**苏州大学本科生毕业设计（论文）任务书**

学院（部）：医学部基础医学与生物科学学院

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 设计（论文）选题：[面向复杂分类体系的多标签情绪分类算法设计与实现](http://bkbylw.suda.edu.cn/Student/ViewReport.aspx?No=809" \o "面向复杂分类体系的多标签情绪分类算法设计与实现) | | | | | |
| 指导教师姓名 | 朱苏阳 | 职 称 | 讲师 | 类 别 | 毕业设计 |
| 学 生 姓 名 | 唐柳健 | 学 号 | 1930401096 | 设计（论文）类型 | 应用型课题 |
| 年 级 | 2019 | 专 业 | 生物信息学 | 是否隶属科研项目 | 否 |
| 1. 设计（论文）的主要任务及目标   从不同情绪之间的层次关系出发，设计一种针对复杂情绪分类体系的多标签情绪分类的算法，以期能够获得比单层次分类方法更高的分类准确率。 | | | | | |
| 1. 设计（论文）的主要内容   采用GoEmotions数据集，该数据集为28类情绪类别的体系。现从粗到细的粒度，将该数据集分成三个层次，依次为由“positive”，“negative”，“ambiuous”够成的3类粗粒度情感层，ekman的6类情感体系层到最细粒度的28类情绪层。从这三个层次的体系出发，先进行粗粒度的分类，再分别进行下一层的细粒度分类直到最后一层。之后将多标签分类任务建模为若干个单标签分类任务，对每个情绪采用二元相关法进行分类，最终使用多标签的Accuracy与F1指标对分类进行评估。  具体为，将数据集文本进行分词，大小写转换，转为嵌入词向量，UNK以及mask后的token进行随机初始化等预处理工作后，将数据集分为训练集，测试集，验证集。使用LSTM等擅长做自然语言处理的分类器为在训练集上为每个层次的每个情绪训练二元分类器。在验证集上进行超参数等调参后，在测试集上查看模型的效果。最后通过对比所设计的多层次分类方法的性能与只使用单层次分类方法的性能，来验证该算法的有效性。 | | | | | |
| 1. 设计（论文）的基本要求   本课题需要一定的自然语言处理知识和机器学习以及深度学习基础。具体来说，需要对自然语言处理领域中的情绪分析任务有一定的了解，并能够使用机器学习，深度学习方法或其他方法解决相关问题。编程语言方面以python为最佳(因为相应的第三方toolkit比较丰富和全面）。另外，需查询及阅读有关文献，适当做一定的探索。定期向导师汇报工作进度和与老师沟通。按时完成相应阶段任务。独立完成毕业设计，实事求是，不抄袭、不代写。毕业答辩后，整理提交相关资料，并做到不外泄。 | | | | | |
| 1. 主要参考文献   [1] Pla F, Hurtado L F. Political tendency identification in twitter using sentiment analysis techniques[C]//Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical Papers. 2014: 183-192.  [2] Desai J, Cao H, Shah R. Attention-based Region of Interest (ROI) Detection for Speech Emotion Recognition[J]. arXiv preprint arXiv:2203.03428, 2022.  [3] Plutchik R. Emotion[J]. A psychoevolutionary synthesis, 1980.  [4] Ekman P. An argument for basic emotions[J]. Cognition & emotion, 1992, 6(3-4): 169-200.  [5] Russell J A, Mehrabian A. Evidence for a three-factor theory of emotions[J]. Journal of research in Personality, 1977, 11(3): 273-294.  [6] Godbole S, Sarawagi S. Discriminative methods for multi-labeled classification[C]//Advances in Knowledge Discovery and Data Mining: 8th Pacific-Asia Conference, PAKDD 2004, Sydney, Australia, May 26-28, 2004. Proceedings 8. Springer Berlin Heidelberg, 2004: 22-30.  [7] Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification[J]. Machine learning, 2011, 85: 333-359.  [8] Tsoumakas G, Katakis I, Vlahavas I. Random k-labelsets for multilabel classification[J]. IEEE transactions on knowledge and data engineering, 2010, 23(7): 1079-1089.  [9] Farruque N, Huang C, Zaiane O, et al. Basic and Depression Specific Emotions Identification in Tweets: Multi-label Classification Experiments[C]//Computational Linguistics and Intelligent Text Processing: 20th International Conference, CICLing 2019, La Rochelle, France, April 7–13, 2019, Revised Selected Papers, Part II. Cham: Springer Nature Switzerland,  2023: 293-306.  [10] He H, Xia R. Joint binary neural network for multi-label learning with applications to emotion classification[C]//Natural Language Processing and Chinese Computing: 7th CCF International Conference, NLPCC 2018, Hohhot, China, August 26– 30, 2018, Proceedings, Part I 7. Springer International Publishing, 2018: 250-259.  [11] Kim Y, Lee H, Jung K. Attnconvnet at semeval-2018 task 1: Attention-based convolutional neural networks for multi-label emotion classification[J]. arXiv preprint arXiv:1804.00831, 2018.  [12] Jabreel M, Moreno A. A deep learning-based approach for multi-label emotion classification in tweets[J]. Applied Sciences, 2019, 9(6): 1123.  [13] Baziotis C, Athanasiou N, Chronopoulou A, et al. Ntua-slp at semeval-2018 task 1: Predicting affective content in tweets with deep attentive rnns and transfer learning[J]. arXiv preprint arXiv:1804.06658, 2018.[14] Kant N, Puri R, Yakovenko N, et al. Practical text classification with large pre-trained language models[J]. arXiv preprint arXiv:1812.01207, 2018.  [15] Desai S, Kshirsagar A, Sidnerlikar A, et al. Leveraging Emotion-specific Features to Improve Transformer Performance for Emotion Classification[J]. arXiv preprint arXiv:2205.00283, 2022.  [16] Hasan M, Rundensteiner E, Agu E. DeepEmotex: Classifying Emotion in Text Messages using Deep Transfer Learning[C]//2021 IEEE International Conference on Big Data (Big Data). IEEE, 2021: 5143-5152.  [17] Huang C, Trabelsi A, Qin X, et al. Seq2emo for multi-label emotion classification based on latent variable chains transformation[J]. arXiv preprint arXiv:1911.02147, 2019.  [18] Fei H, Zhang Y, Ren Y, et al. Latent emotion memory for multi-label emotion classification[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(05):  7692-7699.  [19] Dong Y, Zeng X. Lexicon-Enhanced Multi-Task Convolutional Neural Network for Emotion Distribution Learning[J]. Axioms, 2022, 11(4): 181.  [20] Mukherjee R, Naik A, Poddar S, et al. Understanding the role of affect dimensions in detecting emotions from tweets: A multi-task approach[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021: 2303-2307.  [21] Dorottya Demszky,.Dana Movshovitz-Attias,et al. GoEmotions: A Dataset of Emotions[C]. Fine-Grained Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,2020.4040–4054. | | | | | |
| 5、进度安排   |  |  |  | | --- | --- | --- | |  | 设计（论文）各阶段任务 | 起 止 日 期 | | 1 | 任务书 | 2023-03-01至2023-03-10 | | 2 | 外文翻译和文献综述 | 2023-03-10至2023-03-19 | | 3 | 准备及进行中期检查 | 2023-03-19至2023-04-12 | | 4 | 论文查重及定稿 | 2023-04-12至2023-05-13 | | 5 | 准备上交材料进行答辩 | 2023-05-13至2023-05-19 | | | | | | |

注：1、此表一式三份，学院(部)、指导教师、学生各一份；

2、类别是指毕业论文或毕业设计，类型指应用型、理论研究型和其他；

3、在指导教师的指导下由学生填写。