

# Deep Learning and Continuous Representations for Natural Language Processing

Scott Wen-tau Yih, Xiaodong He & Jianfeng Gao

## Introduction

Deep learning techniques have demonstrated tremendous success in the speech and language processing community in recent years, establishing new state-of-the-art performance in speech recognition, language modeling, and have shown great potential for many other natural language processing tasks. The focus of this tutorial is to provide an extensive overview on recent deep learning approaches to problems in language or text processing, with particular emphasis on important real-world applications including language understanding, semantic representation modeling, question answering and semantic parsing, etc.

In this tutorial, we will first survey the latest deep learning technology, presenting both theoretical and practical perspectives that are most relevant to our topic. We plan to cover common methods of deep neural networks and more advanced methods of recurrent, recursive, stacking and convolutional networks. In addition, we will introduce recently proposed continuous-space representations for both semantic word embedding and knowledge base embedding, which are modeled by either matrix/tensor decomposition or neural networks.

Next, we will review general problems and tasks in text/language processing, and underline the distinct properties that differentiate language processing from other tasks such as speech and image object recognition. More importantly, we highlight the general issues of natural language processing, and elaborate on how new deep learning technologies are proposed and fundamentally address these issues. We then place particular emphasis on several important applications, including (1) machine translation, (2) semantic information retrieval and (3) semantic parsing and question answering. For each task, we will discuss what particular architectures of deep learning models are suitable given the nature of the task, and how learning can be performed efficiently and effectively using end-to-end optimization strategies.

## Outline

### ***Part I. Background of neural network learning architectures***

- Background: A review of deep learning theory and applications in relevant fields
- Advanced architectures for modeling language structure
- Common problems and concepts in language processing:
  - Why deep learning is needed

- Concept of embedding
- Classification/prediction vs. representation/similarity
- Learning techniques: regularization, optimization, GPU, etc.

### ***Part II. Machine translation***

- Overview of Machine Translation
- Deep learning translation models for SMT
- Recurrent neural network for language model for SMT
- Sequence to sequence machine translation

### ***Part III. Learning semantic embedding***

- Semantic embedding: from words to sentences
- The Deep Structured Semantic Model/Deep Semantic Similarity Model (DSSM)
- DSSM in practice: Information Retrieval, Recommendation, Auto image captioning

### ***Part IV. Natural language understanding***

- Continuous Word Representations & Lexical Semantics
- Semantic Parsing & Question Answering
- Knowledge Base Embedding

### ***Part V. Conclusion***

## **Instructor bios**

[Scott Wen-tau Yih](#) is a Senior Researcher in the Machine Learning Group at Microsoft Research Redmond. His research interests include natural language processing, machine learning and information retrieval. Yih received his Ph.D. in computer science at the University of Illinois at Urbana-Champaign. His work on joint inference using integer linear programming (ILP) [Roth & Yih, 2004] helped the UIUC team win the CoNLL-05 shared task on semantic role labeling, and the approach has been widely adopted in the NLP community. After joining MSR in 2005, he has worked on email spam filtering, keyword extraction and search & ad relevance. His recent work focuses on continuous semantic representations using neural networks and matrix/tensor decomposition methods, with applications in lexical semantics, knowledge base embedding and question answering. Yih received the best paper award from CoNLL-2011 and has served as area chairs (HLT-NAACL-12, ACL-14) and program co-chairs (CEAS-09, CoNLL-14) in recent years.

[Xiaodong He](#) is a Researcher of Microsoft Research, Redmond, WA, USA. He is also an Affiliate Professor in Electrical Engineering at the University of Washington, Seattle, WA, USA. His research interests include deep learning, information retrieval, natural language understanding, machine translation, and speech recognition. Dr. He has published a book and more than 70 technical papers in these areas, and has given tutorials at international conferences in these fields. In benchmark evaluations, he and his colleagues have developed entries that obtained No. 1 place in the 2008 NIST

Machine Translation Evaluation (NIST MT) and the 2011 International Workshop on Spoken Language Translation Evaluation (IWSLT), both in Chinese-English translation, respectively. He serves as Associate Editor of IEEE Signal Processing Magazine and IEEE Signal Processing Letters, as Guest Editors of IEEE TASLP for the Special Issue on Continuous-space and related methods in natural language processing, and Area Chair of NAACL2015. He also served as GE for several IEEE Journals, and served in organizing committees and program committees of major speech and language processing conferences in the past. He is a senior member of IEEE and a member of ACL.

[Jianfeng Gao](#) is a Principal Researcher of Microsoft Research, Redmond, WA, USA. His research interests include Web search and information retrieval, natural language processing and statistical machine learning. Dr. Gao is the primary contributor of several key modeling technologies that help significantly boost the relevance of the Bing search engine. His research has also been applied to other MS products including Windows, Office and Ads. In benchmark evaluations, he and his colleagues have developed entries that obtained No. 1 place in the 2008 NIST Machine Translation Evaluation in Chinese-English translation. He was Associate Editor of ACM Trans on Asian Language Information Processing, (2007 to 2010), and was Member of the editorial board of Computational Linguistics (2006 – 2008). He also served as area chairs for ACL-IJCNLP2015, SIGIR2015, SIGIR2014, IJCAI2013, ACL2012, EMNLP2010, ACL-IJCNLP 2009, etc. Dr. Gao recently joined Deep Learning Technology Center at MSR-NEXT, working on Enterprise Intelligence.