

基于双语依存关联图的跨语言情感分类^{*}

白瑞瑞, 王中卿, 周国栋

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

通信作者: 王中卿, E-mail: wangzq@suda.edu.cn



摘 要: 跨语言情感分类在自然语言处理领域非常重要并且已经得到广泛的研究, 因为它可以利用源语言的标签信息构建目标语言的情感分类系统, 从而大大减少目标语言中费时而耗力的标注工作. 不同语言的表达方式存在明显差异是跨语言情感分类的基本挑战, 提出基于双语依存关联图模型的跨语言情感分类方法. 虽然不同语言的表达存在差异, 但是内部的句法依存关系是相似的. 通过在不同语言的词节点之间建边表示双语评论实例的语义相关性, 双语依存关联图能够对不同语言词之间依存关系的相似性进行显式建模, 从而使图神经网络可以在语言内和语言间整合句法结构信息, 进行跨语言情感分类. 利用英文和中文两种语言的数据集进行实验, 实验结果相较于基线方法提高了 3%. 研究表明, 利用双语依存关联图能够对不同语言评论实例之间的关联性进行有效建模, 从而显著提升跨语言情感分类的准确率.

关键词: 图神经网络; 情感分类; 跨语言

中图法分类号: TP18

中文引用格式: 白瑞瑞, 王中卿, 周国栋. 基于双语依存关联图的跨语言情感分类. 软件学报, 2025, 36(6): 2827–2843. <http://www.jos.org.cn/1000-9825/7246.htm>

英文引用格式: Bai RR, Wang ZQ, Zhou GD. Cross-lingual Sentiment Classification Based on Bilingual Dependency Graph. Ruan Jian Xue Bao/Journal of Software, 2025, 36(6): 2827–2843 (in Chinese). <http://www.jos.org.cn/1000-9825/7246.htm>

Cross-lingual Sentiment Classification Based on Bilingual Dependency Graph

BAI Rui-Rui, WANG Zhong-Qing, ZHOU Guo-Dong

(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract: Cross-lingual sentiment classification is very important in natural language processing and has been widely studied. It uses label information from the source language to construct a sentiment classification system for the target language, thereby greatly reducing the laborious labeling work in the target language. A fundamental challenge in cross-lingual sentiment classification is the obvious difference in the expressions of different languages. This study proposes a method for cross-lingual sentiment classification based on a bilingual dependency graph model. Although the expressions in different languages are various, their internal syntactic dependencies are similar. By establishing edges among word nodes in different languages to represent the semantic relevance of bilingual comment instances, the bilingual dependency graph can explicitly model the similarity of the dependency relationships among words in different languages, allowing graph neural networks to integrate syntactic structure information within and across languages for cross-lingual sentiment classification. Experiments conducted on datasets in both English and Chinese show that the proposed method achieves an improvement of 3% over the baseline method. It is proven that bilingual dependency graphs can effectively model the correlation of comment instances in different languages, thereby significantly improving the accuracy of cross-lingual sentiment classification.

Key words: graph neural network (GNN); sentiment classification; cross-lingual

^{*} 基金项目: 国家自然科学基金 (62076175, 61976146); 江苏省双创博士计划

收稿时间: 2023-11-14; 修改时间: 2024-02-14, 2024-05-04; 采用时间: 2024-06-26; jos 在线出版时间: 2024-12-25

CNKI 网络首发时间: 2024-12-26

不同语言的情感标注语料分布不均, 其中资源匮乏语言的情感分类研究的发展受到制约, 催生了跨语言情感分类研究. 它利用源语言的标注语料, 辅助目标语言进行文本情感极性分类. 因此, 需要解决的核心问题是如何有效共享跨语言的情感知识, 并有效构建跨语言情感表达的关联性. 现有的研究工作采取不同策略进行跨语言情感分类. 传统的方法包括机器翻译, 或者使用外部对齐资源 (包括双语词典和平行语料库等) 训练分类器模型. 这些限制阻碍了模型用在没有任何平行资源的语言上. 近期的研究探索了语言之间的语义相似性, 并为不同语言的文本建立共享表示, 包括双语词嵌入^[1,2]、公共子词表示^[3]和多语言预训练语言模型 (例如 mBERT、XLM)^[4-6]等. 其中, 最先进的多语言预训练模型因其跨语言零样本迁移能力而越来越受欢迎^[7-10]. 这些多语言预训练模型虽然在跨语言分类任务中取得了令人印象深刻的结果, 但其泛化能力高度依赖于跨语言的词汇重叠. 由于没有明确考虑语言之间的句法结构差异, 对于与源语言距离较远的目标语言, 可能会导致在目标语言上的泛化性能下降.

对于跨语言情感分类, 通过观察发现, 不同语言的表达方式虽然存在差异, 但是内部的句法依存关系是相似的. 依存树能够捕获词之间的句法关系, 如图 1 所示, 英文表达中的“food-great”“bad-service”“actually-great”在中文表达中都能够找到对应的句法依存关系, 即“食物-棒”“服务-差”“实际上-棒”. 受此启发, 本文提出通过构建双语依存关联图对不同语言之间的句法相似关系进行显式建模. 在双语依存关联图中, 根据句法依存结构在不同语言的词之间建立跨语言依存关系边. 如图所示, 因为英文词“food”和“great”、中文词“食物”和“棒”之间都存在依存关系边, 所以将会分别在英文词“food”和中文词“棒”、中文词“食物”和英文词“great”之间建立依存关系边.

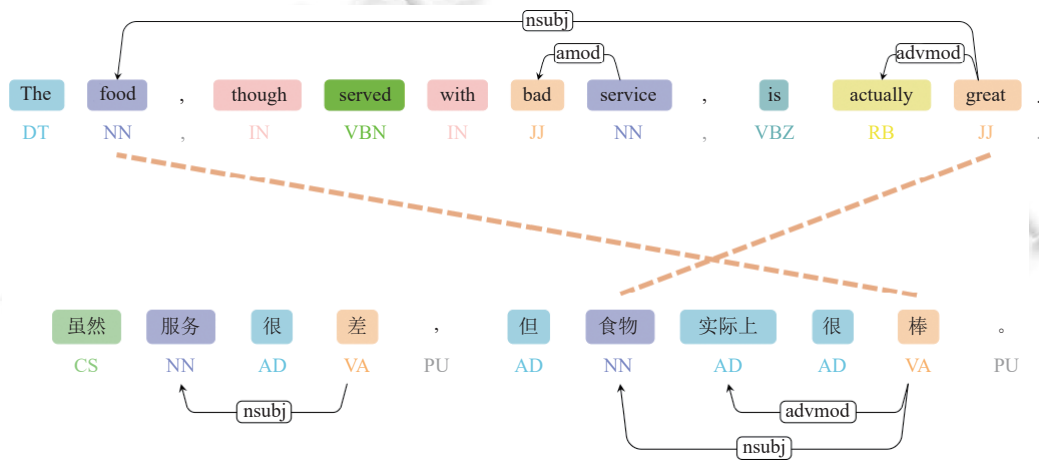


图 1 双语依存关联图示例

句法信息可以在属性词和相应的意见词之间建立关系连接, 基于依存树的图神经网络在属性级情感分类中取得了令人印象深刻的表现^[11]. 图神经网络在面向丰富关系结构的任务中非常有效, 并且可以在图嵌入中保留图的全局结构信息. 因此, 本文采用图神经网络学习双语依存关联图的表示, 在语言内和语言之间整合句法结构信息, 进行跨语言情感分类. 首先, 将单语言句子表示为依存关系图而不是单词序列, 接着对单语依存关系图中的词节点进行双语词对齐, 并且修剪不必要的关系, 这使得双语之间语义上相同的依存关系更加突显. 然后, 将这种相似性通过双语依存关联图进行表示, 图中的节点包括两种语言的词, 边表示不同语言词节点之间的依存关系. 此外, 词对齐信息也被考虑用来增强双语之间的联系, 具体来说, 通过双语文本的词对齐构建双语词-词对齐图, 即在对齐的源语言和目标语言的词之间建立一条边. 文献^[12,13]通过堆叠 GCN 层来提取依存树上的丰富表示. 然而, 这两项研究都是平等地对待图中当前节点的所有邻居节点, 并且缺乏有效的机制来区分邻居节点的重要性, 这可能会降低模型性能^[11]. 因此, 本文引入 Attention 机制来考虑每个邻居节点的语义相关性, 并且分别将图注意力网络和

图卷积网络应用于双语依存关联图和双语词-词对齐图, 最后利用获取到的高阶图表示构建感知器模型进行情感分类.

综合来说, 本文工作的主要贡献可以概括为以下 3 个方面.

- 从句法分析的研究角度提出构建双语依存关联图, 以此整合语言内和语言间的句法知识, 建立双语之间的关联性, 并通过构建图神经网络学习双语嵌入表示用于跨语言情感分类.
- 基于模型集成学习, 对单语和双语情感分类的预测结果进行整合, 并取其平均值作为最终的预测结果, 与只使用双语情感分类相比取得了更好的分类性能.
- 在中英文数据集的多个领域上进行实验, 相比于其他基线模型的准确率达有 3% 的提升.

