



# 基于NLP的新冠相关推特属性判断

Design and demonstration of ultrasonic reflector

■ 汇报人：段皞一

# 目录



01

项目内容简介

02

项目背景

03

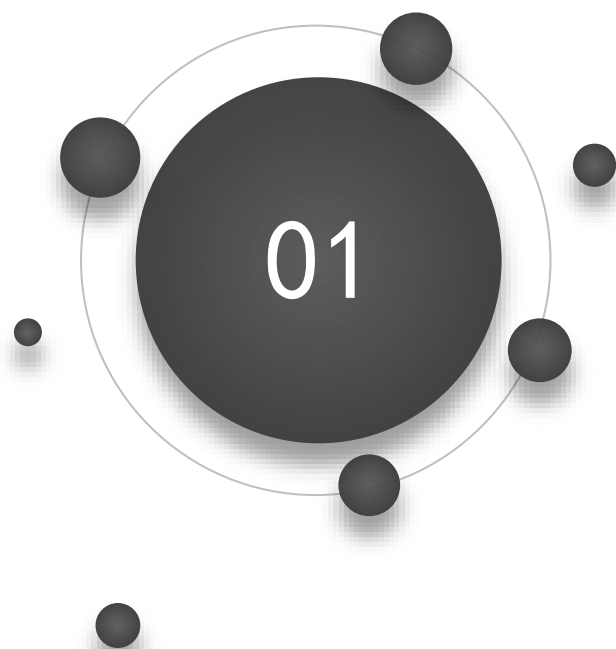
项目研究路线

04

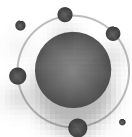
可能的创新方向

05

项目预期成果



## 指导老师

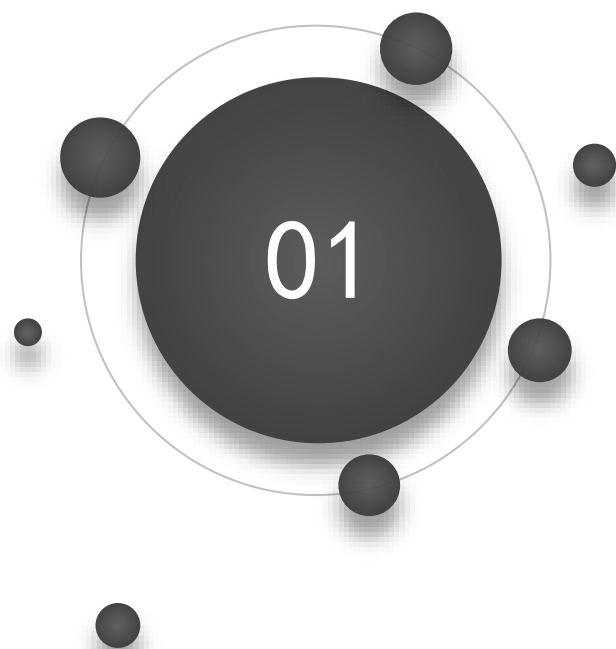


## · 指导老师

### □ Teacher

- Phone/WeChat#: 13588196277
- Email: siliang [AT] zju.edu.cn
- Address: 浙江大学玉泉校区则通楼205-207
- Homepage: <https://person.zju.edu.cn/siliang>
- Research Direction:
  - Information Extraction, Natural Language Processing, Vision and Language Grounding.





## ● 项目内容简介

---



## · 项目内容简介

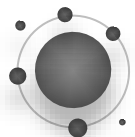
判断关于COVID-19的tweet的一系列二进制属性。

### Task (Binary Properties):

- Verifiable Factual Claim
- False Information
- Interest to General Public
- Harmfulness
- Need of Verification
- Harmful to Society
- Require attention

## 项目背景

02

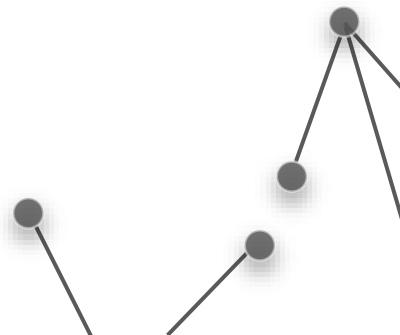


## · 项目背景

在疫情常态化的今天，由于信息墙的存在，观察tweet这样的国外平台能够听到国外网友的真实声音。而一方面媒体在检测不实消息上缺乏足够手段，平台对于消息的推送价值也很难精准判断。

传统的文本分类方法存在一些局限，很多现行的模型都受限于特定场景。近年来研究人员设计了许多模型来增强文本分类模型的准确性。但是，如果数据集中有一些对抗性样本，则模型的性能会大大降低。小的样本噪声可能导致决策置信度发生实质性变化，甚至导致决策逆转。另一方面，尽管一些新的文本分类模型反复擦写了大多数分类任务的准确性指标，但它无法指示模型是否像人类一样从语义层面“理解”文本。

