



北京大学

张博玮



☎ (+86) 18101389470 ✉ bw_zhang@pku.edu.cn
in linkedin.com/in/imbowei
📍 北京市海淀区中关村北大街 128 号
i 邮编:100080 i 在线简历

求职意向：算法工程师

🎓 教育背景

北京大学 硕士 数据科学专业, 自然语言处理方向, 综合排名: 6/67 2017.09 – 2020.07
语言计算与互联网挖掘研究组, 计算语言学教育部重点实验室, 北京大学大数据科学研究中心

中南大学 本科 智能科学与技术专业 2013.09 – 2017.06
综合成绩: 92.11/100 | 排名: 3/59

</> 学术 & 项目

2018.10 中文动词语义网络的半自动扩展 Python(PyTorch)
2018.08 > 半自动标注: 构建半自动化的数据标引体系, 快速高效扩充标注数据, 降低项目成本;
> 语义角色标注: 将原问题抽象为序列标注问题, 构建语义角色自动标注模型;
> 自注意力机制: self-attention 缓解长程语义损失的现象, 获取全局信息编码并且加快训练速度;
语义角色标注 半自动标注 自注意力机制

2018.07 Shared task: 文本溯源技术评测 (SMP2018-ETST) Ranking First Python
2018.05 > 复述检测: 提出一种新的复述检测算法 WFTS(Weighted F-measure-based method for text sourcing);
> 倒排索引: 通过加权的 tf-idf 建立倒排索引, 预筛选候选句子对;
> 模型融合: 运用集成学习方法融合多模型, 辅以少许规则, 提升复述检测准确率;
复述检测 语言模型 句子相似度 无监督学习 tf-idf 倒排索引 集成学习

2018.02 Leveraging Paraphrases to Enhance Semantic Dependency Parsing Python
2017.10 > 复述抽取: 提出一种新的复述抽取算法;
> 复述语料库: 构造首个中文句子级别的大规模复述语料库, 共有 506,610 对复述句子对;
> 语义一致性: 用迁移学习刷新 SDP 的 SOTA, 证明复述句子对的隐含语义一致性;
复述抽取 语义一致性 相似度计算 迁移学习 复述语料库 Co-Training

2017.05 基于扩展主题建模技术的医疗电子病历异常检测工具 SQL/Java/Js
2017.02 > 主题建模: 构建一个扩展的多视图 LDA 模型, 查找同一实例不同类别特征之间的不匹配情况;
> 异常检测: 分析五种相似性度量方法对比实验中的关键特性值, 进而判断两个主题分布的距离;
> 结果展现: 基于 web 的数据可视化。做到既能概览数据整体, 又能按需关注数据细节;
扩展的 LDA 主题模型 词袋模型 无监督学习 异常检测

🎒 校园生活

> 中南大学悦创空间 (中南大学首批创客空间、明星社团) 副会长 (优秀社团干部) 2015—2016
代表中南大学参展湖南省科协的首届“创客·湖南”科技产品、项目展示会;

> 中南大学智能小车协会 副会长 2014—2015
组织 2016 飞思卡尔智能小车华南赛及国赛, 开展一系列赛前培训课程及宣传活动;

> 学院足球队 (全院 3600 人以上) 队员 2013—2015
2013 校级新生赛四强; 2015 校级足球联赛冠军;

奖励 & 荣誉

北京大学三好学生	<10%	2018.10
Shared tasked : 文本溯源技术评测 (SMP2018-ETST)	第一名	2018.07
湖南省优秀毕业生		2017.06
中南大学优秀学生	<10%, 连续三年	2014-2016
中南大学优秀社团干部		2016.10
中南大学镭目奖学金	<0.7‰	2015.10
第五届华为杯中国大学生智能设计大赛	一等奖	2015.08
第十一届湖南省大学生程序设计竞赛 (应用开发类)	一等奖	2015.08
中南大学足球联赛	校级冠军	2015.05
国家励志奖学金	<4%	2014.10
中南大学新华联奖励金	<3‰	2014.10