计算机科学
COMPUTER SCIENCE

# 基于话题注意力和依存句法信息的文本立场分析

康书铭, 朱焱

引用本文

康书铭, 朱焱. 基于话题注意力和依存句法信息的文本立场分析[J]. 计算机科学, 2023, 50(11A): 230200068-5.

KANG Shuming, ZHU Yan. Text Stance Detection Based on Topic Attention and Syntactic Information [J]. Computer Science, 2023, 50(11A): 230200068-5.

---

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230200068

# 基于话题注意力和依存句法信息的文本立场分析

康书铭　朱　焱
西南交通大学计算机与人工智能学院　成都 611756
(ksm0801@qq.com)

**摘　要**　文本立场分析旨在从用户发表的文本中推测其对特定话题的看法,如支持、反对、中立等态度。传统的立场分析研究往往采用卷积神经网络或者长短时记忆网络等深度学习模型学习文本的基本语义信息,忽略了文本蕴含的句法结构信息。针对这一问题,文中设计实现了基于话题注意力和依存句法的文本立场检测模型——AT-BiLSTM-GAT,在BiLSTM提取的文本上下文信息基础上,采用GAT进一步学习文本语言学层次的依存句法信息。同时设计实现一种融合上下文语义信息的话题注意力机制,采用缩放点积注意力学习立场文本中与话题相关的重要内容,在公开数据集上的对比实验证明了AT-BiLSTM-GAT模型的高效性。最后,针对立场分析研究数据集存在规模较小的问题,设计实现了一种基于WordNet同义词库与WebVectors词嵌入模型的同义词替换数据增强方案WWDA,保证了同义词替换过程的词性正确性和语义相似性,通过实验证明其可以生成更多高质量样本,提升模型的检测性能。

**关键词**:立场分析;话题注意力;依存句法;图注意力神经网络;数据增强
**中图法分类号**　TP391

# Text Stance Detection Based on Topic Attention and Syntactic Information

KANG Shuming and ZHU Yan
School of Computing and Artificial Intelligence,Southwest Jiaotong University,Chengdu 611756,China

**Abstract**　Text stance detection aims to infer users' opinions on specific topics,such as supportive,opposing,neutral and other attitudes,from their published texts. Traditional stance detection studies often use deep learning models such as convolutional neural networks or long and short-term memory networks to learn the basic semantic information of the text,ignoring the syntactic structure information embedded in the text. To address this problem,this paper designs and implements a text stance detection model—AT-BiLSTM-GAT based on topic attention and dependent syntax,and on the basis of the text context information extracted by BiLSTM,GAT is used to further learn dependent syntactic information at the text linguistic level. Meanwhile,a topic attention mechanism incorporating contextual semantic information is designed and implemented,and scaled dot product attention is employed to learn the topic-related important content in stance text,and comparative experiments on public datasets prove the efficiency of the designed and implemented AT-BiLSTM-GAT model. Finally,to address the problem of the small size of the stance detection research dataset,a synonym replacement data enhancement scheme based on WordNet synonym database and WebVectors word embedding model—WWDA,which ensures the lexical correctness and semantic similarity of the synonym replacement process,and experiment proves that it can generate more high-quality samples and improve the detection performance of the model.

**Keywords**　Stance detection,Topic attention,Dependency syntax,Graph attention network,Data augmentation

## 1　引言

目前社交网络上针对社会中的热点话题发表评论的数量快速增长。研究社交网络用户对产品、服务、政策和新闻等特定话题的意见反馈,具有重要的舆情分析和热点话题预测价值,有助于提升政务和商务智能化。文本立场分析旨在从用户发表的文本中推测其对特定话题的看法,包括支持、反对、中立等态度[1],表1列出了一些立场分析案例。

立场分析与传统的情感分析存在一定关联,但也存在明显区别。传统的情感分析往往只需要挖掘文本整体表达的情感倾向,比如消极、积极等,而立场分析需要进一步挖掘文本对特定话题的看法。此外,一些用户在相关话题的评论中表达观点的方式比较隐晦[2],导致立场分析对立场文本的语义学习要求更高,这也增加了立场分析研究的难度。相关研究大多采用TextCNN或者BiLSTM等顺序神经网络模型学习立场文本表示,这种学习方式只能学习立场文本的基本语义信息,无法从语言学角度提取更深层次的语义信息。另一方面,已有研究忽略了立场分析数据集规模小导致模型鲁棒性较差的问题。为了更好地学习文本的依存句法信息,提取更高层次的立场文本表征,本文设计实现了一种基于话题

