

基于分层树结构注意力的情感分析模型

罗湔元^{1,2}, 李平杰^{1,2}

¹中交四航工程研究院有限公司, 广东 广州

²中交交通基础工程环保与安全重点实验室, 广东 广州

收稿日期: 2025年4月25日; 录用日期: 2025年5月23日; 发布日期: 2025年5月30日

摘要

现有情感分类模型多采用单一方面词与文本单层粒度的简单匹配策略, 或依赖注意力机制实现特征抽取, 此类方法普遍存在文本结构信息利用率不足、特征筛选针对性弱的缺陷, 导致分析结果粗放化且模型泛化性能受限。针对该问题, 本文提出一种分层树结构注意力模型, 模型首先将已经特征编码后的字向量通过自编码器组合成: 词汇、子句、句子三种粒度的语料, 然后利用树状图神经网络对其进行多层递进的权重分配, 再通过最长路径算法筛选出关键节点, 最后使用由归一化层和卷积层构成的分类网络获得分类结果。经过实验模型在公开数据集上的准确度和F1分数比同类型模型拥有更好的表现。

关键词

深度学习, 方面级情感分析, 图注意力, 树图

A Sentiment Analysis Model Based on Hierarchical Tree-Structured Attention

Haoyuan Luo^{1,2}, Pingjie Li^{1,2}

¹CCCC Fourth Harbor Engineering Institute Co., Ltd., Guangzhou Guangdong

²CCCC Key Lab of Environmental Protection & Safety in Foundation Engineering of Transportation, Guangzhou Guangdong

Received: Apr. 25th, 2025; accepted: May 23rd, 2025; published: May 30th, 2025

Abstract

Existing sentiment classification models mostly use a simple matching strategy of single aspect word and text single-layer granularity, or rely on the attention mechanism to achieve feature extraction. Such methods generally have the defects in insufficient utilization of text structure information and weak targeting of feature screening, resulting in sloppy analysis results and limited model generalization performance. To address this problem, this paper proposes a hierarchical tree-structured

attention model, which first synthesizes the word vectors that have been feature-encoded into three granularity corpus: vocabulary, clauses, and sentences by self-coding group, then uses a tree-graph neural network to assign weights to them in multiple layers progressively, and then filters out the key nodes by the longest-path algorithm, and finally obtains the classification results by using the classification network consisting of the normalization layer and convolutional layer. classification results. The experimental model has better accuracy and F1 score than the same type of model on the public dataset.

Keywords

Deep Learning, Aspect-Level Sentiment Analysis, Graph Attention, Trees

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

情感分析任务旨在分析人们日常话语和文本的情感极性,通过对文本的建模分析,为控制网络舆情、网民情绪、危机事件等提供基础帮助[1],早期基于统计学的情感分析模型是通过将文本中含有明显情感极性的词汇进行概率统计分析,得到文章单一的情感极性[2]。然而在日常交流和阅读过程中,真实的文本往往是含有不止一种情感极性表现[3]。

图注意力机制的引入,使模型能够基于图卷积运算实现词法、句法层面的针对性特征提取,从而使模型能够动态的关注不同的特征信息[4]。

尽管注意力机制能够有效捕捉局部语义关联,但其在跨层级语义依赖场景中常因非相关特征干扰导致聚合偏差,造成模型判别精度下降、训练收敛效率低及抗干扰能力不足等问题。受结构化数据建模思想启发[5],本研究提出一种基于三级语义树的注意力传导方法,通过构建词汇-子句-语句的层级特征依赖图,设计递进式权重分配策略,实现不同粒度语义单元的深度交互。该方法借助文本自身的多级语义结构,在注意力计算过程中建立特征间的因果传导关系,如由“口感”关联“醇厚”,再由“醇厚”指向“陈年红酒”,有效丰富复杂语境下的语义表征。

2. 相关工作

2.1. 模型

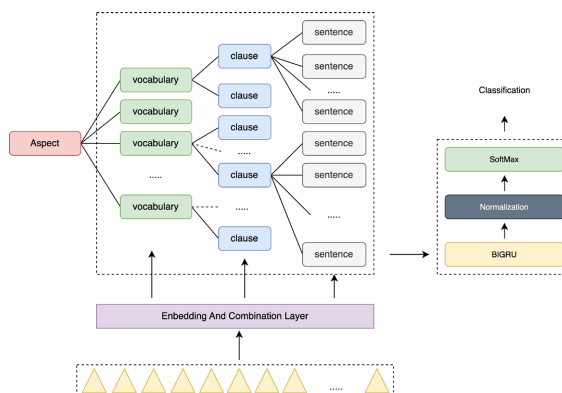


Figure 1. Tree-structured attention network
图 1. TAN 模型

本文提出 TAN (Tree-structured Attention Network, 分层树结构注意力模型)模型如图 1 所示, 主要包括语料组合层、树状图注意力层等模块。