1 相关工作

近年来, 自然语言处理中的情感分析任务越来越受欢迎^[14]. 情感分析严重依赖带注释的标注数据, 对于许多新兴领域来说, 获取这些数据通常既耗时又昂贵. 为了减轻对标注数据的依赖, 先前的工作提出了多种无监督域适应方法^[15], 包括基于枢轴的方法^[16-18]、自动编码器^[19,20]、领域对抗网络^[21-23]和半监督方法^[24,25]. 这些方法主要侧重于学习域不变表示, 以减轻跨领域数据的分布差异. 近期的一些研究探索了新的领域适应框架^[26-29], 即跨领域数据增强 (cross-domain data augmentation, CDDA), 其目的是根据源领域的标注数据直接生成大量目标领域标注数据. 跨语言情感分类可以被看作是一种特殊的跨领域情感分类任务, 相较于跨领域特征空间具有交集, 跨语言特征空间之间并无直接交集, 分布差异更加明显, 因此更具研究难度.

文献 [30] 提出基于协同训练的半监督学习方式利用未标注数据来提升系统性能. 该方法成为跨语言情感分类研究的里程碑. Wan^[30]将训练数据 (英文译成中文) 和测试数据 (中文译成英文) 进行翻译, 将源语言和目标语言作为两个不同的视图, 分别训练分类模型, 协同训练算法用于将双语模型组合在一起并提高分类性能. 除了基于翻译的方法外, 一些研究还利用平行资源, 包括平行语料库 (例如平行维基百科文章)、双语词典等来弥合语言障碍. 文献 [31] 使用机器翻译获取平行语料库, 并对与平行语料库相关的问题进行研究, 包括单语言情感分类和跨语言情感分类. 文献 [32] 利用 WordNet 语义作为监督情感分类的特征, 使用两种语言的 WordNet 链接来弥合语言差距. 文献 [33,34] 利用平行的无标注数据来弥合语言障碍. 文献 [35] 使用双语词典通过结构对应学习算法学习源语言和目标语言之间的映射. 平行语料库通常是利用机器翻译工具进行训练获取的, 而平行语料库的构建首先是捕获跨语言共享的模式, 从而忽略针对特定语言的模式. 换句话说, 在单方面追求跨语言的泛化性能时未能保留特定于语言的情感知识. 因此, 文献 [36] 提出将情感表达中的语言差异建模为内在双语极性相关性, 以更好地进行跨语言情感分类.

近几年学习不同语言文本的共享表示被提出来以解决跨语言情感分类问题. 文献 [37,38] 使用自编码器通过最小化两个平行句子的词袋表示之间的重构误差来获取双语的共享表达. 文献 [39] 提出了双语文档表示学习模型 BiDRL, 通过同时利用双语文档之间的语义和情感相关性, 学习双语的共享嵌入空间. 为了捕捉文本中的长距离依赖性, 文献 [40] 提出了基于注意力的双语 LSTM 模型. 文献 [1] 提出了对抗性深度平均网络 ADAN, 将情感知识从已标注的源语言数据转移到只有未标注数据存在的低资源目标语言, 通过使用双语词嵌入 (BWE)^[41]将源语言和目标语言的文本映射到相同的语义空间. 文献 [42] 通过使用无监督的 BWE^[43]将文献 [1] 提出的跨语言模型扩展到多源语言, 并为每种源语言构建独立的特征提取器, 从而消除对平行语料库的依赖. 文献 [44] 针对目标语言没有标签数据和平行语料库的情况, 提出了多视图编码分类器 MVEC. 通过利用文献 [45] 提出的无监督机器翻译系统的编码器-解码器, 引入语言鉴别器进行对抗性训练, 使得编码器能够将来自两种语言的输入映射到独立于语言的共享特征空间. 文献 [2] 基于枢轴学习和双语嵌入^[46]提出了跨语言跨领域情感分类模型.

近年来, 显式利用句法知识已经在属性级情感分类^[47,48]任务上进行了广泛研究, 并取得了巨大成功. 此类知识有助于在句子中的属性词和其他单词之间建立联系, 以学习属性词的语法感知特征表示. 例如, 文献 [49] 利用句

法相对距离来减少不相关单词的影响. 沿着这个研究思路, 一些工作^[12,13,50]通过句法依存树对 GCN 和 GAT 模型进行了扩展. 这些工作对基于输入文本的依存解析结果构建的图进行学习, 获取相邻节点的节点表示. 文献 [51–53] 组合不同类型的图, 旨在更加准确地提取特定于属性词的情感特征. 文献 [54] 提出了 T-GCN, 利用依存类型并区分依存树中的不同关系. 文献 [11] 提出了 SSEGCN, 将句法结构和语义相关性有效融合以获得更丰富的信息. 这启发我们, 是否可以利用句法结构建立不同语言之间的关联性, 从而进行跨语言情感分类.

在本文的研究中, 同样是利用机器翻译技术获得目标语言带标签的平行语料. 与之前的工作不同, 我们提出构建双语依存关联图模型, 以显式地为不同语言评论实例之间的句法结构相似性建模. 具体来讲, 双语依存关联图依据句法依存关系进行构建, 以此来显式地表达不同语言之间的内在关联性并且可以适应图神经网络. 本文方法首先使用多语言预训练模型获取评论实例的嵌入表示, 然后结合双语依存关联图构建图神经网络模型获取共享的双语特征表示.

2 基于双语依存关联图的跨语言情感分类模型

2.1 问题描述

给定源语言的带标签数据 $D_{src} = \{(x_i, y_i)\}_{i=1}^L$, 其中, x_i 是评论实例, y_i 是极性标签. 目标语言的测试数据表示为 $D_{tgt} = \{(x_i)\}_{i=1}^T$. 为了构建双语依存关联图模型, 本文借助机器翻译技术获取双语伪平行语料, 即 $D_{parl} = \{(x_i, x'_i)\}_{i=1}^P$. 其中, x_i 和 x'_i 是对应源语言和目标语言的平行评论实例.

表 1 给出了本文设定的跨语言情感分类研究场景. 如表所示, 源语言是英文, 目标语言是中文. 表中 ID 编号为 1 和 2 的评价实例是源语言的已标注样本, 对属性词“seafood”和“sashimi”分别表达了正向和负向情感. 表中 ID 编号为 3 和 4 的评论实例是对源语言的英文评论文本进行翻译得到的目标语言的中文评论文本. 需要解决的问题是预测 ID 编号为 5 的目标语言中文评论实例对属性词“服务”表达的情感极性. 本文提出的跨语言情感分类方法旨在使用源语言的带标签训练数据来构建适用于目标语言的分类模型, 具体来讲, 通过联合利用 D_{src} 和 D_{parl} 学习分类模型, 并预测 D_{tgt} 的评论实例的情感极性.

表 1 基于双语平行语料的跨语言情感分类示例

语言	ID	评论实例	属性词	情感极性
英文	1	Great seasonal fish and seafood , with a classy waterfront setting.	seafood	正向
	2	Those rolls were big, but not good and sashimi wasn't fresh.	sashimi	负向
中文	3	很棒的时令鱼和 海鲜 , 拥有优雅的海滨环境.	海鲜	正向
	4	那些面包卷很大, 但不好吃, 而且 生鱼片 不新鲜.	生鱼片	负向
	5	总的来说, 这是一个带几个朋友出去玩的好地方, 而且 服务 很棒.	服务	?

2.2 基于双语依存关联图的跨语言情感分类框架

如图 2 所示, 基于双语依存关联图的跨语言情感分类方法可以分为 3 个阶段. 第 1 个阶段是构建不同的图模型, 基于依存关系结构图得到单语依存关联图和双语依存关联图, 利用句法结构信息显式地对不同语言评论实例之间的关联性进行图建模并且可以适应图神经网络. 此外, 还考虑了词对齐信息, 即通过双语文本的词对齐来构建双语词-词对齐图. 第 2 个阶段是利用图神经网络来训练各种图模型, 并基于学习到的图表示构建用于预测评论实例极性的双语情感分类器和单语情感分类器. 具体而言, 分别将图注意力网络和图卷积网络应用于双语依存关联图和双语词-词对齐图, 并用学习到的图表示作为双语情感分类器的输入. 采用图卷积网络应用于单语依存关联图, 并将学习到的图表示作为单语情感分类器的输入. 在最后一个阶段, 获取目标语言评论实例及其源语言翻译的评论实例的单语情感分类器的预测结果以及双语情感分类器的预测结果, 并基于集成学习对这些预测结果进行联合判断, 得到评论实例最后的情感极性预测标签.

