

Multi-level emotion classification based on BERT

Ruoxin Liao

School of Computer and Artificial Intelligence, Southwest Jiaotong University, Chengdu, Sichuan, 611756, China

Abstract

Aspect-level sentiment analysis hinges on accurately establishing the association between aspect terms and their corresponding opinion words. However, existing methods often fail to fully exploit contextual sentiment knowledge and capture deep semantic features. To address these limitations, this paper proposes a BERT-based multi-feature fusion model. First, a sentiment lexicon is utilized to extract sentiment information from the context, and sentiment knowledge weighting is applied to enhance syntactic dependency relations, thereby constructing a syntactic sentiment graph convolutional network. Second, a dual-layer attention mechanism is designed, integrating self-attention and interactive attention to further capture intra-sentence semantic information and reinforce the interaction between aspect terms and the surrounding context. Additionally, deep features from the second-layer attention mechanism are extracted to improve the accuracy of sentiment polarity prediction. By dynamically integrating syntactic sentiment features, attention-based features, and global features, the proposed model constructs a more comprehensive sentiment representation. Experimental results on three publicly available datasets demonstrate the effectiveness of the proposed approach.

Keywords

Aspect-based Sentiment Analysis; Graph Convolutional Network; Attention Mechanism; Sentiment Knowledge; Multi-Feature Fusion

基于 BERT 的多特征融合方面级情感分类方法

廖若欣

西南交通大学计算机与人工智能学院, 中国·四川成都 611756

摘要

方面级情感分析, 关键在于精准构建方面词与相应意见词间的关联。目前, 现有方法难以充分利用上下文情感知识、挖掘深度语义特征, 为此提出一种基于BERT的多特征融合模型。首先, 利用情感词典挖掘上下文中的情感信息, 并通过情感知识加权增强句法依存关系, 构建句法情感图卷积网络。其次, 设计了一种双层注意力机制, 结合自注意力和交互注意力, 以进一步挖掘句子内部的语义信息, 并强化方面词与上下文之间的交互关系。同时, 进一步提取第二层注意力的深层次特征, 以提升情感极性预测的准确性。该模型通过动态融合句法情感特征、注意力特征和全局特征, 构建更完整的情感表示。3个公开数据集上的实验结果表明, 提出的模型是有效的。

关键词

方面级情感分析; 图卷积网络; 注意力机制; 情感知识; 多特征融合

1 引言

互联网的发展极大改变了信息获取方式, 人们乐于在网络平台表达观点, 这类文本常包含丰富情感。传统情感分析任务聚焦于文档级或句子级的情感极性判定^[1], 但评论往往涉及多个维度, 单纯靠整体情感分析难以满足实际需求。在此背景下, 方面级情感分析 (Aspect-Based Sentiment Analysis, ABSA) 应运而生, 旨在针对句中的特定方面进行情绪极性判定, 而不是对整个句子或篇章进行粗略的整体分析, 提供更细致的情感洞察。

本文针对 ABSA 任务, 提出基于 BERT 的多特征融合

网络模型 MFF-BERT (Multi-Feature Fusion Network based on BERT pretraining model)。主要贡献如下: ①引入外部情感知识加权增强句法依存关系, 通过句法情感图卷积网络提取句法情感信息; ②设计双层注意力机制, 结合自注意力和交互注意力深入挖掘句子内部的语义信息, 强化方面词与上下文之间的交互, 通过第二层注意力提取更深层次的特征; ③在 3 个公开数据集上进行对比实验, 验证模型有效性。

2 相关工作

在 ABSA 的早期研究进程中, 基于情感词典和基于机器学习的方式率先展开探索。基于情感词典的方法实现简单、易于理解, 然而该方法通常依赖于情感词典的搭建。与传统情感词典方法相比, 机器学习方法性能优越, 避免了对情感词典的过度依赖, 但仍需大规模标注数据集和精确的特

