

基于语句结构信息的方面级情感分类^①



李梦磊, 刘新, 赵梦凡, 李聪

(湘潭大学 信息工程学院, 湘潭 411105)
通讯作者: 刘新, E-mail: liuxin@xtu.edu.cn

摘要: 方面级情感分析是情感分析任务中更细粒度的子任务, 目的是预测给定方面的情感倾向. 目前方面级情感分析任务大多采用一定的神经网络提取句子的语义信息, 之后进行情感极性预测. 本文在此基础上, 提出了基于语句结构信息的语义表示方法, 即融合语句词性序列中的句型结构信息. 本文分别使用两个 Bi-LSTM 进行语义特征和语句结构特征的提取, 构建成基于句型结构的语义表示. 然后将给定的方面级向量化, 嵌入到基于语句结构的语义表示中, 再经过 Softmax 层进行情感极性分类. 实验证明, 采用基于语句结构信息的语义表示方法进行方面级情感分析的效果更佳.

关键词: 情感分析; 方面级情感分析; Bi-LSTM; 语句结构特征; 注意力机制

引用格式: 李梦磊, 刘新, 赵梦凡, 李聪. 基于语句结构信息的方面级情感分类. 计算机系统应用, 2020, 29(11): 114-120. <http://www.c-s-a.org.cn/1003-3254/7681.html>

Aspect Level Sentiment Classification Based on Sentence Structure Information

LI Meng-Lei, LIU Xin, ZHAO Meng-Fan, LI Cong

(College of Information Engineering, Xiangtan University, Xiangtan 411105, China)

Abstract: Aspect level sentiment analysis is a more fine-grained sub task of sentiment analysis tasks, the purpose of which is to predict sentiment tendencies of a certain aspect. At present, most aspect level sentiment analysis tasks use neural networks to extract semantic information of sentences, and then predict emotional polarity. Based on this, this study proposes a semantic representation method based on sentence structure information, that is, the fusion of sentence structure information in the part of speech sequence of the statement. In this work, two Bi-LSTM are used to extract the semantic feature and the structural feature of the statement, and the semantic representation based on sentence structure is constructed. Then, the given aspect level vectorization is embedded into the semantic representation based on the sentence structure, and then sent to the Softmax layer for sentiment classification. Experiments show that the semantic representation method based on the information of sentence structure is more effective.

Key words: sentiment analysis; aspect level sentiment analysis; Bi-LSTM; sentence structure feature; attention

在自然语言处理 (NLP) 领域中, 情感分析任务一直都备受关注, 其通过挖掘文本中的信息来分析人们所表达的情感. 情感分析任务本质上是一种情感分类, 通过对语句中的信息进行建模与分析, 从中提取句子的情感倾向 (积极, 消极, 中立). 但是当句子中有多个

目标词汇时, 往往不能准确的表达出句子的情感极性.

目前, 方面级情感分析任务的处理方法主要分为基于机器学习和基于神经网络两类. 基于机器学习的方法主要是对文本进行手动构建特征信息, 再使用机器学习分类器将特征信息进行分类, 常用的特征分类

① 基金项目: 网络犯罪侦查湖南省普通高校重点实验室开放课题 (2018WLFZZC003)

Foundation item: Open Fund of Key Laboratory of Network Crime Investigation of Hunan Provincial Colleges (2018WLFZZC003)

收稿时间: 2020-03-24; 修改时间: 2020-04-24, 2020-05-13; 采用时间: 2020-05-18; csa 在线出版时间: 2020-10-29

器有支持向量机 (Support Vector Machine, SVM)、朴素贝叶斯等. 基于神经网络的方法则是对文本进行深度语义特征向量表示, 之后根据给定方面来预测情感极性. 相比于手动构建特征信息, 使用神经网络进行提取可以更全面, 并且其本身可以从数据中学习到目标特征表示而不需要进行特征工程. 常用的神经网络有循环神经网络 (Recurrent Neural Network, RNN)、长短期记忆网络 (Long Short-Term Memory, LSTM) 等.

