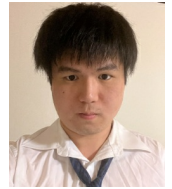


骆子杰

☎ 08078932137 ✉ umikotori233@gmail.com 📍 日立
📞 baika0906 🌐 <https://github.com/Baikabai>
🎂 25岁 🧑 男 🏠 厦门 🌳 汉族 💕 未婚
👤 应届生 📁 自然语言处理



实习经历

大和证券集团本社

2022年03月 - 至今

共同研究研究员 金融分析

日立

- Keywords : Bert, Word2Vec, fasttext, GloVe, MySQL, 信息提取, 文本分类, 网络爬虫, 日文数据预处理等;
- 项目描述: 负责基金搜索系统的开发, 使用几个词语找出与这些词语相近的有价证券报告书, 从而找到和该词语关联的基金和上市公司;
- 追踪最新的处理技术, 将模型落地, 优化算法和提高准确率;
- 使用爬虫技术取得有价证券报告书, 日语训练资料, 基金和股票代码等, 用于训练语言模型和训练;
- 针对遇到的技术提出解决方案, 从而提高了召回率(44%提升至57%);
- 将数据录入MySQL数据库中;
- 总结结果, 将结果汇报并写成论文, 将于2022年8月19日参加日本学术发表, 相关论文《Support for Composition of Thematic Equity Funds Using Text Mining》

项目经历

雅虎FX推特, 广告费, 收益之间的关系

2021年04月 - 2021年10月

共同研究员

日立

- Keywords: 爬虫, Bert, 情感分析, 数据分析
- 此项目是与日本雅虎金融YJFX (Yahoo Japan FX) 进行共同研究, 内容为与YJFX相关的推特和广告费, 收益之间的关系。目的是研究什么样的题材, 什么样的内容能够得到更多的点击率, 并且获得更多的开户;
- 使用推特的API, 但是每日获得数量有限, 并且不能获得全部的推特, 所以使用了selenium进行web自动化爬取;
- 使用数据挖掘技术, 对文本进行分析, 日文与中文相同, 需要使用分词器, 这里使用的是Neologd词典;
- 推测推文的情感值, 使用的是bert模型, bert模型为东京大学和泉研的模型, 然后进行Fine tuning, 用于判别情感值。
- 测定之间的关系, 使用开户返金等促销手段, 进行多个推特账号的宣传的话, 开户数量会增加, 如果赠送礼品的话, 和没活动期间的效果基本一致。如果拥有对FX的利好消息, 开户数量也会增加(例如美联储等大新闻), 而FX相关的知名人士的发言并不能太左右FX开户的数量。

日文亚马逊顾客评价文本分类

2021年03月 - 2021年07月

独立开发

日立

- Keywords: 文本提取, 特征选择, 特征提取, 日文数据预处理, 神经网络, 深度学习, LSTM等
- 项目描述: 负责文本分类系统的开发, 使用已有的日文亚马逊顾客评论数据与星级, 使用各种自然语言模型进行处理;
- 对句子进行Tokenize和padding的处理, 利用Word2Vec预训练词向量构建字典;
- 从Amazonaws取得数据, 将句中日文比例低于0.7的数据全部去除, 提取构造特征(N-gram, Tf-idf特征, 统计特征);
- 最后使用LR, 神经网络, RNN, LSTM, 得到最终准确率(0.8193)

教育经历

茨城大学(QS1000)

2021年04月 - 2023年03月

机械系统(原知能系统) 硕士 工学院

日立

- GPA: 3.62/4.5
- 相关课程: 数据分析, 自然语言处理, 人工智能, 机器学习, 强化学习, 英文学术发表与写作, 人工智能, 数据结构, 计算机基本原理

茨城大学

2017年04月 - 2021年03月

知能系统 本科 工学院

日立

- 相关课程: 数据分析, 高数, 人工智能, 计算机原理, 数据结构, 算法
- 全额学费减免

专业技能

- 熟悉机器学习算法, 具有数据挖掘经验;
- 熟悉自然语言处理(NLP), 具有文本分类, 文本提取, 文本相似度的经验, 熟悉bert, transformer, word2vec, fasttext等自然语言处理模型;
- 熟悉数据结构与算法, 熟悉Python编程语言, 熟悉TensorFlow, Pytorch, Kears, sklearn深度学习框架;
- 熟悉日文数据处理, 有丰富的日文预处理经验;
- 熟悉爬虫技术, 有使用过selenium, request等技术爬取过内容, 了解html5, css, 制作过自己的网页;
- 了解MySQL的构成和使用语法