

* * * PGM: Final Project Presentation * * *

POS-Tagging and NER

C. Biehl, F. Otto, D. Wehner

TU Darmstadt

February 6, 2019



Agenda

① Task Description

② Process and Tools

③ Results and Comparison

Task Description

- The task was to use probabilistic graphical models for:
 - POS-Tagging (Part-of-Speech Tagging)
 - Named Entity Recognition (NER)
- We used the following models:
 - Naïve Bayes (baseline model)
 - HMMs (Hidden Markov models)
 - CRFs (Conditional Random Fields)

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Model Overview

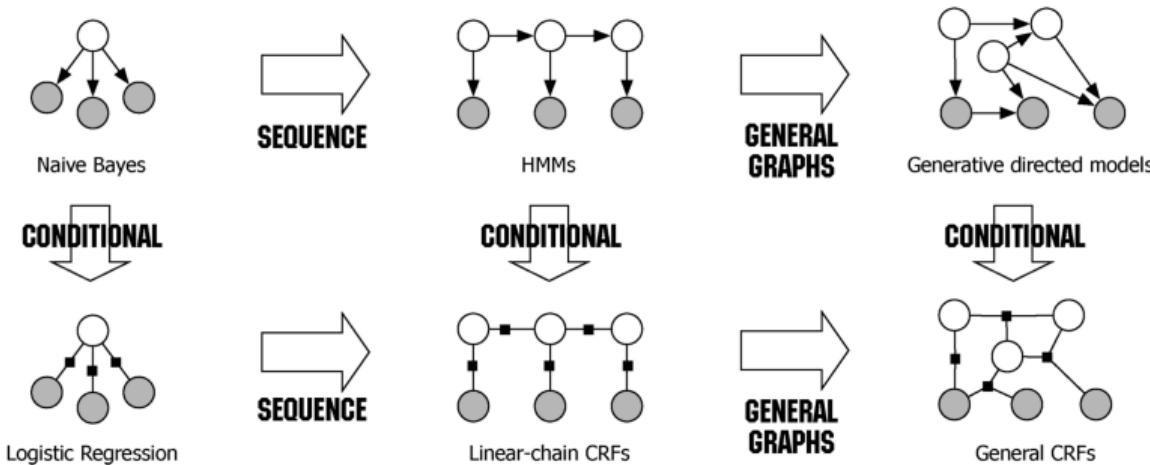


Figure: Model overview

Part of Speech Tagging



Figure: Part of speech tagging

Named Entity Recognition

"There was nothing about this storm that was as expected," said **Jeff Masters**, a meteorologist and founder of **Weather Underground**. "**Irma** could have been so much worse. If it had traveled 20 miles north of the coast of **Cuba**, you'd have been looking at a (Category) 5 instead of a (Category) 3."

Person

Organization

Location

Figure: Named entity recognition

Process and Tools

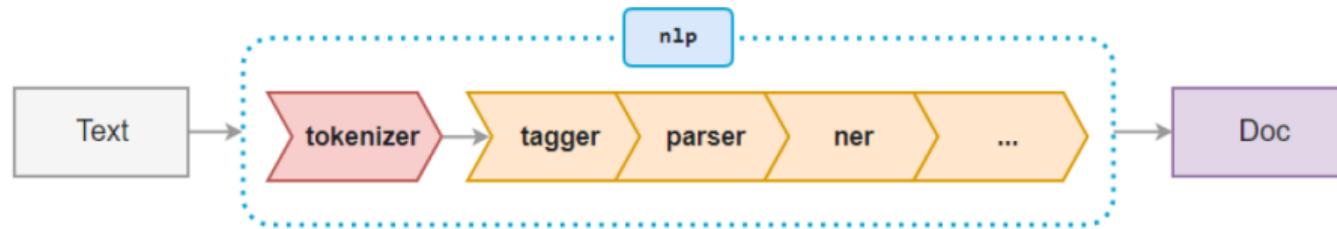


Figure: NLP-Pipeline

- POS tagging is one of the earliest stages in the NLP pipeline and serves as input for most downstream tasks
- E.g. NER: POS tags can be used as a feature

Data Set

- We used the **Groningen Meaning Bank**
- <http://gmb.let.rug.nl/data.php>
- Corpus of public domain texts:
 - 1.4 milion words
 - automatically produced, manually corrected (silver labels)



