

基于多注意力网络的特定目标情感分析^①



宋曙光, 徐迎晓

(复旦大学 计算机科学技术学院, 上海 201203)

通讯作者: 徐迎晓, E-mail: xuyx@fudan.edu.cn

摘要: 作为自然语言处理领域的经典研究方向之一, 特定目标情感分析的任务是根据句子上下文语境判别特定目标的情感极性, 而提升该任务表现的重点在于如何更好地挖掘特定目标和句子上下文的语义表示. 本文提出融合短语特征的多注意力网络 (Phrase-Enabled Multi-Attention Network, PEMAN), 通过引入短语级别语义特征, 构建多粒度特征融合的多注意力网络, 有效提高模型的表达能力. 在 SemEval2014 Task4 Laptop、Restaurant 数据集上的实验结果表明, 与基准模型相比, 本文提出的 PEMAN 模型在准确率上有一定提升.

关键词: 情感分析; 注意力机制; 自然语言处理

引用格式: 宋曙光, 徐迎晓. 基于多注意力网络的特定目标情感分析. 计算机系统应用, 2020, 29(6): 163-168. <http://www.c-s-a.org.cn/1003-3254/7427.html>

Target-Specific Sentiment Analysis Based on Multi-Attention Network

SONG Shu-Guang, XU Ying-Xiao

(School of Computer Science, Fudan University, Shanghai 201203, China)

Abstract: As one of the classic research directions in the field of natural language processing, the task of target-specific sentiment analysis is to determine the sentiment polarity of a specific target based on contexts. The key to improve the performance of this task is how to better mine the semantic representation of specific target and contexts. This study proposes a multi-attention network with phrase features. By introducing phrase-level semantic features, a multi-attention network with multi-granularity features is constructed to improve the expression ability of the model effectively. The experimental results on the SemEval2014 Task4 Laptop and Restaurant datasets show that the PEMAN model proposed in this study has a certain improvement in accuracy compared with the benchmark model.

Key words: sentiment analysis; attention mechanism; natural language processing

1 引言

特定目标情感分析 (Target-Specific Sentiment Analysis, TSSA) 是自然语言处理领域的经典任务, 旨在通过对文本语义的挖掘和分析来判断其所表达的情感极性. 对于该任务, 给定一个完整的句子和其包含的目标词, 应推测出该句包含的各个目标词的情感极性 (包括积极情感、消极情感、中性). 例如, 在句子 “The menu is limited but almost all of the dishes are excellent.”

中, 有两个目标词 “menu” 和 “dishes”, 特定目标情感分析任务的目的就是要分析出目标词 “menu” 的情感极性是消极的, 目标词 “dishes” 的情感极性是积极的.

特定目标情感分析是一种细粒度的情感分析任务, 在国内外有诸多成果涌现. 该任务的研究方法一般分为传统机器学习方法和深度学习方法. 传统机器学习方法使用大量人工设计的特征集合来提高模型性能, 工作量较大. 与传统机器学习方法相比, 深度学习方法

^① 收稿时间: 2019-10-28; 修改时间: 2019-11-20; 采用时间: 2019-11-29; csa 在线出版时间: 2020-06-10

能自动提取文本语义特征,在特定目标情感分析领域取得快速发展.在 LSTM^[1]的基础上,研究者们引入了注意力机制来改善长句表达能力欠佳的问题,取得良好效果.虽然注意力机制能捕捉更多的句子情感信息,但当前研究多采用简单注意力机制,当句子中情感词为多词词组时,简单的注意力机制不能有效提取词组整体的语义特征,容易引起歧义造成模型错判,而引入短语级别语义特征能有效改善这一问题.

基于此,本文提出了融合短语特征的多注意力网络(Phrase-Enabled Multi-Attention Network, PEMAN),通过引入短语级别的语义表示,实现多粒度特征融合的多注意力网络来解决上述问题,并使用该模型在 SemEval2014^[2]的 Laptop 和 Restaurant 两个数据集上进行实验.结果表明,本文提出的 PEMAN 模型相比基线模型有一定提高,准确率分别达到 74.9% 和 80.6%.