本文提出了一种基于语句结构信息的语义表示方法, 用于方面级情感分析任务. 该方法从多方面融合了句子的语义信息和结构信息, 分别从句子的语义序列和词性序列中提取特征向量, 之后采用后文提出的3种方法进行特征向量整合. 采用基于语句结构信息的语义表示方法, 有助于提升方面级情感分析结果的准确性.

1 相关工作

1.1 方面级情感分析

方面级情感分析是情感分析的一个子任务, 目的是分析目标方面在语句中的情感倾向. 方面级情感分析的处理方法主要分为基于机器学习的处理方法和基于深度学习的处理方法. 基于机器学习的方法, 通过将手动构造的特征送入机器学习分类器来处理方面级情感分析. 基于深度学习的方法是将词和词组表示为词向量后送入神经网络学习特征用来分类. 常用的词向量模型有文献 [1] 的 CBOW 和 Skip-Gram. 近年来, 用于处理方面级情感分类任务的神经网络主要是 RNN 网络及其变种. 在此基础上, 文献 [2] 和文献 [3] 将注意力机制首次成功应用到图像领域和自然语言处理领域, 注意力机制开始在各项任务中大放异彩. 文献 [4] 将目标方面与上下文相结合, 通过基于句子的语法结构创建多个与目标相关的特征来改善与目标相关的情感分类. 文献 [5] 提出了一种 Attention-based Long Short-Term Memory Network, 当以不同的方面作为输入时, 使用注意力机制可以注意到语句的不同的部分. 文献 [6] 提出对语句和目标方面分别进行建模再进行处理的方法. 目前的大部分方法都是以语句作为输入, 神经网络的输出作为语句表述, 这种方法主要提取到语句在语义上的信息, 一定程度上忽略了词语在句型结构上的依赖型. 但是语句结构信息包含可以确定与目标方面情感相关词汇的位置信息. 在本文中, 我们尝

试从文本词性序列中学习得到语句结构信息, 并将这些语句结构信息与语义信息融合构建基于语句结构信息的语义表示用于处理方面级情感分析任务.

1.2 双向长短期记忆神经网络

双向长短期记忆神经网络, 简称 Bi-LSTM, 是在长短期记忆神经网络的基础上发展而来的. 双向长短期记忆神经网络是由一个前向的长短期记忆神经网络和一个反向的长短期记忆神经网络所组成, 旨在从两个方向对目标序列进行提取特征. 文献 [7] 长短期记忆神经网络, 简称 LSTM 是循环神经网络的一个变种, 它因克服了循环神经网络存在的梯度消失和梯度爆炸的问题而倍受喜爱. LSTM 在循环神经网络的基础上添加了遗忘门, 输入门和输出门. LSTM 神经元结构如图 1 所示.

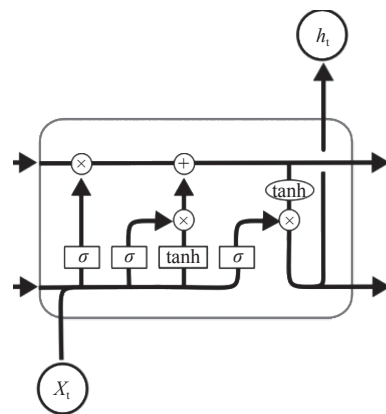


图1 LSTM 神经元结构图

遗忘门通过把上一序列隐藏状态与本神经元的输入送到激活函数决定神经元遗忘或者保留哪些信息.

$$f_t = \sigma(W_f [h_{t-1}, x_t] + b_f) \quad (1)$$

输入门决定神经元中更新哪些新信息, 使用 Sigmoid 函数决定神经元中哪些部分被更新, 然后与通过 tanh 函数的上一序列隐藏状态和本神经元的输入相乘得到新的细胞状态.

$$i_t = \sigma(W_i [h_{t-1}, x_t] + b_i) \quad (2)$$

$$C_t = \tanh(W_c [h_{t-1}, x_t] + b_c) \quad (3)$$

$$S_t = f_t * S_{t-1} + i_t * C_t \quad (4)$$

输出门决定神经元的哪些信息会被输出, 首先使用 Sigmoid 函数决定神经元中哪些部分需要输出, 然后与通过 tanh 函数的细胞状态相乘得到输出的信息.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(S_t) \quad (6)$$

