

基于 BERT 的端到端旅游评论意见挖掘方法

蔡玉舒^{1,2}, 曹 扬^{2,3}, 江 维¹, 詹瑾瑜^{1,2}, 李 响^{2,3}, 杨 瑞^{2,3}

(1. 电子科技大学 信息与软件工程学院, 四川 成都 610054;

2. 中电科大数据研究院有限公司, 贵州 贵阳 550022;

3. 提升政府治理能力大数据应用技术国家工程实验室, 贵州 贵阳 550022)

摘 要:近年来,随着旅游类互联网产品的兴起,网络上产生了大量针对目的景点的主观评论,使用深度学习算法对相关评论进行意见挖掘,帮助游客快速了解景区特点并为旅游监管提供依据,已然成为一个新的趋势。如何将细粒度意见挖掘方法,如方面级情感分析,应用到旅游评论中,成为一个迫切需要解决的问题。针对上述问题,结合方面级情感分析中意见词抽取和类别分类两个子任务,文中提出了一种针对旅游评论的基于 BERT 的端到端意见挖掘方法。首先利用 BERT 对旅游评论进行编码,再经过下游指针网络解码后对相应的旅游评论进行序列标注,得到<意见词,类别>二元组,以形成完整的观点表达。通过实验与现有的序列标注方法相比,该方法具有更好的挖掘效果。此外,构造了针对中文旅游网站评论的数据集,可用于中文方面级情感分析各个子任务的研究。

关键词:旅游评论意见挖掘;细粒度意见挖掘;BERT;序列标注;指针网络

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2021)09-0118-06

doi:10.3969/j.issn.1673-629X.2021.09.020

End to End Opinion Mining Method Based on BERT for Tourism Comments

CAI Yu-shu^{1,2}, CAO Yang^{2,3}, JIANG Wei¹, ZHAN Jin-yu^{1,2}, LI Xiang^{2,3}, YANG Rui^{2,3}

(1. School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China;

2. CETC Big Data Research Institute Co., Ltd., Guiyang 550022, China;

3. Big Data Application on Improving Government Governance Capabilities National Engineering Laboratory, Guiyang 550022, China)

Abstract: In recent years, with the rise of tourism-related Internet products, a large number of subjective comments on specific attractions have been generated on the Internet. It has become a new trend to use deep learning algorithm to mine the opinions of relevant comments, help tourists quickly understand the characteristics of scenic spots and provide basis for tourism supervision. How to apply fine-grained opinion mining methods, such as aspect-level sentiment analysis, on tourism comments has become an urgent problem to be solved. Therefore, we introduce a BERT-based end-to-end opinion mining method for tourism comments, by combining opinion word extraction and category classification, which are the two subtasks of aspect-level sentiment analysis. First, we use BERT to encode the tourism comments and then decode the context representation by passing through the downstream pointer network to do the sequence labeling, which obtains the binary <opinion word, category> to form a complete opinion expression. The experimental results can demonstrate the efficiency of the proposed method compared with the traditional sequence labeling method. In addition, we conduct a data set for comments on Chinese tourism website, which can be used for the research of various subtasks of aspect-level sentiment analysis in Chinese.

Key words: opinion mining of tourism comments; fine grained opinion mining; BERT; sequence labeling; pointer network

收稿日期:2020-11-29

修回日期:2021-03-30

基金项目:提升政府治理能力大数据应用技术国家工程实验室开放基金项目(W-2019007);国家自然科学基金面上项目(62072076);中科院计算机体系结构国家重点实验室开放课题(CARCH201811);中央高校基本科研业务费(ZYGX2019J078)

作者简介:蔡玉舒(1996-),女,硕士研究生,CCF会员(B3080G),研究方向为软件工程与自然语言处理;江 维,博士,副教授,博导,CCF会员(12777M),研究方向为深度学习、大数据处理;詹瑾瑜,博士,副教授,CCF会员(B3099M),通信作者,研究方向为深度学习、大数据处理。

