# Deep Learning in Lexical Analysis and Parsing

**Wanxiang Che**
Harbin Institute of Technology
car@ir.hit.edu.cn

**Yue Zhang**
Singapore University of Technology and Design
yue_zhang@sutd.edu.sg

## Abstract

Lexical analysis and parsing tasks, modeling deeper properties of the words and their relationships to each other, typically involve word segmentation, part-of-speech tagging and parsing. A typical characteristic of such tasks is that the outputs have structured. All of them can fall into three types of structured prediction problems: sequence segmentation, sequence labeling and parsing.

In this tutorial, we will introduce two state-of-the-art methods to solve these structured prediction problems: graph-based and transition-based methods. While, traditional graph-based and transition-based methods depend on "feature engineering" work, which costs lots of human labor and may misses many useful features. Deep learning just right can overcome the above "feature engineering" problem. We will further introduction those deep learning models which have been successfully used for both graph-based and transition-based structured prediction.

## 1 Tutorial Overview

Lexical analysis and parsing tasks, modeling deeper properties of the words and their relationships to each other, typically involve word segmentation, part-of-speech tagging and parsing. A typical characteristic of such tasks is that the outputs have structured. All of them can fall into three types of structured prediction problems: sequence segmentation, sequence labeling and parsing. Because of the pervasive problem of ambiguity, none of the above problems are trivial to predict.

Two different methods have been used to solve these structured prediction problems, including graph-based methods and transition-based methods. The former differentiates output structures based on their characteristics directly, while the latter transforms output construction processes into state transition processes, differentiating sequences of transition actions.

The conditional random fields (CRFs) are typical graph-based methods, which aim to maximize the probability of the correct output structure. The graph-based methods can also be applied to dependency parsing, where the aim change to maximize the score of the correct output structure.

At the beginning, the transition-based methods were applied into dependency parsing. Latter, it was found that sequence segmentation and labeling can also be modeled into a sequence of state transition.

Both graph-based and transition-based methods depend on "feature engineering" work, that is huge hand-crafted features and their combination should be designed according to different tasks. It usually cost lots of human labor. More seriously, many useful features may be missed by human beings. The features extracted from training data also lack in generalization.

Neural networks, also with a fancy name deep learning, just right can overcome the above "feature engineering" problem. In theory, they can use non-linear activation functions and multiple layers to automatically find useful features. The novel network structures, such as convolutional or recurrent, help to reduce the difficulty further.

These deep learning models have been successfully used for both graph-based and transition-based structured prediction. In this tutorial, we will give a review of each line of work, by contrasting them with traditional statistical methods, and organizing them in consistent orders.



*Proceedings of the The 8th International Joint Conference on Natural Language Processing*, pages 1–2,
Taipei, Taiwan, November 27 – December 1, 2017. 

## 2 Outline

1. Typical Lexical Analysis and Parsing Tasks
    - Word segmentation
    - POS tagging
    - Parsing
    - Structured Prediction
        - Sequence Segmentation
        - Sequence Labeling
        - Parsing
2. Deep Learning Background
    - Multilayer Perceptron (MLP)
    - Back-propagation
    - Word Embedding
    - Recurrent Neural Networks (RNNs)
    - Recursive Neural Networks
    - Convolutional Neural Networks (CNNs)
3. State-of-the-art Methods
    - Graph-based Methods
        - Conditional Random Fields
        - Graph-based Dependency Parsing
    - Transition-based Methods
        - Greedy Shift-Reduce Dependency Parsing
        - Greedy Sequence Labeling
        - Beam-search Training and Decoding
4. Neural Graph-based Methods
    - Neural Conditional Random Fields
    - Neural Graph-based Dependency Parsing
5. Neural Transition-based Methods
    - Neural Greedy Shift-Reduce Dependency Parsing
    - Neural Greedy Sequence Labeling
    - Globally Optimized Models

## 3 Instructors

**Wanxiang Che** is currently an associate professor of school of computer science and technology at Harbin Institute of Technology (HIT) and visiting associate professor of Stanford University (at NLP group in 2012). His main research area lies in Natural Language Processing (NLP). He currently leads a national natural science foundation of China, a national 973 and a number of research projects. He has published more than 40 papers in high level journals and conferences, and published two textbooks. He and his team have achieved good results in a number of international technical evaluations, such as the first place of CoNLL 2009 and the fourth place of CoNLL 2017. He was an area co-chair of ACL 2016, publication co-chairs of ACL 2015 and EMNLP 2011. The Language Technology Platform (LTP), an open source Chinese NLP system he leads to develop, has been authorized to more than 600 institutes and individuals including Baidu, Tencent and so on. He achieved the outstanding paper award honorable mention of AAAI 2013, the first prize of technological progress award in Heilongjiang province in 2016, Google focused research award in 2015 and 2016, the first prize of Hanwang youth innovation award and first prize of the Qian Weichan Chinese information processing science and technology award in 2010.

**Yue Zhang** is currently an assistant professor at Singapore University of Technology and Design. Before joining SUTD in July 2012, he worked as a postdoctoral research associate in University of Cambridge, UK. Yue Zhang received his DPhil and MSc degrees from University of Oxford, UK, and his BEng degree from Tsinghua University, China. His research interests include natural language processing, machine learning and artificial Intelligence. He has been working on statistical parsing, parsing, text synthesis, machine translation, sentiment analysis and stock market analysis intensively. Yue Zhang serves as the reviewer for top journals such as Computational Linguistics, Transaction of Association of Computational Linguistics (standing review committee) and Journal of Artificial Intelligence Research. He is the associate editor for ACM Transactions on Asian and Low Resource Language Information Processing. He is also PC member for conferences such as ACL, COLING, EMNLP, NAACL, EACL, AAAI and IJCAI. He was the area chairs of COLING 2014, NAACL 2015, EMNLP 2015, ACL 2017 and EMNLP 2017. He is the TPC chair of IALP 2017.