

# Cong Tan(谭聪)

## Curriculum Vitae

### Natural Language Processing Engineer

#### Education

- 2013.09--2016.01 **Master in Computer Science and Technology**, *Shenzhen Graduate School of HIT*, GPA: 3.516 .  
Advisor: *Prof. Wang Xiaolong*  
Research: Question Answering, Information Retrieval, Machine Learning, Natural Language Processing
- 2009.09--2013.07 **B.S in Computer Science and Technology**, *Hangzhou Dianzi University*, GPA: TOP 15 %.

#### Research Program

- NTCIR-11 Temporal Information Retrieval task's objective is to foster research in temporal information access. The result of document retrieval are content relative , also temporal correlation. Our result get the second among all the teams, and publish a paper[3].  
Temporaliala Major work: First, using Solr to indexing the documents and searching every query; Second, extracting some text features between query and documents searched by Solr; Last, Adopt Learning to Rank algorithm to re-ranking the result to get the final documents list.  
Light spot: 1. classify the query by its temporal property as four categories(Last, recency, future, atemporal); 2. Take the time information of queries into account when re-ranking the candidate result list.
- SemEval-2015 Answer Selection in Community Question Answering is judge the quality of each answers of the questions in task3 CQA. This task include two languages, Arabic and English. Our team got the second and third in English and Arabic task separately. [1].  
Major work: extract text features, includeing: Bag-of-words, tf-idf, co-occurrence words between question and answer and word2vec; then design the classify algorithms.  
Light spot: 1. Against the unbalanced train data (Good, Potential, Bad, Dialogue ), we design different methods to balance it. 2. Design two-level hierarchical classification method, integrate Random Forest and SVM.
- SemEval-2015 QA TempEval is a temporal information extraction task. But use QA to evaluate temporal information under- task5 standing. Major works: 1. extract the temporal expressions and normalized , 2. extract events, 3. extract the temporal relations. [2].  
Light spot: using bagging method to merging the different annotated results.
- Github 1. Chinese Word Segmentation tools based HMM model in python;  
2. Implement the common machine learning algorithms, like SVM, Adaboost, kmeans, LDA and so on;  
3. Pull requests for some projects, like Stanford CoreNLP, timeml-qa, caevo, mainly fix the bugs and improve it.

#### Professional Skills

- 🔧 Master in C++, Java, Python programming language
- 🔧 Master in major Data Structure and algorithms. Awarded the bronze medal of the 8th ACM programming contest in Zhejiang province.
- 🔧 Master in major Machine Learning Algorithms, Nature Language Processing technology and interest in Deep Learning.
- 🔧 Familiar with Linux environment , know about common commands and know well about version manager tool , such as git.

---

## Awards and Grants

- 2014-10 First-class scholarship of Shenzhen Graduate School of HIT
- 2014-10 "Three Good" student of Shenzhen Graduate School of HIT
- 2013-09 School principal level scholarship of Shenzhen Graduate School of HIT

---

## Campus practice

- 2014.12.5-9 As a volunteer of the conference on Natural Language Processing and Chinese Computing(NLPCC) 2014 and CCF Advanced Disciplines Lectures(ADL)
- 2014.01- Minister of MSTC (Microsoft Technology Club) of Shenzhen University Town
- 2014.12 I have organized many activities for club members and take part in Microsoft Summer camp in 2014.
- 2013.09- As the member of the Student Union of Shenzhen Graduate School of HIT, organizes many activities, like debate competitions, science fair and as the host of campus host contest.
- 2014.09

---

## Publications

- [1] Yongshuai Hou, Cong Tan, Qingcai Chen, and Xiaolong Wang. Hitsz-icrc: An integration approach for qa tempeval challenge. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 398--402, Denver, Colorado, June 2015. Association for Computational Linguistics.
- [2] Yongshuai Hou, Cong Tan, Xiaolong Wang, Yaoyun Zhang, Jun Xu, and Qingcai Chen. Hitsz-icrc: Exploiting classification approach for answer selection in community question answering. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 403--409, Denver, Colorado, June 2015. Association for Computational Linguistics.
- [3] Yongshuai Hou, Cong Tan, Jun Xu, Youcheng Pan, Qingcai Chen, and Xiaolong Wang. Hitsz-icrc at ntcir-11 temporalia task. In *Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies*, pages 468--473, 2014.
- [4] Yongshuai Hou, Xiaolong Wang, Qingcai Chen, Man Li, and Cong Tan. User input classification for chinese question answering system. In *Machine Learning and Cybernetics*, volume 481, pages 52--59. Springer Berlin Heidelberg, 2014.