

邵楠

北京大学

18818216357

毕业时间：2019.07

shaonan@pku.edu.cn

硕士研究生（研二）

北京市大兴区金苑路 24 号(102600)

教育背景

2016.09–至今	北京大学	软件与微电子学院	计算机技术	研究生在读
2012.09–2016.07	上海大学	理学院	应用物理学	本科

专业技能

- 熟悉 Java 语言程序设计，能够自行设计与实现常用数据结构及算法，对集合框架、多线程、反射等高级特性及 SSM 框架也有所了解。
- 熟悉 Python 语言程序设计，会使用 pandas、scikit-learn、XGBoost 等常用机器学习库。
- 熟悉常用机器学习/深度学习算法，对算法在机器翻译、NLP 和 CV 领域的应用有所了解。
- 熟悉数据结构、操作系统、计算机网络、组成原理等计算机基础理论知识，长期辅导计算机考研专业课程。

实习经历

2017.12-至今	网易有道事业部	算法工程师实习生
------------	---------	----------

在 mentor 的指导下独立负责有道翻译后端中德/中泰机器翻译模型和系统的研发工作的整个流程，积极推动翻译的系统的研发和最终上线（目前尚未完成）。主要工作包括：

- 德语/泰语语料数据的搜索、爬取和整理，针对不同语料来源的特点进行清洗、句子对齐，抽取高质量句子作为评测集送至标注公司标准。
- 为现有机器翻译系统添加德语/泰语的分词、去分词、语言检测等模型及相关业务代码，修复在模型研发过程中发现的一些 bug。
- 分别训练基于层次短语的统计机器翻译模型（SMT）和基于 Soft Attention 机制的神经机器翻译模型（NMT），实现分流相关业务代码。（进行中）
- 对训练的模型进行持续优化，使模型在使用 BLEU 作为评价指标下与 Google Translate、百度翻译等同类产品相比具有竞争力。（进行中）

项目&比赛

2017.11-2017.12	车辆保险索赔可能性预测	Kaggle 比赛
-----------------	-------------	-----------

- 针对给出的训练数据进行 EDA，完成特征缩放、编码、非平衡类处理和构造新特征等预处理工作。
- 对训练数据尝试 RF、XGBoost、LightGBM 等不同模型进行预测，并使用 Stacking 方案进行模型融合。
- 模型预测得到的概率使用 Gini 系数评价模型质量，最终 Private LB 排名约 8%（铜牌）。

2016.12-2017.10	佛山公交骨干网数据分析系统	实验室项目
-----------------	---------------	-------

- 使用 dataX 将公交车辆 GPS 记录、IC 卡刷卡记录导入至 Hive，并使用 SparkSQL 完成数据的清洗，计算 OD（Origin-Destination）矩阵，并在此基础上计算换乘率、高峰满载率等城市交通网指标。
- 个人在项目中负责开发 web 形式的数据分析系统，连接集群开放的 ThriftServer，为上述数据分析流程提供图形界面，为计算得到的各项交通网指标实现数据可视化。项目中使用到的主要技术包括 Spring、SpringMVC、Bootstrap、ECharts 等。

其他

语言：英语（CET-6）、普通话（母语）

奖学金：2013 上海大学自强不息奖学金、2014 上海大学爱心公益奖学金。

兴趣爱好：剑道、游泳。