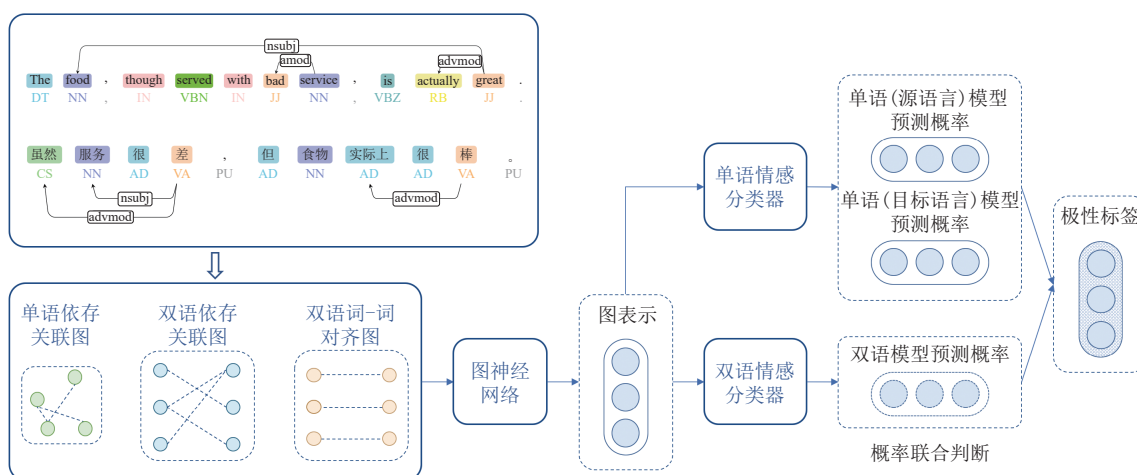


图2 基于双语依存关联图的跨语言情感分类框架图

3 双语依存关联图

为了对不同语言之间的关联性进行显式建模, 本方法从句法结构的研究角度提出双语依存关联图模型. 此外, 本方法构建双语词-词对齐图来表示双语的对齐信息, 目的是增强双语之间的关联. 本节将介绍3种类型的边: 第1种边是根据依存句法结构来构建单语依存关系边; 第2种是双语依存关系边, 与单语依存关系边不同, 它是在两种语言的词节点之间建边; 第3种是用于构建双语词-词对齐图的词-词对齐翻译边.

(1) 单语依存关系边. 句子的句法结构可以通过依存关系图来揭示, 如图3和图4所示. 其中, 图3给出的是一个英文句法分析示例, 图4给出的是一个中文句法分析示例. 基于句法分析的结果, 我们可以在存在依存关系的两个词节点之间建边, 并称这样的边为单语依存关系边, 以此得到单语依存关联图. 例如, 根据图3的分析结果, 将在词节点“not”和“only”之间建立依存关系边“not-only”; 同样地, 根据图4的分析结果, 将在词节点“不仅”和“棒”之间建立依存关系边“不仅-棒”. 可以看到, 构建单语依存关系边的词节点是同语言的.

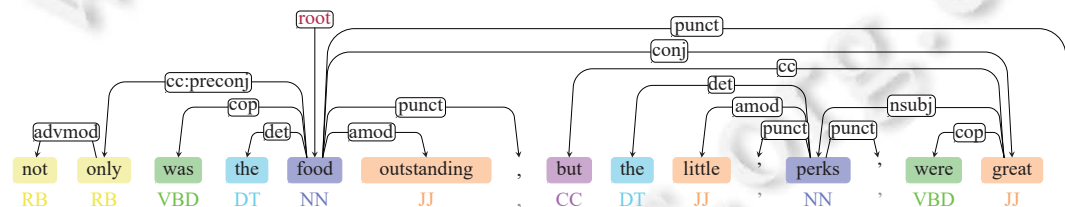


图3 英文(源语言)依存关系结构图示例

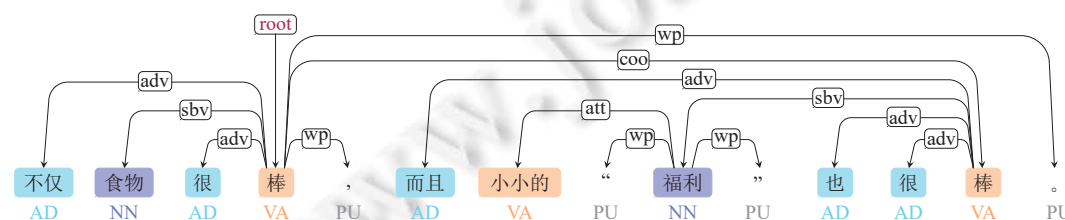


图4 中文(目标语言)依存关系结构图示例

(2) 双语依存关系边. 与单语依存关系边不同, 双语依存关系边是建立在不同语言的词节点之间. 本方法通过3个步骤来构建跨语言依存关系边, 如图5所示. 首先, 将双语评论实例进行文本词对齐. 例如, 在图中, 英文词“not”

和“only”都与中文词“不仅”对齐. 然后, 对原始单语依存关系结构图进行剪枝, 丢弃一些不必要的关系. 例如, 在图中, 因为有英文词“the”“food”与中文词“食物”对齐, 英文词“outstanding”与中文词“很”“棒”对齐, 所以会丢弃“the-food”“很-棒”这些关系. 剪枝之后的结构图使得双语之间在语义上相同的依存关系更加明显. 同时, 为了简化双语依存关联图的结构, 修剪单语依存关系结构图还会丢弃与标点符号相关的依存关系. 最后, 建立双语依存关系边. 例如图 5, 在修剪过的英文依存关系结构图中, 因为“the food”与“not only”之间存在依存关系, 所以会依次在它们的对齐词“the food”与“不仅”、“食物”与“not only”之间建立跨语言依存关系边. 同时, 在修剪过的中文依存关系结构图中, 因为有“很棒”与“食物”之间存在依存关系, 所以会依次在它们的对齐词“outstanding”和“食物”、“很棒”和“the food”之间建立跨语言依存关系边.

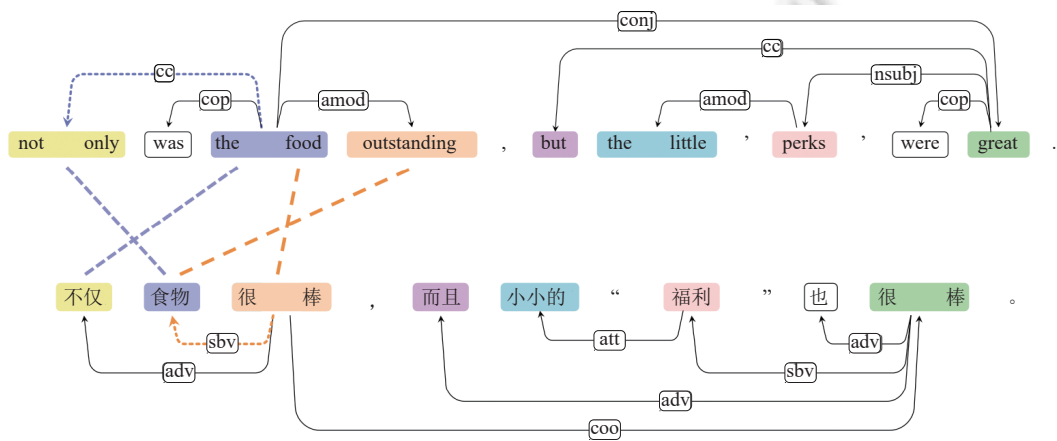


图 5 双语依存关系边示例

(3) 词-词对齐翻译边. 利用双语词对齐信息, 在对齐的单词之间构建词-词对齐翻译边. 如图 6 所示, 由于英文词“not”“only”和中文词“不仅”是对齐的, 因此会分别在词节点“not”与“不仅”以及词节点“only”与“不仅”之间建立对齐翻译边.

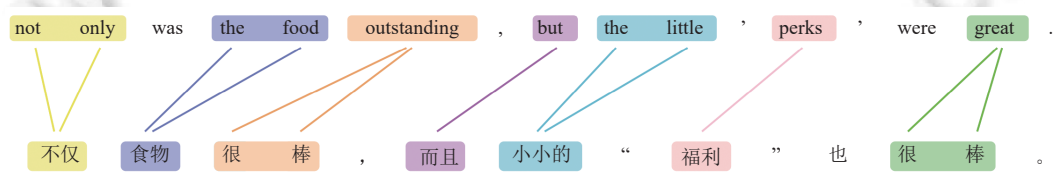


图 6 词-词对齐翻译边示例

4 基于双语依存关联图的情感分类

4.1 文本表示

借助 ELMo^[55]、GPT^[56]和 BERT^[4]等预训练语言模型, 依赖上下文的单词表示学习取得了重大进展. 近期的预训练模型在跨语言方面也进行了扩展. 跨语言预训练模型旨在构建一种跨语言编码器, 从而将不同语言的句子在相同的嵌入空间中进行编码. 目前最常用的跨语言预训练模型包括多语言 BERT、XLM^[5]和 XLM-RoBERTa^[6]等. 最初的多语言 BERT 模型同时支持 104 种语言. 它依赖于跨所有语言的 110k 大小的共享 WordPiece 词汇表, 并在公共多语言空间中生成句子表示, 在自然语言任务的跨语言迁移方面具有良好的表现^[57]. XLM 在多语言 BERT 的基础上增加了一个新的任务, 即翻译语言模型. 它利用有监督的跨语言并行数据, 将机器翻译语料中的双语句对接拼成一个新的句子, 然后在这个句子上进行掩码语言模型的预训练. XLM 可以学习到两个语言之间的对应关

系, 从而获得比多语言 BERT 更好的跨语言迁移能力. XLM 模型作为大规模的跨语言预训练模型, 通常被当作特征抽取器使用. XLM-RoBERTa 基于 RoBERTa 模型框架, 作为 XLM 的进阶版进一步扩大了其跨语言优势.

