

文章编号: 1003-0077 (2020) 00-0000-00

PRBDN: 基于预训练的微博评论情感分类模型

雷鹏斌, 秦斌*, 王志立, 吴宇凡, 梁思怡, 陈宇

(1. 深圳大学 信息中心, 广东 深圳 518061; 2. 深圳大学 电子与信息工程学院, 广东 深圳 518061)

摘要: 文本情感分类是自然语言处理中的经典任务, 在判断文本的情感极性、舆情监控、市场呼声、商品评论等领域有重要的应用价值。本文提出了一种基于预训练模型进行细粒度文本情感分类的新方法。基于文章级别的情感分类任务, 需要模型同时具有较高的语义概括能力和抗噪能力。为此, 我们利用 BiLSTM 网络对预训练模型中每层 Transformer 的权重进行调整, 通过将各层表达的不同粒度的语义表征进行动态融合的方式, 提高模型的语义空间表达能力。为了增强模型的泛化性能, 我们在下游任务结合 BiLSTM 和 BiGRU 等结构对得到的语义向量进行特征过滤。利用该模型, 我们在 CCF2020 年举办的科技战疫·大数据公益挑战赛-疫情期间网民情绪识别赛道中取得第三的成绩, 最终测试集的 F1 值为 0.74537, 与第一名相差仅为 0.0001, 该方法具备可行性与有效性。

关键词: 情感分析; 预训练模型; 动态权重; 长短期记忆网络

中图分类号: TP391

文献标识码: A

PRBDN: A Pretraining-based Model of Weibo Comment Emotion Classification

Pengbin Lei, Bin Qin *, Zhili Wang, Yufan Wu, Siyi Liang, Yu Chen

(1. Information Center, Shenzhen University, Shenzhen, Guangdong 518061, China; 2. School of Electronic and Information Engineering, Shenzhen University, Shenzhen, Guangdong 518061, China)

Abstract: Text sentiment classification is a classic task in natural language processing, and has important application value in the fields of judging text sentiment polarity, public opinion monitoring, market voices, and product reviews. This paper proposes a new method for text sentiment classification based on the pre-training model. By using the BiLSTM network to dynamically adjust the output weight of the Transformer of each layer of the pre-training model, the resulting text representation vectors are filtered using features such as BiLSTM and BiGRU. Finally, the classifier gives the emotional tendency of the text. By using the model, the F1 value of the final test set of the data set of Internet users' emotion recognition during the epidemic period held in 2020 is 0.74537, ranking third, which proved the feasibility and effectiveness of this method.

Key words: text emotion classification; pre-training model; dynamic weight; long short-term memory

0 引言

新型冠状病毒 (COVID-19) 感染的肺炎疫情牵动着全国人民的心, 众多网民参与疫情相关话题的讨论。一个更有效的文本情感分类模型能帮助政府掌握真实社会舆论情况、科学高效地做好防控宣传和舆情引导工作, 具有重要的研究价值和社会意义。

如何对一段文本得到更适合的文本向量表征, 是文本情感分类的重中之重。2014 年 Pennington 等提出了基于词共现矩阵的 Glove^[1]模型, 通过统计词与词共同出现的频次构建相关性矩阵, 利用文本全部信息训练词向量。Huang 等^[2]提出一词多义的训练方法, 分析处理汉字一词多义的情况, 使用位置信息和全局信息学习多义词的词向量表示。Chen 等^[3]在中文上提出加入

字向量的字词联合训练方法,将组成单词的汉字信息联合单词本身一起训练得到中文词向量。Peters 等^[4]使用双向 LSTM 模型学习词的上下文信息,提出了 ELMO 模型,该模型训练的词向量有效的学习了句法、语法信息,提升了各项 NLP 的任务性能。Devlin 等^[5]使用 Transformer 特征提取器,引入 Masked language model 和 Next sentence predict 的预训练策略发布了预训练模型 BERT。在 11 个经典的阅读理解任务里取得了 Sota 的成绩,之后针对中文数据 Liu Y, Ott M, Goyal N 等人改进了 BERT 的预训练策略,发布了中文预训练模型 RoBERTa^[6],在中文数据集上往往有着更好的表现。至今的大部分 NLP 任务都是基于 BERT 的下游任务,但针对不同任务,BERT 的向量表征存在着一定的改进空间。具体而言,本文的贡献可归纳如下:

- 通过对预训练模型 RoBERTa 的多层 Transformer 进行动态加权融合与特征增强,得到不同粒度的文本表征,提高模型的语义空间表达能力;
- 在预训练模型基础上开发了一种面向下游目标任务的新结构进行文本情感分类;
- 通过真实场景下的疫情情绪数据集对提出的模型进行了评估,实验结果证明了本文提出的动态权重策略和下游任务结构的有效性。

1 相关工作

目前文本情感分类技术主要分为三类,基于情感字典的方法,基于机器学习和深度学习的方法。当前最常用的情感词典是 WordNet 网络的 SentiWordNet 组件。该组件支持了包括中文在内的世界主要语言的情感词汇,但其中的英文词汇版本最为强大^[7];国内部分研究人员也建立了中文专用的情感分析词库如 SnowNLP^[8],情感字典为每个情感词建立了三种极性:正面(Positive)、负面(Negative)、中立(Objectivity),通过对文本中出现情感词的个数和极性进行加权计算得到情感类别。

基于机器学习的方法主要使用支持向量机(Support Vector Machine, SVM)、逻辑回归

(Logistic Regression, LR)、贝叶斯网络(Bayesian Network)等传统分类器。Saqib Iqbal 调查了社区媒体数据的情感分析相关研究,结果表明朴素贝叶斯网络、支持向量机等分类器都是非常适用于社区媒体文档的情感分析方法,其中支持向量机的分类准确率最高^[9]。Kathrin Grosse 等提出一个新的用于 Twitter 评论的情感分析框架,该框架能增量式地同时构建一个“观点树”和一个“冲突树”,并解决评论中普遍存在的情感矛盾等问题。该框架的另一大特点是能够增量学习,因此能够用于 Twitter 数据的实时分析^[10]。

近年来,深度学习广泛应用于文本情感分析中。Kim 等人^[11]提出使用卷积神经网络进行句子建模,解决情感分析任务,在多个数据集上取得了很好的结果。Irsoy 等人^[12]提出使用循环神经网络以及长短期记忆网络模型用于情感分析,同时 Tai 等人^[13]证实该模型能够对情感预测分析的触点进行性能提升。Attention 模型由 Bahdanau 等人^[14]提出,最先被用于机器翻译,随后 Attention 的变体形式被广泛应用于 NLP 的各个领域。曲昭伟和王源等人提出一种分层 Attention 网络情感分析算法,相比传统循环神经网络模型该算法提升了将近 8 个百分点。

2018 年后,基于深层神经网络如 ULMFiT^[15]、OpenAI-GPT^[16]、ELMo^[17]和 RoBERTa 等预训练模型被先后提出,它们都是通过上下文语境语境的联合条件和大量的语料预训练进行下游迁移学习任务,在多个公共数据集都取得了更好的成绩,随后成为解决文本情感分类任务的主流方式。在我们的工作主要依赖于 Liu 等人的中文预训练模型 RoBERTa,它为我们提供了一个强大的基线,在此基础上,我们深入探索了新的网络结构和下游任务的方式。

2 疫情期间网民情绪识别

问题定义 疫情期间网民情绪识别是一个文本情感三分类任务,如表 2.1 所示。正文给定了微博 ID 和微博内容,因此其任务就是对微博内容进行情绪分类,判断它的情感倾向。

数据介绍 数据集内容与“新冠肺炎”相关的 230 个主题关键词,时间跨度为 2020 年 1 月 1 日至 2020 年 2 月 20 日期间共计 100 万条微博数据,

