

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

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

2016 年至今，北京大学信息科学技术学院研究生在读（计算语言学方向）
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。**

论文

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”。

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。

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，能独立并开创性工作；