0 引言

以“强通用性、强交互性、高集智性和高增值性”为特征的数字时代的来临,正在激活文化和旅游资源,为文旅产业发展注入新活力。数字技术在文旅产业的应用,正在赋能公共服务与行业监管,形成新服务与新监管,让智能化管理成为可能。

利用文旅产业运行的各类数据,行政管理部门更容易识别差别化个性化的公共服务需求,有助于提升对文旅产业的公共服务效率,同时也为管理部门的监管提供了技术支撑^[1]。分散于各大旅游网站的游客评论包含了丰富的具有情感倾向的观点表达,可以为相关人员提供参考依据。

网站的游客评论数量庞大,可以借助自然语言处理技术进行信息挖掘。方面级情感分析(aspect-based sentiment analysis, ABSA)是细粒度意见挖掘的一种有效方法。ABSA 旨在确定评论对特定方面(包括方面词语和描述类别)的意见(包括意见词语和情感极性)^[2-4]。例如在评论“峨眉山金顶上的景色不错,但这价格真心贵”中,“景色”是方面词,描述类别是风景,“不错”是描述方面词的意见词,涉及的情感极性是正向(如图1所示)。ABSA 包含了许多子任务,现有的研究大部分集中在给定方面词的前提下确定对应的情感极性(aspect-based sentiment classification, ABSC)^[5-7],但实际应用中问题会稍微复杂一些,主要体现在两个方面:一是在实际情况中,由于标注成本比较高,事先给定的方面词往往无法通过人工标注提供。之前的研究试图借助自动化的方法,使用流水线的模式先抽取方面词之后再行情感分析,但会引入错误传播的问题;二是对爬取的游客评论进行分析后发现,中文的口语表达中有省略方面词的现象,例如评论“方便快捷,及时到位”中包含了对网络购票的积极评价,但并没有明确给出方面词。在文中的数据集中,只包含意见词而方面词缺失的情况约占总数的1/5。如果只对方面词进行抽取和情感分析,将会遗漏掉有效评价信息。

峨眉山金顶上景色不错,但这价格真心贵

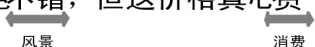


图1 旅游评论意见挖掘示例图

研究者们提出了一些融合词语抽取和分类的多任务方法,试图解决第一个问题。早期的工作^[8-9]尝试验证抽取和分类两个任务之间的关系,并提出更为集成的解决方案,但这些方案的效果并没有优于流水线模型。Li^[10]重新检查了任务,提出了基于两层堆叠LSTM的序列标注模型,其中利用一种简单的门机制来保持标签之间的情感一致性。实验结果证明了两个子任务的相关性,但效果不足以投入到实际应用中。

此后序列标注的方式就成为了融合任务的基本解决思路。BERT 提出之后,率先在 ABSC 领域取得了很好的效果。Li 等人^[11]随后提出了基于 BERT 的序列标注模型,希望用端到端的序列标注方式解决抽取和分类的融合任务,并为之后的基于 BERT 的融合任务研究提供基线。他们利用 BERT 作为嵌入层获得句子表示,并在之上构建了不同的简单解码网络进行了对比实验,结果证明了 BERT 的有效性。但该研究依然以抽取方面词为基础,并没有考虑在实际的中文评论中方面词缺失的情况。同时,在 ABSA 序列标注的问题中,如何保持情感一致性也是解码部分的一个挑战。指针网络最初的设计是简化了 attention 机制,将输出序列对应到输入序列,从而可以适应输入序列长度的变化^[12]。后续应用在机器阅读理解任务(MRC)中的主要方式为设计两个输出长度为1的分类器分别预测开始位置和结束位置,其中1表示句子的长度。对于 ABSA 任务来说,指针网络这种直接预测词语边界的方法,可以解决之前融合任务中情感一致性的问题^[10]。