并对其中 10 万条数据进行人工标注, 标注分为三类, 分别为: 1 (积极), 0 (中性) 和-1 (消极)。

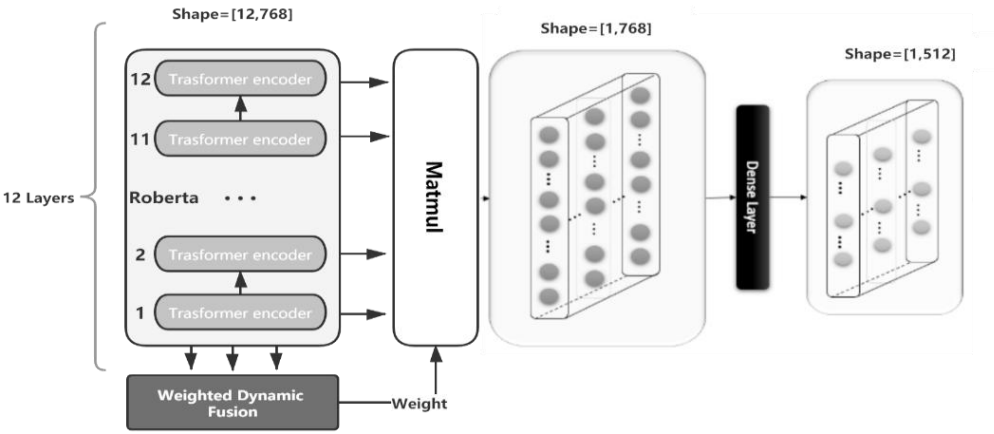


图 3.1.1 PRBDN 多层输出动态表征

表 2.1 疫情期间网民情绪识别训练集实例	
[微博 ID]	4456072029125500
[微博中文内容]	写在年末冬初孩子流感的第五天，我们仍然没有忘记热情拥抱这 2020 年的第一天。带着一丝迷信，早晨给孩子穿上红色的羽绒服羽绒裤，祈祷新的一年，孩子们身体康健。仍然会有一丝焦虑，焦虑我的孩子为什么会过早的懂事，从两岁多开始关注我的情绪，会深沉地说：妈妈，你终于笑了！这句话像刀子一样扎入我?展开全文
[情感倾向]	0

表 2.2 疫情期间网民情绪识别数据集规模

训练集（有标签）	10w（条）
训练集（无标签）	100w（条）
测试集	1w（条）

评估准则 疫情期间网名情绪识别的评测指标为 Macro-F1，其公式如式（1）~（3）所示。其中，TP 是真样例，FP 是假阳例，FN 是假阴例，通过该公式得到该类 F1 值，将每一类 F1 值求平均，即得到 Macro-F1 值。

$$P = \frac{TP}{TP+FP} \tag{1}$$

$$R = \frac{TP}{TP+FN} \tag{2}$$

$$F1 = \frac{2 \times P \times R}{P+R} \tag{3}$$

3 模型

本文改进 RoBERTa 下游任务的方式可以分为两个改进子模块，分别是 RoBERTa 的 Transformer 语义精细化表征和下游任务语境过滤。

针对疫情网民情绪识别的特定下游任务，我们基于预训练模型 RoBERTa 在该任务上的不足进行改进与优化，并提出了新的模型结构，称为 PRBDN(Pretrain RoBERTa Bi-Dynamic Network)。

3.1 Transformer 语义精细化表征

Ganesh Jawahar 等人^[18]曾探寻过 BERT 的多层 Transformer 中提到过 BERT 每一层学到了什么。他通过一系列的实验证明 BERT 学习到了一些结构化的语言信息，如 BERT 的低层网络学习到了短语级别的信息表征，BERT 的中层网络学习到了丰富的语言学特征，而 BERT 的高层网络则学习到了丰富的语义信息特征。而在基于文章级别的情感分析任务中，需要在保留不同粒度语义属性（词、语法、句法）的前提下，对文本内容进行高度概括与抽象表达^[19]。因而，仅使用 Transformer 的最后一层这种偏向于句法粒度的向量输出并不能满足情绪分析任务对内容表征的需求。为此我们提出了一种根据各层过滤信息获取权重的动态输出层的方法，对 RoBERTa 结构进行如下调整，详情如图 3.1.1 所示。