2 基于语句结构信息的表示方法

本文假设语句结构信息由语句词性序列和位置信息所组成, 并试图使用 Bi-LSTM 提取语句结构特征. 模型框架如图 2 所示, 模型的输入有 3 个部分文本序列, 文本词性序列和给定的目标方面. 模型框架的流程

为: 首先, 将文本序列和文本词性序列进行嵌入表示成向量后分别送入到两个 Bi-LSTM 从两个方向进行学习语义特征表示和语句结构特征表示. 接着, 将学习到的语义特征表示和语句结构特征表示进行融合构建成为基于语句结构信息的语义表示. 然后将目标方面的嵌入表示与基于语句结构信息的语义表示相结合得到最终的基于目标方面的语句表示, 最终将语句表示送入到 Softmax 分类器中得到该方面的情感分类.

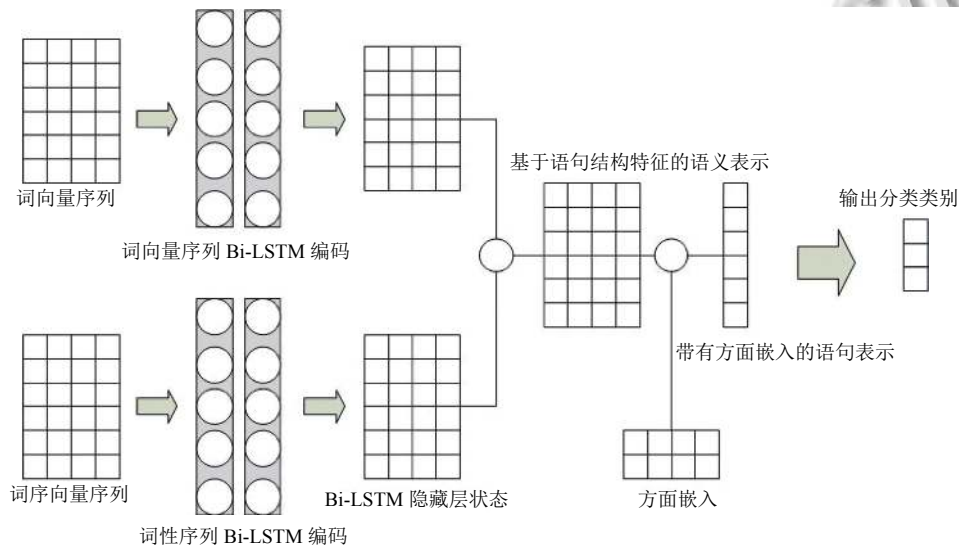


图 2 基于语句结构信息的语义表示模型框架图

2.1 语义表示和语句结构表示

(1) 词性嵌入

对于给定一条长度为 n 的语句 $s=[w_1, w_2, \dots, w_n]$ 和一组长度为 m 的目标方面 $t=[w_i, w_{i+1}, \dots, w_{i+m-1}]$, 将单词在语句中的词性标注映射到一个向量空间中, 我们称其为词性嵌入. 例如, 对于每一个单词 w_i 的词性, 可从 $M^{V \times d_w}$ 中得到 v_i 属于 R^d . 其中 V 是一个词性表的大小, d_w 是 embedding 维度. 在嵌入查找工作完成后, 可以得到语句 s 的词性序列表示和目标方面 t 的词性序列表示.

(2) Bi-LSTM 编码的语义表示和语句结构表示

在这一节中, 我们使用两个 Bi-LSTM 分别提取语义特征和语句结构特征. 输入词向量序列 $s=[w_1, w_2, \dots, w_n]$ 到 Bi-LSTM 中, Bi-LSTM 将从两个方向 (从前向后和从后往前) 提取语义特征. Bi-LSTM 的隐藏单元将生成一个隐藏的状态序列 h 作为由 Bi-LSTM 编码的语义表示. 相同的, 通过将词性向量序列 $pos=[w_1, w_2, \dots, w_n]$ 输入到 Bi-LSTM 中学习到语句结构特征表示.

2.2 基于语句结构信息的语义表示

通过将词向量序列和词性向量序列送入两个 Bi-LSTM 中, 我们将得到语句的语义表示 (表示为 s) 和语句结构特征的表示 (表示为 ss). 接下来, 我们通过将语义表示向量和语句结构表示向量进行融合从而构建基于语句结构信息的语义表示. 在此, 本文提出了 3 种将语义表示向量和语句结构表示向量进行融合的方法.

