# Chao Zhao

Room 803, Zonghe Building, Harbin Institute of Technology, Harbin City, China, 150001

zhaochaocs@gmail.com +86-187-0462-4509 http://wi.hit.edu.cn/dev/chao/

## Education

**Harbin Institute of Technology** Sep 2016 - (Jun 2018)
*M.Sc in Computer Technology, School of Computer Science and Technology* ◦ 2 years, Harbin, China

- GPA: 82.95 Rank: 2/186
- Core courses: Algorithms, Machine learning, Data mining, Natural language processing, Knowledge engineering, Artificial intelligence, Deep learning for NLP

**Harbin Institute of Technology** Sep 2012 - Jun 2016
*B.E. in Flight Vehicle Design and Engineering, School of Astronautics* ◦ 4 years, Harbin, China

- GPA: 92.16 Rank: 5/85

## Research Interest

Natural Language Processing, Knowledge Graph, Health Informatics

## Professional Experience

**Construction of Chinese medical knowledge base** Sep 2017 - Now
*Final year graduate project, Part II* ◦ 3 months

- Design a system to acquire knowledge automatically from medical texts and web sources, and integrate them as the medical knowledge base (consisting of medical ontology and medical knowledge graph).
- Design the descriptive schema of the knowledge base based on the json-ld format, as well as its corresponding storage, search and validation tools.

**Entity classification based on its descriptive sentences** May 2017 - Aug 2017
*Intern at Knowledge Graph Group, Baidu Inc.* ◦ 3 months

- Implemented a system to determine the category of arbitrary entities into one of the 48 pre-defined classes according to their descriptive sentences, with a convolutional neural network-based text classification model.
- Designed a clustering module to filter the noise instances from the training set, as well as to obtain a better down-sampling subset from the majority classes to alleviate the class imbalance problem.
- Applied this system to 2.1 million entities in Baidu Baike, a Wikipedia-like web encyclopedia, and 1.1 million are successfully classified with a precision of 96.2%.

**Concept association from user interest perspective** Feb 2017 - Apr 2017
*Intern at Knowledge Graph Group, Baidu Inc.* ◦ 3 months

- Integrated the user interests in user profiles as a network, and then explored its organizational structure by detecting the network communities.
- Labeled each interest community automatically with several the most relevant concepts (e.g., war, war movie), to depict the interest topic in the concept level.
- Introduced the “related_to” relationship to the “isA”-based taxonomy according to the co-occurrences of concepts in one community, to make the taxonomy more suitable for interest recommendation.

**Error detection and correction of Chinese text** 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 text.
- Corrected about 80% errors of the test data, with only a small corpus (about 4M) to train the language model.

**Removing of protected health information over psychiatric evaluation records** Jun 2016 - Oct 2016
*i2b2 2016 CEGS N-GRID De-identification Task, SUNY at Albany* ◦ 4 months

- Implemented an end2end de-identification system using the character-level bidirectional long short-term memory networks with the enhanced word embeddings.
- Attained the $F_1$ measure of 85.2%, higher than the median of 82.2% among all participants. Subsequent works further improved the $F_1$ measure to 89.98%.

**Construction of consumer-oriented consuming interest knowledge graph** Mar 2016 - May 2016
*With Ricoh Software Research Center Beijing Co.,Ltd.* ◦ 3 months

- Analyzed the activity, similarity, importance, clustering, and purchase possibility of the customers, according to their consuming behavior of each customer in e-commerce sites.

- Developed a demo for knowledge base visualization, with the help of the d3 javascript library.

**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 (e.g., symptoms, diseases) and entity relationships (e.g., cause, alleviate), and then discovered its small-world and scale-free properties, as well as the community structure.
- Regarded entities as random variables and modeled their relationships using Markov random fields (MRF), and then derived the algorithm to infer the states of disease variables, according to the known symptom and test variables.
- Represented the nodes in MRF as entity embeddings to depict their similarities, and then designed two new energy functions accordingly for above inference.

**Biomedical literature retrieving for clinical questions answering** Jun 2015 - Aug 2015
*2015 TREC Clinical Decision Support Track* ◦ 2 months

- Retrieved the relevant medical literatures of the given clinical records from more than 700, 000 articles.
- Built the literature retrieving system using Lucene and Indri.
- Constructed a medical literature network according to the retrieved articles, and then identified the potentially relevant ones from the literature pool.

**Clinical named entity recognition in French** Apr 2015 - May 2015
*2015 CLEF Multilingual Information extraction task* ◦ 2 months

- Recognized ten kinds of clinical named entities from the French medical texts using conditional random fields.

## Publications

**Published**

- **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.
- 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.

**In Progress**

- **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)*. IEEE, 2017. (Accepted)
- 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*, in press, co-first author)
- **Chao Zhao**, Jingchi Jiang, and Yi Guan. "EMR-based medical knowledge representation and inference via Markov random fields and distributed representation learning." (*Artificial Intelligence in Medicine*, under review)
- 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*, under review)

## Skills

| | |
|---|---|
| **Language:** | Chinese(native), English |
| **Programming Language:** | Java, C, C++ |
| **Script Language:** | Python (Tensorflow, theano), Matlab, Shell |
| **Operating System:** | Linux |
| **Others:** | Git, Hadoop, LaTeX, HTML, Photoshop |

## Selected Awards

- 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