【作者简介】廖若欣 (2000-), 女, 中国四川成都人, 在读硕士, 从事自然语言处理研究。

征工程。

伴随深度学习持续发展, 深度神经网络模型凭借其无需繁杂特征工程而备受关注。近几年, 基于图卷积网络 (Graph Convolutional Network, GCN) 和基于注意力机制的模型在 ABSA 研究中得到广泛应用。这两种深度学习方法表现良好, 但仍存在问题。基于 GCN 的方法通常使用句法依存图构建图结构, 但忽略了情感信息, 导致模型在捕捉情感相关句法依赖时表现有限。基于注意力的模型能动态分配全局权重, 但多数仅关注方面词与上下文词的交互, 未能充分捕捉深层语义关联, 也缺乏显式的句法结构约束。因此, 本文提出 MF-F-BERT 模型, 结合句法情感特征、注意力特征和整体特征进行情感极性预测。

3 本文模型

3.1 模型框架

本文提出的 MF-F-BERT 结构如图 1 所示。首先利用 BERT 提取上下文和方面词的词向量, 构建句法依存图。随后, 通过外部情感知识加权句法依存图, 增强情感知识建模。紧接着, 采用图卷积网络对句法情感图进行编码, 以获取丰富的句法情感信息。同时, 方面词和上下文词向量通过双层注意力机制进行深度交互, 既捕捉上下文内部信息, 又提取方面词与上下文的交互特征。通过动态融合策略整合句法情感特征、整体语义特征和注意力特征, 最终用于情感分类。

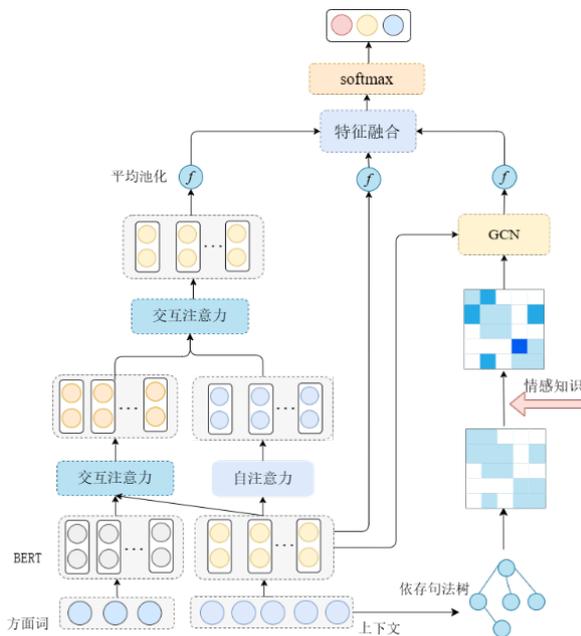


图 1 MF-F-BERT 模型框架

3.2 嵌入层

在嵌入层, 本文采用 BERT 模型。将文本序列输入 BERT 之前, 预处理是一项不可或缺的步骤, 以满足 BERT 的输入格式要求。最终生成上下文的全局表示、方面词的局部表示, 其具体形式如下:

$$H^S = BERT([CLS], Sentence, [SEP], Aspect, [SEP]) \quad (1)$$

$$H^A = BERT([CLS], Aspect, [SEP]) \quad (2)$$

其中, “[CLS]” 用于标志句子序列的开始位置, “[SEP]” 是不同句子间的分割符号。

3.3 句法情感图卷积网络

3.3.1 句法依赖图

在 ABSA 任务中, 方面词的情感通常由上下文中的某些关键词决定, 因此捕捉句子的句法依存关系尤为重要^[2]。因此, 本文以句法依存树为基础, 为输入句子构建句法依赖图。对于评论文本 $S = \{w_1, w_2, \dots, w_n\}$ 构建过程如下:

$$D_{ij} = \begin{cases} 1, & i = j \text{ 或 } w_i \text{ 和 } w_j \text{ 存在依赖关系} \\ 0, & \text{其他} \end{cases} \quad (3)$$

