Seminar Topics: Information Extraction

Matthias Huck, Alexander Fraser

LMU Munich

24 October 2018

Biomedical Named Entity Recognition using Deep Learning & Word Embeddings

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Overview:

- Information extraction systems can be a valuable tool for researchers and practitioners in biomedicine.
- Many state-of-the-art natural language processing systems employ artificial neural networks and word embeddings.
- What impact does this technology have on NER quality? Does it improve specialized NER in the biomedical domain?

Paper:

- **Deep learning with word embeddings improves biomedical named entity recognition**.
  Habibi et al.
  Bioinformatics, 33(14), 2017, pp. i37–i48.
  `https://doi.org/10.1093/bioinformatics/btx228`

Named Entity Recognition for an E-Commerce Use Case

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Overview:

- NER can have applications in e-commerce, e.g. for item categorization on a sales platform.
- eBay has built a statistical NER system for their marketplace.
- The baseline Conditional Random Field (CRF) model is augmented with distributed word representations.
- What are the challenges of NER in an e-commerce scenario? How are word vectors integrated into the statistical system, and do they help in the e-commerce domain?

Paper:

- **Distributed Word Representations Improve NER for e-Commerce**.
  Joshi et al.
  Proc. of the 1st Workshop on Vector Space Modeling for Natural Language Processing, 2015, pp. 160–167.
  `https://aclanthology.info/papers/W15-1522/w15-1522`

Rule-based Relation Extraction

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Overview:

- In older traditional IE systems, handcrafted patterns are used not only for NER, but also to extract relations between entities from unstructured text.
- How exactly is rule-based relation extraction implemented in an actual pattern-based IE system? How does it work in detail?

Paper:

- **REES: A Large-Scale Relation and Event Extraction System**.
  Aone and Ramos-Santacruz.
  Proc. of the Sixth Conference on Applied Natural Language Processing, 2000.
  `https://aclanthology.info/papers/A00-1011/a00-1011`

Optional additional reading:

- Three of the references given in the bibliography of the above paper:
  Aone et al. (1998), Appelt et al. (1995), Yangarber and Grishman (1998).

Relation Extraction using Dependency Parse Trees

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Overview:

- Rule-based relation extraction can benefit from additional annotation such as dependency parses of the text.
- How can manual patterns be specified over dependency parse trees? How well does such a relation extraction engine perform in a narrow (biomedical) domain?

Paper:

- **RelEx—Relation extraction using dependency parse trees**.
  Fundel et al.
  Bioinformatics 23(3), 2007, pp. 365–371.
  `https://doi.org/10.1093/bioinformatics/btl616`

Joint NER and Relation Extraction with Neural Networks

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Overview:

- Tools for NER and for relation extraction are typically built separately. Relation extraction relies on an NER component previously in the IE pipeline.
- To avoid error propagation and capture interactions between the subtasks, can both be done jointly rather than in two separate stages?

Paper:

- **A neural joint model for entity and relation extraction from biomedical text**. Li et al.
  BMC Bioinformatics 18:198, 2017.
  `https://doi.org/10.1186/s12859-017-1609-9`

(Advanced topic.) Recommended prior knowledge:

- Familiarity with artificial neural networks: LSTMs and CNNs.

Automatic Biomedical Knowledge Extraction

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Overview:

- How to automatically discover important facts by mining biomedical literature?
- Named entity extraction, relation extraction, and ranking of extracted insights in the biomedical domain.

Paper:

- **An Insight Extraction System on BioMedical Literature with Deep Neural Networks**.
  He et al.
  Proc. of EMNLP, 2017, pp. 2691–2701.
  `https://aclanthology.info/papers/D17-1285/d17-1285`

Relation Extraction and Scoring for Question Answering

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Overview:

- In 2011, IBM's *Watson* defeated two human champions in the US quiz show *Jeopardy*.
- Watson's *DeepQA* question answering framework includes relation extraction and passage-scoring components.
- What is IBM's approach to relation extraction? How does it utilize both handcrafted patterns and statistical classifiers?

Paper:

- **Relation extraction and scoring in DeepQA**. Wang et al. IBM Journal of Research and Development 56(3/4), 2012. `https://pdfs.semanticscholar.org/88b9/55871e14dabb5e8b132d1dd7c3ea58067eb6.pdf`

Optional additional reading:

- **Building Watson: An Overview of the DeepQA Project**. Ferrucci et al. AI Magazine 31(3), 2010, pp. 59–79. `https://doi.org/10.1609/aimag.v31i3.2303`

Hint: Paywalled Literature

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Access to publications behind a paywall can often be provided via the university library.

Try "**E-Medien-Login**", using your LMU user ID:
`http://www.ub.uni-muenchen.de/ausleihe-online/digitaler-zugriff/e-medien-login/index.html`

Alternatively, search the web for preprint versions.

Questions?

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Thank you for your attention

Matthias Huck

mhuck@cis.lmu.de