在这样的背景之下,文中选择融合意见词抽取(opinion words extraction)和描述类别分类两个任务。通过抽取意见词而不是方面词来弥补方面词缺失导致的有效评价信息遗漏的问题,通过类别分类来弥补方面词缺失导致的描述目标不明确的问题,以形成完整的观点表达,并应用在实际的评论分析中。文中提出了一种基于 BERT 的端到端意见挖掘方法,对上述任务进行分析。首先使用预训练的 BERT 作为嵌入层获得编码后的句子表示,再使用指针网络预测句子中每个位置对应的标签。标签对应着是否处于意见词范围内,以及对应描述的类别,将类别描述和意见词拼接之后可以得到观点表达二元组。此外,由于旅游领域没有相关的数据集,爬取了携程、同城和途牛三个网站关于峨眉山和青城山的评论,并参照 semEval-2014^[2]对评论进行了筛选和手工标注。该数据集除了应用在本任务中,还可作为领域数据集参与 ABSA 其他子任务的研究。在构造的数据集上,文中提出的方法的准确率可达到81.9%,相比现有的序列标注经典方法,效果提升可达到2%。

1 系统模型

文中提供一种基于 BERT 的端到端意见挖掘方法,解决意见词抽取及分类的问题,系统的整体分析框架如图2所示。

将输入的评论句表示为一个长度为L的 token 序列 $x = \{x_1, x_2, \dots, x_L\}$ 。首先使用 BERT 作为嵌入层对输入 token 进行编码处理,得到输入 token 的上下文表

示 $T = \{t_1, t_2, \dots, t_L\} \in R^{L \times \dim_t}$, 其中 \dim_t 是经过 transformer 处理后得到的表示向量的维度。其次将上下文表示输入到下游解码部分, 文中使用指针网络构建下游解码网络, 预测整个句子的标记序列 $y = \{y_1, \dots, y_i, \dots, y_L\}$ 。其中 y_i 的可能取值为 $\{B, I, O\}$ - “景区”, “风景”, “票务”, “体验”, “消费”, “服务”, “旅游团”。其中, “-”之前的字母集中, B 代表当前字处于意见词的开头, I 代表了当前字在意见词中, O 代表了不为意见词。“-”之后为描述类别的集合, 是根据标注经验给定的 7 个预定义类别。

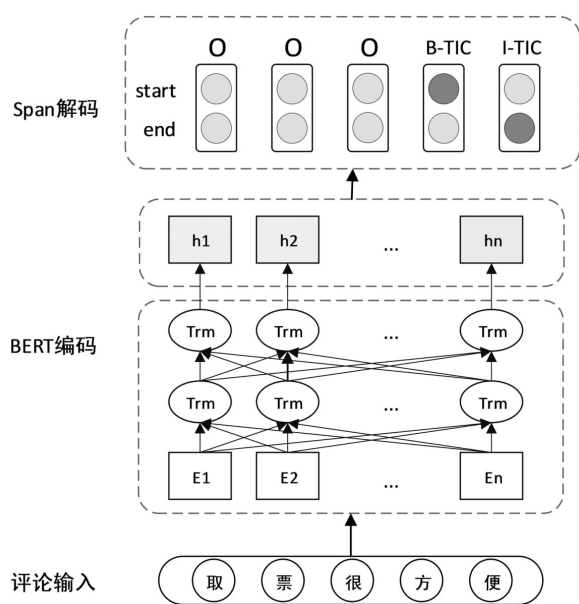


图 2 旅游评论序列标注模型

2 基于 BERT 的端到端旅游评论意见挖掘方法

2.1 编码

文中采取了端到端的方法对输入的旅游评论文本进行处理。其中编码的部分是将输入的中文句 token 转换为可计算的向量形式上下文表示, 用以进行之后的计算。这里采用了预训练的 BERT 模型来编码输入评论的上下文信息。BERT^[13] 是一个基于多层双向 Transformer 的语言表示模型, 相比于传统的 word2vec 编码了更多的上下文和句子信息, 以此增强了表示的能力。首先在输入的部分, 每个单词叠加了三部分的 embedding, 单词 embedding、句子 embedding 以及位置 embedding。之后引入多层 Transformer 对 token 进行优化, 具体的计算方式如下:

$$H^l = \text{Trans}(H^{l-1}) \quad (1)$$

其中, H 表示句子经第 l 层 transformer 处理之后的表示结果, 最后一层的表示计算结果将作为下游任务的输入。

2.2 解码

得到 BERT 层计算的句子表示之后, 文中设计了指针网络为下游任务解决方法。此外, 文中将提出的方法与经典序列标注任务所用解码方式条件随机场 (CRF) 进行对比, 在此也做了简述。

2.2.1 指针网络

(1) 开始结束位置预测。

由于文中的标注任务涉及到多分类的情况, 所以选择使用两个 n 分类的分类器分别预测开始位置和结束位置, 其中 n 表示类别的数量。

$$P_{\text{start}} = \text{softmax}(W^s \cdot H) \in R^{1 \times n} \quad (2)$$

$$P_{\text{end}} = \tanh(W^e \cdot (H; W^s \cdot H)) \in R^{1 \times n} \quad (3)$$

其中, W^s 为计算开始位置的参数矩阵, W^e 为计算结束位置的参数矩阵, H 为 BERT 层输出的句子表示。

(2) 位置匹配。

在一个句子 X 中可能存在多个意见词, 这就意味着可能有多个 start 和 end 标记, 这就涉及到一个位置匹配的问题。因为意见词抽取实际上不存在命名实体识别 (NER) 中的实体嵌套的情况, 所以匹配开始结束位置时使用最近匹配的方式就可以完成。

(3) 训练和测试。

模型构建的过程中, 开始位置的预测结果参与了结束位置的预测, 因而在最后加入了 layer normalization 以降低拟合的难度。两阶段的预测在训练过程中会产生两个 loss, 分别是预测开始位置时的 L_{start} 和预测结束位置时的 L_{end} 。最小化的总体训练目标函数如下:

$$L = \frac{\alpha L_{\text{start}} + \beta L_{\text{end}}}{2} \quad (4)$$

其中, $\alpha, \beta \in [0, 1]$, 为控制 loss 贡献的超参数。这两个 loss 以端到端的方式联合训练。测试时依据分类和抽取是否均一致作为评价标准来计算准确度和召回率。

2.2.2 条件随机场

条件随机场是序列标注任务中常用的解码器, 经常在 NER 等序列标注任务中结合神经网络使用^[14-15]。文中在 BERT 层之上引入一层条件随机场 (CRF) 完成序列标注的任务。

(1) CRF。

CRF 通过引入之前步骤的标注结果进行标签约束, 同样改善了之前的模型中无法保证分类一致性的问题。

$$s(x, y) = \sum_{l=0}^L M^A_{y_l, y_{l+1}} + \sum_{l=1}^L M^P_{l, y_l} \quad (5)$$

其中, s 表示计算的得分, 对这个分数进行 softmax 得到标注序列的概率值; M^A 是随机初始化的转移矩阵,

M^p 是从 BERT 得到的句子表示线性转换而来的发射矩阵。

(2) 训练与测试。

训练时计算 CRF 层的 loss, 并进行优化即可。测试时同样依据的是分类和抽取是否均一致作为评价标准, 实现了对比<类别, 开始位置, 结束位置>三元组的计算方法, 在计算 F1-score 时采用的是 micro-f1。

3 实验分析

3.1 实验设置

本实验的硬件环境: CPU 为 Intel i7 9700K, 内存为 16G DDR4, 显卡为两块 Nvidia RTX 2080ti, 运行环境为 Linux 操作系统 (Ubuntu 16.04.6)。软件编程语言为 python。

本实验使用的预训练模型是哈工大开源的中文预训练 BERT 模型“Chinese-bert-wwm”, 其中 transformer 的层数为 $L=12$, 隐藏层维度为 768, 学习率为 $2e-6$, batch size 为 8。

