存储器状态。图1说明了标准LSTM的体系结构。

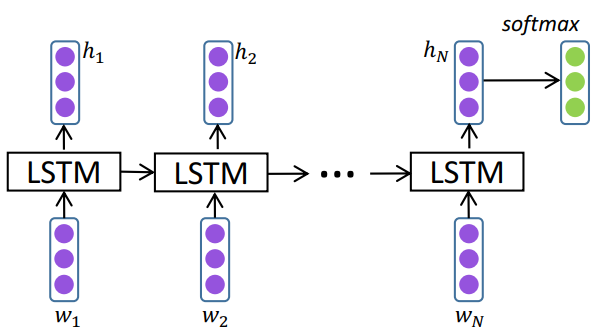
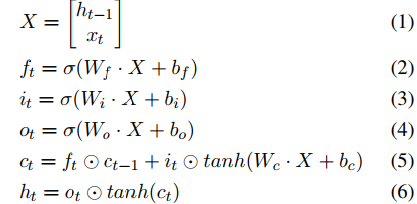


图1：标准LSTM的体系结构。{w1，w2。. . ，wN}表示长度为N的词向量。{h1，h2。. . ，hN}是隐藏向量。

更正式地说，LSTM中的每个单元可以计算如下：



其中Wi，Wf，Wo∈Rd×2d为加权矩阵，bi，bf，bo∈Rd为训练过程中学习的LSTM偏差，分别对输入门、遗忘门和输出门的变换进行参数化。σ是S形函数，表示元素相乘。xt包括LSTM单元的输入，表示图1中的单词嵌入向量wt。隐藏层的矢量是ht。

我们将最后一个隐藏向量hN作为句子的表示形式，将hN线性化为一个长度等于类标签个数的向量，然后将其放入softmax层。在我们的工作中，类标签集是{正、负、中性}。

LSTM with Aspect Embedding(AE-LSTM)

在对一个句子的极性进行分类时，方面信息是至关重要的。如果考虑不同的方面，我们可能会得到相反的极性。为了更好地利用方面信息，我们提出学习每个方面的嵌入向量。

向量vai∈Rda表示方面i的嵌入，其中da是方面嵌入的维数。A∈Rda×| A |由所有方面的嵌入组成。据我们所知，这是第一次提出方面嵌入。

Attention-based LSTM (AT-LSTM)

标准的LSTM无法检测哪一个是方面级情感分类的重要部分。为了解决这一问题，我们提出设计一种注意机制，能够捕捉句子的关键部分，以回应给定的方面。图2表示基于注意力的LSTM（AT-LSTM）的体系结构。

设H∈Rd×N是由隐向量组成的矩阵[h1。，hN]生成的LSTM，其中d是隐藏层的大小，N是给定句子的长度。此外，va表示方面的嵌入，eN∈RN是1s的向量，注意机制将产生注意权重向量α和加权隐藏表示r。

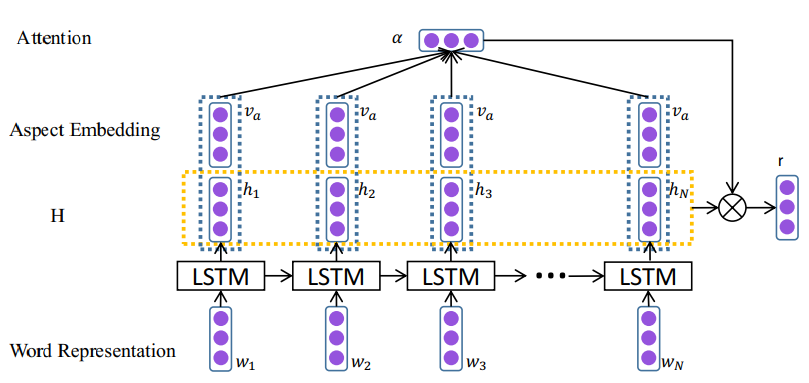
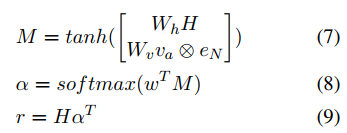


图2：基于注意力的LSTM的体系结构。方面嵌入被用来决定注意权重和句子表示。{w1，w2。. . ，wN}表示长度为N的句子中的词向量。va表示方面嵌入。α是注意力的权重。{h1，h2。. . ，hN}是隐藏向量。