3.3.2 情感知识图

情感知识在情感分析任务中通常用于增强情感特征的表达, 许多学者已证明情感知识对情感分类有着积极作用。为突出单词间的情感依赖关系, 本文引入 SenticNet5 情感词典, 帮助模型更深入地推理文本情感信息并提升情感识别能力。SenticNet 5 为每个单词提供了情感得分, 范围在 -1 到 1 之间。情感越积极得分越接近 1, 越消极得分越接近 -1。

构建情感知识图 S 方式如下:

$$S_{ij} = Score(w_i) + Score(w_j) \quad (4)$$

其中, $Score(w_i)$ 表示单词 w_i 在 SenticNet5 中的情感得分。

3.3.3 句法情感知识图

为突出对特定方面词的关注度, 引入一个方面增强图 T 。

$$T_{ij} = \begin{cases} 1, & \text{如果 } w_i \text{ 或 } w_j \text{ 是方面词} \\ 0, & \text{其他} \end{cases} \quad (5)$$

将前述得到的三个图进行融合, 构建句法情感知识图 R , 该图综合了句法信息、外部情感知识以及方面词自身对情感分类的影响。

$$R_{ij} = D_{ij}(S_{ij} + T_{ij} + 1) \quad (6)$$

3.3.4 多层图卷积

在得到句法情感知识图后, 利用图卷积网络对上下文进行特征提取。GCN 采用逐层传播机制, 每层隐藏状态是基于上一层输出更新, 实现信息逐步聚合与传播。GCN 通过利用邻接节点的信息来更新当前节点的隐藏状态, 更新方式如下:

$$h_j^l = ReLU(\sum_{i=1}^n R_{ij} W^l g_j^{l-1} + b^l) \quad (7)$$

其中 g_j^{l-1} 为节点的层输入; h_j^l 为节点 j 的 l 层输出; W^l 和 b^l 是线性变换权重和偏置项。

为增强模型对上下文词与方面词之间的位置信息的建模能力, 本文对隐藏状态首先进行位置加权处理。以往研究多采用线性加权, 可能导致权重分配不够平滑。本文使用高斯分布加权, 如公式 (8)、(9)、(10) 所示。高斯分布加权通过更平滑的权重分配和自适应的权重范围, 更好地捕捉方面

词及其上下文的关系减少远距离词的干扰。

$$g_j^{l-1} = F(h_j^{l-1}) \quad (8)$$

$$F(h_i) = p_i \cdot h_i \quad (9)$$

$$idx_c = \frac{idx_s + idx_e}{2} \quad (10)$$

$$p_i = \exp\left(-\frac{(i - idx_c)^2}{2len_{aspect}^2}\right) \quad (11)$$

其中, $F(\cdot)$ 是位置权重函数, p_i 是第 i 个词位置权重, idx_s 代表方面词的中心位置, idx_s 和 idx_e 分别是方面词开始和结束索引。

为了更好地学习全局上下文信息, 多层 GCN 以 BERT 生成的隐藏向量 H^S 作为初始节点表示, 即 $H^0 = H^S$ 。第 l 层 GCN 的输出表示如下:

$$H^l = \{h_1^l, h_2^l, \dots, h_n^l\} \quad (12)$$

最终, GCN 最后一层输出 H^G 被用于后续的特征融合。

3.4 双层注意力层

受 Shazeer^[3] 启发, 本层采用一种注意力机制变体, 称为交谈注意力机制 (Talking-Heads Attention)。交谈注意力机制在传统多头注意力机制基础上, 引入两个额外线性投影。logits 投影在 Softmax 操作前对注意力分数进行投影, 允许不同头之间的信息交互。weights 投影则在 Softmax 之后对注意力权重进行投影, 进一步调整每个头的权重分布。计算包括查询 Q、关键字 K 和值 V, 具体计算过程如下:

$$THA(Q, K, V) = \text{Concat}(h_{ead_1}, h_{ead_2}, \dots, h_{ead_n})W^o \quad (13)$$

$$Head_i = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}P_l\right)P_wV \quad (14)$$