2 相关工作

传统机器学习方法利用情感词典、语句分析等获得句子特征,然后利用分类器进行情感预测.例如,Vo 等^[3]提出基于推特语料的情感分析模型,其利用情感词典、多个词嵌入向量来提取语料的语义特征,使得模型的准确率有一定提高.Kiritchenko 等^[4]利用词袋模型、情感词典以及语义解析来构建特征,通过训练支持向量机分类器(Support Vector Machine, SVM)来进行情感分类.以上方法表现不错,但其效果依赖于复杂的特征抽取和设计工作,需要耗费大量人力物力.

近年来,越来越多的学者将深度学习方法用在特定目标情感分析任务中.Tang 等^[5]提出 TD-LSTM、TC-LSTM 两个模型,通过两个 LSTM 分别对特定目标的上下文进行建模,以获得更好的文本表示.Wang 等^[6]提出基于注意力机制的分类模型,引入注意力机制捕获编码后的句子表示中的重要信息.Tang 等^[7]提出了一个深度记忆网络模型,其通过线性方式组合了多个收集目标特征的注意力计算层,以提高注意力网络的准确性.Ma 等^[8]使用两个注意力网络计算特定目标和上下文的语义表示并构建其交互表示,以此来进行情感分类.Chen 等^[9]提出利用循环注意力结构获得多层句子特征,并将其通过 GRU 非线性组合起来,使模型具有更好的表达能力.Huang 等^[10]利用叠加注意力机制对句子和目标词的特征做交互计算,提高了模型的准确率.

在特定目标情感分析任务的已有研究工作中,当特定目标所对应的情感词为多词的时候,可能会因为情感词语义融合不当而导致错误情况.例如,在句子“Great food and the service was not bad.”中,对于目标词“service”来说,表达其情感倾向的是“not bad”这个情感词词组,所表达的是积极情感.在遇到上述例子时,以往的模型可能会聚焦于“bad”这个单词而预测出负面情感,造成分类的错误.针对以上问题,本文提出 PEMAN 模型,通过引入短语级别的语义表示,获取更丰富的句子语义表示,实现多注意力网络的特征融合,以解决注意力分散的问题.

3 PEMAN 模型

本文提出的融合短语特征的多注意力网络 PEMAN 模型的整体框架图如图 1 所示,其结构主要由输入层、编码层、多注意力层、输出层等部分组成:

(1) 输入层:对模型输入部分做处理,进行向量嵌入操作.

(2) 编码层:使用 Bi-LSTM^[11]对输入内容进行编码,并嵌入位置信息.

(3) 多注意力层:使用两个注意力交互矩阵对隐层状态输出进行计算,得到最终的语义表示.

(4) 输出层:使用多注意力层输出的语义表示进行情感分类.

3.1 输入层

本文模型的输入共有 3 个部分:句子表示、目标词表示、短语集合表示.给定一个数据集中长度为 n 的句子 $c = [w_1, w_2, \dots, w_n]$ 和长度为 m 的目标词 $a = [w_i, w_{i+1}, \dots, w_{i+m-1}]$,使用 GloVe 预训练词向量^[12]初始化可得到其词向量表示,其中 d 为词向量维度.

$$C = [W_c^1; W_c^2; \dots; W_c^n] \in R^{n \times d}$$

$$A = [W_a^1; W_a^2; \dots; W_a^m] \in R^{m \times d}$$

构建用于提取短语特征的向量集合.具体方法是:设短语长度为 x ,那么对于长度为 n 的句子 C ,如式(1):

$$p_i = \max[C[i:i+x]], i = 1, 2, \dots, n-x+1 \quad (1)$$

其中, $C[i:i+x]$ 代表从 C 中取第 i 到 $i+x-1$ 个词向量. p_i 表示从 C 中取出 x 个词向量经过最大池化操作得到的第 i 个短语表示. $P = [p_1, p_2, \dots, p_{n-x+1}] \in R^{(n-x+1) \times d}$, 代表从 C 中取出的长度为 x 的短语向量集合.其中,短

语长度的取值范围为 $[1, 2, \dots, n]$, 当 x 取1时, 短语集合即为句子表示, 不能有效挖掘短语语义特征. 当 x 过大时, 将混淆句子中单词的语义表示, 模型的表情能力

将会变差. 短语长度取值的不同将会影响模型的特征抽取能力, 经实验验证, 当 x 取3时, 模型表情能力最强.

