

# CHAO ZHAO

📍 803 Zonghe Building, 92 West Dazhi St., Harbin City, China, 150001

✉️ zhaochaocs@gmail.com

☎️ +86-187-0462-4509

🌐 <https://zhaochaocs.github.io/>

## 🎓 EDUCATION

### University of California, Santa Cruz

*Incoming Ph.D. student in Computer Science*

📅 (Sep 2018) -

◦ Santa Cruz, CA, USA

- Research fields: natural language processing, machine learning

### Harbin Institute of Technology

*M.Sc in Computer Technology, School of Computer Science and Technology*

📅 Sep 2016 - Jun 2018

◦ 2 years, Harbin, China

- Research fields: natural language processing, knowledge graph, health informatics, machine learning

### Harbin Institute of Technology

*B.E. in Flight Vehicle Design and Engineering, School of Astronautics*

📅 Sep 2012 - Jun 2016

◦ 4 years, Harbin, China

## 📖 PUBLICATIONS

- **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*, 87 (2018): 49-59.
- **Chao Zhao**, Jingchi Jiang, Zhiming Xu, and Yi Guan. “A study of EMR-based medical knowledge network and its applications.” *Computer Methods and Programs in Biomedicine* 143 (2017): 13-23.
- **Chao Zhao**, Min Zhao, and Yi Guan. “Constructing a Hierarchical User Interest Structure based on User Profiles.” *2017 IEEE 17th International Conference on Data Mining Workshops (ICDMW)*. pages 156-162, Nov 2017.
- 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*, S75 (2017): S43-S53, **co-first author**
- **Chao Zhao**, Min Zhao, and Yi Guan. “Classification of entities via their descriptive sentences.” arXiv:1711.10317, 2017.
- Jingchi Jiang, Jing Xie, **Chao Zhao**, Jia su, Yi Guan, and Qiubin Yu. “Max-Margin Weight Learning for Medical Knowledge Network.” *Computer Methods and Programs in Biomedicine* 156 (2018): 179-190.
- Jingchi Jiang, Jichuan Zheng, **Chao Zhao**, Jia Su, Yi Guan, and Qiubin Yu. “Clinical-decision support based on medical literature: A complex network approach.” *Physica A: Statistical Mechanics and its Applications* 459 (2016): 42-54.
- Jingchi Jiang, Xueli Li, **Chao Zhao**, Yi Guan, and Qiubin Yu. “Learning and inference in knowledge-based probabilistic model for medical diagnosis.” *Knowledge-Based Systems* 138 (2017): 58-68.
- Jingchi Jiang, Yi Guan, Jia Su, **Chao Zhao**, and Jinfeng Yang. “HIT-WI at TREC 2015 Clinical Decision Support Track.” In *TREC*. 2015.
- Jingchi Jiang, Yi Guan, and **Chao Zhao**. “WI-ENRE in CLEF eHealth Evaluation Lab 2015: Clinical Named Entity Recognition Based on CRF.” In *CLEF (Working Notes)*. 2015.

## 🏢 PROFESSIONAL EXPERIENCE

### Construction of Chinese medical knowledge base

*Final year graduate project, Part II*

📅 Mar 2018 - Jun 2018

◦ 4 months

- Acquiring triple knowledge automatically from medical texts, and integrating them as a medical knowledge base (Including a concept base and a medical knowledge graph).
- Designing the descriptive schema of the knowledge base using JSON language, as well as its corresponding storage, search and validation tools.

**Entity classification based on its descriptive sentences**

📅 May 2017 - Aug 2017

*Internship at Knowledge Graph Group, Baidu Inc.*

◦ 3 months

- Determined the category of arbitrary entities according to their descriptive sentences, with a CNN-based text classification model.
- Designed a clustering module to filter the noisy instances and alleviate the class imbalance problem.
- Applied this system to 2.1 million entities, and 1.1 million are successfully classified with a precision of 99.4%.

**Concept association from user interest perspective**

📅 Feb 2017 - Apr 2017

*Internship at Knowledge Graph Group, Baidu Inc.*

◦ 3 months

- Integrated the interests in user profiles as a network and then explored its structure by community detection.
- Labeled each interest community with relevant concepts, to depict the topics of interest at the concept level.

**Knowledge representation and reasoning from Chinese electronic medical records**

📅 Oct 2015 - Feb 2017

*Final year graduate project, Part I*

◦ 1.3 years

- Constructed a medical knowledge network containing the medical entities and entity relationships
- Regarded the states of medical problems and events as random variables and modeled their relationships via Markov networks and Markov logic networks.
- Represented the medical entities as embeddings to depict their similarities, and then designed new energy functions for probabilistic inference.

**Error detection and correction of Chinese texts**

📅 Nov 2016 - Dec 2016

*With Zhongke Huilian Inc.*

◦ 1 months

- Adopted n-gram language model with Kneser-Ney smoothing to detect and correct typos in Chinese texts.
- Corrected about 80% of the errors of test data, with only a small corpus (about 4M) to train the language model.

**Removing of protected health information from psychiatric evaluation records**

📅 Jun 2016 - Oct 2016

*i2b2 2016 CEGS N-GRID De-identification Task, SUNY at Albany*

◦ 4 months

- Implemented a de-identification system using the character-level bi-LSTM networks with enhanced word embeddings.
- Attained the  $F_1$  measure of 0.899. The best score among the 15 participating teams was 0.914.

**Construction of the knowledge graph on consumer-oriented consuming interest**

📅 Mar 2016 - May 2016

*With Ricoh Software Research Center Beijing Co.,Ltd.*

◦ 3 months

- Analyzed the activities, similarities, importance, clustering, and purchase possibilities of the customers, according to the consumption behavior of each customer in e-commerce sites.

**Biomedical literature retrieving for clinical question answering**

📅 Jun 2015 - Aug 2015

*2015 TREC Clinical Decision Support Track*

◦ 2 months

- Retrieved the relevant medical literature of the given clinical records from 700, 000 articles via Lucene and Indri.
- Constructed a medical literature network and then identified the potentially relevant ones from the literature pool.

**⚙️ SKILLS**

**Language:** Chinese, English  
**Programming:** Python (Tensorflow, theano), Java, C, C++, Matlab, Shell  
**Others:** Git, Hadoop, L<sup>A</sup>T<sub>E</sub>X, HTML, Photoshop

**🏆 SELECTED AWARDS**

- |  |          |
|--|----------|
| ◦ Innovation Scholarship (Top 1% in China)   | Dec 2017 |
| ◦ National Scholarship for Graduate Students (Top 1% in China)                           | Nov 2017 |
| ◦ Outstanding Graduate Award (Top 10% of all graduates in HIT)                           | Jun 2016 |
| ◦ Outstanding Final Year Project Thesis (Top 10% of all graduates in HIT)                | Jun 2016 |
| ◦ Top-grade Scholarship (Top 1% of all students in HIT)                                  | Sep 2015 |
| ◦ First National Prize for China Undergraduate Mathematical Modeling (Top 1.5% in China) | Sep 2015 |
| ◦ National Scholarship for Undergraduate Students (Top 1% in China)                      | Sep 2014 |