3.2 数据集构建

3.2.1 数据采集

本实验中使用的景区评论数据均来自旅游网站的公开评论。通过爬虫采集来自三个旅游网站对于峨眉山和青城山景区的公开评论, 均发表于 2019 年。同时对爬取的原始评论进行了去重和筛选的操作, 去除了无效的评论 (如从携程网站爬取的评论中存在系统默认评论, “用户未点评, 系统默认好评”, 以及毫无含义的评论), 并将评论长度限制在 200 个字之内。

3.2.2 数据标注方法

数据标注参考了英文数据集中 SemEval-2014 提供的标注方法, 将标注分为方面词、意见词、情感分类以及类别四个维度。

关于方面词以及意见词的划分, 由两位专业人员分别对所有的句子独立进行标注, 再讨论并消除有分歧的地方。考虑到 ABSA 其他子任务的情况, 标注方式为, 以方面词为单位标记<方面词, 意见词, 类别, 情感>四元组, 同时提供词语 (包含方面词和意见词) 开始和结束位置标记。一个评论句中可能会有多组标记四元组。我们仅保留具有有效意见词的四元组, 方面词缺失标记为“_”。

关于类别划分, 根据标注经验总结出 7 个预定义的类别, 分别是: 景区、风景、票务、体验、消费、服务以及旅游团。其中“景区”这一类别包含了一些描述范围模糊以及较难分类的情况, 和 SemEval-2014 中的 Others 接近。

3.2.3 数据集分析

标注完成的数据集共有 1 746 条独立的评论, 共

3 155 个标记四元组。其中方面词缺失的四元组约占 20%。情感极性以及类别标记情况如表 1 所示。

表 1 数据标记分析

类别	标记数量
景区	717
旅游团	302
体验	255
服务	192
风景	598
票务	683
消费	408
正向	2 126
负向	1 029

在整个细粒度情感分析的任务中, 文中所解决的问题只是其中一个子任务。构造的标注数据集不仅可以用于当前任务, 也可以用于探索 ABSA 的其他子任务, 因而采取了更为完整的四元组标注方式, 希望可以为之后的研究者提供中文语料。

3.3 模型性能评估

3.3.1 意见挖掘模型中分类任务评估与分析

由于在旅游领域中类别较多, 为了进一步了解模型在分类上的学习情况, 文中对比了同一模型在不同类别上的表现, 结果如表 2 所示, 可以看到同一模型在不同类别上的效果也存在差异。

表 2 模型在不同类别上的表现

类别	acc	recall	F1-score
景区	0.704 9	0.682 5	0.693 5
旅游团	0.750 0	0.576 9	0.652 2
体验	0.846 2	0.733 3	0.785 7
服务	0.800 0	0.923 1	0.857 1
风景	0.890 9	0.765 6	0.823 5
票务	0.824 3	0.924 2	0.871 4
消费	0.961 5	0.806 5	0.877 2

模型在“消费”和“票务”两个类别上表现较好, 得分较低三个分类分别为“景区”、“旅游团”和“体验”。总体而言, 标记数据数量越多, 在模型上的效果越好。“服务”和“景区”两个类别比较特殊, “服务”类别的标记数量最少, 但得分较高, “景区”类别的标记数量最多, 反而得分比较低。观察了测试结果之后发现, “服务”类别虽然数量较少, 但意见词和上下文语境相对比较集中。而“景区”的意见词和上下文语境较为分散, 包含了无法准确分类的部分意见组合, 特征相对其他类别并不明显, 所以各项得分均处于低位。其中意见词和上下文语境特殊性的作用在其他类别的测试结果中也有所体现。

比如最频繁出现的意见词“好”和“棒”,由于指定性不强,会存在被错误分类成数据量更大的上下文类别的情况。但“体验”类别中“不虚此行”和“消费”类别中的“贵”这样比较特殊的意见词,抽取和分类的准确率都更高。