(1) 后注意力机制表示

将得到的语义表示向量 h_s 和语句结构向量 h_{ss} 进行拼接得到向量 h , 并对拼接后的向量使用注意力机制, 计算出语义信息关于语句结构的注意力权重分布 α 和基于语句结构信息的语义表示 γ .

$$e_h = \tanh \begin{pmatrix} W_s h_s \\ W_{ss} h_{ss} \end{pmatrix} \quad (7)$$

$$\alpha = \text{Softmax}(e_h) \quad (8)$$

$$\gamma = S \alpha^T \quad (9)$$

其中, h_s 是由 Bi-LSTM 编码的语义表示, h_{ss} 为由 Bi-LSTM 编码的语句结构表示, W_s 与 W_{ss} 是线性变换函数中的参数, α 是注意力权重分布, γ 是基于语句结构信息的语义表示.

(2) 先注意力机制表示

对语义表示向量 h_s 使用自注意力机制, 以进一步加强词与词在语义上的相关性, 得到向量表述 γ_s . 相同的, 对语句结构表示也使用自注意力机制, 增强词与词在语句结构上的相关性, 得到向量表述 γ_{ss} . 注意力机制的使用可以使得语句在词与词之间在语义和语句结构上的相关性进一步加强, 最终将两个向量进行拼接得到基于语句结构信息的语义表示 γ .

$$e_s = \tanh(W_s h_s) \quad (10)$$

$$\alpha_s = \text{Softmax}(e_s) \quad (11)$$

$$\gamma_s = h_s \alpha_s^T \quad (12)$$

$$e_{ss} = \tanh(W_{ss} h_{ss}) \quad (13)$$

$$\alpha_{ss} = \text{Softmax}(e_{ss}) \quad (14)$$

$$\gamma_{ss} = h_{ss} \alpha_{ss}^T \quad (15)$$

$$\gamma = \begin{pmatrix} \gamma_s \\ \gamma_{ss} \end{pmatrix} \quad (16)$$

其中, h_s 是由 Bi-LSTM 编码的语义表示, W_s 是线性变换函数中的参数, α_s 是语义表示中词与词之间的注意力权重, γ_s 是基于注意力机制的语义表示, h_{ss} 为由 Bi-LSTM 编码的语句结构表示, W_{ss} 是线性变换函数中的参数, α_{ss} 是语义表示中词与词之间的注意力权重, γ_{ss} 是基于注意力机制的语句结构表示, γ 是基于语句结构信息的语义表示.

(3) Attention-Over-Attention (AOA) 表示

受到文献 [8] 的启发, 文献 [8] 使用 AOA 模块计算出目标方面和语义上的互信息, 最终由目标方面关于语义的注意力权重和语义关于目标方面的注意力权重计算出基于目标方面的语句表述. 在此, 我们一样使用 AOA 模块将语义表示和语句结构表示进行点乘计算出互信息矩阵 M , 再对矩阵 M 使用逐行注意力机制和逐列注意力机制, 得到语义关于语句结构的注意力权重和语句结构关于语义信息的注意力权重, 根据权重信息计算出基于语句结构信息的语义表示 γ . AOA 模块使得语句结构特征和语义特征更加充分的融合, 即语句结构特征可以关注到与结构相关的语义特征,

也可以使得语义特征关注到与语义相关语句结构特征.

$$\alpha_{ij} = \frac{M_{ij}}{\sum_i^n M_{ij}} \quad (17)$$

$$\beta_{ij} = \frac{M_{ij}}{\sum_j^n M_{ij}} \quad (18)$$

$$\beta_j^* = \frac{1}{n} \sum_i^n \beta_{ij} \quad (19)$$

$$\gamma = S^T \alpha_{ij} \beta_j^{*T} \quad (20)$$

其中, α 是语句结构关于语义的注意力权重, β 是语义关于语句结构的注意力权重, β^* 是语义关于语句结构的注意力权重的平均值, γ 是基于结构信息的语义表示.

