

徐伟

✉ readonlyfile@hotmail.com · ☎ (+86) 180-0451-5446 · 🌐 <https://memeda.github.io>

🎓 教育背景

哈尔滨工业大学, 哈尔滨, 黑龙江 2015 年 9 月 -- 至今
在读硕士研究生 计算机科学与技术, 社会计算与信息检索研究中心, 预计 2017 年 7 月毕业

哈尔滨工业大学, 哈尔滨, 黑龙江 2011 年 8 月 -- 2015 年 7 月
学士 计算机科学与技术, 学分绩排名前 10%

👨‍💻 项目/实习经历

基于深度学习的中文分词、词性标注和命名实体识别 2016 年 3 月 -- 至今
C++ 主要项目

<https://github.com/memeda/sequence-labeling-by-nn>

以构建准确率高、速度快的工业级序列标注系统为目标。

- 融合深度表示学习 (MLP、LSTM) 与全局结构预测方法 (CRF)。
- 利用 C++ 语言的封装与抽象能力, 编写完成不同输入、不同表示学习、不同输出预测的模型。
- 在与 LTP(www.ltp-cloud.com) 相同的评测数据集上完成大量实验, 目前在词性标注和命名实体识别任务上相比基于传统结构化感知器的 LTP 取得了更好的分数。

面向不均匀类别的文本分类系统设计与实现 2015 年 3 月 -- 2015 年 6 月
Python 本科毕业设计

<https://github.com/memeda/GriduationDesignCodeForTextClassification>

构建了两个文本分类系统, 分别基于提升树 (侧重广适应性) 和支持向量机 (倾向高准确率)。

- 在面向主题文本分类、情感分类的均衡与不均衡数据集上, 完成不同文本表示、不同特征值选择、不同分类器的对比实验。
- 使用 sampling、SMOTE、调整分类器权值的方法提升在不均衡数据集上的文本分类效果。

百度实习, 北京 2014 年 6 月 -- 2014 年 9 月
C/C++, Python, shell, MapReduce 自然语言处理部门

独立完成知识库抽取系统在 Hadoop 集群上的部署。

- 提升脚本处理性能; 验证核心程序正确性。
- 在 5 个集群上同时部署系统, 持续处理 20K 个百灵库。
- 历时一个月, 抽取 200+TB 知识。

🔧 IT 技能

- 编程语言: 偏好 C++, 熟练使用 Python, 有 C/Shell/JavaScript/Java 基础。
- 工程经验: 熟悉基础 Linux 脚本和 WEB (前端) 编写, 有简易 GUI (Qt/MFC/Java) 编写经验。
- 研究方向: 文本分类、序列标注 (结构化预测), 深度学习。

♡ 获奖情况

国家奖学金 2014 年 10 月

📄 其他

- GitHub: <https://github.com/memeda>
- 外语: 英语 6 级