我们通过将每一层的 Transformer 的输出进

行加权相加, 得到一个语义表征向量 R_1 。再将改向量降至 512 维度, 其如公式 (4) 所示, 得到一个相较于经典 RoBERTa 最后一层拥有更丰富的语义信息表征 R_0 。

$$R_0 = Dense_{unit=1}(\sum_{i=0}^n w_i \times Transformer_i) \quad (4)$$

上式 $Transformer_i$ 为 RoBERTa 中每一层的输出, 为我们经过过滤后训练得到的动态权重。如图 3.1.2 所示。

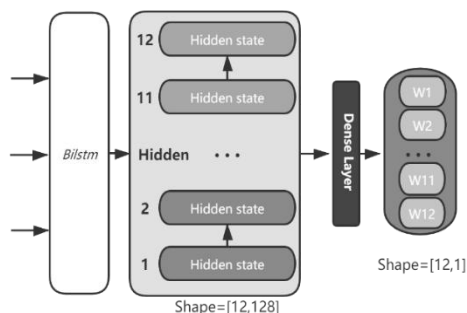


图 3.1.2 RoBERTa 多层输出动态权重 (Weighted Dynamic Fusion)

我们提取每一层的输出输入至双向 BiLSTM 中进行筛选得到精炼信息 H_i 。如公式 (5) 所示, 并将过滤的信息输入至多个全连接层, 得到每一层输出的权重 w_i , 如公式 (6) 所示。

$$H_i = BiLSTM_{unit=64} Transformer_i \quad (5)$$

$$w_i = Dense_{unit=1} H_i \quad (6)$$

3.2 下游任务语境过滤

文本分类中用 TextCNN^[20]、LSTM 等模型结构下接 RoBERTa 往往能提升文本分类模型的性能。在下游任务中, 我们分别外接了 BiLSTM 与 BiGRU 作为语境信息过滤, 得到表达文本情感的特征向量, 如图 3.2.1 所示。

此外对于得到过滤的信息, 我们动态的利用了平均池化和最大池化的结果。对经过 BiLSTM&BiLSTM 层过滤的语义向量, 我们分别将其通过平均池化层和最大池化层。得到了两中不同的特征表达。我们复用了动态获取 Transformer 权重的策略, 动态训练两个权重和来对两种池化策略进行融合, 把这种池化的策略称作 Dympooling, 如公式 (7) 所示, 得到任务最终的情感表征向量, 如公式 (8) 所示, 其中 $R_i \in$

