CE:7455: Deep Learning for Natural Language Processing: From Theory to Practice

Shafiq Joty
School of Computer Science and Engineering
Nanyang Technological University
Singapore

1 Course Objectives

Natural language processing (NLP) is one of the most important fields in **artificial intelligence** (AI). It has become very crucial in the information age because most of the information is in the form of unstructured text. NLP technologies are applied everywhere as people communicate mostly in language: language translation, web search, customer support, emails, forums, advertisement, radiology reports, to name a few.

There are a number of core NLP tasks and machine learning models behind NLP applications. **Deep learning**, a sub-field of machine learning, has recently brought a paradigm shift from traditional task-specific feature engineering to end-to-end systems, and has obtained high performance across many different NLP tasks and downstream applications. Tech companies like Google, Baidu, Alibaba, Apple, Amazon, Facebook, Tencent, and Microsoft are now actively working on deep learning methods to improve their products. For example, Google recently replaced their traditional statistical machine translation and speech recognition systems with systems based on deep learning methods.

1.1 Intended Learning Outcomes

In this course, students will learn state-of-the-art deep learning methods for NLP. Through lectures and practical assignments students will learn the necessary tricks for making their models work on practical problems. They will learn to implement, and possibly to invent their own deep learning models using available deep learning libraries like TensorFlow.

2 Course Outline

2.1 Week 1: Introduction [Lecture: 2 hours + Discussion: 1 hour]

(i) What is Natural Language Processing?
(ii) Why is language understanding difficult?
(iii) What is Deep Learning?
(iv) Deep learning vs. other machine learning methods?
(v) Why deep learning for NLP?
  (a) Continuous encoding like human brains
  (b) Breakthrough in speech and vision, and now in NLP
  (c) Learning representations at different linguistic levels
(vi) Applications of deep learning to NLP
  (a) Morphology and word representation
  (b) Syntax & semantic parsing of sentences

(c) Discourse analysis
(d) Information extraction & search
(e) Classification & ranking
(f) Machine translation (Google Translator)
(g) Question answering (Siri, Google Assistant)
(h) Dialogue systems & chatbots
(i) Speech recognition
(j) NLP in industry

(vii) Knowing the target group (background, field of study, programming experience)

(viii) Expectation from the course (theoretical and practical)

2.2 Week 1: Python & Tensorflow Basics [Supplementary reading]

(i) Programming in Python
   (a) Why Python?
   (b) Control flow, loops
   (c) Numerical programming with numpy/scipy
   (d) Tips & practical issues

(ii) Deep learning with Tensorflow
   (a) Why Tensorflow?
   (b) Tensorflow basics (Variables, Placeholders, Operators)
   (c) Computational graph
   (d) Tensorboard
   (e) Practical tips

2.3 Week 2: Machine Learning Basics [Lecture: 2 hours + Discussion: 1 hour]

(i) What is Machine Learning?

(ii) Supervised vs. unsupervised learning

(iii) Linear Regression

(iv) Logistic Regression: a linear classification model

(v) Multi-class classification

(vi) Parameter estimation (MLE & MAP)

(vii) Gradient-based optimization & SGD

(viii) Practical exercise with Tensorflow
   (a) Email spam filter
   (b) Sentiment classification

2.4 Week 3: Neural Network Basics [Lecture: 2 hours + Discussion: 1 hour]

(i) From Logistic Regression to Feed-forward NN

(ii) Activation functions

(iii) SGD with Backpropagation

(iv) Adaptive SGD (adagrad, adam, RMSProp)

(v) Convolutional neural networks (CNNs)

(vi) Demo and practical exercises
   (a) Text classification

2.5 Week 4: Word Vectors [Lecture: 2 hours + Discussion: 1 hour]

(i) Word meaning
(ii) Distributed representation of words
(iii) Word2Vec models
  (a) Skip-gram
  (b) CBOW
  (c) Glove
  (d) FastText
  (e) ELMo
(iv) Negative sampling
(v) Evaluating word vectors
  (a) Intrinsic evaluation
  (b) Extrinsic evaluation
(vi) Demo and practical exercises