基金项目:四川省科技计划(2019YFSY0032)
This work was supported by the Sichuan Science and Technology Project(2019YFSY0032).
通信作者:朱焱(yzhu@swjtu.edu.cn)

注意力与依存句法信息的立场分析检测模型——AT-BiLSTM-GAT。

表 1 立场分析案例

Table 1 Examples of stance detection

| 立场话题 | 立场文本 | 立场 |
|---|---|---|
| Feminist Movement | the time for gender equality is NOW | 支持 |
| Wearing a Face Mask | If every person in the UK used one single-use mask each day for a year, an extra 66 000 tonnes of contaminated plastic waste would be created | 反对 |
| Hilary Clinton | I would rather choose Donald Trump | 反对 |

本文的主要贡献分为以下 3 点：

1)构建了立场文本的依存句法关系图，采用图注意力网络(Graph Attention Network，GAT)学习文本句法信息，从而在文本上下文信息基础上，从语言学层次学习句法信息。

2)采用 BiLSTM 将立场话题与立场文本进行上下文语境建模，在此基础上采用缩放点积注意力机制进行话题注意力计算，从而更好地学习立场文本中与话题相关的内容。

3)为了缓解公开的立场分析数据集数据集规模小的问题，本文设计实现一种基于 WordNet 与 WebVectors 的同义词替换数据增强方案——WWDA，增加了数据样本，提升了立场分析模型的鲁棒性。

## 2 相关研究

近年来立场分析的研究主要以深度学习模型为主。Du 等[3]采用 BiLSTM 学习立场文本的上下文信息，采用注意力机制计算立场文本中每个词的词向量与立场话题的相关度。Yue 等[4]提出一种基于词向量级别的两阶段注意力机制的立场分析模型，在注意力计算的两个阶段中分别捕捉话题与立场文本的重要词汇，从而实现立场分析。Bai 等[5]采用 TextCNN 与 BiLSTM 分别提取文本的 n-gram 信息与文本时序上下文信息，并结合注意力池化策略实现文本立场分析。Sun 等[6]提出了一个联合神经网络模型 Joint，学习文本的立场和情感表征，利用情感信息辅助立场检测任务。然而以上研究往往只能学习立场文本的基本语义信息，很难从句法结构角度捕捉深层次的语义信息。Wang 等[7]提出了一个基于语言学表征的分层注意力立场检测模型 HAN，从文本中提取依存句法依赖词对，输入 LSTM 模型得到文本依存句法表征，结合情感、论点表征，实现文本立场检测。虽然其考虑了立场文本多层次的语言学信息，但文本的句法关系是图结构信息，采用时序化方式学习不能很好地学习文本的句法信息。

本研究构建立场文本依存句法图，将立场文本中每个词看作句法图的节点，将句法关系看作句法图的边，采用 GAT 学习得到每个词的依存句法表示，在立场文本上下文表示基础上提取文本的依存句法信息，结合设计实现的话题注意力机制，进行文本立场分析。同时本研究针对立场分析数据集规模小、数据增强研究匮乏的问题，设计实现一种数据增强方案 WWDA，用于提升立场模型的鲁棒性和检测性能。

