Subject Review

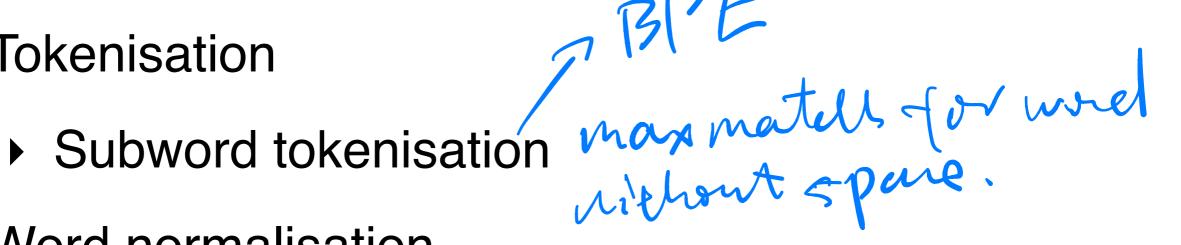
COMP90042
Natural Language Processing
Lecture 22



Preprocessing

- Sentence segmentation
- Tokenisation
- Word normalisation
 - Derivational vs. inflectional morphology
 - Lemmatisation vs. stemming
- Stop words





N-gram Language Models

- Derivation
- Smoothing techniques

 - Take away mass, redistribute evenly

 Absolute discounting

 Katz Backoff

 Me distribute band on this lander
 - Kneser-Ney smoothing
 - Interpolation
- **Evaluation**

erpolation (must sased on versitile.

tion

Pert 20 Pitabet...

COMP90042 L22

* Text Classification

- Building a classification system
- Text classification tasks
 - Topic classification
 - Sentiment analysis
 - Native language identification
 - Automatic fact-checking
- Algorithms
 - Naive-Bayes, logistic regression, SVM
 - kNN. neural networks
- Bias vs. variance
- Evaluation metrics
 - Precision, recall, F1



Part-of-Speech Tagging

- English POS
- Inglish PUS manually annote sentences

 Open vs. closed POS classes
- Tagsets
 - Penn Treebank tags
- Automatic taggers
 - Rule-based
 - Statistical
 - Unigram, classifier-based, HMM

Hidden Markov Models

- Probabilistic formulation
 - Parameters: emission and transition probabilities P(...) P(...)
- Training
- · Viterbi algorithm -> decode. Ineedy decode.
- Generative vs. discriminative models

work through add features performance

DL: Feed-forward Networks

- Formulation
- Designing FF networks for NLP tasks
 - Topic classification
 - Language model
 - POS tagging
- Word embeddings Sparsity. how to fet WE.
 Convolutional networks

DL: Recurrent Networks

- Formulation
- RNN language models
 why
- LSTM
 - Functions of gates
 - Variants
- Designing RNN for NLP tasks
 - Text classification: sentiment analysis
 - POS tagging

Lexical Semantics

- Definition of word senses, glosses
- Lexical relationships
 - Synonymy, antonymy, hypernymy, meronymy
- Structure of WordNet
- Word similarity
 - Path length, depth information, information content
- Word sense disambiguation
 - Supervised vs. unsupervised

Distributional Semantics

- Matrices for distributional semantics
 - VSM, TF-IDF, word-word co-occurrence
- Association measures: PMI, PPMI
- Count-based methods: SVD
- Neural methods: skip-gram, CBOW
- Evaluation
 - Word similarity, analogy

Contextual Representation

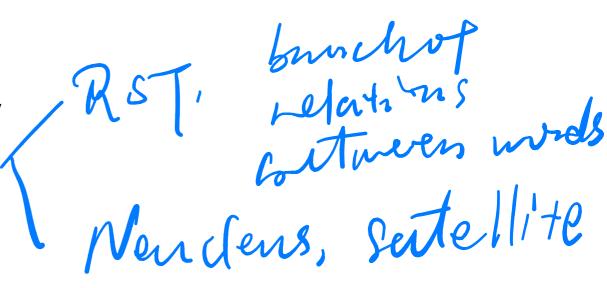
- Formulation with RNN
- ELMo
- BERT
 - Objectives
 - Fine-tuning for downstream tasks
- Transformers
 - Multi-head attention

Discourse

- Motivation for modelling beyond words
- Discourse segmentation



- Discourse parsing
 - Rhetorical structure theory
- Anaphora resolution
 - Centering
 - Supervised models



Formal Language Theory & FST

- Formal language theory as a framework for defining language
- Regular languages
 - Closure properties
- Finite state acceptors

- Word morphology, weighted variant
- N-gram language model as WFSA
- Finite state transducers
 - Weighted variant, edit distance, morphological analysis

Context-Free Grammar

- Center embedding
- Basics of CFG
- Syntactic constituent and its properties
- CFG parsing
 - Chomsky normal form
 - CYK
- English sentence structure (Penn Treebank)

Probabilistic Context-Free Grammar

- Ambiguity in grammars
- Basics of probabilistic CFGs
- Probability of a CFG tree
- Parsing
 - Probabilistic CYK
- Improvements
 - Parent conditioning
 - Head lexicalisation

Dependency Grammar

- Notion of dependency between words
- Universal dependency
- Properties of dependency trees
 - Projectivity
- Parsing
 - Transition-based
 - Graph-based

Machine Translation

- Statistical MT
 - Language + translation model
 - Alignments
- Neural MT
 - Encoder-decoder
 - Beam search decoding
 - Attention mechanism
- Evaluation: BLEU

Information Extraction

- Named entity recognition
 - NER tags, IOB tagging, models
- Relation extraction
 - Rule-based, supervised, semi-supervised, distant supervision
 - Unsupervised: ReVERB
- Temporal expression extraction
- Event extraction

Question Answering

- IR-based QA
 - Question processing, answer type prediction
 - Passage retrieval, answer extraction
- Reading comprehension
 - Models: LSTM-based, BERT
- Knowledge-based QA
- Hybrid QA: IBM Watson

Topic Modelling

- Evolution of topic models
- LDA
 - Sampling-based learning
 - Hyper-parameters
- Evaluation:
 - Word intrusion
 - Topic coherence

Summarisation

- Extractive summarisation
 - Single-document
 - Unsupervised content selection
 - Multi-document
 - Maximum marginal relevance
- Abstractive summarisation
 - Neural models: copy mechanism
- Evaluation: ROUGE

Exam

Exam Structure

- 40 marks
- Gradescope
- 2 hours in total:
 - 1 hour 45 minutes of writing
 - ▶ 15 minutes to upload answers
- 3 parts:
 - A: short answer questions
 - B: method questions
 - C: algorithm questions

Short Answer Questions

- Several short questions
 - 1-2 sentence answers for each
 - Definitional, e.g. what is X?
 - Conceptual, e.g. relate X and Y, purpose of Z?
 - May call for an example illustrating a technique/ problem

Method Questions

- Longer answer
- Focus on analysis and understanding
 - Contrast different methods
 - Outline or analyse an algorithm
 - Motivate a modelling technique
 - Explain or derive mathematical equation

Algorithmic Questions

- Perform algorithmic computations
 - Numerical computations for algorithm on some given example data
 - Present an outline of an algorithm on your own example
- Not required to simplify maths (e.g. leaving fractions as log(5/4) is fine)

What to Expect

- Even coverage of topic from the semester
- Be prepared for concepts that have not yet been assessed by homework / project
- Prescribed reading is fair game for topics mentioned in the lectures and workshops
- Mock exam