

清华-汇联2014-2015合作项目 年度总结

王东

2015/08/13

工作目标

- 增强问答系统的模糊匹配能力和输入容错能力;
- 开发知识库的在线自动学习功能;
- 开发面向语音和文本输入的多模态问答系统;
- 开发与CallCenter集成的电话语音问答系统;
- 开发基于语音的方言检测功能;

参与工作成员

- 赵芳、刘荣、薛陈
- 刑超、洛天一、王晓曦

工作内容 (1)

- 特定领域非结构化QA系统
 - 基于检索的QA方法研究
 - 基于模糊匹配的问答求精
 - 基于ListRank的问答排序
 - 基于贝叶斯的在线学习
 - 指代消解与纠错
 - 句式标注句式推荐

工作内容（2）

- 特定领域结构化QA
 - 基于模板的结构化QA系统

工作内容（3）

- 深度QA
 - 基于RNN的QA向量方法
 - 基于RNN的关系学习方法
 - 基于RNN的自动生成方法

工作内容（4）

- 语音QA

研究成果-论文（标注汇联）

- Chao Liu, Dong Wang, Zhiyong Zhang, "Pruning Deep Neural Networks by Optimal Brain Damage", Interspeech 2014
- Xuwei Zhang, Yiye Lin, Dong Wang, "LASSO-BASED REVERBERATION SUPPRESSION IN AUTOMATIC SPEECH RECOGNITION", ICASSP 2015.
- Xi Ma, Xiaoxi Wang, Dong Wang, Low-Frequency Word Enhancement with Similar Pairs in Speech Recognition, ChinaSIP 2015.
- Xiangyu Zeng, Shi Yin, Dong Wang, Zhiyong Zhang, "Learning Spectrum Distortion Caused by Speech Rate in Speech Recognition", interspeech 2015
- Xi Ma, Xiaoxi Wang, Dong Wang, "Recognize Foreign New Words With Similar Words", interspeech 2015
- Noisy Training for Deep Neural Networks in Speech Recognition, Shi Yin, Chao Liu, Zhiyong Zhang, Yiye Lin, Dong Wang, Javier Tejedor, Thomas Fang Zheng, Yiguo Li, EURASIP Journal on Audio, Speech, and Music Processing, 2015(2).
- Rong Liu, Dong Wang, Chao Xing, Document classification based on word vectors, ISCSLP 2014 (in Chinese).
- Chao Xing, Dong Wang, Xuwei Zhang, Chao Liu, document classification based on i-vector distributions, APSIPA 2014.
- Chao Xing, Dong Wang, Chao Liu, Yiye Lin, "Normalized Word Embedding and Orthogonal Transform for Bilingual Word Translation", NAACL 2015
- **Dongxu Zhang, Dong Wang, Rong Liu, "Joint Semantic Relevance Learning with Text Data and Graph Knowledge", ACL 2015, workshop CVSC.**
- **Tianyi Luo, Dong Wang, Rong Liu and Yiqiao Pan, "Stochastic Top-k ListNet", EMNLP 2015**

研究成果-专利

- 基于多层贝叶斯网络的在线自学习问答系统
- 多模态问答系统

未来研究方向

- RNN S2S QA
 - 去掉对传统NLP的依赖，特别是语法分析
 - 解决人工标注问题
 - 解决词义歧义问题
 - 解决顺序问答
 - 解决拼写错误、语音识别产生的噪声
 - 部分解决前后相关性问题
 - 统一结构化QA和非结构化QA

未来研究方向

- **Memory Machine**
 - 解决长时记忆问题
 - 从学习知识到学习策略

未来研究方向

- 阅读理解 (Reading comprehension)
- 重复学习 (Reinforcement learning)
- 大数据智能 (big data intelligence)

几点感想

- 这一年的合作是有成效的，共赢的
- 让专业的人做专业的事至关重要
- 分工清晰有利于合作推进
- 脚踏实地，立足长远