给定评论文本 r 和属性词 a , 例如“食物很美味, 酒吧的氛围很棒”和“食物”, 本方法选用 XLM-RoBERTa 模型作为跨语言编码器, 学习由属性词和评论文本拼接构建的输入单词序列的向量表示, 例如:

$$W = [\text{CLS}] \text{食物} [\text{SEP}] \text{食物很美味, 酒吧的氛围很棒} [\text{SEP}].$$

如图 7 所示, 将输入序列中特殊符号 [CLS] 的编码向量 $h_{[\text{CLS}]}$ 作为整个序列的向量表示 H , 后面每个位置的编码向量则是每个词的对应编码表示.

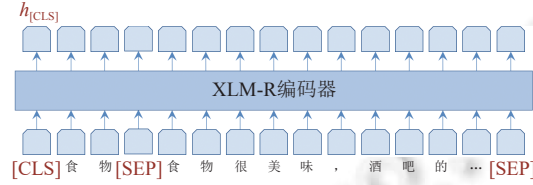


图 7 多语言预训练模型编码示例

4.2 基于双语依存关联图的情感分类模型

如图 8 所示, 在得到评论实例的编码表示之后, 本方法分别通过图注意力网络和图卷积网络对双语依存关联图和双语词-词对齐图进行建模, 计算并获取评论实例的高阶图表示, 最后将图表示作为双语情感分类器的输入进行情感极性预测.

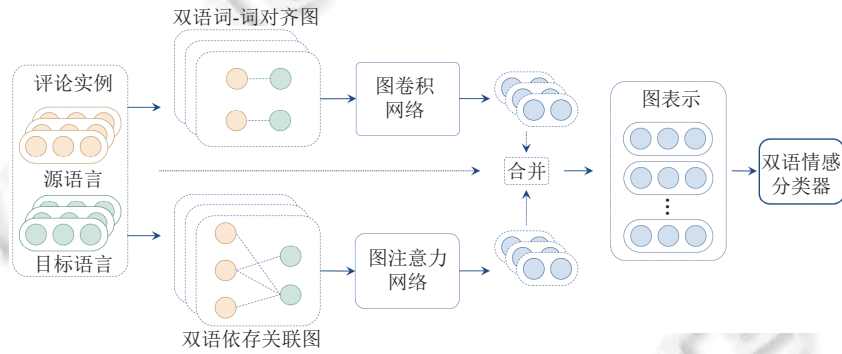


图 8 基于双语依存关联图的双语情感分类模型

4.2.1 图卷积网络

GCN 是一种卷积神经网络, 专为图结构数据的建模而设计, 能够利用图的结构信息, 通过其邻域的属性得到每个节点的嵌入向量. 给定一个图 $G = (V, E)$, 其中, V 和 E 分别表示节点和边的集合; 令 $X \in \mathbb{R}^{|V| \times d}$ 是一个矩阵, 包含节点及其特征 $x_i \in \mathbb{R}^d$ (每一行 x_i 是节点 i 的特征向量). 借助图的拉普拉斯矩阵的特征值和特征向量来定义卷积算子的方法, GCN 逐层传播的规则如下:

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in N(i)} \frac{1}{c_{ij}} h_j^{(l)} W^{(l)} \right) \quad (1)$$

其中, $h_i^{(l+1)}$ 表示第 $l+1$ 层节点 i 的特征; c 是边的归一化系数, $c_{ij} = \sqrt{\deg(i)\deg(j)}$, $j \in N(i)$, $N(i)$ 是节点 i 的邻接节点集合, $\deg(i)$ 和 $\deg(j)$ 分别为节点 i 和 j 的度, 也就是度矩阵 D 在节点 i 和 j 处的值; $W^{(l)}$ 是一个特定于层可训练的变换矩阵; $\sigma(\cdot)$ 表示非线性激活函数.

4.2.2 图注意力网络

GAT 与 GCN 相比, 都是通过相邻节点来更新每个节点得到新的嵌入向量. 不同的是, GAT 利用注意力机制

来强调每个节点与多个相邻节点之间关系的差异性,从而将模型的注意力分配给对学习当前节点特征更有帮助的邻接节点. GAT^[58]通过使用注意力机制聚合迭代更新每个节点的嵌入向量,计算公式如下:

$$h_{(att)_i}^{(l+1)} = \sigma \left(\sum_{j \in N(i)} a_{ij}^{(l)} W^{(l)} h_j^{(l)} \right) \quad (2)$$

$$a_{ij}^{(l)} = \text{attention}(i, j) \quad (3)$$

其中, $a_{ij}^{(l)}$ 是节点 i 与其邻接节点 j 之间在第 l 层的注意力系数.

GAT 通过归一化节点之间的相关系数 e 来计算注意力权重 a , 其相关系数计算公式如下:

$$e_{ij}^{(l)} = \alpha(W h_i^{(l)}, W h_j^{(l)}) \quad (4)$$

$$a_{ij}^{(l)} = \frac{\exp(\text{LeakyReLU}(e_{ij}^{(l)}))}{\sum_{k \in N(i)} \exp(\text{LeakyReLU}(e_{ik}^{(l)}))} \quad (5)$$

其中, 计算 $e_{ij}^{(l)}$ 使用的是加型注意力机制, W 和 a 是需要学习的参数.

为了学习不同子空间中的注意力权重, GAT 还可以使用多头注意力:

$$h_i^{(l+1)} = \parallel_{k=1}^K \sigma \left(\sum_{j \in N(i)} a_k^{(l)} W_k^{(l)} h_j^{(l)} \right) h_i^{(l+1)} \quad (6)$$

4.2.3 基于图模型的表示

给定一条英文 (源语言) 评论实例 s 及其平行中文 (目标语言) 评论实例 s' , 例如:

s : “Not only was the food outstanding, but the little ‘perks’ were great.”

s' : “不仅食物很棒, 而且小小的‘福利’也很棒.”

在源语言单语依存关联图 $G_s = (V, E)$ 中, 节点集 V 是英文评论文本 s 中的所有词, 边集 E 是该句子中所有的依存关系边, 关联图 G_s 可以用邻接矩阵 A_s 来表示, 如图 9 所示. 需要注意的是, 构建 G_s 使用的是剪枝之后的单语依存关系结构图. 例如, 因为词“the food”和“outstanding”之间存在依存关系, 所以在 A_s 有 $\text{edge}(\text{the food}, \text{outstanding}) = 1$. 在目标语言单语依存关联图 $G_r = (V, E)$ 中, 节点集 V 是中文评论文本 s' 中的所有词, 边集 E 是该句子中所有的依存关系边, 图 10 是 G_r 的邻接矩阵 A_r . 同样地, 构建 G_r 使用的也是剪枝之后的单语依存关系结构图. 例如, 词“食物”和“很棒”之间存在依存关系, 则有 $\text{edge}(\text{食物}, \text{很棒}) = 1$. 由于本文关注的是词之间的关联性, 因此, 图 G_s 和 G_r 都是无向图, 在建边过程中忽略依存关系的方向.

	not only	was	the food	outstanding	but	th little	perks	were	great
not only	0	0	1	0	0	0	0	0	0
was	0	0	1	0	0	0	0	0	0
the food	1	1	0	1	0	0	0	0	1
outstanding	0	0	1	0	0	0	0	0	0
but	0	0	0	0	0	0	0	0	1
the little	0	0	0	0	0	0	1	0	0
perks	0	0	0	0	0	1	0	0	1
were	0	0	0	0	0	0	0	0	1
great	0	0	1	0	1	0	1	1	0

图 9 单语 (源语言) 依存关联图邻接矩阵示例

	不仅	食物	很棒	而且	小小的	福利	也	很棒
不仅	0	0	1	0	0	0	0	0
食物	0	0	1	0	0	0	0	0
很棒	1	1	0	0	0	0	0	1
而且	0	0	0	0	0	0	0	1
小小的	0	0	0	0	0	1	0	0
福利	0	0	0	0	1	0	0	1
也	0	0	0	0	0	0	0	1
很棒	0	0	1	1	0	1	1	0

图 10 单语 (目标语言) 依存关联图邻接矩阵示例

在基于 s 和 s' 构造的双语依存关联图 $G_r = (V, E)$ 中, 节点集 V 包括两种语言评论文本中的所有词, 与图 G_s 和 G_r 中的边类型不同, G_r 的边集 E 中都是双语依存关系边. 邻接矩阵 A_r 用于表示图 G_r 的结构, 如图 11 所示, 因为

在 A_r 中有 $edge(\text{食物}, \text{很棒}) = 1$, 并且“食物”和“很棒”的对齐词分别是“the food”和“outstanding”, 所以在 A_r 中会有 $edge(\text{the food}, \text{很棒}) = 1$ 和 $edge(\text{食物}, \text{outstanding}) = 1$.

	not only	was	the food	outstanding	but	the little	perks	were	great
不仅	0	0	1	1	0	0	0	0	0
食物	1	1	0	1	0	0	0	0	1
很棒	1	0	1	0	0	0	0	0	1
而且	0	0	0	0	0	0	0	0	1
小小的	0	0	0	0	0	0	1	0	0
福利	0	0	0	0	0	1	0	0	1
也	0	0	0	0	0	0	0	0	1
很棒	0	0	0	1	1	0	1	1	1

图 11 双语依存关联图邻接矩阵示例

此外, 在基于 s 和 s' 构造的双语词-词对齐图 $G_a = (V, E)$ 中, 节点集 V 包括两种语言评论文本中的所有词, 边集 E 中是词-词对齐翻译边. 邻接矩阵 A_a 用于表示图 G_a 的结构, 如果节点 i 和节点 j 是词对齐, 则在 A_a 中有 $edge(i, j) = 1$. 图 G_a 同样是无向图.

