BIN HE

Web Intelligent Laboratory, Language Technology Research Center

hebin_hit@hotmail.com | binherunning.github.io/ | BinHeRunning

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

EDUCATION

Ph.D. candidate, Harbin Institute of Technology Sep. 2013 - PRESENT

Computer Science and Technology, School of Computer Science and Technology Harbin, China

M.S., Harbin Institute of Technology Aug. 2011 - Jul. 2013

Computational Mathematics, School of Science Harbin, China

B.S., Harbin Institute of Technology Aug. 2007 - Jul. 2011

Information and Computing Science, School of Science Harbin, China

PUBLICATIONS

- **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)
- **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)
- 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)
- 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)
- 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.

In progress

- **Bin He**, Yi Guan, and Rui Dai. "Classifying medical relations in clinical text via convolutional neural networks." (*Artificial Intelligence in Medicine*, IF: 2.009, under review)
- **Bin He**, Yi Guan, and Rui Dai. "Combining Convolutional and Recurrent Neural Network for Relation Classification in Clinical Text." (under review)

REVIEWER

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

EMPLOYMENT

Indentify should-not-be-translated text from Wikipedia Aug. 2014 - Nov. 2014

Internship in natural language computing (NLC) group, Microsoft Research Asia Beijing, China

- If the translation of a phrase in the source text will bring difficulties (confused about the meaning of the text or do not know the object described) in understanding the translated text or make information in the translated text redundant, we call the portion "should-not-be-translated text".

PROJECTS

Online demo for natural language processing of Chinese clinical text Sep. 2015 - PRESENT

- Modules: word segmentation, part-of-speech tagging, chunk tagging, named entity recognition and relation extraction

Corpus construction on Chinese clinical text Dec. 2013 - PRESENT

- Annotations: word segmentation, part-of-speech tags, chunk tags, entities, assertions and relations
- Annotation guidelines and annotation tools were developed for this project

Corpus construction on Chinese clinical text (Risk fators for cardiovascular disease) Oct. 2015 - PRESENT

- Annotations: risk factors, indicators, time attributes and assertions
- Annotation guidelines and annotation tools were developed for this project

Natural language processing system on Chinese EMR and its related resource construction May 2017 - Dec. 2017

With Beijing Unisound Information Technology Co.,Ltd.

- Modules: word segmentation, part-of-speech tagging, chunk tagging, named entity recognition and relation extraction

Intelligent entity recognition project based on medical and social insurance data Mar. 2016 - Dec. 2016

With Ebaonet Medical Information Technology (Beijing) Co., Ltd.

- Method: a character-based CRF model
- 17 medical entity types and 1 social insurance entity type

Removing protected health information from psychiatric intake records Jun. 2016 - Nov. 2016

2016 CEGS N-GRID challenge, De-identification Task, SUNY at Albany

- Method: a character-level bi-directional LSTM-CRF network
- 28 sub-categories

Entity relation extraction system on Chinese electronic medical records Dec. 2015 - Feb. 2016

With Ricoh Software Research Center(Beijing) Co., Ltd.

- Method: combination of feature-based and kernel-based SVM
- 15 entity relation types

Removing protected health information from diabetic patient records Mar. 2014 - Sep. 2014

2014 i2b2/UTHealth challenge, De-identification Task, SUNY at Albany

- Method: rule-based preprocessing + CRF
- 28 sub-categories

TEACHING EXPERIENCE

TA, Natural language processing Fall 2012, Fall 2013, Fall 2014

SKILLS

Languages	Chinese, English
Programming	Python (Tensorflow, theano), Shell, Matlab, C#/C++
Others	LaTeX, Git, HTML

Awards

Nov. 2009 **Second Class Award, Top 1%**, The first mathematics competition of Chinese college students

Dec. 2010 **Merit Student**, Harbin Institute of Technology