Statistical Natural Language Processing

An overview of NLP applications: some topics not covered during the course

Çağrı Çöltekin

University of Tübingen
Seminar für Sprachwissenschaft

Summer Semester 2018

Some remarks on the exam

first things first

- Exam is scheduled on Wed July 25, at 14:00 (sharp)
- The duration is 2 hours
- The exam (type of questions, length) will be similar to last year's exam
- Topics may shift, covering anything we studied during the course
- You can bring a 'cheat sheet':
  - Single a4 paper with anything that you want to remember
  - You can use both sides
  - You can hand-write/print as small as you like, but should be legible with bare eye

Questions?

Resit

nobody will need it, but just in case...

- Note that your final score is combination of
  - Exam (70 %)
  - Assignments (30 % + 5 %)
  - Attendance (+ 5 %)
  - Easter-egg bonus
- The exam scores will be announced (latest) the week after the exam
- Last two assignments will be graded during second week of August
- You can take a resit exam if you fail (<60 % of total)
- Resit will be scheduled before the beginning of the winter semester. Likely first (maybe second) week of October

Assignment 6: clarification

- You are not required implement full Good-Turing discounting
- You only need to estimate a single number (in two places) ‘the probability of unobserved n-grams’ (p_0) according to Good-Turing
- For the rest of ‘discounting’ you need to only adjust your probability estimates so that probabilities of observed n-grams sum to $(1 - p_0)$, and probabilities of unobserved n-grams sum to p_0

A quick summary so far

Part I Background & machine learning

- Math: linear algebra, probability & information theory
- Supervised methods: regression / classification
- How evaluate machine learning methods
- Sequence learning
- Unsupervised learning
- Neural networks: MLP, CNN, RNN

Part II NLP methods

- Tokenization / segmentation
- N-gram language models
- Statistical parsing
- Vector representations / vector semantics

Part III (would be) NLP applications

Machine translation

what & why

- Motivation for MT does not need many words: it is the example you give to your grandmother when she asks 'what does a computational linguist do?'
- Rule-based machine translation is difficult
- Most modern MT systems are statistical

Machine translation

how: basic idea

$$\arg\max_{e} p(e|f) = \arg\max_{e} p(f|e)p(e)$$

- The above defines a noisy-channel model
- $p(f|e)$ estimated with the noisy channel idea
- $p(e)$ is a language model

Machine translation

how: phrase-based MT

$$\arg\max_e p(e|f) = \arg\max_e p(f|e)p(e)$$

Using a parallel corpus,

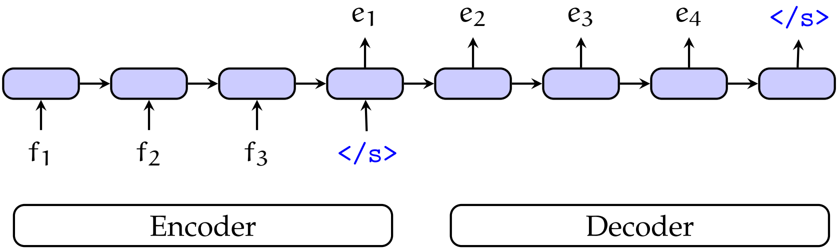
- Align sentences, estimate $p(f|e)$
- We can estimate $p(e)$ even from a (larger) mono-lingual corpus

Machine translation

how: end-to-end systems (mostly neural)

$$\arg\max_{e} p(e|f) = \arg\max_{e} p(f|e)p(e)$$

Estimate $p(e|f)$ directly, typically with a recurrent neural network

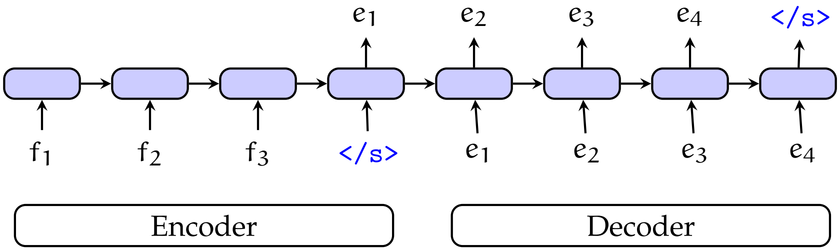


Machine translation

how: end-to-end systems (mostly neural)

$$\arg\max_{e} p(e|f) = \arg\max_{e} p(f|e)p(e)$$

Estimate $p(e|f)$ directly, typically with a recurrent neural network



Machine translation

How does it work? (1)