给定双语依存关联图的邻接矩阵 A_r , 本方法使用图注意力网络来计算节点的高阶表示. 网络的输入是两种语言评论实例中的单词和双语依存关联图, 并且所有的单词都已经被编码为向量, 输出 H_R 作为评论实例的向量表示, H_R 的计算公式如下:

$$H_R = GAT(W_s, W_t, A_r) \quad (7)$$

给定双语词-词对齐图的邻接矩阵 A_a , 本方法使用图卷积网络来计算节点的高阶表示. 网络的输入是两种语言评论实例中的单词和双语词-词对齐图, 输出 H_A 作为评论实例的向量表示, H_A 的计算公式如下:

$$H_A = GCN(W_s, W_t, A_a) \quad (8)$$

此外, 给定单语依存关联图的邻接矩阵 A , 本方法使用图卷积网络来获取评论实例的向量表示. 我们用 H_S 和 H_T 分别代表源语言和目标语言的向量表示, 即:

$$H_S = GCN(W_s, A_s) \quad (9)$$

$$H_T = GCN(W_t, A_t) \quad (10)$$

其中, W_s 是每个源语言评论实例中的所有词, W_t 是目标语言的平行评论实例中的所有词.

4.2.4 情感分类模型

(1) 双语情感分类模型

当分别从双语依存关联图和双语词-词对齐图中得到评论实例的向量表示 H_R 和 H_A 之后, 本方法首先利用拼接操作融合不同方式得到的向量表示, 然后构建感知器模型预测评论实例的情感极性.

双语情感分类器的输入向量表示 I 由 3 部分构成:

$$I = [H_E || H_R || H_A] \quad (11)$$

然后, 将 I 输入 *Softmax* 输出层用于分类:

$$O = \text{Softmax}(W_p I + b_p) \quad (12)$$

其中, H_E 是评论实例的编码向量, $||$ 表示拼接操作, W_p 和 b_p 是输出层的参数.

(2) 单语情感分类模型

此外, 本方法还考虑利用源语言和目标语言的向量表示 H_S 和 H_T 获取单语情感分类器, 具体做法是: 首先与评论实例的编码向量 H_E 进行拼接, 然后采用感知器模型预测评论实例的情感极性.

对于源语言, 单语情感分类器的输入向量表示 I_s 为:

$$I_s = [H_E \| H_S] \quad (13)$$

然后, 将 I_s 输入 *Softmax* 输出层用于分类:

$$O_s = \text{Softmax}(W_s I_s + b_s) \quad (14)$$

对于目标语言, 单语情感分类器的输入向量表示 I_t 为:

$$I_t = [H_E \| H_T] \quad (15)$$

然后, 将 I_t 输入 *Softmax* 输出层用于分类:

$$O_t = \text{Softmax}(W_t I_t + b_t) \quad (16)$$

其中, W_s 、 b_s 、 W_t 和 b_t 是需要学习的参数.

(3) 模型训练

给定评论实例和目标属性词, 情感分类器的任务是预测评论实例针对目标属性词表达的情感极性. 因此, 训练目标是最小化一组训练样本的真实标签和预测标签之间的交叉熵损失, 计算公式如下:

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^M p(y_{ic}) \log p(\hat{y}_{ic}) + \lambda \|\Theta\|^2 \quad (17)$$

其中, N 是样本总数, M 是类别数量, y_{ic} 是样本 i 的真实标签 y_i 经过 one-hot 编码后在第 c 个类别上的取值, $\hat{y}_{ic}$ 是样本 i 属于第 c 个类别的预测概率, Θ 是模型参数集合, λ 是 l_2 正则化项的系数. 虽然单语情感分类器和双语情感分类器的输入向量表示不同, 但是使用的真实标签是相同的.

4.3 概率联合判断

为了提高模型分类性能, 在预测阶段, 本方法采用集成学习的思想, 对单语和双语情感分类器的预测结果进行联合判断, 获取最终的预测概率.

给定目标语言的测试样本 x , 我们利用翻译工具获取对应的源语言测试样本 x' , 接着利用两个单语情感分类器得到测试样本 x 和 x' 的预测概率: $O_t(x)$ 和 $O_s(x')$, 并利用双语情感分类器得到 (x', x) 的预测概率 $O(x', x)$, 则测试样本 x 最终的情感极性分布 $O(x)$ 的计算公式如下:

$$O(x) = \text{avg}[O(x', x), O_s(x'), O_t(x)] \quad (18)$$

其中, *avg* 表示对不同分类器的预测概率结果取平均值.

5 实验与分析

本节首先介绍实验中使用到的数据集和用于性能比较的基线方法, 然后报告从不同角度进行实验的结果, 并分析所提出模型的有效性.

5.1 实验设置

5.1.1 数据

实验中使用的两种语言分别是英文和中文, 数据集的下载地址是: https://drive.google.com/drive/folders/1w5ramwgbEP7GS0jo9oxKrMnjQJPN_tJp?usp=sharing. 英文数据集包括 2 个领域, 分别是来自 SemEval-14 数据集的餐馆 (Res) 和笔记本电脑 (Lap) 的评论数据. 并且, 将评分 > 3 的标记为正样本, 将评分 < 3 的标记为负样本. 中文数据集包括 5 个领域: 家具、家纺、厨房、百货、玩具, 是从淘宝网收集和标注的.

我们分别进行了两个方向的实验. 如表 2 所示, 在正向实验中, 使用英文作为源语言, 预测目标语言中文的评论实例的情感极性. 训练集包含 350 条餐馆评论样本和 350 条笔记本电脑评论样本, 总共 700 条评论样本. 测试集分为 5 个部分, 即家具、家纺、厨房、百货、玩具, 每个领域选取 400 条评论样本. 另外抽取 100 条餐馆评论样本和 100 条笔记本电脑评论样本作为验证集. 在逆向实验中, 分别使用中文和英文作为源语言和目标语言. 训练集

由 5 部分组成, 即家具、家纺、厨房、百货、玩具, 每个领域选取 150 条评论样本, 总共 750 条评论样本. 验证集同样是从 5 个中文数据集进行抽取, 总共包括 250 条评论样本. 测试集分为 2 个部分, 即在餐馆和笔记本电脑两个领域分别抽取 400 条评论样本. 本实验使用的训练集都是正负样本平衡的.

表 2 双语数据集样本数统计

语言	数据集	领域	总样本数
英文→中文	训练集	Res+Lap	350+350
	验证集	Res+Lap	100+100
	测试集	家具	400
		家纺	400
		厨房	400
		百货	400
		玩具	400
中文→英文	训练集	家具+家纺+厨房+百货+玩具	150+150+150+150+150
	验证集	家具+家纺+厨房+百货+玩具	50+50+50+50+50
	测试集	Res Lap	400 400

5.1.2 评价标准

本实验使用准确率 (*Accuracy*) 来衡量模型的分类性能, 其计算公式如下:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (19)$$

其中, *TP* 为正类被分类器正确地预测为正类的样本数目; *FP* 为正类被分类器错误地预测为负类的样本数目; *FN* 为负类被分类器错误地预测为正类的样本数目; *TN* 为负类被分类器正确地预测为负类的样本数目.

5.1.3 参 数

本实验使用 Hugging Face 开源的预训练语言模型 xlm-roberta-base (<https://huggingface.co/xlm-roberta-base>) 作为编码器, 对 xlm-roberta-base 的微调是在英文训练集上以 Batch 大小为 32, 学习率为 1E-3, 进行 3 次迭代完成, 进而分别获取源语言和目标语言评论文本的向量表示. 在训练情感分类器时, 本实验通过验证集调整模型参数. 模型的参数设置如表 3 所示, 选用的优化器是 Adam. 所有实验均在 1 张 Nvidia RTX 2080S GPU 上完成.