2.2. 多维度语义建构模块

2.2.1. 编码层

对字粒度文本进行 CBOW 预训练编码, 基于文本句法结构将字级嵌入向量聚合成词汇级、子句级、语句级三类语义单元, 构建多维度特征颗粒体系, 如图 2:

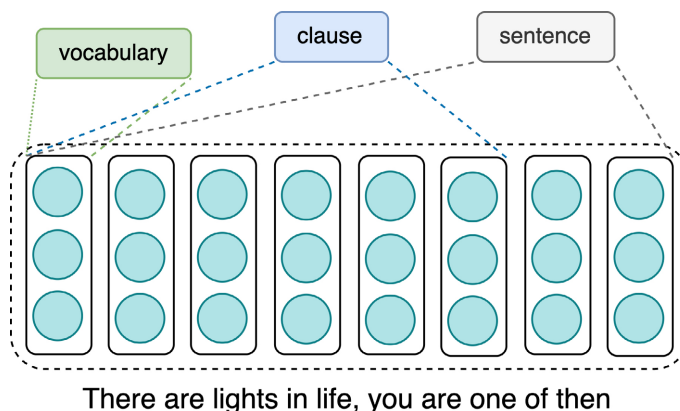


Figure 2. Corpus Combination

图 2. 语料组合

假定文本有 k 个词汇、 n 个子句、 m 个句子, 则生成的语料可表示为: $V = [v_1, v_2, v_3, \dots, v_k]$, $C = [c_1, c_2, c_3, \dots, c_n]$, $S = [s_1, s_2, s_3, \dots, s_m]$ 。

2.2.2. 自编码层

针对三级语义单元的特征空间异构问题, 通过编码 - 解码过程实现特征维度压缩与信息重构。编码器对输入特征进行压缩编码, 解码器重构原始输入, 通过最小化重构误差获取最优抽象语义表示, 消除跨粒度特征差异。

设定输入为 $X \in \mathbb{R}^{d \times n}$, (n 为所含字向量个数, d 为字向量的维度), 编解码权重矩阵为 W , 激活函数为 $g = (*)$, 编码过程通过线性变换与非线性激活实现, 如所示:

$$H = g(WX)$$

解码过程将编码特征映射回特征空间, 得到重构 $\hat{X}$ 作为输入预测, 解码过程如所示:

$$\hat{X} = g(W^T H)$$

模型训练旨在最小化损失函数 $j(X, \hat{X})$ 达到最小值, 可表示为:

$$j(X, \hat{X}) = -\sum_{i=1}^n [x_i \log(\hat{x}_i) + (1 - x_i) \log(1 - \hat{x}_i)]$$

2.3. 层级注意力传播机制

2.3.1. 多头注意力

考虑模块权重递进深入而导致的整体泛化问题, 使用单一维度的参数矩阵将难以对多重极性进行权重分配。例如: “一直都爱吃她家的泰国菜, 回味无穷”, 若方面词为“味道”, 则会出现“爱吃”和

“回味无穷”，两个修饰性的极性词，因此多头参数矩阵的加入，能够增加模型的不多维特征提取能力。

目前，多层参数矩阵的设计已经广泛应用于特征提取领域[6]，本文设计了含有 h 层权重矩阵的注意力层来构建路径，假定由自编码器输出的特征矩阵为 E ，方面词的特征向量为 t ，则表达式如所示：

$$Q = tW_q$$

$$V = EW_v$$

$$K = EW_k$$

其中 Q 为查询向量， V 为值矩阵， K 为键值矩阵， W_i^q 、 W_i^v 、 W_i^k 分别代表第 i 个的权重矩阵。

此时进行 h 次注意力并行计算，其中 Q 和 K 之间进行点积运算获得第 i 层注意力所对应的权重矩阵 a_i ，将最终获得的 h 个权重矩阵进行平均，而后与 V 进行加权计算，公式如所示：

$$a = \text{mean}(a_1, a_2, \dots, a_h)$$

$$c = a^T V$$

其中， c 为 h 层权重矩阵权重分配后结果。

2.3.2. 树状图生成机制

模块依据权重系数对三级粒度语料实施跨层级权重再分配，每层树节点完成权重计算后，按权重比例与上层节点进行加权聚合，为后续层级的注意力计算提供特征输入。路径生成单元通过这种层级关联机制，高效提取词汇语义、文本结构及同义短语等多维度信息，如图 3：