## 3 模型设计

本文设计实现的基于话题注意力和依存句法信息的立场分析模型——AT-BiLSTM-GAT 如图 1 所示。

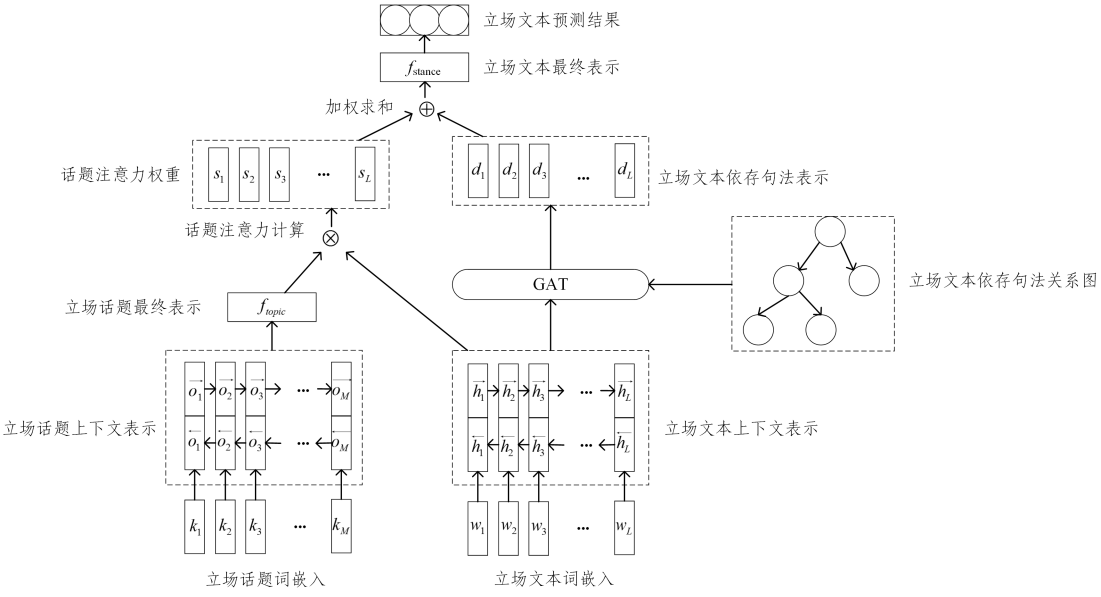


图 1 AT-BiLSTM-GAT 立场分析模型框架图

Fig. 1 Framework of AT-BiLSTM-GAT stance detection model

### 3.1 上下文表示学习

本文采用 $\boldsymbol{w}=\{\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_L\}$ 表示立场文本的初始词嵌入矩阵，$L$ 表示立场文本长度，$\boldsymbol{k}=\{\boldsymbol{k}_1,\boldsymbol{k}_2,\cdots,\boldsymbol{k}_M\}$ 表示立场话题的初始词嵌入矩阵，$M$ 表示立场话题长度。本文采用 BiLSTM 学习得到立场文本和立场话题的上下文表示，分别表示为 $\boldsymbol{h}_{\text{text}}=\{\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_L\}$ 与 $\boldsymbol{o}_{\text{topic}}=\{\boldsymbol{o}_1,\boldsymbol{o}_2,\cdots,\boldsymbol{o}_M\}$。

### 3.2 依存句法表示学习

已有研究大多采用顺序化处理的方式捕获相邻单词的信息[1]，很难从语言学角度学习深层次的语义信息。图 2 给出了立场文本“A face mask is necessary”的依存句法关系图，其中，“necessary”是依存句法图的中心词，“mask”是修饰“necessary”的名词性主语，“is”是修饰“necessary”的系动词，“A”是修饰“mask”的限定词，“face”是修饰“mask”的复合词。文献[7]在之前研究的基础上引入依存句法信息，它可以很好地剖析文本中各个组成成分之间的句法依赖关系，有助于更好地让模型理解和学习自然语言[8]，从语言

学层次捕捉更高级的语义信息。

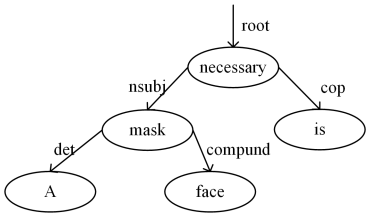


图 2　依存句法关系示例图

Fig. 2　Example of dependency syntactic relationship diagram

文本的依存句法关系是图结构信息，文献[7]采用 LSTM 无法很好地学习句法信息。为此，本研究采用图神经网络相关技术学习文本的依存句法信息。GAT[9] 是在 GCN 的基础上改进并实现的，它针对 GCN 节点聚合采用拉普拉斯矩阵静态分配权重可能导致过平滑的问题，采用注意力机制的思想，根据当前节点与邻居节点的属性关系差异，分配不同的权值，是一种优秀的图节点表征学习技术。

本研究采用 GAT 学习立场文本的依存句法表征，立场文本中每个词被看作句法图的一个节点。针对立场依存句法图中某个词 $i$，在第 $l$ 层聚合时，通过 GAT 学习得到的依存句法表示 $\boldsymbol{d}_i^l$ 代表词 $i$ 通过注意力机制融合邻居依赖词信息后的表征，可以表示为：

$$\boldsymbol{d}_i^l=\sigma\left(\sum_{j\in N_i}\alpha_{ij}\boldsymbol{W}_d\boldsymbol{d}_j^{l-1}\right) \tag{1}$$

