

# 何 彬

语言技术研究中心 网络智能研究室

✉ hebin\_hit@hotmail.com | 🏠 binherunning.github.io/ | 🌐 BinHeRunning

“If you don't focus on strengths today, you're playing a losing game.”

## 教育经历

- 博士研究生, 哈尔滨工业大学 2013.09 - 现在  
专业: 计算机应用技术
- 理学硕士, 哈尔滨工业大学 2011.08 - 2013.07  
专业: 计算数学
- 理学学士, 哈尔滨工业大学 2007.08 - 2011.07  
专业: 信息与计算科学

## 发表论文

- **Bin He**, Yi Guan, and Rui Dai. “Classifying medical relations in clinical text via convolutional neural networks.” *Artificial Intelligence in Medicine* (2018). (IF: 2.009, In Press)
- Chao Zhao, Jingchi Jiang, Yi Guan, Xitong Guo, and **Bin He**. “EMR-based medical knowledge representation and inference via Markov random fields and distributed representation learning.” *Artificial Intelligence in Medicine* (2018). (IF: 2.009, In Press)
- Jia Su, **Bin He**, Yi Guan, Jingchi Jiang, and Jinfeng Yang. “Developing a Cardiovascular Disease Risk Factor Annotated Corpus of Chinese Electronic Medical Records.” *BMC Medical Informatics and Decision Making* 17, no. 1 (2017). (IF: 1.643)
- Zhipeng Jiang, Chao Zhao, **Bin He**, Yi Guan, and Jingchi Jiang. “De-Identification of Medical Records Using Conditional Random Fields and Long Short-Term Memory Networks.” *Journal of Biomedical Informatics* 75 (2017): S43–53. (IF: 2.753)
- **Bin He**, Bin Dong, Yi Guan, Jinfeng Yang, Zhipeng Jiang, Qiubin Yu, Jianyi Cheng, and Chunyan Qu. “Building a Comprehensive Syntactic and Semantic Corpus of Chinese Clinical Texts.” *Journal of Biomedical Informatics* 69 (2017): 203–17. (IF: 2.753)
- Runqi Wang, Xueli Li, Yuli Huang, **Bin He**, and Yi Guan. “Research on the Active Learning Method for Named Entity Recognition of Chinese Electronic Medical Records.” *China Digital Medicine* 12 (10) (2017): 51–53. (In Chinese)
- Jinfeng Yang, Yi Guan, **Bin He**, Chunyan Qu, Qiubin Yu, Yaxin Liu, and Yongjie Zhao. “Corpus Construction for Named Entities and Entity Relations on Chinese Electronic Medical Records.” *Journal of Software* 27, no. 11 (2016): 2725–46. (In Chinese)
- **Bin He**, Yi Guan, Jianyi Cheng, Keting Cen, and Wenlan Hua. “CRFs Based de-Identification of Medical Records.” *Journal of Biomedical Informatics* 58 (2015): S39–46. (IF: 2.753)
- Jinfeng Yang, Yi Guan, Xishuang Dong, and **Bin He**. “Representing Words as Lymphocytes.” *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* 4 (2014): 3146–47.

## 审稿中

- **Bin He**, Yi Guan, and Rui Dai. “Combining Convolutional and Recurrent Neural Network for Relation Classification in Clinical Text.” (under review)





## 审稿人

- Journal of Biomedical Informatics (IF: 2.753) (2017, 2018)

## 工作经历

- 识别英文维基百科翻译时不应被翻译的文本 📄 2014.08 - 2014.11  
微软亚洲研究院 (MSRA), 自然语言计算组 (NLC), 实习生 北京
- 把满足以下两个条件之一的文本串叫做 “不应被翻译的文本 (Should-not-be-translated text)” : 1. 翻译后带来理解困难 (翻译文本难以理解或指代不明); 2. 翻译后带来文本上的信息冗余。

## 参与项目

<b>中文临床文本上的自然语言处理在线演示系统</b> 	2015.09 - 现在
• 模块：分词、词性标注、浅层句法分析、命名实体识别和实体关系抽取	
<b>中文临床文本上的语料库构建</b> 	2013.12 - 现在
• 标注内容：分词、词性标注、浅层句法标注、实体标注、实体修饰标注和实体关系标注	
• 标注规范制定，标注工具开发	
<b>中文临床文本上的心血管疾病风险因素语料库构建</b> 	2015.10 - 现在
• 标注内容：风险因素及其指针、时间属性和修饰信息	
• 标注规范制定，标注工具开发	
<b>中文电子病历上的自然语言处理系统及其语料库构建项目</b>	2017.05 - 2017.12
北京云知声信息技术有限公司	
• 模块：分词、词性标注、浅层句法分析、命名实体识别和实体关系抽取	
<b>基于社保和医疗数据的智能实体识别项目</b>	2016.03 - 2016.12
易保互联医疗信息科技(北京)有限公司	
• 模型：基于字的条件随机场模型	
• 17 种医疗实体和 1 种社保实体	
<b>精神病的摄入记录中的隐私信息识别任务</b>	2016.06 - 2016.11
2016 CEGS N-GRID challenge, De-identification Task, SUNY at Albany	
• 模型：基于字符级表示的 bi-LSTM-CRF 模型	
• 28 种隐私信息类型	
<b>中文电子病历上的实体关系抽取引擎研究</b>	2015.12 - 2016.02
理光软件研究所(北京)有限公司	
• 模型：特征核与树核函数相结合的支持向量机模型	
• 15 种实体关系	
<b>糖尿病患者记录中的隐私信息识别任务</b> 	2014.03 - 2014.09
2014 i2b2/UTHealth challenge, De-identification Task, SUNY at Albany	
• 模型：基于正则表达式的预处理 + 条件随机场模型	
• 28 种隐私信息类型	

## 教学经历

助教，自然语言处理	2012、2013、2014 秋季
-----------	-------------------

## 技能情况

语言	英语
编程	Python (Tensorflow, theano), Shell, Matlab, C#/C++
其他	LaTeX, Git, HTML

## 获奖情况

- 2012.05 优秀团员, 哈尔滨工业大学 2011-2012 年度
- 2010.12 “三好学生”, 哈尔滨工业大学 2009-2010 年度
- 2010.10 优秀团干部, 哈尔滨工业大学 2009-2010 年度
- 2009.11 二等奖, 第一届全国大学生数学竞赛