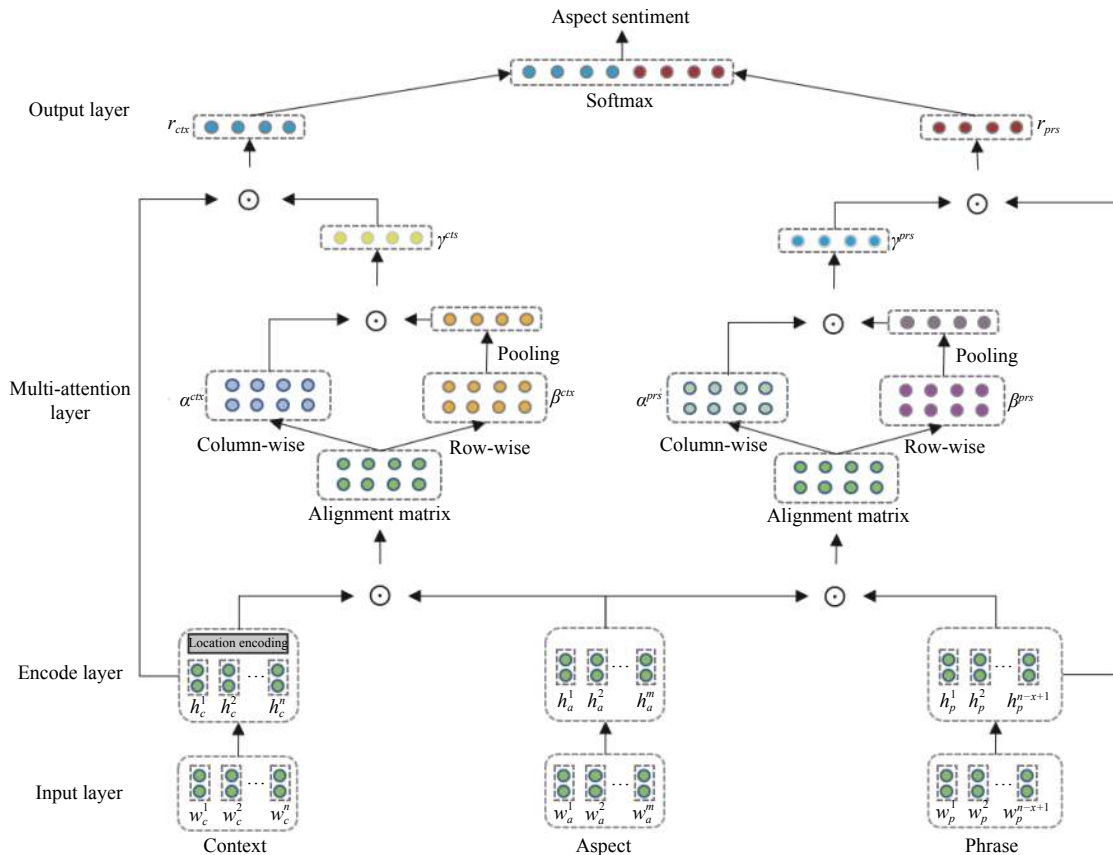


图1 PEMAN 模型结构图

3.2 编码层

模型将输入层得到的句子、目标词、短语集合的词向量表示分别送入到3个Bi-LSTM中, 分别学习整个句子、目标词、短语集合的隐藏语义信息.

