Foundations for Natural Language Processing

Revision

Ivan Titov

School of informatics

Recall: class goals

▸ We focused mainly on core ideas and methods needed for language technologies (and eventually for applications).

  ▸ Linguistic facts and issues

  ▸ Computational models and algorithms

▸ We provided also a preview to (nearly) state-of-the-art methods

▸ … but more advanced methods and specific application areas are covered in 4/5th year courses:

  ▸ Natural Language Understanding, Generation and Machine Translation (NLU+)

  ▸ Automatic Speech Recognition

Today

- We will go <u>briefly</u> over the material and I will highlight some topics and ideas …
    - Important: do not assume that if I have not highlighted a certain topic, it is not going to be examined
- Ask questions about the class
    - Especially about my part but I will try my best to answer questions about other parts as well

Revision

▸ Do readings from Jurafsky & Martin (and other sources we pointed out)

   ▸ We won't examine what has not been at all discussed in the class …

   ▸ but the the materia covers much of it in more detail, with more examples…

      ☐ It will definitely help you prepare better to the exam

▸ Use lecture slides / recording and quizzes

Ambiguities / Challenges in NLP

▸ Remember different challenges for NLP (e.g., sparsity, ambiguity, robustness, …)

  ▸ Be ready to identify the key challenges for a given specific setting

  ▸ (and ways to circumvent these challenges)

▸ Ambiguity

  ▸ A topic we discussed a lot in the class

  ▸ Why a problem? How to deal with it? Types of ambiguities?

Annotation and evaluation

▸ Make sure you understand challenges and considerations in defining annotation guidelines and organizing annotation

  ▸ E.g., be ready to examine two potential annotation schemas and discuss pros and cons

▸ Make sure you understand how to evaluate different types of NLP models and hypotheses

  ▸ E.g., you can be asked to consider a specific setting / application and come up with a way of evaluating NLP tools, or discuss advantages / disadvantages of alternatives

  ▸ We also talked about the evaluation in the context of seq2seq models (ngram overlap metrics, BLEU)

Language Models

▸ **Ngram language models**

 ▸ How to estimate? Limitations?

▸ **Smoothing (for LM and more generally)**

 ▸ Methods you discussed with Alex

▸ **Neural language models**

 ▸ Contrast them with ngram models (will get back to it)

▸ **Evaluation of language models**

ML techniques

▸ Naïve Bayes and Logistic Regression

  ▸ incl. their pros / cons, contrast with <u>Neural Classifiers</u>

  ▸ Estimation, Inference ('the use')

▸ Tagging / Part-of-speech tagging

  ▸ Hidden Markov Models

▸ Algorithms:

  ▸ Viterbi (edit distance, HMM)

  ▸ Expectation Maximization (informally, the way we introduced it in the class)

  ▸ Forward algorithm for HMMs

Be able to deal with novel problems

▸ Methods are your toolbox

▸ Be prepared to reduce a given problem to the modeling set-ups we discussed

  ▸ How can we convert a problem into a (set of) classification problems? Sequence labelling problem?

▸ The same for algorithms, evaluation, …

Morphology

- Morphology
  - across languages
- Parsing / generation
- Finite state transducers

Syntactic parsing

▸ **Syntactic ambiguity**

　▸ Types, challenges

▸ **(Probabilistic) CFGs**

　▸ Estimation for PCFGs

　▸ CKY for CFGs and PCFGs

　▸ Weaknesses of treebank PCFGs

　▸ Producing more powerful PCFGs

　▸ Evaluation

▸ **Dependency parsing**

　▸ Relations to constituency syntax / head rules

　▸ Transition-based parsing

　▸ Non-projectivity

　▸ Intuition for Graph-based vs Transition-based

Compositional semantics

▸ First order logic vs. propositional logic

  ▸ Why / when one and another

▸ Principle of compositionality

  ▸ Augmenting grammars with logic

  ▸ Scope ambiguity

More semantics / lexical semantics

▸ Word senses

  ▸ Relations / classes

  ▸ The generative lexicon

  ▸ Word sense disambiguation

▸ Distribution semantics

  ▸ Understand the underlying assumption (what it can do, what it cannot do…)

  ▸ Count-based methods and LSA

  ▸ Neural Embeddings

    ▸ SkipGram, including negative sampling

Neural networks for classification

▸ Neural classification and connection to logistic regression

▸ Bags of embeddings

▸ Convolutional neural networks

  ▸ their relations to ngram models

  ▸ interpretability

▸ Recurrent Neural Networks

▸ Differences in their expressivity, multilayer/multidirectional architectures

Neural networks for language modeling

▸ Reduction to classification

▸ Relations to smoothing in ngram models

▸ Estimation

▸ Decoding (greedy, sampling with temperature..)

Sequence-to-Sequence modeling

▸ From language modeling to seq2seq

▸ Vanilla Encoder-Decoder and its weakness

▸ Attention (please remember scoring functions; but don't have to remember details of Luong / Bahdanau architecture)

▸ Training / decoding

▸ Issues (e.g., hallucination)

▸ Evaluation for text generation

Transformers

▸ QKV attention

▸ Multi-head attention model

▸ Key modules and how they fit together

▸ Masked attention

▸ Linearities and Interpretability

Transfer learning

▸ Transfer learning with word embedding, its version and limitations

▸ Word-in-context (ELMO)

▸ Pretrain and fine-tune (BERT)

▸ Masked language modelling objective

This is all!

Thank you for attending the class and good luck at the exam