3.3.2 旅游评论意见挖掘模型评估与分析

本节将提出方法与现有方法相比较,使用准确率

表 3 不同模型效果比较

模型	Precision	Recall	F1-score
BiLSTM+CRF	0.791 0	0.768 1	0.779 4
BERT+CRF	0.778 5	0.787 0	0.782 7
BERT+Span(文中)	0.819 5	0.774 7	0.796 5

设计了两组对比实验进行评估。首先将经典序列标注模型 BiLSTM 结合 CRF 作为基线,由图 3 可以看出,同样以 CRF 作为解码结构,即使在领域语料规模较小的情况下,使用 BERT 作为编码模块的效果依然

(precision)、召回率(recall)和 F1-score 指标评估模型性能。

如表 3 所示,在使用构建的旅游数据集的情况下,文中提出的基于 BERT 和指针网络的方法在测试集上准确率可达到 81.9%,F1-score 可达到 79.6%。

要好于结构相对简单的 BiLSTM。表现在最终的 F1 值上,以 BERT 为编码模块的模型得分高出将近 0.3%,这证明了 BERT 作为编码模块相比 word2vec 与 BiLSTM 编码可以生成更有效的句子表示。

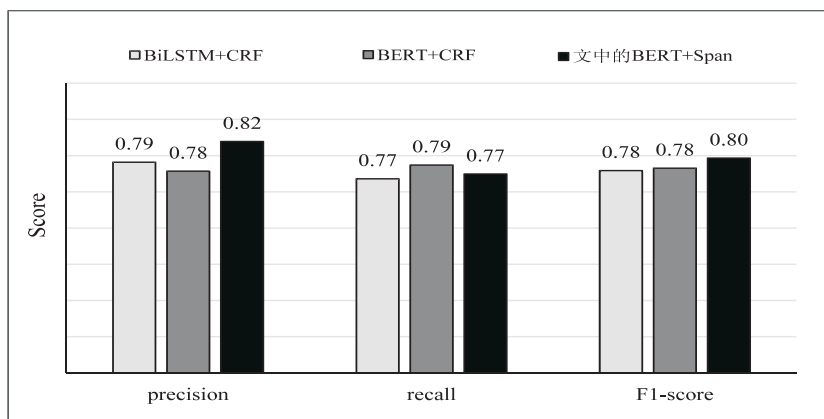


图 3 旅游评论意见挖掘不同模型比较

在解码方式的对比上,本实验设计了序列标注任务中经典的解码方式 CRF,与提出的指针网络进行对比。从图中可以看出,从实验最终的效果来看,指针网络模型最终的 F1 得分要稍高于 CRF,高出约 1%。同时,在训练的过程中,观察最终不同类别各自的得分,CRF 会出现由于训练数据不够充分,导致得分偏高的情况,而指针网络相比而言则要更加稳定。

由上述分析可知,文中提出的方法相比于经典序列标注方法,在解码上使用 BERT 生成了更为有效的句子表示。在编码上,以指针网络作为解码网络相比 CRF,在准确率和 F1-score 上均有提升。

4 结束语

随着旅游类互联网产品的兴起,网络上产生了大量针对目的景点的主观评论,通过对评论进行意见挖掘可为旅游监管提供重要依据。在此背景下,文中结合方面级情感分析中意见词抽取和类别分类两个子任务,提出了基于 BERT 的端到端评论意见挖掘模型,得

到<意见词,类别>二元组,以形成完整的观点表达,并应用于实际的评论分析中。和现有的序列标注模型相比,文中提出的模型具备更好的性能。此外,还构建了中文旅游评论数据集,供此后相关任务的研究。