$$\vec{h}_c = \overrightarrow{LSTM}(W_c^1; W_c^2; \dots; W_c^n) \quad (2)$$

$$\overleftarrow{h}_c = \overleftarrow{LSTM}(W_c^1; W_c^2; \dots; W_c^n) \quad (3)$$

其中, 式(2)表示Bi-LSTM正向的隐藏状态输出, 用于提取句子的正向语义特征. 式(3)表示Bi-LSTM反向的隐藏状态输出, 用于提取句子反向的语义特征. 二者拼接起来得到句子表示的隐藏状态输出 $h_c = [h_c^1, h_c^2, \dots, h_c^n] \in R^{n \times 2d_h}$, 如式(4)所示.

$$h_c = [\vec{h}_c, \overleftarrow{h}_c] \quad (4)$$

同理可得目标词表示的隐藏状态输出 $h_a = [h_a^1, h_a^2, \dots, h_a^m] \in R^{m \times 2d_h}$, 以及短语集合表示的隐藏状态输

出 $h_p = [h_p^1, h_p^2, \dots, h_p^{n-x+1}] \in R^{(n-x+1) \times 2d_h}$, 如式(5)、式(6)所示.

$$h_a = [\vec{h}_a, \overleftarrow{h}_a] \quad (5)$$

$$h_p = [\vec{h}_p, \overleftarrow{h}_p] \quad (6)$$

其中, d_h 为Bi-LSTM隐藏状态的维度, n 为句子长度, m 为目标词长度, $n-x+1$ 为短语集合的向量个数.

句子中离目标词越近的情感词更可能表达目标词的情感极性. 因此, 模型在得到句子表示之后, 还应充分考虑目标词和上下文单词的位置信息.

给定长度为 n 的句子和长度为 m 的目标词, 对于句中任意一个单词 w_t , 设 w_t 和目标词的相对距离是 l (目标词可能是单词或多词, 这里统一把目标词当做一个短语整体), 则单词 w_t 的位置权重 v_t 的计算方式如式(7):

$$v_t = 1 - \frac{l}{n - m + 1} \quad (7)$$

得到位置权重后,可由式(8)计算包含位置信息权重的句子最终语义表示:

$$h_c = [h_c^1 * v_1, h_c^2 * v_2, \dots, h_c^n * v_n] \quad (8)$$

3.3 多注意力层

在分别得到句子、目标词、短语集合的语义表示 h_c 、 h_a 、 h_p 之后,借助以上语义信息,引入多注意力机制进行特征融合,具体做法如下:给定句子表示 $h_c \in R^{n \times 2d_h}$ 和目标词表示 $h_a \in R^{m \times 2d_h}$,首先计算交互矩阵 $I^{ctx} \in R^{n \times m}$,计算方式如式(9)所示.其中,交互矩阵中每个元素表示句子中每个词和目标词之间的相关度.

$$I^{ctx} = h_c \cdot h_a^T \quad (9)$$

对于交互矩阵 I^{ctx} ,分别按列、行进行归一化处理,记为 $\alpha^{ctx} \in R^{n \times m}$ 、 $\beta^{ctx} \in R^{n \times m}$,矩阵 α^{ctx} 和 β^{ctx} 中每个元素分别表示目标词中每个词对于句子中各个词的权重和句子中每个词对于目标词中各个词的权重,如式(10)所示.将 β^{ctx} 按列取平均,可得目标词中每个词所占的权重,记为 β_j^{ctx} ,如式(11)所示,由此可得所有目标词的权重为 $\overline{\beta^{ctx}} \in R^m$.

