Retrieving Multi-Entity Associations: An Evaluation of Combination Modes for Word Embeddings

Gloria Feher, Andreas Spitz and Michael Gertz

Database Systems Research Group, Heidelberg University, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

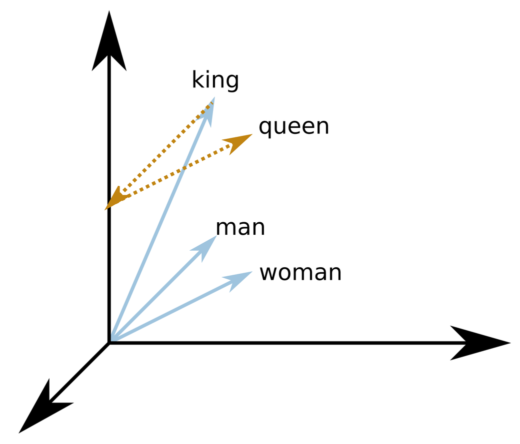
Problem Statement

While word embeddings have been shown to improve the results of many NLP and IR applications, little effort has been devoted to using embeddings for the **retrieval of associations between multiple entities**. We use several embedding methods to generate word representations from entity-annotated news data and evaluate them against a word cooccurrence network for the task of predicting entity participation in events to answer the questions:

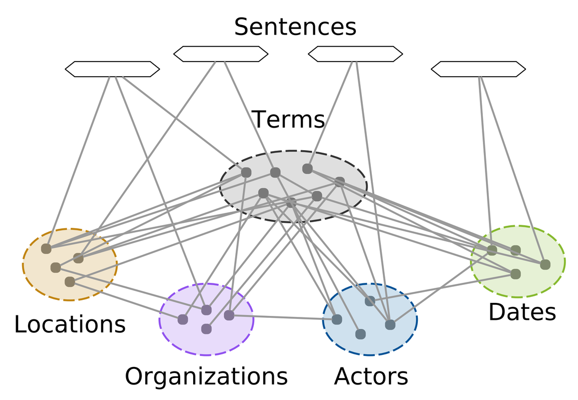
- **How should multiple entity embedding vectors be combined?**
- **How important is the frequency of entities for the performance?**

Entity Relation Models

- Embeddings:
  - word2vec: skip-gram, CBOW [1]
  - GloVe [2]



- Implicit Networks [3]



Task: Predicting Entity Participation

- Event: Set of k participating entities.
- Query generation:
  - Input: $k-1$ entities
  - Output: 1 entity
  - Example: $\{e_1, e_2, \textbf{?}\} \quad \rightarrow \quad \{t\}$

News Article Data

- 127.5k news articles (June – Nov 2016)
- Entities replaced by Wikidata IDs
- Ground truth: Wikipedia Current Events descriptions of events from data set

Evaluation: Combining Embedding Vectors

$$t_{\text{MINMAX}} = \underset{x \in X}{\text{argmin}} \;\; \underset{q \in Q}{\text{argmax}} \;\; cosdist(\theta_q, \theta_x)$$

$$t_{\text{SUM}} = \underset{x \in X}{\text{argmin}} \sum_{q \in Q} cosdist(\theta_q, \theta_x)$$

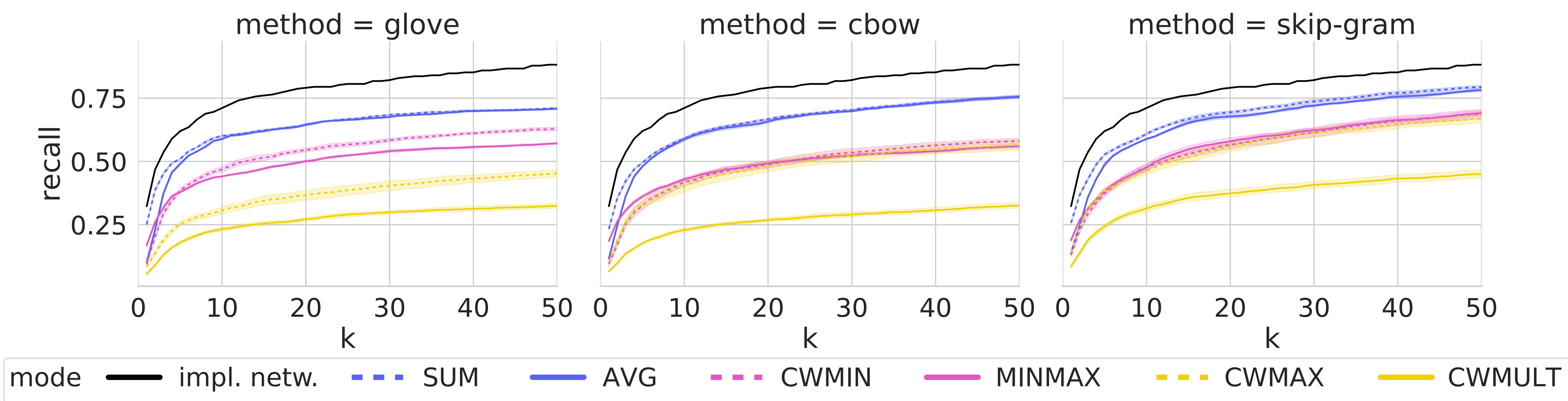
$$t_{\text{AVG}} = \underset{x \in X}{\text{argmin}} \;\; cosdist\left(\frac{1}{|Q|}\sum_{q \in Q} \theta_q, \theta_x\right)$$