Tag Distribution

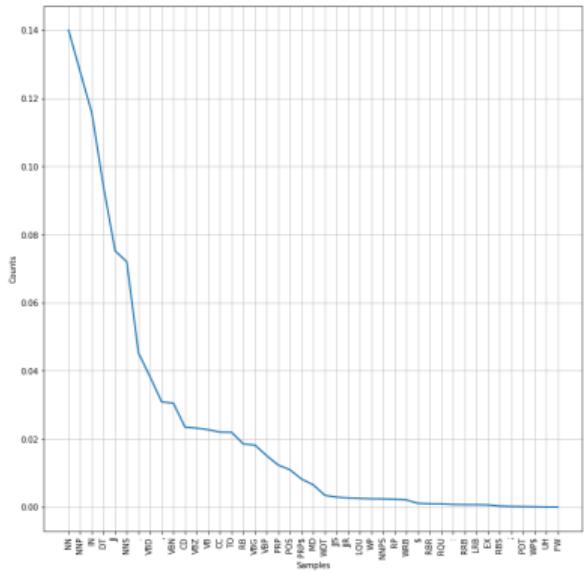


Figure: POS tag distribution

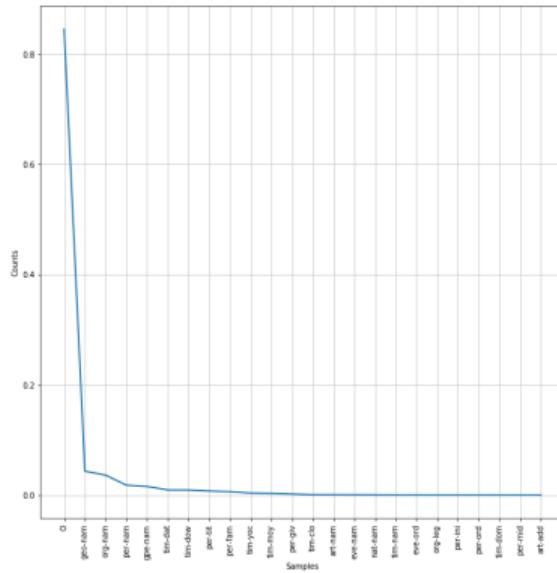


Figure: NER tag distribution

Results

POS Tagging

| Naïve Bayes [POS] | | |
|-------------------|----------|---------------|
| Features | F1 score | Acc (w / s) |
| w, l, s | 0.927 | 0.927 / 0.240 |
| w, l, s, uwf | 0.928 | 0.928 / 0.241 |
| w, l, s, bf | 0.957 | 0.957 / 0.467 |
| w, l, s, uwf, bf | 0.949 | 0.949 / 0.369 |

| HMM [POS] | | |
|---------------|----------|---------------|
| Features | F1 score | Acc (w / s) |
| word sequence | 0.837 | 0.837 / 0.496 |

Legend:

w: word, l: lowercase word, s: stem, uwf: unknown word features, bf: bigram features

Bi-LSTM Baseline: 0.950 [Bjerva et al., 2016]

Results (Ctd.)

POS Tagging

| CRF [POS] | | |
|------------------|----------|------------------|
| Features | F1 score | Accuracy (w / s) |
| w, l, s | 0.974 | 0.974 / 0.607 |
| w, l, s, uwf | 0.980 | 0.980 / 0.749 |
| w, l, s, bf | 0.978 | 0.978 / 0.669 |
| w, l, s, uwf, bf | 0.985 | 0.985 / 0.740 |

Legend:

w: word, l: lowercase word, s: stem, uwf: unknown word features, bf: bigram features

POS: Confusion Matrix CRF (all features)

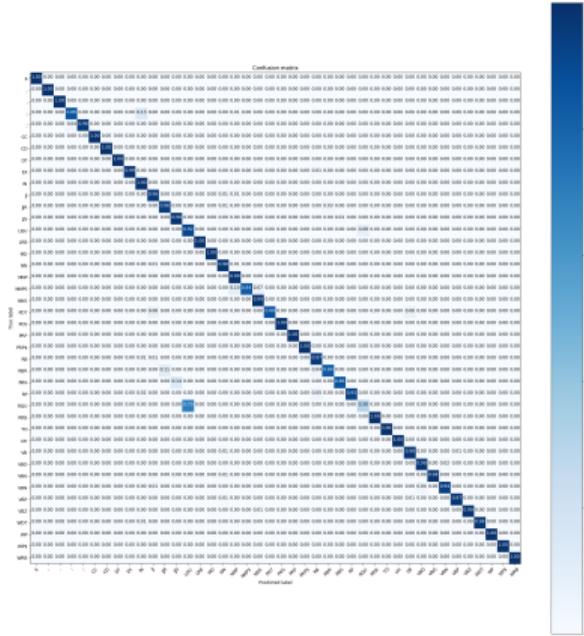


Figure: Named entity recognition

Results (Ctd.)

Named Entity Recognition

| Naïve Bayes [NER] | | |
|-----------------------|----------|---------------|
| Features | F1 score | Acc (w / s) |
| w, l, s | 0.921 | 0.921 / 0.262 |
| w, l, s, uwf | 0.922 | 0.922 / 0.237 |
| w, l, s, uwf, pos | 0.921 | 0.921 / 0.240 |
| w, l, s, uwf, pos, bf | 0.929 | 0.929 / 0.288 |

| HMM [NER] | | |
|---------------|----------|---------------|
| Features | F1 score | Acc (w / s) |
| word sequence | 0.946 | 0.946 / 0.523 |

Legend:

w: word, l: lowercase word, s: stem, uwf: unknown word features, bf: bigram features, pos: part of speech tags

LSTM Baseline: 0.931 [Akbik et al., 2018], different data set: CoNLL03

Results (Ctd.)