其中, d_k 为键向量维度; P_l 为 logits 投影矩阵; P_w 为 weights 投影矩阵; W^o 为输出投影矩阵。

在 MFF-BERT 中, 引入双层注意力机制。第一层交互式注意力用于捕捉上下文和方面词之间的交互信息; 第一层自注意力计算上下文中各词之间的相关性, 增强模型对关键信息的关注能力。最终获得交互特征表示 H^{IA} , 自注意力的特征表示 H^{SA} , 计算如下:

$$H^{IA} = THA(H^A, H^S, H^S) \quad (15)$$

$$H^{SA} = THA(H^S, H^S, H^S) \quad (16)$$

第二层交互注意力进一步增强特征表达能力, 形成一个更加丰富和深入的语义理解 H^{FA} , 计算方式如下:

$$H^{FA} = THA(H^{IA}, H^{SA}, H^{SA}) \quad (17)$$

3.5 多特征融合

在进行特征融合之前, 对双层注意层提取语义特征 H^{FA} 、GCN 层的最终提取特征 H^G 和整体上下文编码 H^S 进

行平均池化处理, 以降低特征空间的维度, 保留关键特征信息。

$$H^{Sem} = \text{average}(H^{FA}) \quad (18)$$

$$H^{GCN} = \text{average}(H^G) \quad (19)$$

$$H^{Global} = \text{average}(H^S) \quad (20)$$

通过融合不同层的特征表示, 可以充分发挥各类特征之间的互补性。在面对结构复杂、语义模糊或者情感表达隐晦的样本时, 不同特征能够从各自的独特角度提取信息, 弥补单一特征的局限性。特征融合的计算过程如下:

$$g_i = \sigma(W_g[H^{Sem}; H^{GCN}; H^{Global}] + b_g) \quad (21)$$

$$H^F = g_1H^{Sem} + g_2H^{GCN} + g_3H^{Global} \quad (22)$$

其中 H^F 为最终表征, g_1, g_2, g_3 是可学习的权重参数, σ 表示 sigmoid 激活函数, W_g 是可学习的权重矩阵, b_g 为偏置项。

3.6 输出层

输出层利用全连接层和 Softmax 归一化, 获得模型的最最终分类概率分布 p , 计算方式如下:

$$p = \text{Softmax}(W_pH^F + b_p) \quad (23)$$

其中, W_p 为可学习的参数矩阵, b_p 为偏置项。

模型参数采用标准梯度下降法进行优化, 旨在最小化带 L2 正则化的交叉熵损失, 该损失函数定义如下:

$$\text{Loss} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m (y_{ij} \log \hat{y}_{ij}) + \lambda \|\theta\|^2 \quad (24)$$

其中, n 为样本数量; m 为类别数量; y_{ij} 为样本 i 实际标签值; \hat{y}_{ij} 是模型对样本 i 属于类别 j 的预测概率; $\lambda \|\theta\|^2$ 表示 L2 正则化项。

4 实验结果与分析

4.1 实验数据集

为验证 MFF-BERT 有效性, 选取 Twitter 和 Restaurant、Laptop 公开数据集进行实验。Restaurant 数据集由餐馆相关的用户评论构成; Laptop 数据集为笔记本电脑评价; Twitter 数据集为推文数据。这些数据集的每条评论都标注了方面词及其极性。

4.2 实验参数设置与评价指标

实验中, BERT 词嵌入维度设为 768, 采用 Adam 作为优化器, 学习率为 $2e-5$, 批处理大小为 16, GCN 层数为 2, 最大序列长度设定为 85。同时, 采用准确率和 F1 值作为评价指标。

4.3 对比实验

为了全面评价 MFF-BERT 表现效能, 选取以下基准模型:

