

## Recent Progress in Deep Learning for NLP

**Instructors:** Zhengdong Lu and Hang Li

### **Abstract:**

Neural network-based methods have been viewed as one of the major driving force in the recent development of natural language processing (NLP). We all have witnessed with great excitement how this subfield advances: new ideas emerge at an unprecedented speed and old ideas resurge in unexpected ways. In a nutshell, there are two major trends:

- Ideas and techniques from other fields of machine learning and artificial intelligence (A.I.) have increasing impact on neural network-based NLP methods.
- With end-to-end models taking on more complex tasks, the design of architecture and mechanisms often needs more domain knowledge from linguists and other domain experts.

Both trends are important to researchers in the computational linguistics community. Fundamental ideas like external memory or reinforcement learning, although introduced to NLP only recently, have quickly lead to significant improvement on tasks like natural language generation and question answering. On the other hand, with complicated neural systems with many cooperating components, it calls for linguistic knowledge in designing the right mechanism, architecture, and sometimes training setting.

As a simple example, the introducing of automatic alignment in neural machine translation, has quickly led to the state-of-the-art performance in machine translation and triggered a large body of sequence-to-sequence models. It is therefore important to get the researchers in computational linguistics community acquainted with the recent progress in deep learning for NLP.

We will focus on the work and ideas strongly related to the core of natural language and yet not so familiar to the majority of the community, which can be roughly categorized into: 1) the differentiable data-structures, and 2) the learning paradigms for NLP.

Differentiable data-structures, starting with the memory equipped with continuous operations in Neural Turing Machine, have been the foundation of deep models with sophisticated operations. Some members of it, such as Memory Network, have become famous on tasks like question answering and machine translation, while other development in this direction, including those with clear and important application in NLP, are relatively new to this community.

Deep learning, with its promise on end-to-end learning, not only enables the training of complex NLP models from scratch, but also extends the training setting to include remote and indirect supervision. We will introduce not only the end-to-end learning in its general notion, but also newly emerged

topics in formulating learning objective, taming the non-differentiable operations, and designing the learning system for getting supervision signal from real world.

### **Outline:**

- Part I: Overview (30 minutes)
  - Basic concepts and techniques in deep learning for NLP
  - A brief summary of recent progress
- Part II: Differentiable data-structures for NLP (one hour)
  - External memory, reading and writing, attention
  - Short-term memory, episodic memory, and long-term memory
  - Structures in memory, deep memory-based architectures
  - Application on machine translation and dialogue
  - Other differentiable data-structures
  - Neural stack and queue for sequence processing
  - Neural pointers for nonlinear data-structures, e.g., trees
- Part III: Learning paradigms (one hour)
  - End-to-end learning in general
  - Grounded language learning in neural models
  - Learning from execution: e.g., for semantic parsing in neural QA models
  - Grounding to other modalities, e.g., images
  - Reinforcement Learning in deep NLP models
  - For learning dialogue policy
  - For language generation
  - For other discrete choices in neural language models, e.g., memory-access
  - Other non-differentiable choices in learning deep NLP models, eg. Minimum risk training
- Part IV: Conclusion

### **About the Instructors:**

Zhengdong Lu is a senior researcher at Noah's Ark Lab, Huawei Technologies. His research interests are neural network-based methods for natural language processing, including dialogue, machine translation, semantic parsing, and reasoning. Previously he was an associate researcher at Microsoft Research Asia and a postdoctoral researcher at University of Texas at Austin, after receiving his Ph.D. degree from Oregon Health and Science University in 2008 in computer science. He has published over 30 papers in prestigious journals and conferences, including NIPS, ICML, ACL, KDD, IJCAI and AAAI, including over 10 recent papers on deep learning methods for NLP and AI.

Hang Li is director of the Noah's Ark Lab of Huawei Technologies, adjunct professors of Peking University and Nanjing University. His research areas include information retrieval, natural language processing, statistical machine learning, and data mining. He is ACM Distinguished Scientist. Hang

graduated from Kyoto University in 1988 and earned his PhD from the University of Tokyo in 1998. He worked at the NEC lab as researcher during 1991 and 2001, and Microsoft Research Asia as senior researcher and research manager during 2001 and 2012. He joined Huawei Technologies in 2012. Hang has published three technical books, and more than 100 technical papers at top international conferences including SIGIR, WWW, WSDM, ACL, EMNLP, ICML, NIPS, SIGKDD and top international journals including CL, NLE, JMLR, TOIS, IRJ, IPM, TKDE, TWEB, TIST. He and his colleague's papers received the SIGKDD-08 best application paper award, the SIGIR-08 best student paper award, the ACL-12 best student paper award. Hang worked on the development of several products such as Microsoft SQL Server 2005, Office 2007, Office 2010, Live Search 2008, Bing 2009, Bing 2010, Huawei AppStore, Huawei Phones. He has 40 granted US patents. Hang is also very active in the research communities and has served or is serving top international conferences as PC chair, Senior PC member, or PC member, including SIGIR, WWW, WSDM, ACL, EMNLP, NIPS, SIGKDD, ICDM, ACML, and top international journals as associate editor or editorial board member, including CL, IRJ, TIST, JASIST, JCST. Hang Li gave a number of tutorials on various topics about machine learning for natural language processing and information retrieval, including a tutorial on learning to rank at ACL 2009.