

许士亭

✉ [xvshiting@live.com](mailto:xvshiting@live.com)  
🏠 15600355778  
🌐 <https://github.com/xvshiting>

## Interests

自然语言处理  
深度学习

## Skills

Programming: Python, Java, C

## Education

2014.9-2017.3 北京邮电大学, 研究生, 信息安全  
2010.9-2014.7 山东理工大学, 本科, 计算机科学与技术

## Work & Research Experience

2022.08-至今 山东政法学院网络空间安全学院-专职教师  
教授计算机专业相关课程, 担任 2023 届毕业生辅导员, 负责网安学院毕业生就业等工作。

2021.09-2022.07 度小满-数据智能部-NLP 研究员

**1. 基础分词**  
用户 Query 文本分词优化, 基于预训练模型训练 teacher 模型, 预测 BMES 标签对文本进行分词, F1 优化至 95%。利用有标签和无标签数据对 teacher 模型进行蒸馏, 训练 CNN 结构的 student 模型, 并添加 Early Exit 机制。对 student 模型进行剪枝, 进一步优化速度。最终转化成 ONNX model, 速度相比 teacher 模型优化百倍速度, 单核 cpu 速度达到平均 1ms/条。

**2. 用户画像**  
对用户搜索历史, app 使用历史以及网址浏览历史进行建模, 预测用户收入、投资客、小微等标签。

2019.08-2021.09 北京好未来科技有限公司-AI 中台-自然语言处理高级工程师

**1. 英文作文批改**  
负责英文作文批改算法接口设计和语法错误模块, 使用基于 Roberta 的拼写纠错模型和基于 transformer 的语法纠错模型, 在业务数据集上语法检测 F0.5 达到 81%, 纠错 F0.5 达到 73%。优化图片端到端批改效果, 优化后端到端语法检测 Precision 达到 70%, recall 50%。

**2. 中文作文批改**  
主要负责中文拼写纠错。伪造拼写纠错训练预料, 实现并训练改进的 Bert 结构模型, 最终在业务数据上错别字检测 F0.5 达到 80%, 并进行了相关端到端优化。

**3. 要点识别**  
判断学生作文是否对题目中的要点进行写作。牵头完成要点训练数据集的标注和整理工作。利用 QA 模型实现学生作文中要点回答的自动识别功能。最终在 WH 类问题上 Accuracy 为 0.89, 答案 overlap F1 为 0.84, 对接到英文作文批改项目, 并在 2021 AIED 上发表相关 paper。

2017.9-2018.11 普强信息技术科技有限公司 - 数据挖掘工程师

**1. 角色切割**  
使用 seq2seq 模型, 利用语义信息对 ASR 识别的客服客户识别文本进行重新进行角色分割。

**2. 客户信息抽取**  
构建分类模型, 从客户和客服的对话文本中提取客户标签。提取包含是否有车、有房、有无保险之类的几十中与保险购买意识和能力相关标签。

## Publication

---

1. **Xu, Shiting** , et al. "Malicious Application Dynamic Detection in Real-Time API Analysis ." IEEE International Conference on Internet of Things IEEE, 2017.

2. **Xu, Shiting** , W. Ding , and Z. Liu . Automatic Dialogic Instruction Detection for K-12 Online One-on-One Classes. Artificial Intelligence in Education. 2020.

3. **Shiting Xu, Guowei Xu, Peilei Jia, Wenbiao Ding, Zhongqin Wu and Zitao Liu.** Automatic Task Requirements Writing Evaluation via Machine Reading Comprehension. Artificial Intelligence in Education. 2021.