$(R_{max}, R_{average})$ 。

$$w_i = Dense_{unit=1} H_i \quad (7)$$

$$R_e = w_i \times R_{max} + w_2 \times R_{average} \quad (8)$$

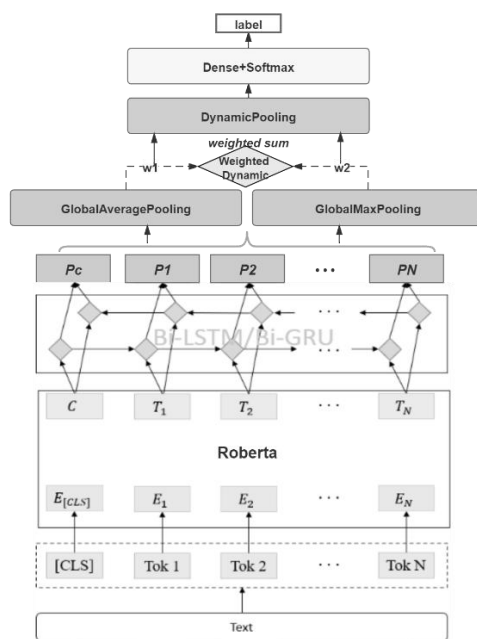


图 3.2.2 RoBERTa BiLSTM & BiGRU-Dympooling

4 实验

4.1 数据处理

实验所使用的数据为上文提及的疫情期间网民情绪识别比赛提供的 10w 条已标注数据 100w 条未标注数据。其中已标注数据文本长度如下图 4.1 所示。

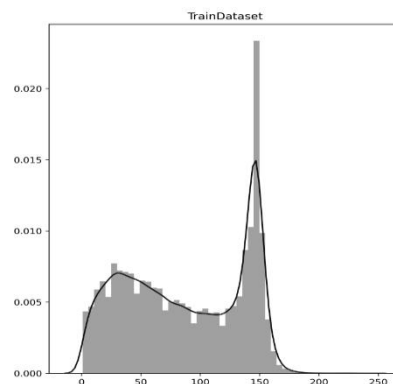


图 4.1-a 已标注测试集文本长度分布

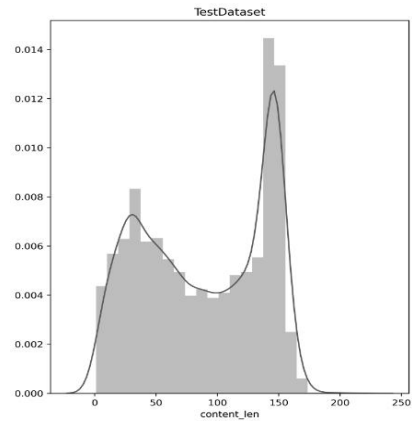


图 4.1-b 已标注训练集文本长度分布

我们将 100w 条的无标签数据设置最大长度为 256 的文本进行分割, 切分成更多的语料继续预训练 RoBERTa_Base 模型, 以加强预训练模型的泛化能力。

因为训练集和测试集的文本经过去噪处理之后, 得到的数据正文文本长度都在 200 以内, 根据上图 4.1 的长度分布, 我们将 Max Len 调整为 150, 加快模型的训练与推断的速度。

4.2 参数设置

实验设置的超参数如表 4.2 所示, 每一个双向 LSTM 的隐藏节点均为 256, BERT 微调 Adam^[21]使用的学习率为 0.00002, 下游结构的学习率为 0.0001。Dropout^[22]层的丢弃率为 0.1, 训练的 Batch_Size 为 32, 经实验表明, 我们的模型能在前四个 epoch 内快速收敛。

表 4.2 模型参数设置

参数	值
Batch Size	32
Epoch	4
Max Length	150
Dropout Rate	0.1
BiLSTM HiddenSize -Transformer 权重	64
BiLSTM&BiGRU HiddenSize-后接结构	256
BiLSTM Layer Numberts	2
RoBERTa-finetune-learning Rate	2e-5
下游结构 Learning Rate	1e-4

4.3 多分类阈值搜索

经统计, 训练集中标签情感倾向为消极、中性、积极的比例为 3:12:5, 存在样本不均衡问题。我们的模型结构在训练过程中反向传播优化 loss 时, 会给标签为 0 (中性) 的样本更高的分数从而导致模型会更倾向于给出更多预测为中性情感的标签。为此我们通过多分类 F1 搜索, 将训练得到的模型对验证集进行线下预测, 学习出三个类别的权重比例使得预测值与真实值之间的 F1 最大化, 并将此权重用于测试集的预测, 得到了更好的结果。

4.4 评测结果

对于模型的验证, 我们使用 RoBERTa 作为基线, 对比了文本分类常用的下接 TextCNN 与我们提供的模型以比赛方评定的 A 版 F1 作为指标。最终的结果对比如下表 4.3 所示。实验表明我们提出的动态权重策略与下游结构在疫情期间网民情绪识别中相较于直接使用 RoBERTa 的 CLS 分类向量单模提高了 2.9 个百分点, 模型融合的效果较 RoBERTa 单模提高了 3.6 个百分点, 有着更好的表现。

表 4.3 结果对比表

模型	F1 (比赛 A 榜测试)
RoBERTa	0.706
RoBERTa-TextCNN	0.708
RoBERTa-BiLSTM	0.710
PRBDN (动态融合)	0.728
PRBDN (动态融合)+ 阈值搜索	0.739
PRBDN 多模融合	0.746