其中，$\boldsymbol{W}_d$ 表示参数矩阵，$\sigma$ 为非线性激活函数，$N_i$ 表示与单词 $i$ 有依存句法关系的邻居依赖词集合（包括自身），$j$ 为和单词 $i$ 有依存句法关系的某邻居依赖词，$l$ 表示当前聚合层数。当 $l=0$ 时，词的初始表征为上阶段学习的 BiLSTM 隐藏层向量 $\boldsymbol{h}_i$。$\alpha_{ij}$ 表示当前词 $i$ 与邻居依赖词 $j$ 在第 $l$ 层聚合的注意力得分，代表邻居依赖词 $j$ 的权重，在 GAT 中的计算公式如下：

$$\alpha_{ij}=\frac{\exp(\mathrm{LeakyReLU}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{W}_d\boldsymbol{d}_i^{l-1}\oplus\boldsymbol{W}_d\boldsymbol{d}_j^{l-1}]))}{\sum\limits_{k\in N_i}\exp(\mathrm{LeakyReLU}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{W}_d\boldsymbol{d}_i^{l-1}\oplus\boldsymbol{W}_d\boldsymbol{d}_k^{l-1}]))} \tag{2}$$

其中，$\oplus$ 表示特征向量的拼接操作，$\boldsymbol{W}_d$ 与 $\boldsymbol{a}^{\mathrm{T}}$ 都是参数矩阵，LeakyReLU 为非线性激活函数。

### 3.3　结合话题信息的依存句法嵌入表示

立场文本中每个词对于立场话题有着不同的重要性，文献[3-4]采用话题注意力机制捕捉立场文本每个词与立场话题表示之间的关系，从而生成更好的立场文本表征，然而它们直接进行词向量级别的注意力计算，缺乏语义上下文语境信息。为此，本研究采用 BiLSTM 学习立场话题与立场文本的上下文信息，将话题上下文表示的平均池化输出作为最终立场话题的上下文表示 $\boldsymbol{f}_{\mathrm{topic}}$，如式(3)所示。

$$\boldsymbol{f}_{\mathrm{topic}}=pooling(\boldsymbol{o}_1,\boldsymbol{o}_2,\cdots,\boldsymbol{o}_M) \tag{3}$$

在上下文编码的基础上，本文采用缩放点积注意力[10]来实现话题注意力计算，该机制通过将向量点积的结果除以一个输入向量维度的平方根，缓解特征维度较高时直接点积计算结果太大或者太小的问题，提高计算数值的稳定性。立场话题与立场文本每个词 $i(1\leqslant i\leqslant L)$ 上下文表示之间的缩放点积为 $\alpha_i$，其计算公式如式(4)所示。

$$\alpha_i=\frac{\boldsymbol{f}_{\mathrm{topic}}^{\mathrm{T}}\boldsymbol{h}_i}{\sqrt{d_{h_i}}} \tag{4}$$

其中，$d_{h_i}$ 表示 $\boldsymbol{h}_i$ 的特征维度。

将立场文本得到的每个词的缩放点积 $\alpha_i$ 输入 softmax 函数进行归一化处理，得到文本中词 $i$ 的话题注意力得分 $s_i$，其计算公式如下：

$$s_i=\mathrm{softmax}(\alpha_i)=\frac{\mathrm{e}^{\alpha_i}}{\sum\limits_{j=1}^{L}\mathrm{e}^{\alpha_j}} \tag{5}$$

采用立场文本中每个词的依存句法表征与对应的话题注意力乘积的加权和表示立场文本最终的表征 $\boldsymbol{f}_{\mathrm{stance}}$，其计算公式如式(6)所示：