目前该研究还有可以改进的地方。比如,由于方面级情感分析领域数据标记成本较高的原因,标记数据的数量往往非常少,文中提出的数据集规模也并不大,在这种情况下模型受数据约束的影响比较大。在实验中,BERT 由于复杂度较高,在较小规模的领域数据集上,整体效果相比 BiLSTM 难以有很大的提升。利用已有的领域数据集进行知识迁移可能是一个比较有效的解决小数据量问题的思路。因此在未来的研究中,可考虑使用知识迁移的方法,解决方面级情感分析中领域数据缺少的问题。

参考文献:

- [1] 夏杰长,徐金海.以数字化推动文旅产业融合发展[EB/OL].(2020-03-31).<http://www.xinhuanet.com/tech/>

- 2020-03/31/c_1125791048.htm.
- [2] PONTIKI M, GALANIS D, PAVLOPOULOS J, et al. SemEval-2014 task 4: aspect based sentiment analysis[C]//Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). Dublin, Ireland: ACL, 2014: 27-35.
 - [3] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2015 task 12: aspect based sentiment analysis[C]//Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015). Denver, Colorado: ACL, 2015: 486-495.
 - [4] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2016 task 5: aspect based sentiment analysis[C]//Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016). San Diego, California, USA: ACL, 2016: 19-30.
 - [5] TANG Duyu, QIN Bing, LIU Ting. Aspect level sentiment classification with deep memory network[C]//Conference on empirical methods in natural language processing (EMNLP). Austin, Texas, USA: [s. n.], 2016: 214-224.
 - [6] XU Hu, LIU Bing, SHU Lei, et al. BERT post-training for review reading comprehension and aspect-based sentiment analysis[C]//Proceedings of the 2019 conference of the North American Chapter of the association for computational linguistics: human language technologies (NAACL-HLT). Minneapolis, Minnesota: [s. n.], 2019: 2324-2335.
 - [7] LI Zheng, WEI Ying, ZHANG Yu, et al. Exploiting coarse-to-fine task transfer for aspect-level sentiment classification[C]//The thirty-third AAAI conference on artificial intelligence (AAAI-19). Honolulu, Hawaii, USA: AAAI, 2019: 4253-4260.
 - [8] MITCHELL M, AGUILAR J, WILSON T, et al. Open domain targeted sentiment[C]//Proceedings of the 2013 conference on empirical methods in natural language processing. Seattle, Washington, USA: [s. n.], 2013: 1643-1654.
 - [9] ZHANG M, ZHANG Y, VO D T. Neural networks for open domain targeted sentiment[C]//Proceedings of the 2015 conference on empirical methods in natural language processing (EMNLP). Lisbon, Portugal: [s. n.], 2015: 612-621.
 - [10] LI Xin, BING Lidong, LI Piji, et al. A unified model for opinion target extraction and target sentiment prediction[C]//The thirty-third AAAI conference on artificial intelligence (AAAI-19). Honolulu, Hawaii, USA: AAAI, 2019: 6714-6721.
 - [11] LI Xin, BING Lidong, ZHANG Wenxuan, et al. Exploiting BERT for end-to-end aspect-based sentiment analysis[C]//Proceedings of the 2019 EMNLP workshop W-NUT: the 5th workshop on noisy user-generated text. Hong Kong, China: [s. n.], 2019: 34-41.
 - [12] VINYALS O, FORTUNATO M, JAITLY N. Pointer networks[C]//Proceedings of the 28th international conference on neural information processing system (NIPS). Montreal, Quebec, Canada: [s. n.], 2015: 2692-2700.
 - [13] DEVLIN J, CHANG Mingwei, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 conference of the North American Chapter of the association for computational linguistics: human language technologies (NAACL-HLT). Minneapolis, Minnesota: [s. n.], 2019: 4171-4186.
 - [14] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition[C]//Proceedings of the 2016 conference of the North American Chapter of the association for computational linguistics: human language technologies (NAACL-HLT). San Diego, California, USA: [s. n.], 2016: 260-270.
 - [15] MA Xuezhe, HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF[C]//Proceedings of the 54th annual meeting of the association for computational linguistics (ACL). Berlin, Germany: ACL, 2016: 1064-1074.