表 3 模型参数设置

参数描述	参数值
句子最大长度	100
Batch大小	32
隐层维数	128
学习率	1E-3
L2正则	0.001

5.2 基线方法

为了评估所提出模型在跨语言情感分类任务上的鲁棒性, 本实验考虑了以下的基线方法.

- CLD-based^[59]: 利用模型蒸馏和平行语料库将模型从源语言转换为目标语言. 其中, CLD-KCNN 利用平行语料库上源语言的软预测来训练目标语言的分类模型. CLDFA-KCNN 是一种改进的变体, 它利用对抗训练在目标语言中进行领域适应.

- Adversarial DAN (ADAN)^[1]: 是另一种通过对抗训练提高跨语言泛化能力的方法. 它使用单词的平均双语嵌入作为输入, 并调整特征提取器以在两种语言中产生相似的特征.

- 多语言预训练模型: 我们比较了多语言预训练模型 XLM-RoBERTa. 在本实验中, 首先利用源语言训练集对 xlm-roberta-base 进行微调, 然后对目标语言评论实例直接预测其情感极性.
- CLHG^[60]: 使用 xlm-roberta-base 作为编码器, 并在目标语言和源语言之间构建跨语言图. 具体来讲, 把两种语言的文档和单词视为节点, 将具有不同关系 (包括词性角色、语义相似性和文档翻译) 的节点链接起来构建异构图, 利用不同的子图来表示不同的边关系, 然后训练异构图神经网络输出分类结果. 与 CLHG 方法不同, 本文提出的方法将两种语言的单词视为节点, 在不同语言的词节点之间分别构造双语依存关系边和对齐翻译边, 得到双语依存关系图和双语词-词对齐图.

5.3 实验结果与分析

本节首先将所提出方法与基线方法进行比较分析, 然后通过消融实验对不同图结构的影响进行详细分析, 最后通过实例分析展示本模型的有效性.

5.3.1 与基线方法对比

表 4–表 6 给出了正向实验的比较结果. 其中, 方法 XLM-R 的训练集使用的是源语言语料, 方法 XLM-R (Trans)、CLD-based (包括 CLD-KCNN 和 CLDFA-KCNN)、ADAN、CLHG 和本文的方法使用的训练集都是双语平行语料. 为了进行统一比较, 本实验使用 xlm-roberta-base 编码器的向量表示作为 CLD-based 和 ADAN 方法的词嵌入. 在预测阶段, 首先获取双语测试集 (包括目标语言测试集和利用机器翻译得到的源语言测试集) 的单语情感分类器和双语情感分类器的预测结果, 然后对单语和双语情感分类器的预测结果进行联合判断, 提高分类准确率. 表 4 给出了目标语言单语情感分类器在目标语言测试集上的预测结果, 表 5 给出了源语言单语情感分类器在翻译的源语言测试集上的预测结果, 表 6 给出了在双语测试集上利用集成学习进行概率联合判断的实验结果.

表 4 正向实验单语 (中文) 情感分类的准确率 (%)

领域	XLM-R	XLM-R (Trans)	CLD-KCNN	CLDFA-KCNN	ADAN	CLHG	本文方法
家具	75.5	76.5	78.2	79.4	75.5	81.5	83.1
家纺	70.6	71.3	76.1	76.8	77.9	76.7	78.2
厨房	72.8	74.1	77.6	78.5	80.8	80.9	84.3
百货	74.7	75.6	76.5	77.2	78.4	80.7	84.3
玩具	77.1	78.8	79.7	80.5	81.6	83.4	84.6
平均	74.1	75.3	77.6	78.5	78.8	80.6	82.9

表 5 正向实验单语 (英文) 情感分类的准确率 (%)

领域	XLM-R	XLM-R (Trans)	CLD-KCNN	CLDFA-KCNN	ADAN	CLHG	本文方法
家具	76.2	77.2	76.6	77.3	74.1	80.2	82.8
家纺	71.4	70.5	74.5	75.3	78.2	76.1	77.9
厨房	73.1	73.7	76.8	76.1	78.6	79.7	83.9
百货	74.3	74.8	75.7	77.7	79.3	81.1	83.6
玩具	75.9	76.6	78.3	79.8	80.8	82.5	84.2
平均	74.2	74.6	76.4	77.2	78.2	79.9	82.5

从上述的正向实验结果可以看出, XLM-R (Trans) 使用双语平行语料要比 XLM-R 只使用单一语言的语料效果好. 同时, 可以发现 CLD-based 和 ADAN 方法优于 mPLM 方法 (包括 XLM-R 和 XLM-R (Trans)), 这说明在 mPLM 的基础上, 无论是通过跨语言知识蒸馏还是跨语言对抗训练, 显式考虑双语之间的关联性进行跨语言情感分类都是有效的. 此外, 基于图的方法 (包括 CLHG 和本文方法) 取得了最好的性能, 而且本文的方法明显优于其他基线方法, 并实现了对 CLHG 的持续改进, 这表明本文从句法分析角度构建的双语依存关联图和双语词-词对齐图能够更好地表达双语之间的关联关系.

表 7 给出了逆向实验的比较结果. 其中, “EN”表示目标语言单语情感分类器的预测结果, “ZH”表示源语言单语情感分类器的预测结果, “UNI”表示对单语和双语情感分类器的预测概率进行联合判断的结果. 如表 7 所示, 本文提出的方法同样取得了最好的性能, 进一步表明结合句法结构信息基于图方法显式考虑双语之间的关联性是很有必要的, 而且也证明了构建的双语依存关联图和双语词-词对齐图的有效性.

表 6 正向实验在测试集上的概率联合判断的准确率 (%)

领域	XLM-R	XLM-R (Trans)	CLD-KCNN	CLDFA-KCNN	ADAN	CLHG	本文方法
家具	77.8	79.4	78.9	81.1	76.2	82.3	85.9
家纺	72.2	71.8	77.3	78.4	79.8	79.4	80.1
厨房	74.5	75.6	79.1	79.6	82.4	81.1	85.5
百货	76.3	76.2	78.2	79.3	80.7	81.5	86.4
玩具	78.3	79.7	80.6	81.2	82.5	84.6	86.8
平均	75.8	76.5	78.8	79.9	80.3	81.8	84.9

表 7 逆向实验在测试集上的准确率 (%)

领域		XLM-R (Trans)	CLD-KCNN	CLDFA-KCNN	ADAN	CLHG	本文方法
Res	EN	74.6	75.4	76.8	77.3	80.3	81.5
	ZH	73.5	74.8	76.1	75.6	78.9	80.3
	UNI	75.2	76.3	77.6	78.5	81.2	82.6
Lap	EN	72.1	74.2	75.1	75.5	76.9	78.2
	ZH	71.6	73.6	73.2	74.6	75.3	77.6
	UNI	73.4	75.1	75.5	76.7	78.4	79.8
平均	EN	73.4	74.8	76.0	76.4	78.6	79.9
	ZH	72.6	74.2	74.7	75.1	77.1	79.0
	UNI	74.3	75.7	76.6	77.6	79.8	81.2

5.3.2 图结构的影响

通过消融实验分析不同图结构的有效性的结果如表 8 所示, 并且在后文表 9 中描述了不同图结构的组合. 可以看到, 无论是哪种方式的图结构组合, 模型的准确率都超过了基线方法 XLM-R 的 75.8%, 这说明考虑双语之间的语义关联性是有效的. 在表 8 中, 当去掉“双语词-词对齐图”时, 组合方式 1 的准确率下降到了 77.4%; 当去掉“双语依存关联图”时, 组合方式 2 的准确率下降到了 80.2%. 相比之下, 当去掉“单语依存关联图”时, 模型准确率下降不多. 这说明, 无论是“双语词-词对齐图”还是“双语依存关联图”, 在表达双语之间的关联性方面都发挥了重要作用. 并且, 组合方式 Full 同时使用“双语词-词对齐图”和“双语依存关联图”, 使得模型的分类性能比 XLM-R 有了明显的提升.

表 8 图结构对于模型性能 (准确率) 的影响 (%)

领域	XLM-R	1	2	3	Full
家具	77.8	78.7	80.5	83.7	85.9
家纺	72.2	73.7	76.5	79.2	80.1
厨房	74.5	76.2	80.7	83.4	85.5
百货	76.3	78.3	81.2	85.6	86.4
玩具	78.3	80.3	82.2	84.9	86.8
平均	75.8	77.4	80.2	83.4	84.9

5.3.3 实例分析

表 10 是具体的实例分析, 选取两条目标语言待测试评论文本, 对 XLM-R、CLHG 和本文方法进行具体的对比分析.