其中，*M*∈R(*d*+*da*)*×N*, *α* ∈R*N*, *r* ∈R*d*, *Wh* ∈R*d×d*, *Wv* ∈R*da×da* 和 *w* ∈R*d*+*da*具有投影参数。a是由注意权重组成的向量，r是具有给定方面的句子的加权表示。7中的运算符（内有乘法符号的圆，这里简称OP）表示：*va⊗eN* = [*v*; *v*; *. . .* ; *v*]，即，运算符重复连接v N次，其中eN是具有N 1s的列向量。*Wvva ⊗ eN*表示重复线性转换的va的次数与句子中有单词的次数相同。

最后一句陈述如下：



式中，*h∗* ∈R*d*, *Wp* and *Wx*是训练过程中需要学习的投影参数。我们发现，如果将*WxhN*添加到句子的最终表示中（Rocktaschel et al.，2015），这实际上更有效。

注意机制允许模型在考虑不同方面时捕捉句子中最重要的部分。

h∗被认为是给定输入方面的句子的特征表示。我们添加一个线性层将句子向量转换为e，e是一个长度等于类种数*|C|*的实值向量。然后，跟随一个softmax层将e转换为条件概率分布。



其中Ws和bs是softmax层的参数

Attention-based LSTM with AspectEmbedding (ATAE-LSTM)

在AE-LSTM中使用方面信息的方法是让方面嵌入在计算注意权重中发挥作用。为了更好地利用方面信息，我们将输入方面嵌入到每个单词输入向量中。该模型的结构如图3所示。这样，输出隐藏表示（h1、h2、…、hN）可以具有来自输入方面（va）的信息。因此，在接下来计算注意权重的步骤中，可以模拟单词和输入方面之间的相互依赖关系

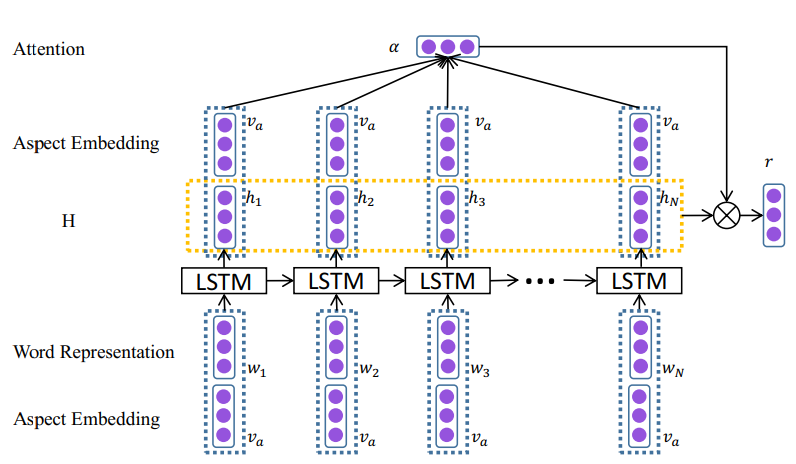


图3：具有方面嵌入的基于注意的LSTM的体系结构。方面嵌入和单词嵌入一起作为输入。{w1，w2。. . ，wN}表示长度为N的句子中的词向量。va表示方面嵌入。α是注意的权重。{h1，h2。. . ，hN}是隐藏向量。

Model Training

该模型可以通过反向传播进行端到端的训练，目标函数（损失函数）为交叉熵损失。设y为句子的目标分布，yˆ为预测的情绪分布。训练的目的是最小化所有句子y和yˆ之间的交叉熵误差



其中i是句子的索引，j是类的索引。我们的分类是三种。λ是L2-正则项。θ是参数集。

与标准LSTM类似，参数集是{Wi、bi、Wf、bf、Wo、bo、Wc、bc、Ws、bs}。此外，单词嵌入也是参数。请注意，Wi、Wf、Wo、Wc的维度随模型的不同而变化。如果在LSTM单元的输入端加入方面嵌入，Wi、Wf、Wo、Wc的维度将相应增大。其他参数如下：