Machine translation

How does it work? (2)

Machine translation

How does it work? (seriously)

- Works fine if you have *lots* of parallel text
- A lot of work remains in:
  - Solving issues with ambiguities, idioms, special/rare constructions
  - Low resource languages

Entity recognition

what & why

UN	Secretary-General	Antonio	Guterres
ORG	NONE	PER	PER
plans	to	visit	Ukraine
NONE	NONE	NONE	GEO

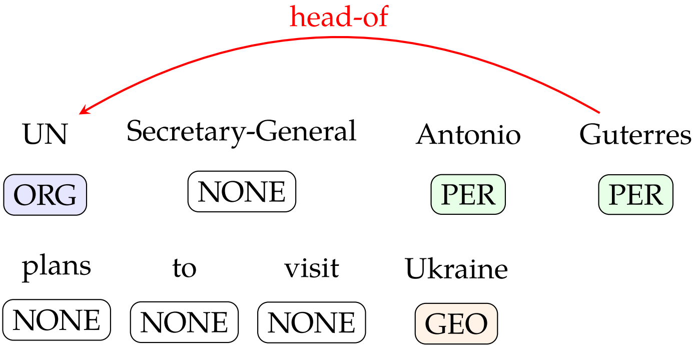
- Many other applications depend on locating certain entities in text
- Typical entities interest include: people, organizations, locations
- Can be application specific too: e.g., drug/disease names

Entity recognition
how

- Generally viewed as a typical sequence learning task
- Any sequence learning model applies: e.g., HMMs, RNNs
- Some linguistic processing is often helpful (e.g., POS tagging)

Relation extraction

what & why



- For many other tasks, we do not only need entities, but the relations between them

Relation extraction

how

- Many approaches rely on patterns
- Using classifiers on annotated data is also popular
  1. Extract all pairs of entities of interest
  2. Train the classifier, to predict whether the entities are related
- Semi-supervised learning methods are common
- Does it also look like dependency parsing?

Summarization

what & why

- We have lots, lots of text on any subject of choice
- Probably you use them daily (e.g., news aggregators), but applications of summarization are much wider
- Summarization
  - reduces the reading time
  - helps selecting right documents to read
  - may improve/help with
    - indexing
    - storing/processing/searching large document collections
    - other applications like question answering

Summarization

how

Extractive summarization selects important sentences from the text.

- The task is binary classification (paying attention to the sequence)
- Classifier decides whether to keep or discard the sentence in the summary

Abstractive summarization fuses sentences, combining and re-structuring them

How about treating it like a machine translation problem?

- RNNs of the sort used in MT have lately been popular for summarization too

Question answering

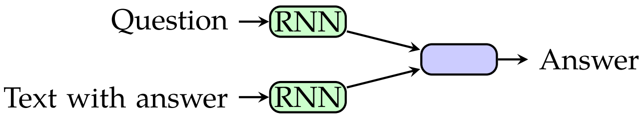
what & why

- QA is another NLP application that needs little explanation
- The task is given a question find the answer in a database, or a unstructured document collection
- Domain specific specific are common
- More general QA systems can perform well, sometimes better than humans (e.g., IBM Watson)
- Also an important part of for modern personal assistant systems
- Most systems are complex, combining many of the methods we discussed in the class (and more)

Question answering

how

- The natural language questions are turned int formal queries, searched in a database
  - linguistic processing (parsing) helps
  - Supervised methods can learn queries from natural language questions
- Again, RNNs have been recent popular approach



More…

- Topic modeling / text mining
- Information extraction
- Coreference resolution
- Semantic role labeling
- Dialog systems
- Speech recognition
- Speech synthesis
- Spelling correction
- Text normalization

Summary

- Many other problems/applications in NLP can be solved with the methods we studied in this course
- Most of the real-world problems require a combination of multiple methods

Summary

- Many other problems/applications in NLP can be solved with the methods we studied in this course
- Most of the real-world problems require a combination of multiple methods

Next:

Mon	Summary & your questions
Wed	Exam
Fri	Exam discussion

Additional reading, references, credits

- The textbook (Jurafsky and Martin 2009) includes detailed information on many of these problems/applications (more on the 3rd edition draft)