

焦雪莹 (Xueying Jiao)

基本信息

性别: 女 民族: 汉
出生年月: 1998-09
常住地: 北京
电话: (+86)18710212152
邮箱: ajslpzcd@gmail.com
微信: aj0020120101

教育经历

清华大学 2015.09 - 2019.06
本科 清华大学工程物理系
专业方向: 工程物理

实习经历

基于 Hierarchical Transformer 的生成式文本摘要模型 @ 北大计算机科学技术研究所 2019.07 - 2019.08
职务: 暑研实习生
主要工作: 对生成式文本摘要方向进行了调研, 主要研究多层 Transformer 在长文本摘要任务上的表现, 以Hierarchical Transformers for Multi-Document Summarization 一文的源码为基础, 尝试复现了一个长文档生成摘要的模型, 并在 WikiSum database 上进行训练和测试。结果表明 Hierarchical Transformer 比 Flat Transformer 在生成式文本摘要任务的 ROUGE L 指标上表现更好。

基于 Faster-RCNN 的行人感知模型调试 @ 北京 Momenta 科技 . 2018.10 - 2018.11
职务: 行人感知组算法实习生
主要工作: 研究无人驾驶技术领域中, 检测和感知车辆周围行人的算法。用试验车采集车辆四周的行人运动数据, 在基于 Faster-RCNN 的行人感知模型上进行测试和调优, 选取测试结果中的片段作为 CES 大会展示的 demo。

模型与算法构建 @ 辰安科技 2018.07 - 2018.09
职务: 中央研究院实习生
主要工作: 地震应急需求计算模型的校验性分析与修正、模型算法说明书撰写。

项目经历

实现一个 Seq2Seq 的闲聊对话系统 2020.02 - 2020.03
基于 seq2seq 结构, 使用 LSTM 模型和 Luong attention 机制实现一个简单的 DL 闲聊聊天机器人, 参考 Pytorch 官网 Chatbot Tutorial, 训练数据来自 Cornell 电影对话语料库。该对话系统可以解析用户输入的语义, 并给出生成式的 (generative) 回复。

本科毕设 2019.02 - 2019.06
毕设题目《高密度人群运动的风险评估方法》, 使用 DPIV 算法从视频中提取行人速度场。使用 Multi-Column CNN 模型从识别图片中的行人, 从而获得视频环境中行人密度分布。计算行人分布随时间的变化, 并用连续行人模型计算预测行人运动风险的时空分布。

Kaggle 比赛: Comment Toxicity 分类任务并消除意外偏差 2019.04 - 2019.05
建立一个模型来判断社交网站中评论是否为恶意评论, 并将与身份相关的无意识偏见最小化, 使用 Pytorch pretrained BERT model, 用比赛提供的训练集集做 fine-tuning, 并在测试集上测试评估。

团队项目：在线社交媒体中热点的预测

2018.11 -2018.12

在 AMA Team 公开的数据集上（在 Twitter 网站统计涉及某话题，在一周内讨论数量和质量的变化的变化，以预测该话题是否是热点话题），使用 scikit-learn 包，用 LR, KNN, 决策树，随机森林，贝叶斯，LDA, SVM 等机器学习模型，对话题能否成为热点（Buzz or non-Buzz）做了预测，并对不同方法做出的结果做了评价。

清华大学学生研究训练项目 SRT

2017.07 - 2018.09

导师：清华大学公共安全研究院 研究员 刘奕

SRT(Student Research Training) 项目是为本科生提前进入实验室开展项目研究而设立的。

轨迹数据挖掘 @SRT

行人轨迹数据挖掘项目，利用校园服务类 APP 后台轨迹数据，挖掘出服务于校园公共安全的有效信息。在项目初期，主要进行实验设计、算法调研、数据预处理等工作。

网站前端 @ SRT:

对实验室应急交通疏散仿真系统的网站进行升级改造，完成了网站的视觉设计、交互设计，使用 Bootstrap 框架进行前端搭建，并与已有的后端代码以及数据库进行连接。

技能

- 编程语言：Python, C++, Matlab
- DL 框架：Pytorch(prefered), TensorFlow
- 其他：Linux, MySQL
- 英语：CET-6
- 日语：JLPT N3