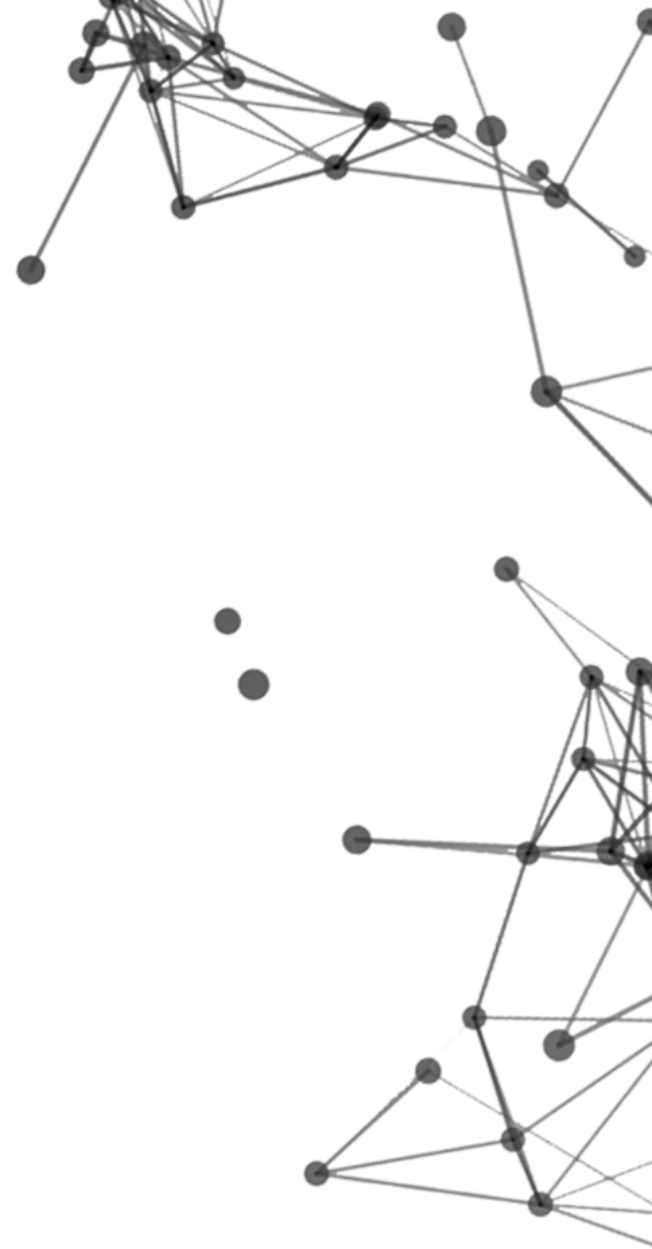
因此，如何改进模型以适应更广泛的场景并具有更加精准可控的结果，成为我们关注的重点。