AT-LSTM：嵌入A的方面自然地添加到参数集中。此外，Wh，Wv，Wp，Wx，w是注意的参数。因此，ATLSTM的附加参数集是{A，Wh，Wv，Wp，Wx，w}。

AE-LSTM：参数包括方面嵌入A。此外，由于方面向量被连接，Wi、Wf、Wo、Wc的维数将被扩展。因此，附加参数集由{A}组成。

ATAE-LSTM：参数集由{A，Wh，Wv，Wp，Wx，w}组成。此外，Wi、Wf、Wo、Wc的维度将随着方面嵌入的级联而扩展。

在训练过程中对单词嵌入和方面嵌入进行了优化。词汇量不足的词的百分比约为5%，它们是从U（－ϵ，ϵ）随机初始化的，其中ϵ=0.01。

在我们的实验中，我们使用AdaGrad（Duchi等人，2011）作为优化方法，在分布式环境下显著提高了SGD对大规模学习任务的鲁棒性（Dean等人，2012）。AdaGrad根据参数调整学习速率，对不频繁的参数执行较大的更新，对频繁的参数执行较小的更新。

**Experiment**

我们将所提出的模型应用于层面情绪分类。在我们的实验中，所有词向量都由Glove1初始化（Pennington等人，2014）。在一个约8400亿大小的未标记语料库上，对嵌入向量进行预训练。其他参数通过从均匀分布U（－ϵ，ϵ）采样来初始化。字向量的维数、方面嵌入和隐藏层的大小为300。注意权重的长度与句子的长度相同。Theano（Bastien等人，2012）用于实现我们的神经网络模型。我们训练的所有模型的批量大小为25例，动量为0.9，L2正则化权重为0.001，初始学习率为0.01。

Dataset

我们在SemEval 2014任务4的数据集上进行实验（Pontiki等人，2014）。数据集由客户评论组成。每个评论都包含一个方面和相应极性的列表。我们的目的是用对应的方面来识别句子的方面极性。统计数据见表1。

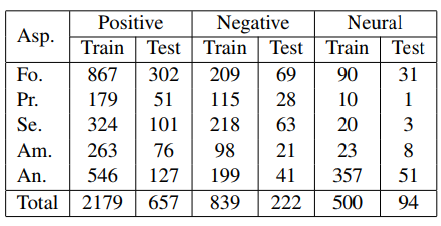


表1：每个情感类的方面分布。{Fo.，Pr.，Se，Am.，An.}指{食物、价格、服务、氛围、轶事/杂项}。“Asp”指的是方面。

Task Definition

Aspect-level Classification

给定一组预先确定的方面，这个任务是确定每个方面的极性。例如，有一句话，“这家餐馆太贵了”，有一个方面的价格是负的。一组方面是{食物、价格、服务、氛围、轶事/杂项}。在SemEval 2014任务4的数据集中，只有餐厅数据具有方面特定的极性。表2说明了比较结果。

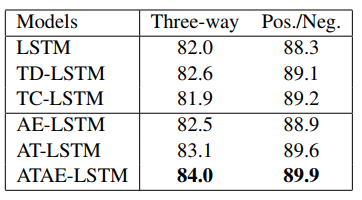


表2：餐馆方面极性分类的准确性。三向代表三级预测。Pos/Neg指示忽略所有中性实例的二进制预测。最好的分数是粗体的。

Aspect-Term-level Classification

对于句子中给定的一组方面术语，这项任务是确定每个方面术语的极性是正、负还是中性。我们在2014年SemEval任务4的数据集上进行了实验。在餐馆和笔记本电脑数据集的句子中，每个方面词的出现都有位置和情感的极性。例如，有一个方面术语fajitas，它的极性在句子“我爱他们的fajitas”中是否定的。

实验结果见表3和表4。与方面级分类的实验类似，我们的模型达到了最新的性能。

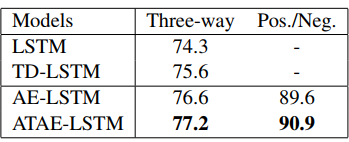


表3：餐馆方面术语极性分类的准确性。三向代表三级预测。Pos/Neg指示忽略所有中性实例的二进制预测。最好的分数是粗体的.