表 9 不同图结构的组合

图结构	1	2	3	Full
双语词-词对齐图	√	—	√	√
双语依存关联图	—	√	√	√
单语依存关联图	√	√	—	√

表 10 实例分析

评论文本	情感极性	XLM-R	CLHG	本文方法
自己弟弟装的,除了龙头质量感觉差点,其他无可挑剔,很美,打算再买两个.	正向	×	√	√
东西还不错,有点不耐脏,拿回家刚拆开就有点脏,但是价格便宜,真的解放双手,平时都要抱着,现在放腰凳上轻松点了.	正向	×	×	√

注:“√”表示预测正确,“×”表示预测错误

从表 10 中可以看到,选取的两条评论文本的句法结构都有一定的复杂性.对于第 1 条评论文本,CLHG 和本文提出的方法都能够正确预测.其原因是这两种方法都对不同语言之间的关联性进行了显式建模,说明构建双语依存关联图对于跨语言情感分类是有效的.然而对于第 2 条评论文本,与第 1 条评论文本相比,它的句法结构更加复杂,只有本文提出的方法预测到正确的情感标签.这是因为依存关联图能够捕获词之间的句法依存关系,使得模型可以结合句法结构信息进行情感极性预测.以上实例也证明了本文提出的双语依存关联图模型确实能够在语言内和语言间整合丰富的句法结构信息,使得分类模型可以对目标语言具有复杂句法结构的评论文本做出正确的情感极性判断.

6 结 语

为了对不同语言评论实例之间的句法结构相似性进行显式建模,克服语言鸿沟,本文提出基于双语依存关联图的跨语言情感分类.在本文提出的方法中,首先依据句法依存关系构建单语和双语依存关联图,用于显式地表达不同语言之间的内在关联并且可以适应图神经网络.此外,利用双语文本的词对齐构建双语词-词对齐图.然后分别将图注意力网络应用于双语依存关联图、图卷积网络应用于双语词-词对齐图,获取评论实例的图表示,并分别构建单语和双语情感分类器预测评论实例的情感极性.最后利用集成学习的思想,对不同情感分类器的预测结果进行联合判断,得到目标语言评论实例的最终情感标签.实验结果表明,双语依存关联图能够有效地表达不同语言评论实例之间的语义相关性,从而显著提升目标语言情感分类的准确率.

在接下来的工作中,我们将从以下两方面开展研究.(1)进一步拉远源语言和目标语言之间的距离,探究更加有效的跨语言模型来解决情感分类任务.(2)本文的研究表明了跨语言图模型的有效性,而且考虑到了句法信息,后面会深入探索如何将该研究思路与生成式大语言模型结合,从而提高跨语言情感分类的性能.

References:

[1] Chen XL, Sun Y, Athiwaratkun B, Cardie C, Weinberger K. Adversarial deep averaging networks for cross-lingual sentiment classification. Trans. of the Association for Computational Linguistics, 2018, 6: 557–570. [doi: 10.1162/tac1_a_00039]

[2] Ziser Y, Reichart R. Deep pivot-based modeling for cross-language cross-domain transfer with minimal guidance. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: ACL, 2018. 238–249. [doi: 10.18653/v1/D18-1022]

[3] Zhang MZ, Fujinuma Y, Boyd-Graber J. Exploiting cross-lingual subword similarities in low-resource document classification. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence. New York: AAAI, 2020. 9547–9554. [doi: 10.1609/aaai.v34i05.6500]

[4] Devlin J, Chang M W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1: Long Papers). Minneapolis: ACL, 2019. 4171–4186. [doi: 10.18653/v1/N19-1423]

[5] Lample G, Conneau A. Cross-lingual language model pretraining. arXiv:1901.07291, 2019.

[6] Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, Grave E, Ott M, Zettlemoyer L, Stoyanov V. Unsupervised

- cross-lingual representation learning at scale. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 8440–8451. [doi: [10.18653/v1/2020.acl-main.747](https://doi.org/10.18653/v1/2020.acl-main.747)]
- [7] Conneau A, Wu SJ, Li HR, Zettlemoyer L, Stoyanov V. Emerging cross-lingual structure in pretrained language models. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 6022–6034. [doi: [10.18653/v1/2020.acl-main.536](https://doi.org/10.18653/v1/2020.acl-main.536)]
 - [8] Zhao MJ, Schütze H. Discrete and soft prompting for multilingual models. In: Proc. of the 2021 Conf. on Empirical Methods in Natural Language Processing. ACL, 2021. 8547–8555. [doi: [10.18653/v1/2021.emnlp-main.672](https://doi.org/10.18653/v1/2021.emnlp-main.672)]
 - [9] Lin XV, Mihaylov T, Artetxe M, Wang TL, Chen SH, Simig D, Ott M, Goyal N, Bhosale S, Du JF, Pasunuru R, Shleifer S, Koura PS, Chaudhary V, O'Horo B, Wang J, Zettlemoyer L, Kozareva Z, Diab M, Stoyanov V, Li X. Few-shot learning with multilingual language models. arXiv:2112.10668, 2022.
 - [10] Huang LZ, Ma SM, Zhang DD, Wei FR, Wang HF. Zero-shot cross-lingual transfer of prompt-based tuning with a unified multilingual prompt. In: Proc. of the 2022 Conf. on Empirical Methods in Natural Language Processing. Abu Dhabi: ACL, 2022. 11488–11497. [doi: [10.18653/v1/2022.emnlp-main.790](https://doi.org/10.18653/v1/2022.emnlp-main.790)]
 - [11] Zhang Z, Zhou ZL, Wang YN. SSEGCN: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis. In: Proc. of the 2022 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle: ACL, 2022. 4916–4925. [doi: [10.18653/v1/2022.naacl-main.362](https://doi.org/10.18653/v1/2022.naacl-main.362)]
 - [12] Zhang C, Li QC, Song DW. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: ACL, 2019. 4568–4578. [doi: [10.18653/v1/D19-1464](https://doi.org/10.18653/v1/D19-1464)]
 - [13] Sun K, Zhang RH, Mensah S, Mao YY, Liu XD. Aspect-level sentiment analysis via convolution over dependency tree. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: ACL, 2019. 5679–5688. [doi: [10.18653/v1/D19-1569](https://doi.org/10.18653/v1/D19-1569)]
 - [14] Birjali M, Kasri M, Beni-Hssane A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-based Systems, 2021, 226: 107134. [doi: [10.1016/j.knosys.2021.107134](https://doi.org/10.1016/j.knosys.2021.107134)]
 - [15] Zhuang FZ, Qi ZY, Duan KY, Xi DB, Zhu YC, Zhu HS, Xiong H, He Q. A comprehensive survey on transfer learning. Proc. of the IEEE, 2021, 109(1): 43–76. [doi: [10.1109/JPROC.2020.3004555](https://doi.org/10.1109/JPROC.2020.3004555)]
 - [16] Blitzer J, Dredze M, Pereira F. Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: Proc. of the 45th Annual Meeting of the Association of Computational Linguistics. Prague: ACL, 2007. 440–447.
 - [17] Yu JF, Jiang J. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Austin: ACL, 2016. 236–246. [doi: [10.18653/v1/D16-1023](https://doi.org/10.18653/v1/D16-1023)]
 - [18] Ziser Y, Reichart R. Pivot based language modeling for improved neural domain adaptation. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1: Long Papers). New Orleans: ACL, 2018. 1241–1251. [doi: [10.18653/v1/N18-1112](https://doi.org/10.18653/v1/N18-1112)]
 - [19] Glorot X, Bordes A, Bengio Y. Domain adaptation for large-scale sentiment classification: A deep learning approach. In: Proc. of the 28th Int'l Conf. on Machine Learning. Bellevue: Omnipress, 2011. 513–520.
 - [20] Zhou GY, Xie ZW, Huang JX, He TT. Bi-transferring deep neural networks for domain adaptation. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Berlin: ACL, 2016. 322–332. [doi: [10.18653/v1/P16-1031](https://doi.org/10.18653/v1/P16-1031)]
 - [21] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. In: Proc. of the 32nd Int'l Conf. on Machine Learning. 2015. 1180–1189.
 - [22] Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky V. Domain-adversarial training of neural networks. The Journal of Machine Learning Research, 2016, 17(1): 2096–2030.
 - [23] Li Z, Wei Y, Zhang Y, Yang Q. Hierarchical attention transfer network for cross-domain sentiment classification. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. New Orleans: AAAI, 2018. 5852–5859. [doi: [10.1609/aaai.v32i1.12055](https://doi.org/10.1609/aaai.v32i1.12055)]
 - [24] He RD, Lee WS, Ng HT, Dahlmeier D. Adaptive semi-supervised learning for cross-domain sentiment classification. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: ACL, 2018. 3467–3476. [doi: [10.18653/v1/D18-1383](https://doi.org/10.18653/v1/D18-1383)]
 - [25] Ye H, Tan QY, He RD, Li JT, Ng HT, Bing LD. Feature adaptation of pre-trained language models across languages and domains with robust self-training. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP). ACL, 2020. 7386–7399. [doi: [10.18653/v1/2020.emnlp-main.599](https://doi.org/10.18653/v1/2020.emnlp-main.599)]
 - [26] Yu JF, Gong CG, Xia R. Cross-domain review generation for aspect-based sentiment analysis. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing (Findings). ACL, 2021. 4767–4777. [doi: [10.18653/v1/2021.findings-acl.421](https://doi.org/10.18653/v1/2021.findings-acl.421)]