ATAE-LSTM: 将方面词嵌入向量和上下文嵌入向量相连接, 通过注意力机制加权输入的词向量。

AOA: 利用“上下文注意力”聚焦于输入文本的关键信息,通过“方面注意力”将情感信息与目标方面对齐。

ASGCN: 首次引入特定于方面的 GCN,充分利用句子的依存关系进行情感分类。

BiGCN: 从语法层次和词汇层次捕捉文本中的情感信息,采用图卷积网络对这些层次图和词汇图进行卷积操作。

InterGCN: 针对方面词设计方面间邻接矩阵,提供方面之间的关系信息,进而联合上下文邻接矩阵更全面理解文本中的情感信息。

SK-GCN: 通过图卷积网络结合句法信息和知识图谱,利用语法结构和外部知识增强文本表示。

T-MGAN: 通过对文本进行词粒度、短语粒度等多粒度分析,捕捉不同层次的情感信息。

BERT-SPC: Google AI 团队研发的预训练语言模型,用于各种文本处理任务。

AEN-BERT: 通过 BERT 学习文本的上下文信息,并利用双重注意力机制关注输入文本中的关键词以及聚焦于目标情感与上下文之间的关联。

BERT4GCN: 通过提取 BERT 模型的中间层信息,捕捉深层语义特征,利用 GCN 建模词汇间的关系和上下文信息,融合 BERT 的语言理解能力与 GCN 的结构建模优势。

T-GCN-BERT: 引入类型感知机制,同时在 GCN 结构的多个层次上进行特征提取,并采用层集成方法,使得不同层次的信息可以互相补充。

表 1 呈现了实验结果。

由表 3 可知:

① MFF-BERT 在 Restaurant 和 Twitter 数据集上准确率和 F1 值均最佳,在 Laptop 数据集上,准确率略低于 T-GCN-BERT 和 AEN-BERT,但 F1 值接近最优。

② BERT 预训练模型的引入显著提升分类性能。非 BERT 模型主要依赖静态词向量,难以充分捕捉上下文信息,而 BERT 通过动态词向量生成机制,有效提升了方面级情感分析的词表示能力和分类效果。

③句法结构信息与注意力机制结合有效提升分类效果。对比仅依赖注意力机制的模型,GCN 类模型通过依存句法结构更好地捕捉方面词与上下文关系,提升分类性能。

④ MFF-BERT 通过多源信息融合提高对复杂句子的理解。与传统 GCN 模型不同,MFF-BERT 结合全局特征、句

法情感特征和注意力特征,增强了模型对多方面词的感知能力。

表 1 MFF-BERT 与基线模型对比结果 (%)

模型	Restaurant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
ATAE-LSTM	77.32	66.57	69.14	63.18	69.65	67.40
AOA	79.97	70.42	72.62	67.52	72.30	70.20
ASGCN	80.77	72.02	75.55	71.87	72.15	70.40
BiGCN	81.97	73.48	74.49	71.84	74.16	73.35
InterGCN	82.23	74.01	77.86	74.32	-	-
SK-GCN	80.54	70.14	73.04	68.41	71.82	69.56
T-MGAN	82.06	72.65	76.38	73.02	71.23	70.63
BERT-SPC	84.11	76.68	77.59	75.03	75.92	75.18
AEN-BERT	83.12	73.76	79.93	76.31	74.54	73.26
BERT4GCN	84.75	77.11	77.49	73.01	74.73	73.76
T-GCN-BERT	86.16	79.95	80.88	77.03	76.45	75.25
Our MFF-BERT	86.27	79.99	79.36	75.52	76.97	75.88

5 结语

本文提出了一种基于多特征融合的方面级情感分析模型(MFF-BERT),通过融合句法结构、情感知识和语义信息,提升模型对方面级情感的理解能力。模型采用双层注意力机制来建模方面词与上下文的关系,并强化句子内部的语义关联。此外,结合外部情感知识的句法增强图卷积网络在捕捉句法依赖的同时提升情感理解。最后,设计动态特征融合机制,融合语义、句法情感和全局特征,优化模型性能。实验结果表明,MFF-BERT 在三个公开数据集上超越了大多数现有模型。

参考文献

- [1] 汪红松,李嘉展,曾碧卿.文本方面级情感分析方法研究综述[J].软件导刊,2023,22(09):1-8.
- [2] Zhang C, Li Q, Song D. Aspect-based sentiment classification with aspect-specific graph convolutional networks[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019: 4568-4578
- [3] Shazeer N, Lan Z, Cheng Y, et al. Talking-heads attention[J]. arXiv preprint arXiv:2003.02436, 2020.