Named Entity Recognition

| CRF [NER] | | |
|-----------------------|----------|------------------|
| Features | F1 score | Accuracy (w / s) |
| w, l, s | 0.965 | 0.965 / 0.604 |
| w, l, s, uwf | 0.969 | 0.969 / 0.639 |
| w, l, s, uwf, pos | 0.969 | 0.969 / 0.645 |
| w, l, s, uwf, pos, bf | 0.974 | 0.974 / 0.693 |

Legend:

w: word, l: lowercase word, s: stem, uwf: unknown word features, bf: bigram features, pos: part of speech tags

NER: Confusion Matrix CRF (all features)

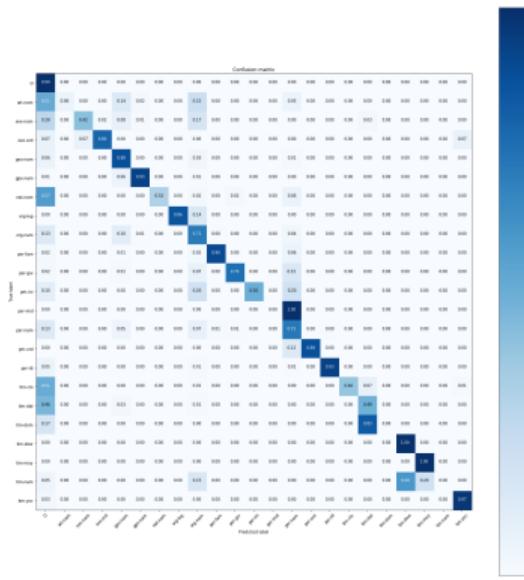
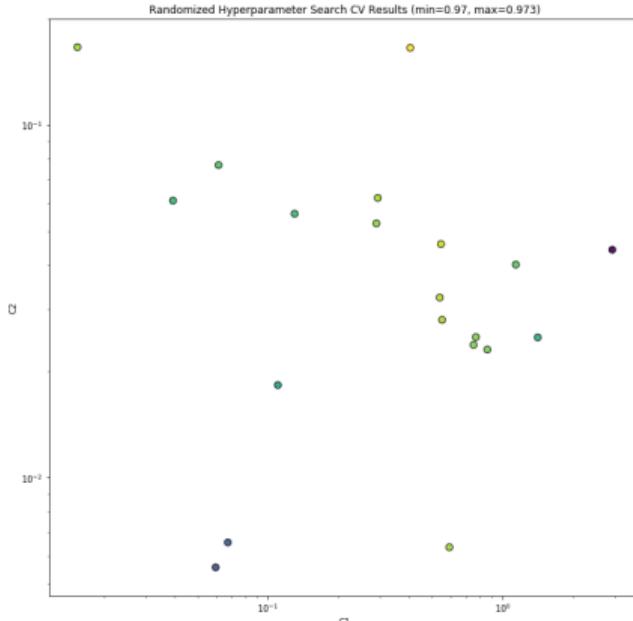


Figure: Named entity recognition

NER: Random Search for CRF



- C1: L_1 -regularization
- C2: L_2 -regularization
- Best parameters found:
 - c1: 0.40431785916906465
 - c2: 0.16525290148249140

Figure: Random Search CRF

Result Comparison

- Naive Bayes confuses per-ord with eve-ord, but HMM does not
⇒ context knowledge!
- Naive Bayes predicts per-tit very often
- Naive Bayes confuses tim-clo and tim-dat often
- HMM incorrectly predicts DET tag (POS) and O tag (NER) very often
- **POS tagging**
 - Likely transitions: MD ⇒ VB; NNS ⇒ VBP
 - Unlikely transitions: DT ⇒ DT; VBD ⇒ VBD
- **NER**
 - Likely transitions: per-giv ⇒ per-fam
 - Unlikely transitions: per-giv ⇒ tim-dat

What we have learned...

- Feature engineering is 'dirty work' but **crucial** for later model performance
- POS tagging is **easier** than NER
- Using less training data **hurts performance**, especially on sentence level (e.g. POS: foreign words misclassified, mistakes interjection with adjective)
- CRF mostly works '**out-of-the-box**'; we conducted a random search and noticed only tiny differences (≤ 0.03)
- We tried word embeddings (GloVe) for POS tagging with limited success
- **Little data:** On approx. 1,000 (= 2 %) training sentences, we got a word accuracy of 0.952 (frequency of majority tag NN ≈ 0.14)
- Sentence level accuracy volatile (≈ 0.50)

Thank you very much for the attention!

Topic: * * * PGM: Final Project Presentation * * * POS-Tagging and NER
Date: February 6, 2019

Contact:

C. Biehl, F. Otto, D. Wehner
TU Darmstadt

Do you have any questions?