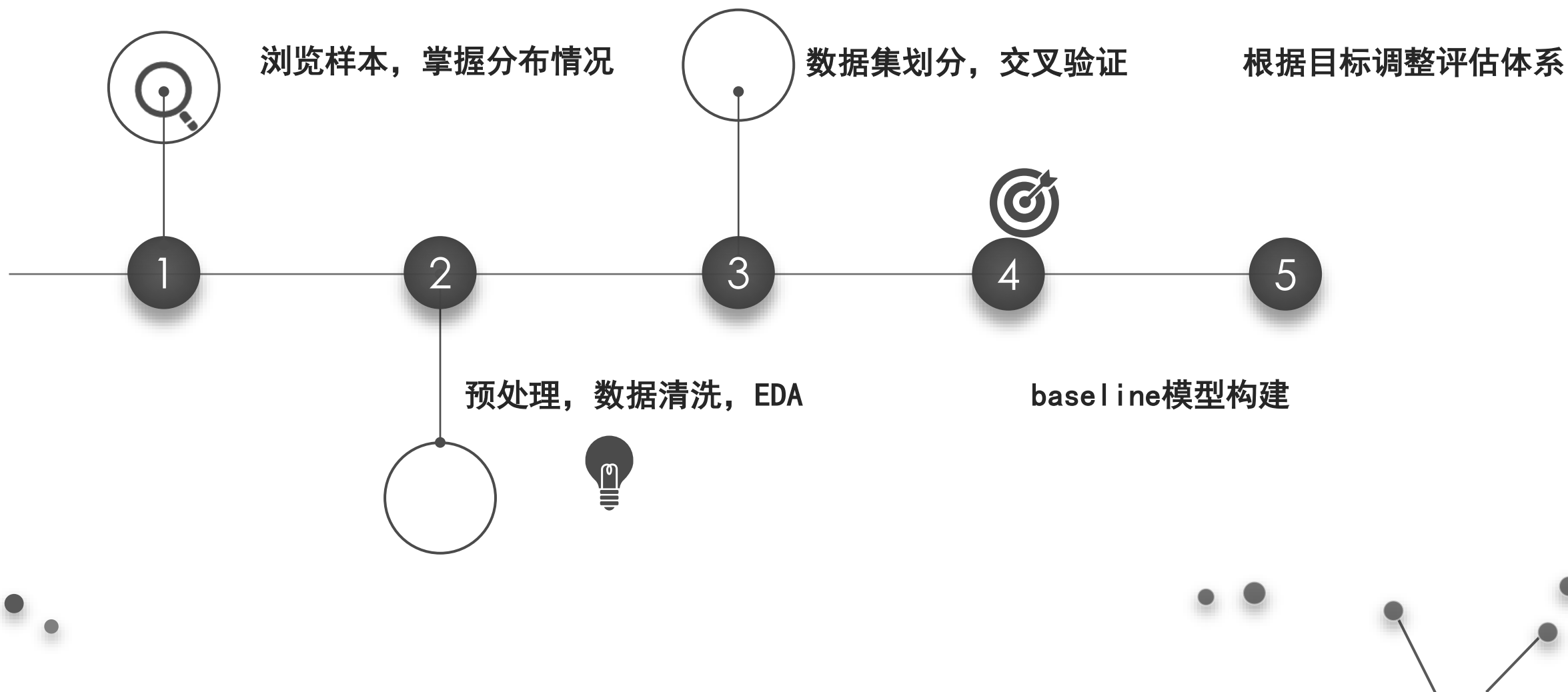


## 项目研究路线





# 项目研究方案





## · 进度安排

- 2021年4月-6月 相关领域入门，了解机器学习，词向量，词嵌入等知识，运用PyTorch搭建神经网络
- 2021年6月-7月 领域文献综述；读懂相关模型算法
- 2021年7月-10月 复现主流的两层词义分类模型(如Bert、TextCNN、TextRNN等)
- 2021年10月-11月 通过分割的数据集对搭建的模型进行调试，分析其可靠性
- 2021年11月-2021年1月 在之前的测试基础上，对模型进行算法上的优化
- 2022年1月-4月 整理项目成果，进行论文的撰写



可能的创新方向

04

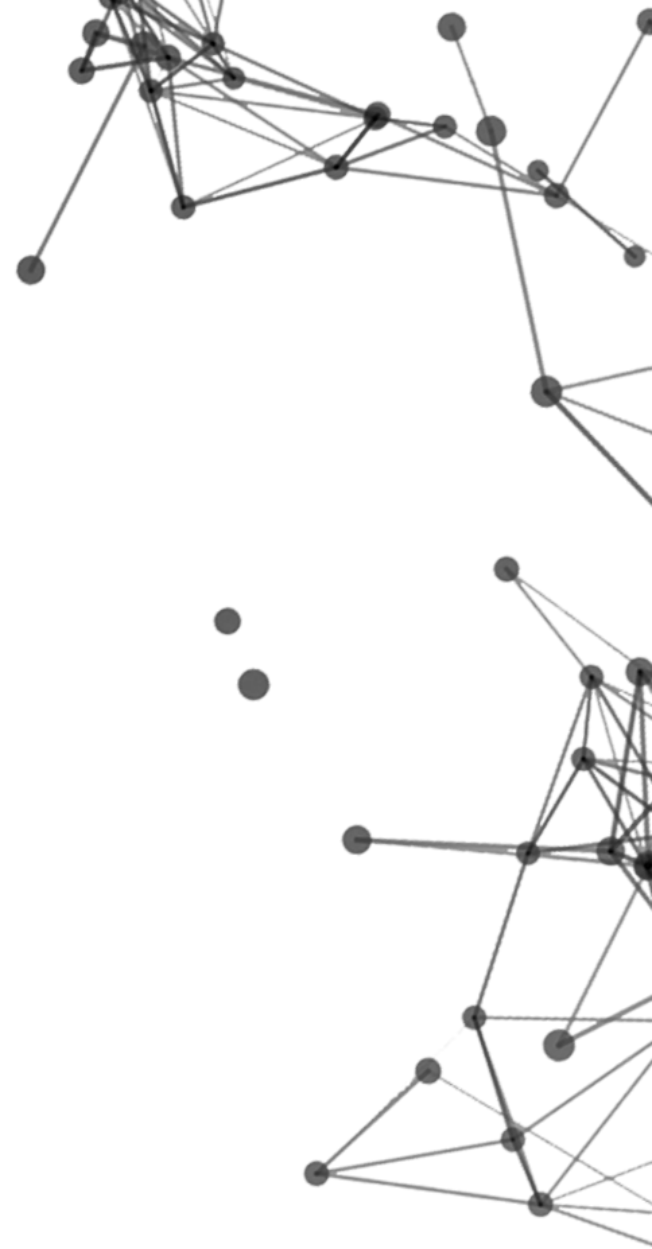


## 可能的创新方向

- 由词向量表示的预训练语义表示模型通常可以提高下游NLP任务的性能。基于这一点，改进模型可以尝试从采用更好的关于上下文无关单词向量的传输策略上入手。
- 多标签文本分类需要充分考虑标签之间的语义关系，并且模型的嵌入和编码是有损压缩的过程。因此，改进模型可以着眼于减少训练过程中层次语义的丢失以及保留丰富而复杂的文档语义信息。



## 项目预期成果





## · 项目预期成果

- 实现基于NLP的判断模型;
- 根据数据集测试的准确率改进模型;
- 进行优化, 提高计算速度。



# Thanks

敬 请 指 导