- [27] Li JJ, Yu JF, Xia R. Generative cross-domain data augmentation for aspect and opinion co-extraction. In: Proc. of the 2022 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle: ACL, 2022. 4219–4229. [doi: [10.18653/v1/2022.naacl-main.312](https://doi.org/10.18653/v1/2022.naacl-main.312)]
- [28] Yu JF, Zhao QK, Xia R. Cross-domain data augmentation with domain-adaptive language modeling for aspect-based sentiment analysis. In: Proc. of the 61st Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Toronto: ACL, 2023. 1456–1470. [doi: [10.18653/v1/2023.acl-long.81](https://doi.org/10.18653/v1/2023.acl-long.81)]
- [29] Deng Y, Zhang WX, Pan SJ, Bing LD. Bidirectional generative framework for cross-domain aspect-based sentiment analysis. In: Proc. of the 61st Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Toronto: ACL, 2023. 12272–12285. [doi: [10.18653/v1/2023.acl-long.686](https://doi.org/10.18653/v1/2023.acl-long.686)]
- [30] Wan XJ. Co-training for cross-lingual sentiment classification. In: Proc. of the Joint Conf. of the 47th Annual Meeting of the ACL and the 4th Int'l Joint Conf. on Natural Language Processing of the AFNLP. Suntec: ACL, 2009. 235–243.
- [31] Banea C, Mihalcea R, Wiebe J. Multilingual subjectivity: Are more languages better? In: Proc. of the 23rd Int'l Conf. on Computational Linguistics. Beijing: ACL, 2010. 28–36.
- [32] Balamurali AR, Joshi A, Bhattacharyya P. Cross-lingual sentiment analysis for Indian languages using linked WordNets. In: Proc. of the 24th Int'l Conf. on Computational Linguistics (Posters). Mumbai: Indian Institute of Technology Bombay, 2012. 73–82.
- [33] Meng XF, Wei FR, Liu XH, Zhou M, Xu G, Wang HF. Cross-lingual mixture model for sentiment classification. In: Proc. of the 50th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Jeju Island: ACL, 2012. 572–581.
- [34] Lu B, Tan CH, Cardie C, Tsou BK. Joint bilingual sentiment classification with unlabeled parallel corpora. In: Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland: ACL, 2011. 320–330.
- [35] Prettenhofer P, Stein B. Cross-language text classification using structural correspondence learning. In: Proc. of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala: ACL, 2010. 1118–1127.
- [36] Chen Q, Li CL, Li WJ. Modeling language discrepancy for cross-lingual sentiment analysis. In: Proc. of the 2017 ACM on Conf. on Information and Knowledge Management. Singapore: ACM, 2017. 117–126. [doi: [10.1145/3132847.3132915](https://doi.org/10.1145/3132847.3132915)]
- [37] Chandar APS, Lauly S, Larochelle H, *et al.* An autoencoder approach to learning bilingual word representations. In: Proc. of the 27th Int'l Conf. on Neural Information Processing Systems, Vol. 2. Montreal: MIT Press, 2014. 1853–1861.
- [38] Zhou HW, Chen L, Shi FL, Huang DG. Learning bilingual sentiment word embeddings for cross-language sentiment classification. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing (Vol. 1: Long Papers). Beijing: ACL, 2015. 430–440. [doi: [10.3115/v1/P15-1042](https://doi.org/10.3115/v1/P15-1042)]
- [39] Zhou XJ, Wan XJ, Xiao JG. Cross-lingual sentiment classification with bilingual document representation learning. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Berlin: ACL, 2016. 1403–1412. [doi: [10.18653/v1/P16-1133](https://doi.org/10.18653/v1/P16-1133)]
- [40] Zhou XJ, Wan XJ, Xiao JG. Attention-based LSTM network for cross-lingual sentiment classification. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Austin: ACL, 2016. 247–256. [doi: [10.18653/v1/D16-1024](https://doi.org/10.18653/v1/D16-1024)]
- [41] Zou WY, Socher R, Cer D, Manning CD. Bilingual word embeddings for phrase-based machine translation. In: Proc. of the 2013 Conf. on Empirical Methods in Natural Language Processing. Seattle: ACL, 2013. 1393–1398.
- [42] Chen Z, Qian TY. Transfer capsule network for aspect level sentiment classification. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 547–556. [doi: [10.18653/v1/P19-1052](https://doi.org/10.18653/v1/P19-1052)]
- [43] Conneau A, Lample G, Ranzato MA, Denoyer L, Jégou H. Word translation without parallel data. arXiv:1710.04087, 2018.
- [44] Fei HL, Li P. Cross-lingual unsupervised sentiment classification with multi-view transfer learning. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 5759–5771. [doi: [10.18653/v1/2020.acl-main.510](https://doi.org/10.18653/v1/2020.acl-main.510)]
- [45] Lample G, Conneau A, Denoyer L, Ranzato MA. Unsupervised machine translation using monolingual corpora only. arXiv:1711.00043, 2018.
- [46] Smith SL, Turban DHP, Hamblin S, Hammerla NY. Offline bilingual word vectors, orthogonal transformations and the inverted Softmax. arXiv:1702.03859, 2017.
- [47] Wang K, Shen WZ, Yang YY, Quan XJ, Wang R. Relational graph attention network for aspect-based sentiment analysis. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 3229–3238. [doi: [10.18653/v1/2020.acl-main.295](https://doi.org/10.18653/v1/2020.acl-main.295)]
- [48] Tang H, Ji DH, Li CL, Zhou QJ. Dependency graph enhanced dual-Transformer structure for aspect-based sentiment classification. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 6578–6588. [doi: [10.18653/v1/2020.acl-main.588](https://doi.org/10.18653/v1/2020.acl-main.588)]
- [49] Phan MH, Ogunbona PO. Modelling context and syntactical features for aspect-based sentiment analysis. In: Proc. of the 58th Annual

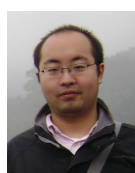
- Meeting of the Association for Computational Linguistics. ACL, 2020. 3211–3220. [doi: [10.18653/v1/2020.acl-main.293](https://doi.org/10.18653/v1/2020.acl-main.293)]
- [50] Huang BX, Carley K. Syntax-aware aspect level sentiment classification with graph attention networks. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: ACL, 2019. 5469–5477. [doi: [10.18653/v1/D19-1549](https://doi.org/10.18653/v1/D19-1549)]
- [51] Chen CH, Teng ZY, Zhang Y. Inducing target-specific latent structures for aspect sentiment classification. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP). ACL, 2020. 5596–5607. [doi: [10.18653/v1/2020.emnlp-main.451](https://doi.org/10.18653/v1/2020.emnlp-main.451)]
- [52] Liang B, Yin RD, Gui L, Du JC, Xu RF. Jointly learning aspect-focused and inter-aspect relations with graph convolutional networks for aspect sentiment analysis. In: Proc. of the 28th Int'l Conf. on Computational Linguistics. Barcelona: Int'l Committee on Computational Linguistics, 2020. 150–161. [doi: [10.18653/v1/2020.coling-main.13](https://doi.org/10.18653/v1/2020.coling-main.13)]
- [53] Zhang M, Qian TY. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP). ACL, 2020. 3540–3549. [doi: [10.18653/v1/2020.emnlp-main.286](https://doi.org/10.18653/v1/2020.emnlp-main.286)]
- [54] Tian YH, Chen GM, Song Y. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In: Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. ACL, 2021. 2910–2922. [doi: [10.18653/v1/2021.naacl-main.231](https://doi.org/10.18653/v1/2021.naacl-main.231)]
- [55] Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L. Deep contextualized word representations. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1: Long Papers). New Orleans: ACL, 2018. 2227–2237. [doi: [10.18653/v1/N18-1202](https://doi.org/10.18653/v1/N18-1202)]
- [56] Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding with unsupervised learning. Technical Report, OpenAI, 2018. <https://openai.com/index/language-unsupervised/>
- [57] Wu SJ, Dredze M. Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: ACL, 2019. 833–844. [doi: [10.18653/v1/D19-1077](https://doi.org/10.18653/v1/D19-1077)]
- [58] Velicković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. arXiv:1710.10903, 2018.
- [59] Xu RC, Yang YM. Cross-lingual distillation for text classification. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Vancouver: ACL, 2017. 1415–1425. [doi: [10.18653/v1/P17-1130](https://doi.org/10.18653/v1/P17-1130)]
- [60] Wang ZY, Liu X, Yang PJ, Liu SX, Wang ZS. Cross-lingual text classification with heterogeneous graph neural network. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing (Vol. 2: Short Papers). ACL, 2021. 612–620. [doi: [10.18653/v1/2021.acl-short.78](https://doi.org/10.18653/v1/2021.acl-short.78)]



白瑞瑞(1989—), 女, 博士, CCF 学生会员, 主要研究领域为自然语言处理, 情感分析。



周国栋(1967—), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为自然语言处理。



王中卿(1987—), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为自然语言处理, 情感分析。