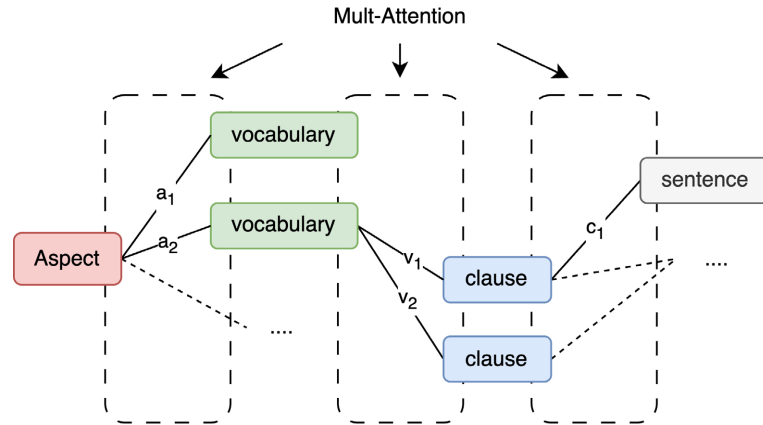


Figure 3. Path generation module
图 3. 基于多头注意力的路径生成模块

借助文本内在语义关联进行关键信息筛选，规避了单一词汇匹配导致的外延信息丢失问题，有效缓解过拟合并提升分类准确率，假定多层递进注意力中三种粒度语料的权重系数为 a_i 、 v_i 、 c_i ，则计算公式如所示：

$$a_i = f(w_{ai}, w_{vi})$$

$$v_i = f(a_i \cdot w_{vi}, w_{ci})$$

$$c_i = f(v_i \cdot w_{ci}, w_{si})$$

其中 $f(Q, K) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$ ，词汇、子句、句子的向量表示为 w_{vi} 、 w_{ci} 、 w_{si} ，方面词向量为 w_{ai} 。

为了最大化利用上层权重分配结果, 树节点在每层注意力权重更新阶段, 会执行上层节点特征的迁移学习操作, 使得模型的整体训练过程更加稳定, 形成类似语义关联挖掘的学习模式, 其隐状态更新过程如图 4 所示:

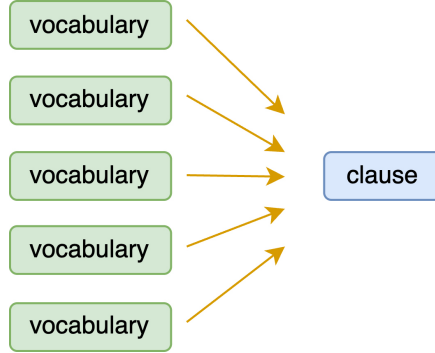


Figure 4. Hidden State Update Process
图 4. 隐状态更新过程

为了传递三种粒度特征之间的关键信息, 通过构建多头自注意力机制, 假定 α_{ij}^k 为上层节点对下层节点的注意力系数, V_j 为上层节点特征表示, $\sigma(*)$ 表示激活函数。通过均值法实现多节点的特征融合, 具体公式如下:

$$C' = \sigma \left(\frac{\sum_{k=1}^K \sum_{j \in N_i} \alpha_{ij}^k V_j C}{K} \right)$$

同时, 为了解决多头注意力可能产生的权重冗余问题, 引入正交化来约束, 保证权重矩阵之间的差异性。

3. 实验结果与分析

3.1. 评价指标

准确率: 衡量模型预测进度的指标, 定义为正确分类样本数占总样本数的比例, 如所示:

$$accuracy = \frac{\sum_{i=1}^m TP_i}{\sum_{i=1}^m (TP_i + FP_i + TN_i)}$$

F1 分数: 评估模型查准率与查全率的复合指标, 由精确率和召回率加权平均获得, 公式如所示:

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

3.2. 数据集的说明

将 SemEval2016 - Task5 和 MAMS 数据集按 8:2 的比例划分为训练集与测试集, 正向、负向、中立三

种标签分别用 1、-1、0 表示, 其数据分布如图 5~6 所示:

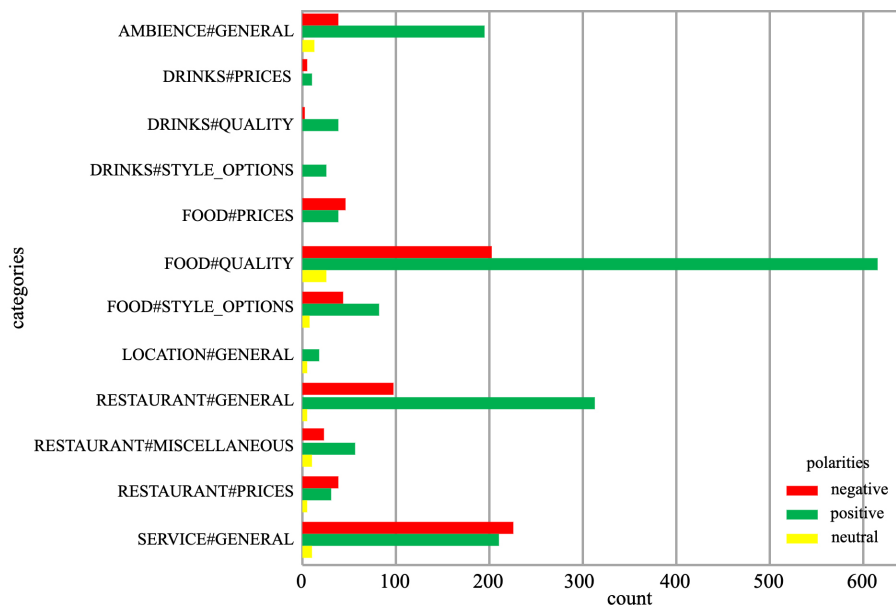


Figure 5. SemEval2016-Task5 database

图 5. SemEval2016-Task5 数据集

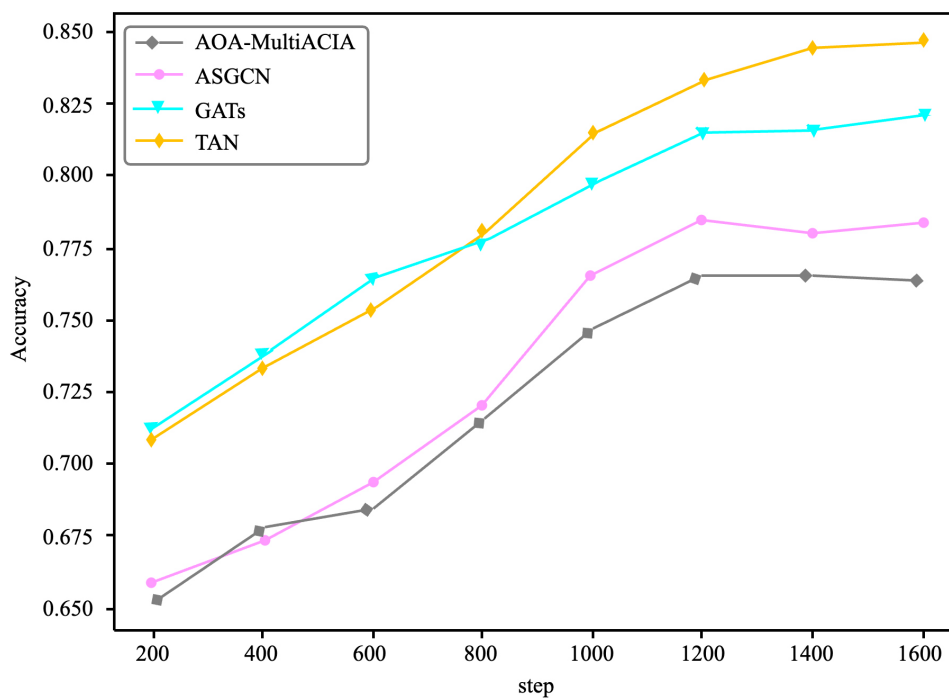


Figure 6. MAMS database

图 6. MAMS 数据集

在 Chinese Review Dataset 数据集中, 有多个极性词修饰同一个评论文本, 正向极性设置为 1, 负向极性设置为 0。同时实验按 7:3 划分训练集和测试集, 如表 1 所示:

Table 1. Chinese Review Dataset database**表 1.** Chinese Review Dataset 数据集

DataSet	Positive		Negative	
	Test	Train	Test	Train
Notebook	201	247	123	83
Car	531	354	200	86
Camera	1090	469	400	200

3.3. 实验参数设计

卷积核以及全链接层神经元的初始化方法为均方分布, 模型参数设置如表 2 所示:

Table 2. Experimental parameters**表 2.** 实验参数配置

模型模块	Chinese Review Dataset	SemEval2016Task5&MAMS
嵌入层维度	300	300
自编码器压缩维度	200	200
注意力头数	3	3
Dropout	0.2	0.2
优化器	RMS	RMS
BatchSize	15	25
Epoch	5	7

3.4. 对比算法

AOA-MultiACIA [7]: 模型构建联合分析, 处理方面词与句子, 通过融合两者语义关系, 挖掘方面词与文本间的潜在影响机制, 实现对文本情感倾向的捕捉。

HN-ATT [8]: 该模型基于多向传播的图网络架构, 对文本进行深度语义建模, 提取双向语境的语义表征。借助多层模块堆叠的方式, 逐步细化语义特征, 完成情感极性分类任务。

ASGCN [9]: 针对句子的句法依存树结构实施特征抽取, 有效捕获句法层面的语义关联信息, 同时结合注意力机制对关键情感特征进行加权, 通过计算实现对文本情感的精准分析。

GATs [10]: 采用多头注意力策略对邻居节点的特征信息进行聚合, 依据语义关联动态调整不同节点的权重分配, 增强模型对多样化语义单元的代表能力, 提升整体性能表现。

3.5. 实验结果分析

模型对比分析

四个对比模型和 TAN 模型的准确度和 F1 分数在 SemEval2016-Task5 和 MAMS 数据集的表现如表 3 所示:

Table 3. Experimental results**表 3.** 实验结果

Model	SemEval2016		MAMS	
	ACC	F1	ACC	F1
AOA-MultiACIA	0.832	0.716	0.746	0.717
ASGCN	0.816	0.729	0.735	0.720
GATs	0.824	0.741	0.716	0.715
TAN	0.896	0.764	0.812	0.722

在 SemEval2016 和 MAMS 数据集的实验结果中, TAN 模型的准确率、召回率均领先于其他模型, 充分证明了其在方面级情感分析任务中的高效分类能力。相比之下, ASGCN 模型通过融合句法结构图与图卷积网络(GCN), 使特征向量能够深度吸收句法信息, 从而增强了模型适应性。GATs 模型则通过引入注意力权重机制, 将传统 GCN 的拉普拉斯矩阵替换为动态注意力系数, 不仅强化了节点间的关联强度, 还在处理句法有向图时展现出更高的灵活性。AOA-MultiACIA 模型提出创新的交互架构, 通过长距离残差保留文本原始特征, 并采用多层堆叠策略优化匹配精度。TAN 模型采用多层递进融合策略, 使模型能够处理多粒度有向图结构。在针对不同方面词建模时, 能够更高效地挖掘目标文本的内在信息, 最终在各类数据集中均有较好表现, 其 F1 分数与分类准确率均达到最优水平。

Chinese Review Datasets 数据集 Notebook、Car、Camera、Phone 领域的准确度以及 F1 分数的实验结果如表 4 所示:

Table 4. Experimental results of each model on Chinese Review Datasets
表 4. Chinese Review Datasets 上各模型实验结果

Model	Notebook		Car		Camera	
	ACC	F1	ACC	F1	ACC	F1
AOA-MultiACIA	0.794	0.612	0.801	0.654	0.769	0.673
ASGCN	0.812	0.601	0.825	0.667	0.784	0.717
GATs	0.831	0.634	0.831	0.701	0.821	0.721
TAN	0.855	0.671	0.859	0.732	0.871	0.762

