# Assignment 9 + Word Sense Disambiguation

## (SNLP Tutorial 10)

Vilém Zouhar, Awantee Deshpande, Julius Steuer

29th June, 1st July

# Assignment 9

- Exercise 1: Feature Engineering, Classification
- Bonus: Support Vector Machines

# Word Sense Disambiguation

Apple is full of vitamins.
Apple was struggling last quarter.
Apple was thrown away from the meeting.

$f(w, C) = s \in S_w$
$f(\text{Apple}, * \text{ was thrown away from the meeting}) \in \{\text{fruit}, \text{company}\}$

# Word Sense Disambiguation

Machine translation:

- Apfel ist voller Vitamine.
- Apple ist voller Vitamine.
- Apfel hatte im letzten Quartal Probleme.
- Apple hatte im letzten Quartal Probleme.

Information retrieval:

- Query: Apple vitamins
- Relevant document: benefits of eating apples

Dialogue systems

Spelling correction

# One sense per . . .

One sense per discourse

- One meaning per word+document

One sense per collocation

- Nearby words help determine the sense

# Dictionary

- Dictionary/Thesaurus: $\forall w, s \in S_w : D(s) =$ description of sense $s$
- Context: $\forall w, C(w) =$ context of word $w$ in a specific occurence

## Lesk's Algorithm

- Idea: Sense $s_i$ of ambiguous word $w$ is likely to be the correct sense if many of the words used in the dictionary definition of $s_i$ are also used in the definitions of words in the ambiguous word's context.

$$s_{opt} = \underset{s_k}{\operatorname{argmax}}\, sim\left(D(s_k), \bigcup_{v_j \in C} E(v_j)\right)$$

## Similarity

$$\frac{2|X \cap Y|}{|X| + |Y|} \qquad \frac{2|X \cap Y|}{|X \cup Y|} \qquad \frac{|X \cap Y|}{\sqrt{|X| \cdot |Y|}}$$

- Advantages? Disadvantages?

# Supervised Disambiguation

- Sequence Labelling / Classification

## Bayes Decision

$$\hat{s} = \arg\max_s p(s|C) = \arg\max_s \frac{p(C|s) \cdot (p(s))}{p(C)}$$
$$= \arg\max_s p(C|s) \cdot (p(s))$$

## Naïve Bayes

$$p(C|s) = \prod_{x \in C} p(x|s)$$

- Estimate by MLE counts (+ smoothing)
- Independence within context
- Position in context does not matter
- Advantages? Disadvantages?

# Unsupervised Disambiguation

- Machine translation is able to choose the right sense (assuming different senses have different translations)
- MT is trained on unsupervised data
- Apple was struggling last quarter.
  Apple hatte im letzten Quartal Probleme.
- Apple is full of vitamins.
  Apfel ist voller Vitamine.
- Translations (in German): {Apfel, Äpfel, Apple}
- Indicator words: {struggling, quarter, full, vitamins} (stopwords removed)

Partition translated words ($\{Q_1, Q_2\}$) and indicator words ($\{P_1, P_2\}$) to maximize:

$$I(P; Q) = \sum_{i \in Q, t \in P} \log \frac{p(i,t)}{p(i) \cdot p(t)}$$

# Flip-Flop Algorithm

1. find random partition P = $\{P_1, P_2\}$ of $t_1, ..., t_m$
2. while improving I(P;Q) do
3. - find partition Q = $\{Q_1, Q_2\}$ of $x_1, ..., x_n$ that maximises I(P;Q)
4. - find partition P = $\{P_1, P_2\}$ of $t_1, ..., t_m$ that maximises I(P;Q)
5. end

- $t_i$ : translations of the ambiguous word
- $x_i$ : indicator words
- I(P;Q) monotonically increases until convergence

- Disambiguation
  Determine $x_i$

if $x_i \in Q_1$ assign sense 1

if $x_i \in Q_2$ assign sense 2

# EM Algorithm

- Idea: Random initialisation followed by parameter estimation
- Paramaters? $P(v_j|s_k)$ and $P(s_k)$
- Maximise log-likelihood $\log \prod_i \sum_k P(c_i|s_k)P(s_k)$
- E step: $h_{ik} = \frac{P(c_i|s_k)P(s_k)}{\sum_l P(c_i|s_l)P(s_l)}$
- M step: $P(v_j|s_k) = \frac{\sum_i C(v_j \in c_i) \cdot h_{ik}}{\sum_j \sum_i C(v_j \in c_i) \cdot h_{ik}}$

  $P(s_k) = \frac{\sum_i h_{ik}}{\sum_k \sum_i h_{ik}}$
- Disambiguation: $s_{opt} = argmax_{s_k}[\log P(s_k) + \sum_{v_j \in C} \log P(v_j|S_k)]$

# Yarowsky Algorithm

TODO

# Resources

1. UdS SNLP Class, WSD: https://teaching.lsv.uni-saarland.de/snlp/
2. Classical Statistical WSD: https://www.aclweb.org/anthology/P91-1034.pdf
3. WSD: https://www.cs.toronto.edu/~frank/csc2501/Lectures/8%20Word%20sense%20disambiguation.pdf
4. Lesk Algorithm: https://www.c-sharpcorner.com/article/lesk-algorithm-in-python-to-remove-word-ambiguity/