$$\boldsymbol{f}_{\mathrm{stance}}=\sum_{i=1}^{L}s_i\boldsymbol{d}_i \tag{6}$$

将立场文本表征输入分类函数 softmax，得到模型的预测概率向量。采用交叉熵作为模型训练的损失函数，用于度量预测标签与真实标签的差异。

## 4　立场文本数据增强

深度学习模型的训练需要较多的训练数据，但是公开的立场话题研究数据集普遍规模较小，每个话题普遍只含有不到 2000 条数据样本[11-12]，导致模型鲁棒性较差，影响模型的性能。数据增强是一种在不直接收集数据的情况下，通过将原数据进行一些变换或者重组来生成新样本的技术。Wei 等[13]提出了同义词替换（随机将部分句中单词进行同义词替换）、随机插入（随机在句中任意位置插入非停用词的同义词）、随机交换（随机交换两个单词在句中的顺序）和随机删除（按一定概率删除句中单词）4 种常见的文本数据增强。后 3 种方式引入的噪声信息可能导致立场文本中与话题相关的内容出现语义损失，导致文本立场发生变化，不适用于立场数据增强；而同义词替换通过大型同义词库将文本中部分单词进行同义词替换，在生成更多样本，在丰富数据集的同时，尽量保证文本的语义完整性，适用于文本立场分析数据增强。然而传统的同义词替换技术随机从同义词库中抽取同义词进行替换，没有考虑到语义正确性的问题，盲目将不同词性的同义词对原词进行替换可能会导致语义的错误，比如单词“kind”在英文中根据上下文有名词和形容词两种词性，如果将名词词性下的原词替换为形容词词性下的同义词，就不符合语义。本文将文本进行词性分析，然后采用 WordNet[14] 作为同义词库，它是一个公开的大型英文词汇同义词库，提供了大量英文单词在不同词性下的同义词集合，可以保证同义词替换的词性正确性。此外，不同同义词和原词的语义相似度也不同，WebVectors[15] 可以提供某个英文单词在某种词性限制下的词嵌入表示，本文采用它来计算不同同义词与原词的语义相似度，选取语义相似度最高的同义词对原词进行替换，保证替换同义词的高质量。最终，本文设计实现了一种基于 WordNet 同义词库与 WebVectors 词嵌入模型的同义词替换数据增强方案——WWDA，保证了同义词替换过程的词性正确性和语义相似性。

首先通过 StanfordCoreNLP 对立场文本进行词性标注，找出动词、名词、形容词、副词词性下的所有单词。随机从非停用词中抽取 $p\%$ 的单词作为同义词替换候选集 $Candidates=\{word_1,word_2,\cdots,word_{CL}\}$，其中 $CL$ 表示候选集的长度。针对 $Candidates$ 中某个替换候选词 $word_i(1\leqslant i\leqslant CL)$，采用 WordNet 找到其在对应词性下的同义词集 $SimWords=\{simword_1,simword_2,\cdots,simword_{SL}\}$，其中 $SL$ 表示同义词集的长度。采用 WebVectors 获得 $word_i$ 和同义词集中每个同义词的词向量表示，$WV(word_i)$ 代表 $word_i$ 的词向量表示，

$WV(simword_j)$代表$simword_j$的词向量表示，并采用向量点积计算它们的语义相似度，用相似度最高的同义词对原词进行替换，并依次对剩余的候选词进行操作，最终生成新的数据增强句子。

## 5 验证实验和结果分析

### 5.1 实验设计

#### 5.1.1 数据集与数据预处理

实验采用 SemEval2016[11] 和 Covid-19[12] 两个立场数据集进行性能对比。SemEval2016 数据集采用的 4 个话题分别为 Atheism(AT)，Feminist Movement(FM)，Legalization of Abortion(LA)和 Hillary Clinton(HC)，共含有 3599 个样例。Covid-19 数据集采用的 4 个话题分别为 Fauci(FC)，Keeping Schools Closed(KC)，Stay at Home Orders(SO)和 Wearing a Face Mask(WM)，共含有 6133 个样例。文本的立场标签包括“FAVOR”“AGAINST”“NONE”。

数据集采用的评估指标为 $F_1$ 值，但具体计算有一定差异。Covid-19 数据集发布者考虑了 3 类标签的算术平均值[12]，SemEval2016 数据集发布者只考虑“FAVOR”和“AGAINST”两者的 $F_1$ 值的均值作为最终评估指标[11]。

实验过滤立场文本中的特殊符号、URL 链接、多余空格等冗余信息，把大写字母转化为小写字母。本文针对英文中常见的缩写单词和含有连续重复字母的单词进行还原，最终得到更高质量的立场文本。

#### 5.1.2 参数设置

实验选用的 CPU 为 Intel i7-10875H，GPU 为 NVIDIA GeForce RTX 2060 6G，运行内存为 16GB，实验语言为 Python，选用 Pytorch 深度学习框架。本文选用的词嵌入模型是谷歌公司提供的 Word2Vec 预训练模型，其中每个词向量维度为 300。在模型训练过程中，batch size 设置为 16，学习率设置为 0.0005，dropout 设置为 0.5，LSTM 隐藏层单元数设置为 200，GAT 层数设置为 2。采用 Stanford CoreNLP 提取文本的依存句法信息。