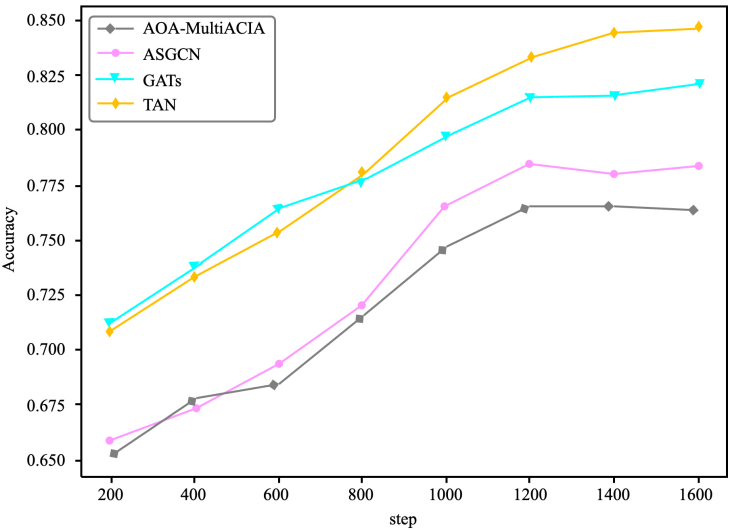


Figure 7. Mainstream model training comparison under MSMA data
图 7. 基于 MSMA 数据的主流模型训练状况对比

根据表 4 和图 7 的分析结果, TAN 模型在处理复杂语境的中文数据集时表现出色, 在训练性能和准确率均优于其他模型。尽管在训练过程中对每条评论的极性判断仅依赖于单一的方面词, 但数据集存在一定数量的多个极性词修饰的评论文本, 因此对同一评论中不同方面词进行特定建模的能力对最终的判别结果非常重要。

TAN 模型的数据预处理阶段, 采用了自编码机制对三种语料进行筛选与关键特征提取。通过路径构造模块生成特定路径, 判别向量不仅融合了文本结构特征与句法隐状态信息, 还显著提升了模型对情感

信息的捕捉效率。这一过程有效增强了模型在复杂语境下的适应性与精准度, 使其在多方面词共存的评论中能够更准确地识别与分类情感倾向。

4. 结语

基于深度学习思想的模型在训练之初, 初始参数的设定往往要借鉴于以往的经验, 并没有相关的理论可作为支持, 需要不断地进行尝试才能够获得模型的最好效果。因此, 还需要对相关深度学习的参数设置方面做进一步的研究和证明, 以便对灵活多变的模型做进一步的改进和优化。

参考文献

- [1] 陈亚茹, 陈世平. 融合自注意力机制和 BiGRU 网络的微博情感分析模型[J]. 小型微型计算机系统, 2020, 41(8): 1590-1595.
- [2] Zheng, L., Wang, H. and Gao, S. (2015) Sentimental Feature Selection for Sentiment Analysis of Chinese Online Reviews. *International Journal of Machine Learning and Cybernetics*, 9, 75-84. <https://doi.org/10.1007/s13042-015-0347-4>
- [3] 王春东, 张卉, 莫秀良, 等. 微博情感分析综述[J]. 计算机工程与科学, 2022, 44(1): 165-175.
- [4] Zhang, S., Tong, H., Xu, J. and Maciejewski, R. (2019) Graph Convolutional Networks: A Comprehensive Review. *Computational Social Networks*, 6, 1-23. <https://doi.org/10.1186/s40649-019-0069-y>
- [5] Zhang, C., Song, D., Huang, C., Swami, A. and Chawla, N.V. (2019) Heterogeneous Graph Neural Network. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, 4-8 August 2019, 793-803. <https://doi.org/10.1145/3292500.3330961>
- [6] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [7] Wu, Z., Li, Y., Liao, J., Li, D., Li, X. and Wang, S. (2020) Aspect-Context Interactive Attention Representation for Aspect-Level Sentiment Classification. *IEEE Access*, 8, 29238-29248. <https://doi.org/10.1109/access.2020.2972697>
- [8] Li, L., Liu, Y. and Zhou, A. (2018) Hierarchical Attention Based Position-Aware Network for Aspect-Level Sentiment Analysis. *Proceedings of the 22nd Conference on Computational Natural Language Learning*, Brussels, 31 October-1 November 2018, 181-189. <https://doi.org/10.18653/v1/k18-1018>
- [9] Zhang, C., Li, Q. and Song, D. (2019) Aspect-Based Sentiment Classification with Aspect-Specific Graph Convolutional Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, November 2019, 4568-4578. <https://doi.org/10.18653/v1/d19-1464>
- [10] Veličković, P., Cucurull, G., Casanova, A., et al. (2017) Graph Attention Networks.