2.3 带有方面嵌入的语句表示

由于基于语句结构信息的语义表述 γ 不能够充分利用目标方面的信息, 如果对基于结构信息的语义表示直接分类, 得出的结果近似于语句的情感分类. 因此, 我们利用注意力机制计算出目标方面在语句描述 γ 上的注意力分布 β , 并根据该注意力分布 β 得到可以用于分类的带有方面嵌入的语句表示 h^* .

$$e_{asp} = \tanh \begin{pmatrix} W_\gamma \gamma \\ W_{asp} t_{asp} \end{pmatrix} \quad (21)$$

$$\alpha_{asp} = \text{Softmax}(e_{asp}) \quad (22)$$

$$h^* = \gamma \alpha^T \quad (23)$$

其中, γ 是基于语句结构信息的语义表示, t_{asp} 是目标方面的嵌入表示, e_{asp} 是将目标方面嵌入到语义信息中每个单词的向量表示中, α_{asp} 是 γ 关于给定方面的注意力权重分布, h^* 是带有方面嵌入的基于语句结构的语义表示.

带有方面嵌入的基于语句结构的语义表示就是本文所提取的最终用于情感分类的语句表示. 我们使用线性层把该语句表示从语句空间转化到情感分类的类别空间中后使用 *Softmax* 函数将其转换成各个类别的概率.

$$y = \text{Softmax}(Wh^* + b) \quad (24)$$

其中, 概率最大的就是关于目标方面的情感分类类别.

2.4 模型训练

由于任务为分类任务, 所以模型的损失函数采用交叉熵损失函数, 模型使用反向传播算法通过端到端的方式进行训练. 损失函数如下,

$$y = \sum_i^n \sum_j^m y_i^j \log y_i^j + \lambda \|w\|_2^2 \quad (25)$$

其中, y 为目标函数, n 是语句的数目, i 为语句的索引值, m 为类别数, j 为类别的索引, y 为第 i 条语句关于目标方面的真实类别, y^j 为第 i 条语句预测的为第 j 个类别的概率, $\lambda \|w\|_2^2$ 为损失函数的正则项用以降低过拟合现象的发生, 提高模型的泛化能力, λ 为正则化系数. 模型通过最小化损失函数进行更新训练.

3 实验分析

3.1 实验数据集

实验的数据集来自文献 [9], 本数据集是方面级情感分析任务研究中经常使用的数据集. 本数据集含有手提电脑评论和餐厅评论两个领域的子数据集. 数据集中已经标注好了每个语句中的方面和目标方面所呈现的情感倾向. 任务目标是预测出给定方面在本条语句中的情感倾向. 数据集中数据的分布在表 1 中给出.

表 1 SemEval 2014 Task 4 数据集分布

数据集	积极	中立	消极
电脑训练集	994	464	870
电脑测试集	341	169	128
餐厅训练集	2164	637	807
餐厅测试集	728	196	196

3.2 超参数设置

在实验中, 词向量由文献 [10] 中的 GloVe1 进行初始化, 每个词向量维度为 300. 我们使用 Batch size 为 25, 初始学习率为 0.01 的 Adam 优化器来训练模型. 其他模型参数采用随机初始化, 初始化的概率分布服从均匀分布. 实验中, Bi-LSTM 的隐藏单元大小维度为 150 维.

3.3 词性嵌入维度实验

词性的表示与词性的嵌入维度息息相关, 一般来说, 向量维度的大小决定了向量所包含信息量的多少. 所以词性嵌入的维度过小不足以完全表示词性的信息, 但是词性向量维度过多会使网络中含有过多的冗余信息干扰, 使得模型效果不佳. 因此, 我们以先注意力机制表示模型作为标准, 使用对比实验来证明哪种词性嵌入维度最适合模型. 由于词向量维度等其他网络参数的限制, 将词性的嵌入维度设定为 50, 100, 150, 200, 250, 300. Batch size 设置为 25, 模型运行 10 个 epoch 的实验结果来进行对比.

从表 2 可以看出将词性嵌入维度可以在一定程度上

影响模型的效果, 当词性嵌入维度较小时, 模型的效果会随着词性维度的增大而逐渐增大, 但是当词性嵌入维度变得过大时, 模型的效果甚至出现了下降的趋势. 因此, 为了词性嵌入表示更多的信息同时又兼顾模型的效果, 本文将词性嵌入维度设置为 150.