2.6 Week 5: Window-based Approach [Lecture: 1 hour + Discussion: 30 min]

(i) Language models
(ii) Neural language model
(iii) Chunking and NER with window-based approach
(iv) Pairwise training and max-margin loss
(v) CNNs for NLP
(vi) Demo and practical exercises

2.7 Week 5: Recurrent Neural Nets [Lecture: 1 hour + Discussion: 30 min]

(i) Basic RNN structures
(ii) Language modeling with RNNs
(iii) Backpropagation through time
(iv) Text generation with RNN LM
(v) Sequence labeling with RNNs
  (a) Named Entity Recognition (NER)
  (b) Part-of-speech (POS) tagging
  (c) Opinion analysis
(vi) Sequence classification with RNNs
  (a) Sentiment classification
  (b) Dialogue act classification
(vii) Demo and practical exercises
(viii) **Assignment 1 out**

2.8 Week 6: More Recurrent Neural Nets [Lecture: 2 hrs + Discussion: 1 hr]

(i) Issues with Vanilla RNNs
  (a) Information decay
  (b) Vanishing gradient
  (c) Exploding gradient
(ii) Gradient clipping
(iii) Gated Recurrent Units (GRUs) and LSTMs
(iv) Bidirectional RNNs

(v) Multi-layer RNNs
(vi) Fun with RNNs
(vii) Demo and practical exercises

2.9 Week 7: Recursive Neural Nets & Parsing [Lec: 2 hrs + Discussion: 1 hr]

(i) Compositionality in language & recursion
(ii) Recursive vs. recurrent NN
(iii) Parsing with tree-structured recursive NN
(iv) Tree LSTMs
(v) Backpropagation through tree
(vi) Other applications of recursive NN
  (a) Scene parsing
  (b) Fine-grained sentiment analysis
  (c) Semantic relationship identification
(vii) Neural dependency parsing
  (a) Transition-based neural parsing
  (b) Graph-based neural parsing
(viii) Demo and practical exercises
(ix) **Assignment 1 in**

2.10 Week 8: Neural Nets for Discourse Analysis [Lec: 2 hrs + Disc.: 1 hr]

(i) Types of discourse
(ii) Discourse analysis tasks
  (a) Topic analysis
  (b) Coreference resolution
  (c) Coherence modeling
  (d) Discourse parsing
  (e) Conversational structures
(iii) Neural topic segmenter
(iv) Neural coreference resolution
(v) Neural discourse parsing
(vi) Neural coherence modeling
(vii) Neural dialogue act modeling
(viii) Demo and practical exercises

2.11 Week 9: Machine translation & Seq2Seg Models [Lec: 2 hrs + Disc: 1 hr]

(i) Machine translation
  (a) Early days (1950s)
  (b) Statistical machine translation or SMT (1990 - 2010)
  (c) Alignment in SMT
  (d) Decoding in SMT
(ii) Neural machine translation or NMT (2014 -)
(iii) Encoder-decoder model for NMT
(iv) Advantages and disadvantages of NMT
(v) Greedy vs. beam-search decoding
(vi) MT evaluation
(vii) Demo and practical exercises

2.12 Week 10: Seq2Seg Models with Attentions [Lec: 2 hrs + Disc: 1 hr]

(i) Information bottleneck issue with vanilla Seq2Seq
(ii) Attention to the rescue
(iii) Details of attention mechanism
(iv) Attention variants
(v) Soft vs. hard attention
(vi) Local vs. global attention
(vii) NMT results
(viii) Demo and practical exercises