$$\alpha_{ij}^{ctx} = \frac{\exp(I_{ij}^{ctx})}{\sum_i \exp(I_{ij}^{ctx})}, \beta_{ij}^{ctx} = \frac{\exp(I_{ij}^{ctx})}{\sum_j \exp(I_{ij}^{ctx})} \quad (10)$$

$$\beta_j^{ctx} = \frac{1}{n} \sum_i \beta_{ij}^{ctx} \quad (11)$$

句子中所有词的注意力权重分布 γ^{ctx} 由 α^{ctx} 和 $\overline{\beta^{ctx}}^T$ 相乘得到,如式(12)所示:

$$\gamma^{ctx} = \alpha^{ctx} \cdot \overline{\beta^{ctx}}^T \quad (12)$$

同理,对短语集合表示 h_p 和目标词表示 h_a 做同样的注意力权重计算可得短语集合中各短语的注意力权重分布 γ^{prs} ,方法同上.计算公式如式(13)~式(16)所示:

$$I^{prs} = h_p \cdot h_a^T \quad (13)$$

$$\alpha_{ij}^{prs} = \frac{\exp(I_{ij}^{prs})}{\sum_i \exp(I_{ij}^{prs})}, \beta_{ij}^{prs} = \frac{\exp(I_{ij}^{prs})}{\sum_j \exp(I_{ij}^{prs})} \quad (14)$$

$$\beta_j^{prs} = \frac{1}{n} \sum_i \beta_{ij}^{prs} \quad (15)$$

$$\gamma^{prs} = \alpha^{prs} \cdot \overline{\beta^{prs}}^T \quad (16)$$

3.4 输出层

在输出层,句子表示 r_{ctx} 由句子多个隐藏状态的输出加权求和得到,短语集合表示 r_{prs} 由短语集合多个隐藏状态的输出加权求和得到,计算公式如下所示:

$$r_{ctx} = h_c^T \cdot \gamma^{ctx} \quad (17)$$

$$r_{prs} = h_p^T \cdot \gamma^{prs} \quad (18)$$

句子最终的语义表示 r 由二者拼接得到,当作最终的句子特征送入到 Softmax 层中,得到结果的概率分布,计算公式如下所示:

$$r = [r_{ctx}; r_{prs}] \quad (19)$$

$$y = L_{\text{Softmax}}(W_l * r + b_l) \quad (20)$$

其中, W_l 和 b_l 是 Softmax 层的参数,结果 y 表示模型预测结果.

模型通过端到端反向传播的方式进行训练,以交叉熵^[13]作为损失函数,并加入正则化项^[14]减少过拟合,如式(21)所示:

$$loss = \sum_k \sum_{i \in C} y_i^g \cdot \log(y_i) + \lambda \|\theta\|^2 \quad (21)$$

其中, k 表示训练数据集中的样本, C 表示分类的类别,本实验中 $C=3$. λ 是 L_2 正则化的参数.在分类器得到的结果 y 中,概率最大的 y_i 作为模型预测得到的标签.同时,为了进一步防止模型过拟合,引入 dropout^[15].

4 实验与分析

4.1 数据集

本文在 SemEval2014^[1]比赛的 Restaurant、Laptop 数据集上验证模型的效果.目标词的情感极性分为3类:积极情感、消极情感、中性.数据集的统计情况如表1所示.

表1 数据集总体统计

数据集		积极情感	消极情感	中性
餐馆数据	训练集	2164	807	637
	测试集	728	196	196
笔记本数据	训练集	994	870	464
	测试集	341	128	169

4.2 超参数设置

实验过程中,文本中的单词采用 300 维的 GloVe 预训练词向量^[12]进行初始化,所有不在词向量词典中的词,随机初始化为服从 $[-0.25, 0.25]$ 均匀分布的 300 维随机向量.所有的权重矩阵被初始化为服从 $[-0.01, 0.01]$ 的均匀分布,所有偏置量全都置为 0 向量.

本文模型使用 Pytorch 实现,模型训练过程中采用随机梯度下降法^[16]进行参数更新,实验中使用的超参数值如表2所示.

表2 超参数设置

超参数	PEMAN 模型
词向量维度	300
隐藏状态维度	300
Batch size	64
Dropout	0.5
学习率	0.001
L2 正则化参数	10 ⁻⁵

4.3 结果讨论

基线模型和本文模型的实验结果如表3所示. 由表3可知, 本文提出的PEMAN模型在特定目标情感分析任务中, 相对于诸多基线模型均有不同程度提升.