4.5 模型效果

我们进一步分析了本文提出的模型效果, 并可视化文本向量以显示 PRBDN 的优势。如图 4.5 经过基线 RoBERTa 和 PRBDN 分别输出的文本向量利用 Tsne^[23]降维后可视化。我们随机选择了 900 个训练集数据, 分别提取了使用 RoBERTa 的 CLS 分类的特征向量与利用 PRBDN 分类模型 MLP 前的特征向量, 使用 Tsne 降至二维向量绘制散点图进行了可视化对比。可以看到, 由 PRBDN 输出的文本向量具有更高的区分度, 充分证明了我们的模型能够更有效的提炼出与情感任务相关的高层语义向量。

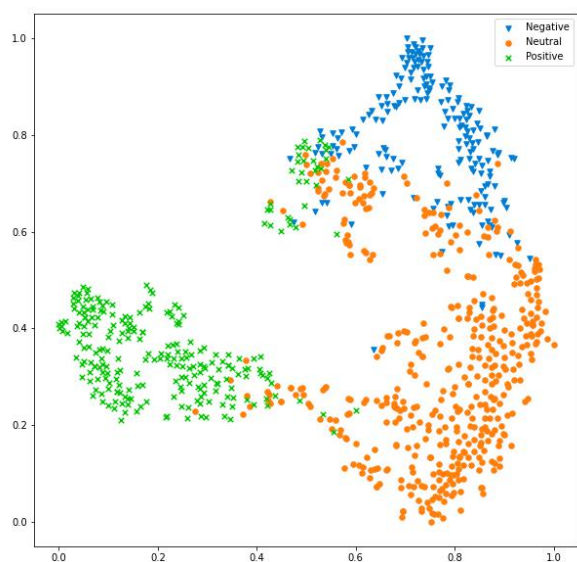


图 4.5-a RoBERTa CLS 特征向量可视化

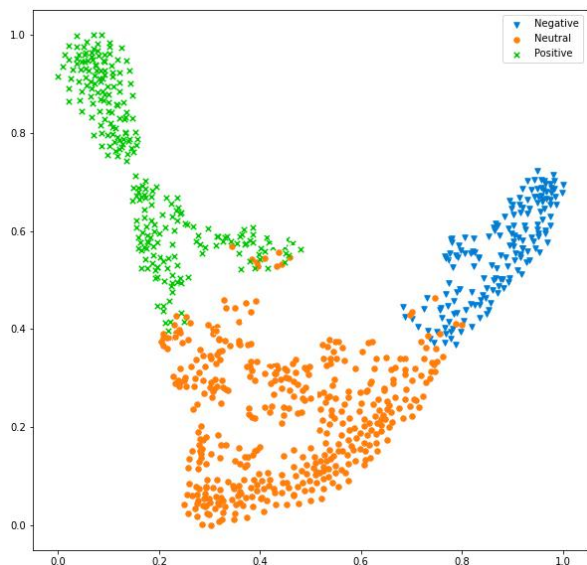


图 4.5-b PRBDN MLP 前特征向量可视化

5 结论

针对疫情情绪识别这一任务, 本文提出了一种新的基于 RoBERTa 预训练模型的下游任务的文本情感分类模型, 该模型能每一层的 Transformer_Encoder 的输出进行信息筛选然后再进行语义整合, 形成一种更精细的语境化语言表示。之后再通过双向 LSTM&GRU 和动态的池化层进行信息过滤, 得到和每一句话的情感语义表征。实验表明, 在现有语料和超参数空间下, 本模型能达到不错的效果。同时除了模型本身外, 针对特定任务的数据处理, 结果优化等方法对任务性能的提升也同样不可忽视。

6 不足与展望

大量实验^[24, 25]证明无监督数据的有效利用会大大提高预训练模型的语义理解与推理能力。本文中未对无监督的数据使用伪标签策略, 即把预测结果中多数模型都一致认为的类别当做其正确的标签, 然后将该标签及其对应的内容进行增量训练, 因此在性能上并未达到最优。此外对 BERT 模型进行剪枝和蒸馏, 以降低模型的时间与空间复杂度, 也是我们未来的研究目标之一。

