

应聘职位

自然语言处理实习生

基本信息

姓名：李伟康 学校：北京大学 英语水平：CET 6
籍贯：河南洛阳 学历：研二在读（学硕） 出生年月：1993 年 10 月
性别：男 专业：计算语言学 爱好：篮球，乒乓球
联系电话：13146038034 邮箱：wave.jkd@pku.edu.cn 邮编：100871
通信地址：北京市海淀区颐和园路 5 号北京大学



教育经历

本科阶段-中央民族大学（计算机科学与技术, 2012.9 - 2016.7）

2012-2014 年，连续两年获得专业一等奖学金

2014-2015 年，获得国家奖学金

2012-2014 年，连续两年获得国家励志奖学金

2012-2013 年，获得三菱日产奖学金

2013-2014 年，获得院电子作品大赛编程类二等奖（获奖作品：基于 Heritrix 和 lucene 的手机领域的垂直搜索引擎）

2014-2015 年，参加学院 T. B2 创新比赛（作品名称：体型测试仪 APP），获得三等奖

2012-2014 年，连续两年获得校数学建模比赛三等奖

2015 年 9 月，本科时期成绩在所在专业位列第二，获得北京大学学术硕士研究生的保送资格

2016 年 7 月，获得北京市优秀毕业生称号

研究生阶段-北京大学（计算语言学, 2016.9 - 2019.7）

2017 年 9 月，获得学业专项奖学金

科研项目经历

2016 年 3-6 月，课题：问句理解的话题和焦点信息的识别。

通过基于树切分的 MDL 方法将问题表示为包含话题信息和焦点信息的数据结构，然后利用语言模型框架对这些信息进行建模，从而解决实际问题中的问题检索问题。这个课题主要通过传统的机器学习方法进行，过程中实现了一些简单模型（VSM、LMIR、BM25、Lucene）等。

2016 年 10-12 月，课题：通过深度学习方法进行问答系统中的答案选择工作（Answer Selection）的研究。

这个过程主要借助 tensorflow 深度学习框架对 CNN、RNN、CNN-RNN、RNN-CNN、RNN-attention 等模型在问答系统中的答案选择问题上进行了研究，并尝试了将问题和答案间的一些结构信息引入分类前的向量表示中。目前取得的结果相对可观。

2017 年 3-6 月，课题：深度学习中汉语字向量和词向量结合方式探究。

相比于英语，汉语中的字和词均包含很多有意义的信息。我们借助字和词的基础信息探究了汉语中的最佳语义单元的构建方式。通过实验，我们发现以字作为基础、辅之词的信息，构建出的语义单元效果最好，提升了基于篇章的问答任务的模型表现。

2017 年 6-7 月，任务：NLPCC 2017 基于篇章的问答测评比赛。

我们对 compare-aggregate 网络进行了改进，增加了其对篇章连贯性的建模能力。最终，我们的队伍在所有队伍中位列前 1/3。

2017 年 7 月-现在，课题：通过深度学习方法进行问答系统中的基于篇章的答案选择工作（Document-based Answer Selection）的研究。

工作实践经历

2014 学年 9 月 2015 年 6 月 **地点：**国家语言资源监测与研究中心少数民族语言分中心 **内容：**项目实习
参与实验室和清华合作的新华社项目，主要负责通过 Crawler4j 的数据抓取及 oracle 数据库的任务。熟悉了项目的开发流程，强化了自己的科研能力，对自然语言处理有了较深入的理解。

2015 学年 7 月至 2015 年 9 月 **地点：**浙江省杭州市阿里巴巴集团 **内容：**研发工程师实习
参与公司“村淘”项目，主要通过公司内部框架 WebX3.0 从事服务端的开发工作，并借助公司内部框架 BUI 进行部分网页设计。此次经历使我熟悉了部门内的协同工作和部门间的交流沟通，了解了团队合作。

2016 学年 7 月至 2016 年 9 月 **地点：**江苏省苏州市金童技术公司 **内容：**自然语言处理实习
主要从事财经数据方面的“实体关系”提取的研究工作，在实体识别的基础上，通过选取实体内、外、间的特征进行分析并构建 SVM 进行关系的初步提取（分类）。此次经历使我熟悉了如何将所学用于实际的市场需求之中，提升了知识运用能力。

2017 学年 3 月至 2017 年 8 月 **地点：**北京市腾讯视频 **内容：**自然语言处理实习
参与腾讯视频评论及天天快报的数据迁移工作，将存储于 Mysql 及 Mongodb 的数据导入到内部的分布式数据平台。参与部门“较真”产品的数据研发工作，完成了数据收集、分析及对上层接口的数据支持功能。此次经历使我了解了如何将实验室研究的内容和具体产品进行结合。

专业技能

- 1、熟悉自然语言处理的智能问答领域；
- 2、了解自然语言处理的相关领域及机器学习算法；
- 3、了解 python 语言及 tensorflow 深度学习框架；
- 4、了解 linux 操作系统基本知识与基本命令。

自我评价

- 1、工作态度严谨，责任心强，不轻言放弃，善于沟通，理解能力较强，具有较强的团队协作精神；
- 2、学习能力强，喜欢钻研技术问题；
- 3、有较强的专业基础知识，能独立地进行开创性的工作；

论文

研究生阶段

2017 年 8 月，以第一作者身份在国内自然语言处理核心会议 CCL 2017 发表论文“深度学习中汉语字向量和词向量结合方式探究”，获得最佳展示论文奖。收录于国内核心期刊《中文信息学报》2017 年第六期。

2017 年 11 月，以第一作者身份在国际人工智能顶级会议 AAAI 2018 发表论文“A Unified Model for Document-based Question Answering based on Human-like Reading Strategy”。

2018 年 1 月，以第一作者身份完成论文“An End-to-End Human-like Reader for Document-based Answer Selection and Triggering”，投稿于国际人工智能顶级会议 IJCAI 2018，目前正在审稿中。

2018 年 2 月，以第二作者身份完成论文“Learning Generic Sentence Representations with Mean-Max Attention Autoencoder”，投稿于国际自然语言处理顶会 ACL 2018，目前正在审稿中。

本科阶段

2013-2014 年，参加国家级**创新性科研项目**（GCCX），项目名称：基于视觉技术的四旋翼飞行器的目标定位系统的设计与实现，基于该项目以第一作者身份在 2015 应用科学与工程国际学术会议上发表论文“Design of Four-rotary-wing Aircraft Targeting System”，被 EI 检索

2014-2015 年，参加国家级**创新性科研项目**（GCCX），项目名称：面向藏文网络流通语料的语义分类研究，基于该项目以第二作者身份在 2015 计算机科学与通信工程国际会议上发表论文“Semantic Classification Method for Network Tibetan Corpus”，被 EI 检索

2014-2015 年，在老师的指导下进行藏文语义相似度的研究，以第二作者的身份在 2015 第六届 IEEE 软件工程与服务科学国际会议上发表论文“Research on Text Similarity Computing Based on Word Vector Model of Neural Networks”

2016 年 4 月，在导师指导下参加华为在日本举办的 NTCIR-12 比赛，参赛系统位列所有系统的前三分之一，赛会的论文被邀请做 poster。