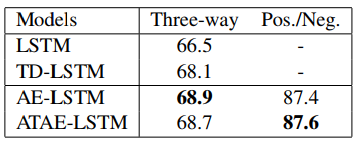


表4：笔记本电脑方面术语极性分类的准确性。三向代表三级预测。Pos/Neg指示忽略所有中性实例的二进制预测。最好的分数是粗体的。

Comparison with baseline methods

我们将我们的模型与几个基线进行比较，包括LSTM、TD-LSTM和TC-LSTM。

LSTM：标准的LSTM不能捕捉句子中的任何方面信息，因此它必须得到相同的情感极性，尽管有不同的方面。因为它不能利用方面信息，所以毫不奇怪模型的性能最差。

TD-LSTM：TD-LSTM将一个方面作为目标，可以提高情感分类器的性能。由于TD-LSTM中没有注意机制，它无法“知道”哪些单词对于给定的方面是重要的。

TC-LSTM：TC-LSTM扩展了TD-LSTM，它将目标合并到句子的表示中。值得注意的是，在表2中，TC-LSTM的性能比LSTM和TD-LSTM差。TC-LSTM在LSTM单元的输入中加入了从字向量中获得的目标表示。

在我们的模型中，我们将方面嵌入到另一个向量空间中。在训练过程中可以很好地学习各方面的嵌入向量。ATAELSTM不仅解决了词向量与方面嵌入不一致的缺点，而且可以根据给定的方面捕获最重要的信息。此外，当给定不同的方面时，ATAE-LSTM可以捕获句子的重要部分和不同部分。

Qualitative Analysis

分析哪一个词在某一方面决定了句子的情感极性是有启发性的。我们可以得到等式8中的注意权重α，并相应地可视化注意权重。

图4展示了注意如何集中在具有给定方面影响的单词上。我们使用可视化工具Heml（Deng等人，2014）对句子进行可视化。颜色深度表示注意向量α中权重的重要程度，颜色越深越重要。图4中的句子是“我不得不说，他们的送货时间是城里最快的。”还有“我们试过的法吉塔酒味道很差，烧得很烂，鼹鼠酱太甜了。”相应的方面分别是服务和食品。显然，注意可以动态地从整个句子中获得重要的部分。在图4（a）中，“最快交付时间”是一个多词短语，但是如果服务可以作为输入方面，我们的基于注意力的模型可以检测出这样的短语。此外，如果存在不止一个关键字，则注意可以检测多个关键字。在图4（b）中，检测到无味和太甜。

