

# 袁 搏

+86 13369069688 | [354151531@qq.com](mailto:354151531@qq.com) | 浙江省杭州市西湖区

## 教育背景

---

- 浙江大学（博士） — 计算机学院
- 四川大学（本科） — 计算机学院

## 科研论文

---

- ① Yulin Chen, **Bo Yuan**, Beishui Liao et al. A self-explanatory contrastive logical knowledge learning method for sentiment analysis
  - 已被 *Knowledge-Based Systems* 期刊发表（中科院 SCI 一区，Top 期刊）
- ② Yulin Chen, **Bo Yuan**, Beishui Liao. KDL: Knowledge-aware deep logic learning method for interpretable named entity recognition
  - 已投稿至 *Information fusion* 期刊，小修（中科院 SCI 一区，Top 期刊）
- ③ **Bo Yuan**, Yulin Chen, Zhen Tan, Lu Cheng, Huan Liu, and Yin Zhang. Label-Guided Self-Knowledge Distillation for Multi-Class Text Classification
  - 已投稿到 *Knowledge-Based Systems* 期刊（中科院 SCI 一区，Top 期刊）
- ④ **Bo Yuan**, Yulin Chen et al. Label Distribution Learning Can Improve Dual-KNN for Text Classification
  - 已被国际顶级会议 SDM 2024 接收（Oral, CCF-B 类会议）
- ⑤ **Bo Yuan**, Yulin Chen et al. Hide and Seek in Noise Labels: Noise-Robust Collaborative Active Learning with LLM-Powered Assistance
  - 已被国际顶级会议 ACL 2024 接收（CCF-A 类会议）
- ⑥ Zhen Tan, Lu Cheng, Song Wang, **Yuan Bo** et al. Interpreting Pretrained Language Models via Concept Bottlenecks.
  - 已被国际顶级会议 PAKDD 2024 接收（Best Paper CCF-C 类会议）
- ⑦ Yulin Chen, Zelai Yao, Haixiao Chi, D Gabbay, **Bo Yuan**, et al. BTPK-based learning: An Interpretable Method for Named Entity Recognition.
  - 已被国际会议 LNGAI 2023 会议发表
- ⑧ Yunyun Niu, Yulin Chen, Detian Kong, **Bo Yuan**, et al. Strategy evolution of panic pedestrians in emergent evacuation with assailants based on susceptible-infected-susceptible model[J]. *Information Sciences*, 2021, 570:105-123.

- 已被 *Information Sciences* 期刊发表（中科院 SCI 一区，Top 期刊）

⑨ Yunyun Niu, Yulin Chen, **Bo Yuan**, et al. Dynamic evolution model of pedestrian cooperation behavior based on coordination game[J]. *Expert Systems with Applications*, 2020, 167(3-1):114173.

- 已被 *Expert Systems with Applications* 期刊发表（中科院 SCI 一区，Top 期刊）

## 实习经历

---

### 一. 百度

多行业分类

数据处理：在 Spark 平台上从服务器取数据并进行处理

训练模型：用 Ernie 模型，组合不同的文本特征（标题，摘要，资质，创意等）和不同月份的数据分别做多组对比实验进行筛选

提升模型：对 badcase 人工 recheck，如果存疑则记录下来返回给 PM 重新标注  
如果是模型本身判别不出来，则对这些 badcase 寻找规律添加规则  
在全量数据集上验证准确率和召回率，确保对一类提升后不影响其他类的判定

### 二. DTL (Dynamic Technology Lab)

文本信息和股票量化

数据处理：从原始数据中抽取会议文本信息和股票信息进行清洗，匹配，合并

文本建模：1. 基于频率方面：筛选反映情绪的词汇，构建情感词典，计算情绪指数

2. 基于预测方面：基于 Bert 等模型进行情感分类，文本相似度计算等提取特征

模型预测：搭建模型对提取特征进行训练，预测股票的趋势，确定基于文本指标的交易策略

### 三. 浙江大学求是高等研究院

全诊通医学辅助机器人项目

数据处理：抽取药品描述数据文本

搭建模型：提出信息抽取和分类两种解决方案去提取药品文本中用药剂量

提升模型：根据模型返回的结果增加正则去返回需要的结果，准确率达到 90% 以上

### 四. 蚂蚁金服

研究型实习生

对内输出：基于业务中的意图识别场景，解决如何从原始的意图数据中（含有噪音标签）选择高质量的样本数据进行模型训练

对外输出：基于实际问题，完成一篇论文并投稿 EMNLP2024

## 主要竞赛

---

### 一. 百度竞赛《新闻标题多分类》第 7 名

模型搭建：搭建 Bert, Ernie, Robert 等模型进行训练

模型提升：文本增强，处理分类不均衡问题，进行模型融合，  
多任务学习提升分数

### 二. 百度竞赛项目《问答摘要与推理》第 3 名

数据处理：过滤低标签词，停用词，标点符号等

文本表示：对比 Word2vec, Fasttext 去训练词向量

模型训练：搭建 Seq2seq 模型，并使用 Beam Search，指针网络等去改进预测结果

## 项目经历

---

### 一. 基于学术知识图谱的构建与应用

数据处理：对数据进行统计和清洗

机构实体匹配：由于机构名称有的是缩写，有的是全称，有的位置精确程度不同，所以采用两种方法判断是否为一个机构

- 1.数据清洗：主要是通过分词将公司全称拆解并对可能的简称形式进行组合
- 2.模糊匹配：主要是基于最短编辑距离算法计算猜测的简称和需要匹配的简称的匹配分数，然后通过筛选最高匹配分数来找到最佳的匹配结果

图谱构建：确定实体和关系，搭建学术知识图谱可以进行实体查询，关系查询，还可以看出学者随着时间演进而产生的学术成果变化，还可以进行论文查询，合作者查询（包括合作次数，合作年份），知识问答

学者消歧：1.论文关系表征学习：先预构建异质网络，然后基于原路径进行随机游走得到节点关系向量，最后根据关系相似度微调  
2.论文语义表示学习：先用 Bert 得到论文文本特征的语义表征向量，然后构建正负三元组样本进行微调  
3.结合全局语义特征和局部关系特征，并使用图自动编码器获得论文的最终表示向量  
4.使用 HAC 聚类算法把论文聚类到多个簇中，其中每个簇中的论文表示为同一学者所写

### 二. 中药处方分类

处方建模：

- 1.把处方看作成药物的集合，利用集合的相似性度量去建模药物与处方类别之间的关系
- 2.引入外部知识（药物的描述, label 的描述），例如方剂含义，性能特点，功效

## 学生工作

---

党素中心, 浙江大学

浙江大学党素中心主任团

代领党素中心实践部同学一起完成学生党建工作, 主要负责以下几项工作：

1. 发展对象培训班通知、报名、组织
2. 预备党员培训班通知、报名、组织
3. 三会一课督查、年底民主评议情况督查、年底对标争先述职
4. 学院特色党建活动组织、实施

学生会女工部, 四川大学

部长

1. 主要负责学生会大型活动、会议的礼仪、接待、颁奖工作及学院礼仪队的选拔、培训和管理等工作
2. 负责对接各班的心理委员，开展心理健康活动，及时了解各位同学的心理状态工作