7 致谢

衷心感谢 CCF 以及赛事方提供的疫情期间网民情绪的微博文本数据和广东省深圳大学-深信服联合培养研究生基地提供的实验环境。

参考文献

- [1] Pennington J, Socher R, Manning C D, et al. Glove: Global Vectors for Word Representation[C]// Empirical methods in natural language processing, 2014: 1532-1543.
- [2] Huang E H, Socher R, Manning C D, et al. Improving Word Representations via Global Context and Multiple Word Prototypes[C]// Meeting of the association for computational linguistics, 2012: 873-882.
- [3] Chen X, Xu L, Liu Z, et al. Joint learning of character and word embeddings[C]// Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015.
- [4] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[J]. North American Chapter of the Association for Computational Linguistics, 2018, (1):2227-2237.
- [5] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [6] Liu Y, Ott M, Goyal N, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach[J]. 2019.
- [7] Learning lexical scales: wordnet and sentiwordnet [EB/OL]. <http://compprag.christopherpotts.net/wordnet.html>, June 5, 2017
- [8] SnowNLP: Python library for processing Chinese text [EB/OL]. <https://pypi.python.org/pypi/snownlp>, June 13, 2017
- [9] Iqbal S, Zulqurnain A, Wani Y, et al. The survey of sentiment and opinion mining for behavior analysis of social media[J]. arXiv preprint arXiv:1610.06085, 2015.
- [10] Grosse K, González M P, Chesñevar C I, et al. Integrating argumentation and sentiment analysis for mining opinions from Twitter[J]. AI Communications, 2015, 28(3): 387-401.

- [11] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.
- [12] Irsoy O, Cardie C. Deep recursive neural networks for compositionality in language[C]//Advances in neural information processing systems. 2014: 2096-2104.
- [13] Chakraborty K, Bhattacharyya S, Bag R, et al. Comparative sentiment analysis on a set of movie reviews using deep learning approach[C]//International Conference on Advanced Machine Learning Technologies and Applications. Springer, Cham, 2018: 311-318.
- [14] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [15] Howard J, Ruder S. Universal Language Model Fine-tuning for Text Classification[J]. 2018.
- [16] Radford A, Marasimhan K. Improving Language Understanding By Generative Pre-Training[J]. 2018.
- [17] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[J]. 2018.
- [18] Ganesh Jawahar, Benoît Sagot, Djamé Seddah. What does BERT learn about the structure of language?. ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Jul 2019, Florence, Italy. fhal-02131630
- [19] Wang Z, Wu Y, Lei P, et al. Named Entity Recognition Method of Brazilian Legal Text based on pre-training model[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1550(3): 032149.
- [20] Zheng S., Yang M. (2019) A New Method of Improving BERT for Text Classification. In: Cui Z., Pan J., Zhang S., Xiao L., Yang J. (eds) Intelligence Science and Big Data Engineering. Big Data and Machine Learning. ISIDE 2019. Lecture Notes in Computer Science, vol 11936. Springer, Cham
- [21] Kingma D, Ba J. Adam: A Method for Stochastic Optimization[J]. Computer ence, 2014.
- [22] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting, 2014
- [23] Laurens V D M, Hinton G. Visualizing Data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(2605):2579-2605.
- [24] Ben Athiwaratkun, Marc Finzi, Pavel Izmailov, and Andrew Gordon Wilson. There are many consistent explanations of unlabeled data: Why you should average. 2018.
- [25] Xie Q, Dai Z, Hovy E, et al. Unsupervised Data Augmentation for Consistency Training[J]. 2019.



雷鹏斌(1997—), 硕士, 主要研究领域为大数据与网络空间安全。
E-mail: szuleipengbin@foxmail.com



秦斌(1972—), 通信作者, 高工, 主要研究领域为大数据技术。
E-mail: qinbin@szu.edu.cn



王志立(1995—), 硕士, 主要研究领域为大数据与网络空间安全。
E-mail: realchilewang@foxmail.com