表 2 词性嵌入维度对比实验

词性嵌入维度	餐厅	词性嵌入维度	餐厅
50	0.7358	200	0.7704
100	0.7543	250	0.7682
150	0.7741	300	0.7659

3.4 融合表示向量对比实验

本文通过将语义表示向量和语句结构表示向量进行融合构建出基于语句结构的语义表示向量, 并将该向量与方面嵌入结合后用于最终的情感分类. 为了证明本文所提出的 3 种融合方法对本模型情感分类的结果的贡献程度. 因此本节设置对比实验, 将使用 3 种融合方法网络模型与不使用 3 种融合方法的网络模型 (即使用语义表示向量和语句结构表示向量简单堆叠后的向量作为基于语句结构的语义表示向量, 其余模块不做任何变动) 进行对比, 以证明 3 种方法的有效性. 对比结果如表 3 所示, 对比实验选用餐厅子数据集, 词性嵌入维度为 150, 其余超参数设置与 3.2 节中描述相同.

表 3 融合表示向量对比实验

融合向量构建方法	餐厅
简单堆叠	0.7493
后注意力机制表示	0.7759
先注意力机制表示	0.7875
AOA表示	0.8085

由表 3 可知, 语义表示与语句结构表示的融合方法会很大程度上影响模型的效果. 其中, 不使用任何融合方法, 直接将语义表示向量和语句结构表示向量简单堆叠后作为基于语句结构的语义表示向量, 这种方法最为简单, 但是同时模型的效果也很一般. 这主要是由于语义信息与语句信息没有充分融合所导致的. 而在其余 3 种融合向量构建方法中, 基于 AOA 的融合表示效果最佳, AOA 可以从语句结构与语义信息的互信息中提取相关性较强的特征, 因此可以使得模型效果更好一些.

3.5 对比实验与分析

将本文提出的模型与未使用语句结构信息的基线

模型进行比较以证明本文模型结构的有效性. 待比较的模型包括: LSTM, TD-LSTM, ATAE-LSTM. 我们使用准确度和 $F1$ 值作为衡量模型的指标.

SVM^[11]: 使用 N-gram 特征, 解析特征等信息使用 SVM 分类器进行情感分类.

LSTM^[5]: 把文本数据送入到 LSTM 网络中, 使用 LSTM 的输出作为文本的语句表示, 然后直接使用该语句表示进行情感分类.

TD-LSTM^[12]: 根据目标方面把原始语句分为前后两个语句, 使用两个 LSTM 分别对前后语句进行建模, 最终将两个 LSTM 的输出拼接起来进行最后情感分类.

ATAE-LSTM^[5]: 用 LSTM 对输出的语句进行建模得到语句表示, 然后将语句表示和方面嵌入表示结合后使用注意力机制得到带有方面嵌入的语句表示, 最后根据该语句表示进行情感分类.

Memnet^[13]: 在词嵌入的基础上多次应用注意力机制, 最终选择最顶层的语句表示进行情感分类.

IAN^[6]: 使用两个 LSTM 网络对语句和方面进行建模, 分别使用两个 LSTM 的隐藏状态生成注意力向量, 最终基于这两个向量生成语句表示和方面表示实现最后的分类.

由表 4 中数据可知, 与基线模型相比, 本文的模型在该数据集上的表现更佳一些. 在本文提出的 3 种关于语义表示与语句结构表示的融合方法中, 使用 AOA 方法的效果更佳些, 这可能是由于 AOA 模型分别考虑了语句结构关于语义信息的特征和语义信息关于语句结构的特征, 使得语句结构特征与语义信息能够充分融合. 因此, 带有语句结构信息的网络模型比不带有语句结构信息的网络模型表现更佳.