表3 不同模型的情感分类准确率

模型	餐馆数据集	笔记本数据集
TD-LSTM ^[5]	0.756	0.681
ATAE-LSTM ^[6]	0.772	0.687
IAN ^[8]	0.786	0.721
AOA ^[10]	0.797	0.726
MemNet ^[7]	0.803	0.723
PEMAN	0.806	0.749

实验结果表明, 本文提出的PEMAN模型在餐馆数据集和笔记本数据集上的准确率分别达到了80.6%和74.9%, 相比基线模型有了明显提高. PEMAN模型借助句子、短语和目标词表示, 构建两个注意力网络, 有效融合句子上下文语义信息和短语级别特征, 在一定程度上解决了注意力分散等问题, 使得模型的表达能力相对基线模型有一定提升. 针对情感词为多词的情况时, 引入短语特征的PEMAN模型能更准确地挖掘词组语义, 具有更好的理解能力, 有效避免歧义.

另外, 本文选择短语长度为2~5进行对比实验, 来

验证短语长度的取值对PEMAN模型表达能力的影响. 结果如表4所示.

表4 短语长度的取值对本文模型效果的影响

短语长度	餐馆数据集	笔记本数据集
2	0.804	0.743
3	0.806	0.749
4	0.795	0.744
5	0.792	0.735

由表4中结果可知, 当短语长度取3时, PEMAN模型在两个数据集上准确率分别为80.6%和74.9%, 达到了最好的效果. 当短语长度取2时, 模型准确率比大部分基线模型高. 这证明了引入的短语级别特征使得PEMAN模型的表达能力更强. 然而, 当短语长度取4或5时, 模型的准确率有明显下降, 说明当短语长度过长时, 可能导致语义特征过于抽象而降低准确率.

4.4 样本分析

本节通过样本分析来验证PEMAN模型的改进之处, 分析模型准确率提高的原因. 在例句“The appetizers are good and you will not be disappointed if you are a Tequila fan.”中, 包含与餐馆相关的两个目标词: “appetizers”和“Tequila”. 表5展示了AOA^[10]模型、PEMAN模型针对这两个目标词推理得到的情感倾向.

表5 两种模型在该实例中的结果对比

目标词	实际情感倾向	AOA ^[10] 预测结果	PEMAN 预测结果
Appetizer	正向	正向	正向
Tequila	正向	负向	正向

在两种模型中, 针对不同目标词上下文信息的注意力权重分布情况如图2所示. 图中每个格子的颜色表示模型给句中每个词的权重分配情况, 颜色越深代表权重越大.