2.13 Week 10: More Advanced Attentions & Applications [supp. reading]

(i) Advanced attentions & pointer nets
(ii) Self attention
(iii) Dealing with open and large vocabulary
    (a) Byte-pair encoding
    (b) Sub-word NMT
    (c) Hybrid NMT
(iv) Open issues with NMT
(v) Other applications of attentional Seq2Seq model
    (a) Summarization
    (b) Dialogue
    (c) Parsing
    (d) Multimodal (image and video captioning)
(vi) Transformer Networks
(vii) CNN-based encoder-decoder model
(viii) Demo and practical exercises
(ix) **Assignment 2 out**

2.14 Week 11: Semisupervised Deep Learning for NLP [Lec: 2 hr + Dis: 1 hr]

(i) Why semi-supervsied?
(ii) Pre-training methods (Word embeddings, CoVe, ELMo)
(iii) Consistency regularization
(iv) Cross-view consistency
(v) Generative adversarial nets (GANs)
(vi) Training with adversarial examples
(vii) Domain adversarial nets (DANs)
(viii) Transfer learning with DANs
(ix) Domain adaptation with DANs

2.15 Week 12: Deep Reinforcement Learning for NLP [Lec: 2 hr + Dis: 1 hr]

(i) What is RL?
(ii) Key concepts: Rewards, Policy, Value Function
(iii) What is Deep RL?
(iv) Policy-based Deep RL
    (a) Deep Policy Network

(b) Policy Gradient

(v) Deep Q-Learning

(vi) Applications of Deep RL in NLP

(a) Abstractive summarization
(b) Dialogue generation
(c) Question answering
(d) Multimodal (image and video captioning)
(e) Machine translation

(vii) Conclusions

2.16 Week 12: Variational Methods for Deep NLP [supp. reading]

(i) Deep learning meets graphical models

(ii) Variational autoencoders

(iii) Variational Generative adversarial nets (GANs)

(iv) Applications

(v) **Assignment 2 in**

2.17 Week 13: Limits & Future of Deep NLP [supp. reading]

(i) Multi-sentence processing

(ii) Multi-task learning

(iii) Multimodal learning

(iv) Model interpretability

2.18 Week 13: In-class Project presentation [3 hour]

(i) Project presentation: 15 min/group

3 Course Prerequisites

(a) Proficiency in Python (using numpy and tensorflow). There is a lecture for those who are not familiar with Python.

(b) College Calculus, Linear Algebra

(c) Basic Probability and Statistics

(d) Machine Learning basics

4 Teaching and Assessment Approach

Weekly Workload

- Every **two-hour** lecture will be accompanied with practice problems implemented in TensorFlow, a popular Python framework for deep learning.
- There will be one **one-hour** office hour per week to discuss assignments and project.

Assignments (individually graded)

- There will be **two (2) assignments** contributing to 50% of the total assessment.
- Students will be graded individually on the assignments. They will be allowed to discuss with each other on the homework assignments, but they are required to submit individual write-ups and coding exercises.

Final Project (Group work but individually graded)

- There will be a **final project** contributing to the remaining 50% of the total coursework assessment.
- The project will be a group or individual work depending on the student's preference. Students will be graded individually. The final project presentation will ensure the student's understanding of the project.

5 Course Instructor

Dr. Shafiq Joty (srjoty@ntu.edu.sg)
Assistant Professor
NTU-NLP Group
School of Computer Science & Engineering
Nanyang Technological University
Singapore

Dr. Shafiq Joty is an Assistant professor at the School of Computer Science and Engineering, NTU. He holds a PhD in Computer Science from the University of British Columbia. His work has primarily focused on developing NLP tools, and exploiting these tools effectively in downstream NLP applications like machine translation, summarization, sentiment analysis, question answering, and multi-modal applications like image/video captioning. His work has appeared in major journals and conferences such as CL, JAIR, TACL, ACL, EMNLP, NAACL, CVPR, ECCV, IJCAI, IUI, and ICWSM. Shafiq is a recipient of NSERC CGS-D scholarship and Microsoft Research Excellent Intern award.