3.6 样例分析

在图 3 中列举了测试集中的部分样例, 为了分析每个单词对于目标方面情感分类的重要程度, 本文将最终的带有方面嵌入的语句表示进行可视化. 图中 x 轴方向为语句中的每个单词, y 轴方向为目标方面, 颜色的深浅代表了本单词对于目标方面最终分类的重要程度, 单词对于最终分类影响程度越大, 则颜色越深. 例如: 在“The food is good but the price is so expensive.”这条语句中, 一共有“food”和“price”两个方面, 而且在本条语句中这两个方面呈现了相反的情感极性. 但在图 3 中, 可以看到当给定的目标方面为“food”时, 本文的网络模型会关注到“good”而不是“expensive”, 当给定

目标方面为“price”时, 网络模型关注到“expensive”, 从而给出正确的情感极性判断. 这一个特点得益于在语句结构中“good”用来修饰“food”而不是“price”, 本文的网络模型通过学习这一个特点来给出正确的情感极性判断. 因此, 本文的网络模型在给出不同的方面时, 可以关注到语句中与目标方面相关的部分而不会被其他无关词语干扰.

表 4 模型对比实验结果

模型	餐厅		手提电脑	
	ACC	F1	ACC	F1
SVM	0.8016	-	0.7049	-
LSTM	0.7430	-	0.6653	-
TD-LSTM	0.7563	-	0.6813	-
ATAE-LSTM	0.7720	-	0.6870	-
MemNet	0.7816	0.6583	0.7033	0.6409
IAN	0.7860	-	0.7210	-
后注意力机制表示SS-LSTM (本文)	0.7759	0.6923	0.6991	0.6420
先注意力机制表示SS-LSTM (本文)	0.7875	0.6832	0.7038	0.6465
AOA表示SS-LSTM (本文)	0.8085	0.7104	0.7342	0.6732

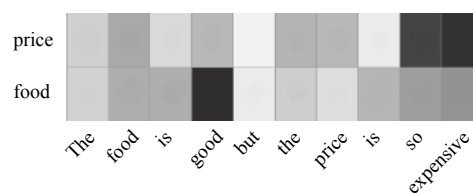


图 3 最终语句表示的注意力权重分布图

4 结论与展望

在本文中, 我们提出了一种基于语句结构的语义表示来解决方面级的情感分析任务. 本文提出了一种从词性序列提取语句结构特征的新方法以及 3 种关于语句结构特征和语义特征相融合的方法, 最终通过基于语句结构特征的语义表示去预测目标方面的情感倾向. 实验证明, 带有语句结构信息的模型比不带有语句结构信息的模型效果更好一些. 在接下来的工作中, 可以进一步的完善词性嵌入的表示来完善文本结构信息的提取.

参考文献

- 1 Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. arXiv: 1301.3781, 2013.

- 2 Mnih V, Heess N, Graves A, *et al.* Recurrent models of visual attention. Annual Conference on Neural Information Processing Systems 2014. Montreal, QC, Canada. 2014. 2204–2212.
- 3 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv: 1409.0473, 2014.
- 4 Jiang L, Yu M, Zhou M, *et al.* Target-dependent twitter sentiment classification. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, OR, USA. 2011. 151–160.
- 5 Wang YY, Huang ML, Zhao LY, *et al.* Attention-based LSTM for aspect-level sentiment classification. Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing. Austin, TX, USA. 2016. 606–615.
- 6 Ma DH, Li SJ, Zhang XD, *et al.* Interactive attention networks for aspect-level sentiment classification. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia. 2017. 4068–4074.
- 7 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)]
- 8 Huang BX, Ou YL, Carley KM. Aspect level sentiment classification with attention-over-attention neural networks. arXiv: 1804.06536, 2018.
- 9 Wagner J, Arora P, Cortes S, *et al.* DCU: Aspect-based polarity classification for SemeVal task 4. Proceedings of the 8th International Workshop on Semantic Evaluation. Dublin, Ireland. 2014. 223–229.
- 10 Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar. 2014. 1532–1543.
- 11 Kiritchenko S, Zhu XD, Cherry C, *et al.* NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. Proceedings of the 8th International Workshop on Semantic Evaluation. Dublin, Ireland. 2014. 437–442.
- 12 Tang DY, Qin B, Feng XC, *et al.* Effective LSTMs for target-dependent sentiment classification. arXiv: 1512.01100, 2015.
- 13 Tang DY, Qin B, Liu T. Aspect level sentiment classification with deep memory network. Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing. Austin, TX, USA. 2016. 214–224.