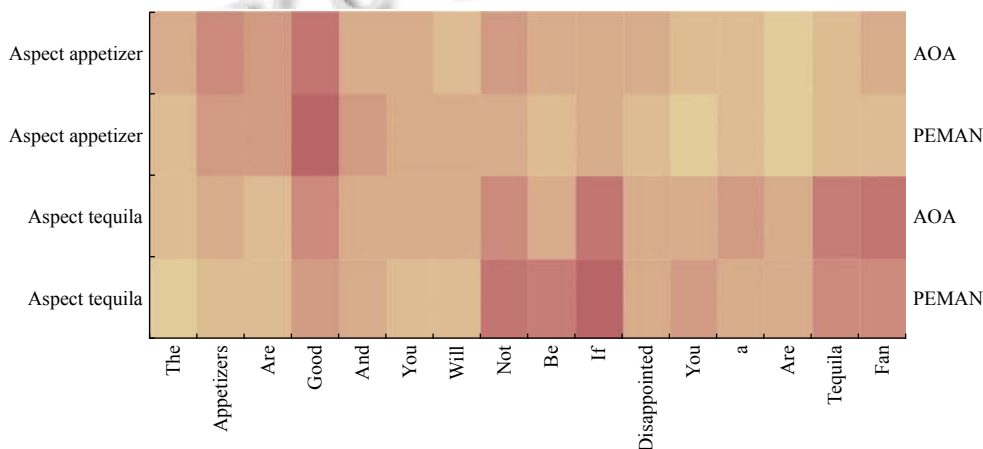


图2 AOA^[10]和PEMAN模型中句子的注意力权重分布

例句中针对“Tequila”这个目标词,表达其情感倾向的是词组“not be disappointed”,表达了正向的情感倾向. AOA^[10]模型对于表示消极倾向的情感词“disappointed”和表示积极倾向的情感词“fan”都有较大权重,然而其对于词组“not be disappointed”整体没有给予足够高的权重分配,所以 AOA 模型对目标词“Tequila”预测了负面的情感倾向,得到了错误的判别结果. 而 PEMAN 模型通过引入短语级别特征语义,更准确地捕捉到词组“not be disappointed”所表达的积极情感倾向,因此能够正确分类. 另外,针对目标词“appetizers”,PEMAN 模型给对应的情感词“good”分配了更高的权重,证明短语级别特征的融合能捕捉到更多的句子语义信息,模型的表达能力更强.

5 总结与展望

特定目标情感分析是一种细粒度的情感分析任务,旨在分析句子中特定目标的情感极性. 本文提出了融合短语特征的多注意力网络 PEMAN 模型. PEMAN 模型通过引入短语级别特征,构建多粒度特征融合的多注意力机制,有效提高表达能力. 实验结果表明,本文提出的 PEMAN 模型在特定目标情感分析任务的准确率有一定提升.

尽管本文的工作相比诸多基线模型有了一定进步,但仍存在一些问题有待探索:(1) 针对特定目标情感分析任务,目前的研究工作在训练过程中同时只能对一个目标词进行计算,未来考虑如何对多个目标词同时进行计算.(2) 针对数据中可能出现的成语或口语化表达,尝试探索如何将先验的语言学知识补充到神经网络模型中,使模型理解能力得到进一步提升.

参考文献

- 1 Hochreiter S, Schmidhuber J. LSTM can solve hard long time lag problems. Proceedings of the 9th International Conference on Neural Information Processing Systems. Cambridge, UK. 1996. 473–479.
- 2 Pontiki M, Galanis D, Pavlopoulos J, *et al.* SemEval-2014 task 4: Aspect based sentiment analysis. Proceedings of the 8th International Workshop on Semantic Evaluation. Dublin, Ireland. 2014. 27–35.
- 3 Vo DT, Zhang Y. Target-dependent twitter sentiment classification with rich automatic features. Proceedings of the 24th International Conference on Artificial Intelligence. Denver, CO, USA. 2015. 1347–1353.
- 4 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.
- 5 Tang DY, Qin B, Feng XC, *et al.* Effective LSTMs for target-dependent sentiment classification. arXiv: 1512.01100, 2015.
- 6 Wang YQ, Huang ML, Zhao L, *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.
- 7 Tang DY, Qin B, Liu T. Aspect level sentiment classification with deep memory network. arXiv: 1605.08900, 2016.
- 8 Ma DH, Li SJ, Zhang XD, *et al.* Interactive attention networks for aspect-level sentiment classification. arXiv: 1709.00893, 2017.
- 9 Chen P, Sun ZQ, Bing LD, *et al.* Recurrent attention network on memory for aspect sentiment analysis. Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark. 2017. 452–461.
- 10 Huang BX, Ou YL, Carley KM. Aspect level sentiment classification with attention-over-attention neural networks. Proceedings of the 11th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. Washington, WA, USA. 2018. 197–206.
- 11 Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 2005, 18(5–6): 602–610.
- 12 Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar. 2014. 1532–1543.
- 13 Deng LY. The cross-entropy method: A unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning. Technometrics, 2006, 48(1): 147–148. [doi: 10.1198/tech.2006.s353]
- 14 Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille, France. 2015. 448–456.
- 15 Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 2014, 15(56): 1929–1958.
- 16 Bottou L. Large-scale machine learning with stochastic gradient descent. Proceedings of the 19th International Conference on Computational Statistics. Paris, France. 2010. 177–186.