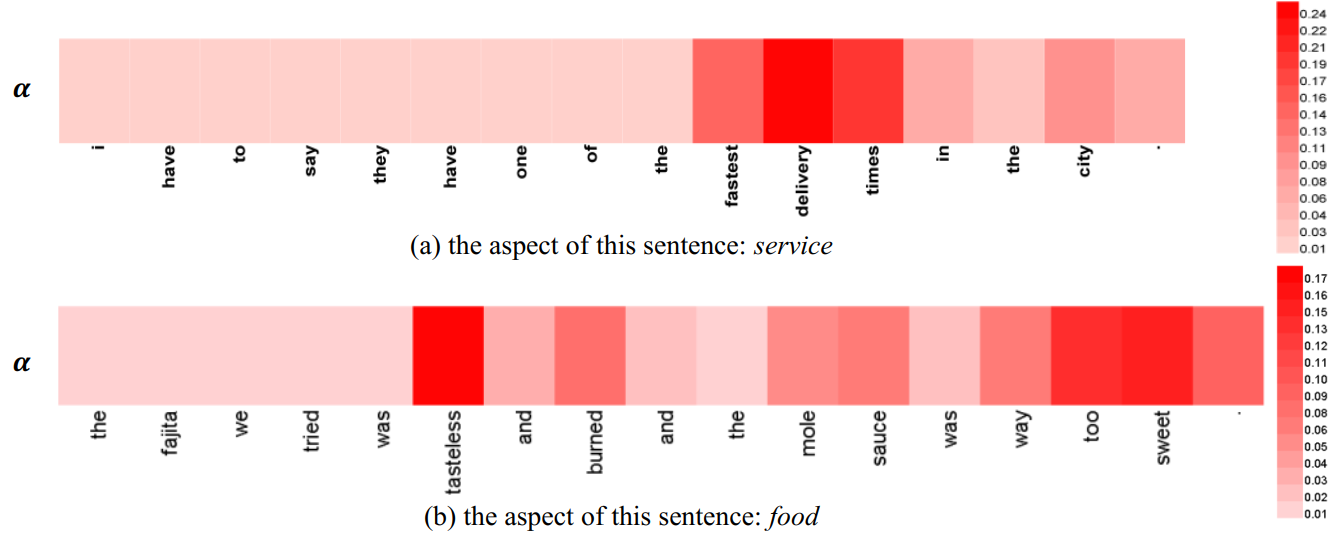


图4：注意力可视化。（a）和（b）分别是服务和食品。颜色深度表示注意向量α中权重的重要程度。从（a）开始，注意可以从整个句子中动态地检测出重要的词，即使是可以在其他领域使用的多语义短语，如“最快的传递时间”。从（b），注意可以知道多个关键点，如果存在一个以上的关键点。

Case Study

正如我们所展示的，我们的模型获得了最先进的性能。在这一部分中，我们将通过一些典型的例子进一步说明我们的建议的优点。

在图5中，我们列出了来自测试集的一些示例，这些示例具有典型的特性，并且不能由LSTM推断。在（a）句中，“开胃菜没问题，但服务很慢。”，有两个方面：食物和服务。我们的模型可以从不同的方面区分不同的情绪极性。在（b）句中，“我强烈推荐它不仅是因为它的美味佳肴，也因为它友好的主人和员工。”，有一个否定词“不”。我们的模型可以得到正确的极性，不受不代表否定的否定词的影响。在最后一个例子（c）中，“服务，然而，是一个或两个低于食物质量的指标（可怕的调酒师），而客户，在很大程度上，是吵闹的，大声说话的通勤者（这可以解释来自工作人员的坏态度）得到一个AC/DC音乐会或尼克斯比赛加载。”，这句话有一个漫长而复杂的结构因此，现有的解析器很难获得正确的解析树。因此，基于树的神经网络模型很难正确预测极性。而基于注意的LSTM则借助于注意机制和方面嵌入在句子中起到了很好的作用。

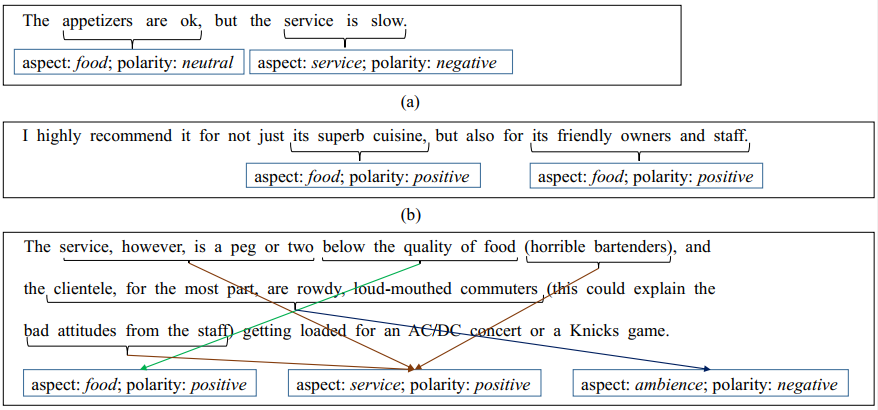


图5：分类示例。（a） 是一个具有不同方面的实例。（b） 表示我们的模型可以集中在关键点所在的位置，而不受私有词not的干扰。（c） 代表长而复杂的句子。我们的模型可以得到正确的情感极性

**Conclusion and Future Work**

在本文中，我们提出了一种基于注意的LSTMs，用于方面层次情感分类。这些建议的核心思想是学习方面嵌入并让方面参与计算注意权重。我们提出的模型可以在给出不同方面时集中于句子的不同部分，从而使它们在方面级别分类方面更具竞争力。实验结果表明，本文提出的模型AE-LSTM和ATAE-LSTM在性能上优于基线模型。

虽然这些建议显示了方面层面情绪分析的潜力，但不同的方面是分别输入的。作为未来的工作，一个有趣的和可能的方向将是与注意机制同时建模多个方面。