### 5.2 实验结果与分析

#### 5.2.1 对比实验

本文模型与以下出色的基线模型在预处理后的数据集上进行对比实验，性能对比结果如表 2 所列。

1)TAN[3]：基于词向量级别的话题注意力机制与 BiLSTM 的立场分析模型。

2)ATA[4]：基于词向量级别的两段注意力机制与 BiLSTM 立场分析模型。

3)BiLSTM-CNN-ATT[5]：基于 BiLSTM-CNN 和注意力池化的立场分析模型。

4)Joint[6]：基于多任务学习的立场分析模型，将情感分析与立场分析共同训练，将文本的情感信息融入立场分析模型中。

5)HAN[7]：基于多语言学表征的分层注意力立场分析模型。

表 2 对比实验

Table 2 Comparison experiments

| 模型 | SemEval2016 数据集 | | | | Covid-19 数据集 | | | |
|---|---|---|---|---|---|---|---|---|
| | AT | FM | HC | LA | WM | FC | KC | SO |
| TAN | 59.30 | 56.03 | 62.36 | 64.72 | 58.92 | 65.32 | 55.36 | 61.20 |
| ATA | 60.29 | 57.55 | 63.30 | 64.89 | 59.64 | 65.73 | 54.63 | 64.18 |
| BiLSTM-CNN-ATT | 65.49 | 56.46 | 65.26 | 65.69 | 64.44 | 66.03 | 54.87 | 65.46 |
| Joint | 66.43 | 59.31 | 63.70 | 64.98 | 65.53 | 65.59 | 55.95 | 68.15 |
| HAN | **69.49** | 57.40 | 62.79 | 67.04 | 65.99 | 65.75 | 56.38 | 68.06 |
| AT-BiLSTM-GAT | 68.53 | **59.55** | **67.72** | **68.73** | **67.09** | **66.84** | **57.99** | **70.82** |

#### 5.2.2 对比实验结果分析

如表 2 所列，本文设计实现的 AT-BiLSTM-GAT 综合表现相比相关研究取得了更好的性能。与 TAN，ATA 相比，AT-BiLSTM-GAT 在话题注意力计算过程中采用 BiLSTM 建模立场话题与立场文本的上下文语义信息，采用缩放点积注意力学习立场文本中的重要内容，同时采用 GAT 学习文本依存句法信息，从而提升了综合性能。虽然 BiLSTM-CNN-ATT 融合了 TextCNN 的局部信息与 BiLSTM 的时序上下文特征，但未能从句法结构角度提取文本语义信息，综合表现不如文本模型。Joint 模型学习了文本的情感信息，但本文设计实现的 AT-BiLSTM-GAT 模型从句法结构角度让模型捕捉了语言学关联的单词信息，同时结合话题注意力机制，实现了更优的检测性能。HAN 模型学习了立场文本的情感、句法与辩论关系，相比本文模型提取了更丰富的特征信息，额外提取文本的情感与辩论关系信息，在 AT 话题上性能最佳。但 HAN 采用 LSTM 不适合提取图结构的句法信息，而本文采用 GAT 可以很好地建模图结构的文本依存句法信息，此外考虑实现了话题注意力机制，综合性能更优。

#### 5.2.3 数据增强方案测试

本文设计了一个实验来检验 WWDA 数据增强方案的有效性，$p$ 设置为 50。在 SemEval2016 数据集和 Covid-19 数据集进行数据增强，生成多一倍的增强样例，应用 AT-BiLSTM-GAT 模型分析文本立场，性能如表 3、表 4 所列，可以看到 WWDA 数据增强方案提升了立场分析的性能。

表 3 SemEval2016 数据集上 WWDA 方案测试($F_1$)

Table 3 WWDA tested on SemEval2016 dataset($F_1$)

| 方案 | AT | FM | HC | LA |
|---|---|---|---|---|
| 数据增强前 | 68.53 | 59.55 | 67.72 | 68.73 |
| 数据增强后 | **69.11** | **59.96** | **68.32** | **69.19** |

表 4 Covid-19 数据集上 WWDA 方案测试($F_1$)

