

Annotated datasets for NER

TOPIC: Training data for Named Entity Recognition

- Give a brief overview of available annotated datasets for NER
 - I.e. the data we need to train models with full supervision
- Do you think this is enough data to train good supervised models?
 - Give us some results that support your answer
 - What about using unsupervised learning?

Sources and possible papers:

- Nadeau and Sekin, *A survey of named entity recognition and classification*, *Linguisticae Investigationes* 30, 2007, pp. 3–26.

Annotated data for Medical NER

TOPIC: Named Entities in the CLEF-eHEALTH challenge

- Give an overview of the CLEF-eHEALTH challenge
- Talk about NER in this challenge (Task 1)
- Present the training data provided for medical NER
 - Which set of classes are annotated?
 - How can you use this data to train a classifier (e.g. a linear model)?

Sources and possible papers:

- <https://sites.google.com/site/clefehealth2016/>
- <https://sites.google.com/site/clefehealth2015/>

Supervised NER

TOPIC: Linear models for Named Entity Recognition

- Get a training set for a NER task (e.g. CLEF e-health)
- Model the problem as a multi-class classification task
- Consider the following methods:
 - (non-sequential) Linear models
 - Linear-chain conditional random fields
- Which one do you think will work better? and why?

Sources and possible papers:

- <https://sites.google.com/site/clefehealth2016/>
- Nadeau and Sekin, *A survey of named entity recognition and classification*, *Linguisticae Investigationes* 30, 2007, pp. 3–26.

Supervised NER

TOPIC: Neural Networks for Named Entity Recognition

- What are the advantages of neural networks over linear models?
 - What do the non-linear „activations“ do?
- Present a neural network for the NER task
- Should we use neural networks instead of linear models for NER
 - Give us some results that support your answer

Sources and possible papers:

- Collobert et al., *Natural Language Processing (Almost) from Scratch*, Journal of Machine Learning Research, 2011, pp. 2493–2537.

Supervised NER

TOPIC: Weakly Supervised Named Entity Recognition

- Starting from a few examples ("seed examples"), how do you automatically build a named entity classifier?
 - This is sometimes referred to as "bootstrapping"
- What are the problems with this approach?
 - How do you block the process from generalizing too much?
- Should we use weak supervision instead of (full) supervision for NER
 - Give us some results that support your answer

Sources and possible papers:

- Nadeau and Sekin, *A survey of named entity recognition and classification*, *Linguisticae Investigationes* 30, 2007, pp. 3–26.

NER Domain Adaptation

TOPIC: Domain adaptation and failure to adapt

- What is the problem of domain adaptation?
- How is it addressed in statistical classification approaches to NER?
- How well does it work

Sources and possible papers:

- Daume III, *Frustratingly Easy Domain Adaptation*, ACL, 2007.

Classification-based Citation Parsing

TOPIC: Parsing citations using classifiers

- How is the citation parsing problem formulated using classifiers ?
- What sort of information is available?
- What does the training data look like?
- What sorts of downstream applications are based on citation parsing ?

Sources and possible papers:

- Peng et al., *Information extraction from research papers using conditional random fields*, Information Processing & Management, 2006, pp. 963–979.

Question Answering

TOPIC: Information Extraction for Question Answering

- In 2011, IBM's *Watson* defeated two human champions in the US quiz show *Jeopardy*
- Give an overview of *Watson's* question answering engine *DeepQA*
- Highlight how information extraction techniques are used in a complex pipeline for this application

Sources and possible papers:

- Ferrucci et al., *An Overview of the DeepQA Project*, *AI Magazine*, 2010, pp. 59–79.

Reading Comprehension

TOPIC: Natural Language Comprehension with Neural Networks

- A machine reading system can answer queries about the content of natural language documents
- Which resources are required to build a system that is able to solve real-world tasks?
- How would we design and train a system based on Artificial Neural Networks?

Sources and possible papers:

- Hermann et al., *Teaching Machines to Read and Comprehend*, NIPS, 2015, pp. 1693–1701.

Event Detection

TOPIC: Event Detection in Social Media

- Activity in social media (e.g., *Twitter*) can be monitored and analyzed to spot events
- Use cases: natural disasters, epidemics, stock market, ...
- What are the challenges and which information extraction techniques can be employed?
- Give a high-level sketch of the overall pipeline

Sources and possible papers:

- Yin et al., *Using Social Media to Enhance Emergency Situation Awareness*, IEEE Intelligent Systems, November/December 2012, pp. 52–59.
- Sakaki et al., *Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors*, WWW, 2010, pp. 851–860.

Sentiment Analysis

TOPIC: Applications of Sentiment Analysis:
Political Opinion and Customer Suggestions

- Sentiment analysis and opinion mining: Capturing public opinion in forums, blogs, social networks, ...
- Automatic classification of sentiment
- Describe possible applications of sentiment analysis, e.g. for election prediction, product preferences, marketing, ...

Sources and possible papers:

- Wang et al., A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle, ACL System Demonstrations, 2012, pp. 115–120
- Negi and Buitelaar, *Towards the Extraction of Customer-to-Customer Suggestions from Reviews*, EMNLP, 2015, pp. 2159–2167.

IE and Computer Vision (ADVANCED!)

TOPIC: Cross-modal Information Extraction

- Detecting objects in the visual world (in images) and mapping them to words
- Possible applications: caption generation, event detection based on multi-modal input, image search, ...
- Are methods from natural language processing helpful?
- Distributional semantics, with a projection between an image-based semantic space and a word-based semantic space
- How to learn new concepts?

Sources and possible papers:

- Lazaridou et al., *Is this a wampimuk? Cross-modal mapping between distributional semantics and the visual world*, ACL, 2014, pp. 1403–1414.