Table 4 WWDA tested on Covid-19 dataset($F_1$)

| 方案 | WM | FC | KC | SO |
|---|---|---|---|---|
| 数据增强前 | 67.09 | 66.84 | 57.99 | 70.82 |
| 数据增强后 | **67.35** | **67.74** | **58.33** | **71.67** |

**结束语** 本文设计实现了 AT-BiLSTM-GAT 立场分析模型，针对已有研究不能很好地捕捉文本的依存句法结构信息的问题，本文构建文本的依存句法关系图，在文本上下文

语义基础上,采用 GAT 模型从句法层次提取了更优秀的立场文本表示。针对当前研究话题注意力机制缺乏上下文语义建模的问题,本文采用 BiLSTM 建模立场话题和立场文本的上下文语义信息,通过缩放点积注意力计算话题与文本各个词的上下文表示之间的关系,从而更好地捕捉立场文本中的重要内容。针对当前立场分析研究存在数据集规模小导致模型鲁棒性较弱的问题,设计实现一种数据增强方案 WWDA,保证了同义词替换过程的词性正确性和语义相似性。实验结果表明,本文设计实现的 AT-BiLSTM-GAT 立场分析模型相比以往研究有一定的性能提升,设计实现的 WWDA 数据增强方案可以提升立场分析模型的鲁棒性,从而提升模型的检测性能。

后续研究可构建合适的多模态立场分析数据集,集成图片、音频和文字特征,促进多模态立场分析的研究。同时本研究设计实现的数据增强方案未来可以考虑引入更多大型同义词库,构建覆盖率更高的同义词集合,在替换过程中找到更合适的替换词,并将数据增强技术运用到其他研究上。

## 参 考 文 献


[1] LI Y,SUN Y Q,JING W P. Summary of Text Stance Detection [J]. Journal of Computer Research and Development, 2021, 58(11):2538-2557.

[2] LIU W,PENG X,LI C,et al. A Survey on Stance Detection[J]. Journal of Chinese Information,2020,34(12):1-8.

[3] DU J,XU R,HE Y,et al. Stance classification with target-specific neural attention networks[C]// International Joint Conferences on Artificial Intelligence. 2017.

[4] YUE T C,ZHANG S W,YANG L,et al. A stance detection method based on two-stage attention mechanism[J]. Journal of Guangxi Normal University(Natural Science Edition), 2019, 37(1):42-49.

[5] BAI J,LI F,JI D H. Attention-based BiLSTM - CNN Chinese Weibo Stance Detection Model [J]. Computer Applications and Software,2018,35(3):266r274.

[6] SUN Q,WANG Z,LI S,et al. Stance detection via sentiment information and neural network model[J]. Frontiers of Computer Science,2019,13(1):127-138.

[7] WANG Z,SUN Q,LI S,et al. Neural Stance Detection With Hierarchical Linguistic Representations [J/OL]. IEEE/ACM Transactions on Audio,Speech,and Language Processing,2020, 28. https://ieeexplore.ieee.org/abstract/document/8949710.

[8] WU L,CHEN Y,SHEN K,et al. Graph neural networks for natural language processing:A survey[J]. arXiv:2106.06090, 2021.

[9] VELICKOVIC P,CUCURULL G,CASANOVA A,et al. Graph attention networks[J]. arXiv:1710.10903,2017.

[10] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[J/OL]. Advances in Neural Information Processing Systems,2017,30. https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

[11] MOHAMMAD S,KIRITCHENKO S,SOBHANI P,et al. Semeval-2016 task 6:Detecting stance in tweets[C]// Proceedings of the 10th international workshop on semantic evaluation(SemEval-2016). 2016:31-41.

[12] GLANDT K,KHANAL S,LI Y,et al. Stance detection in COVID-19 tweets[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1:Long Papers). 2021:1596-1611.

[13] WEI J,ZOU K. EDA:Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019:6382-6388.

[14] MILLER G A. WordNet:a lexical database for English [J]. Communications of the ACM,1995,38(11):39-41.

[15] KUTUZOV A,FARES M,OEPEN S,et al. Word vectors,reuse,and replicability:Towards a community repository of large-text resources[C]// Proceedings of the 58th Conference on Simulation and Modelling. Linköping University Electronic Press, 2017:271-276.


**KANG Shuming**, born in 1998, postgraduate. His main research interests include stance detection and natural language processing.

**ZHU Yan**, born in 1965, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include data mining, computational network analysis, and big data.