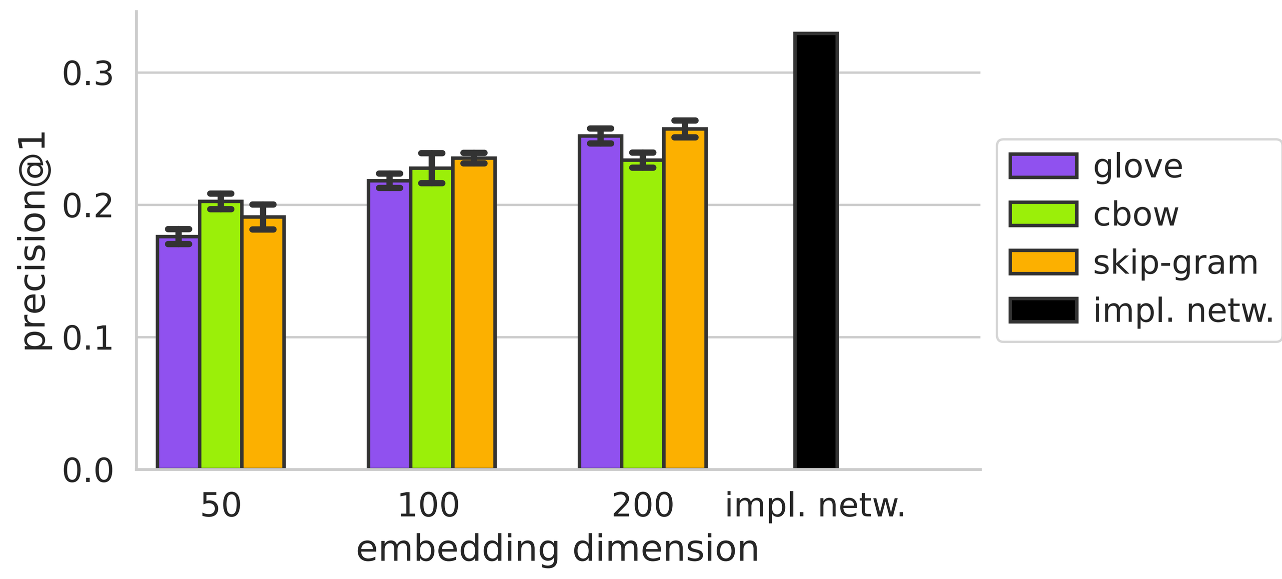
$$t_{\text{CWMIN}} = \underset{x \in X}{\text{argmin}} \;\; cosdist\left([\min(\Theta_1^Q), \cdots, \min(\Theta_{|Q|}^Q)]^T, \theta_x\right)$$

$$t_{\text{CWMULT}} = \underset{x \in X}{\text{argmin}} \;\; cosdist\left(\theta_{q_1} \odot \cdots \odot \theta_{q_{|Q|}}, \;\; \theta_x\right), q_i \in Q$$

Precision@1:

	skip-gram	CBOW	GloVe
MINMAX	0.189	0.186	0.168
SUM	**0.257**	**0.234**	**0.252**
AVG	0.140	0.116	0.101
CWMIN	0.130	0.095	0.095
CWMAX	0.140	0.102	0.085
CWMULT	0.085	0.066	0.056

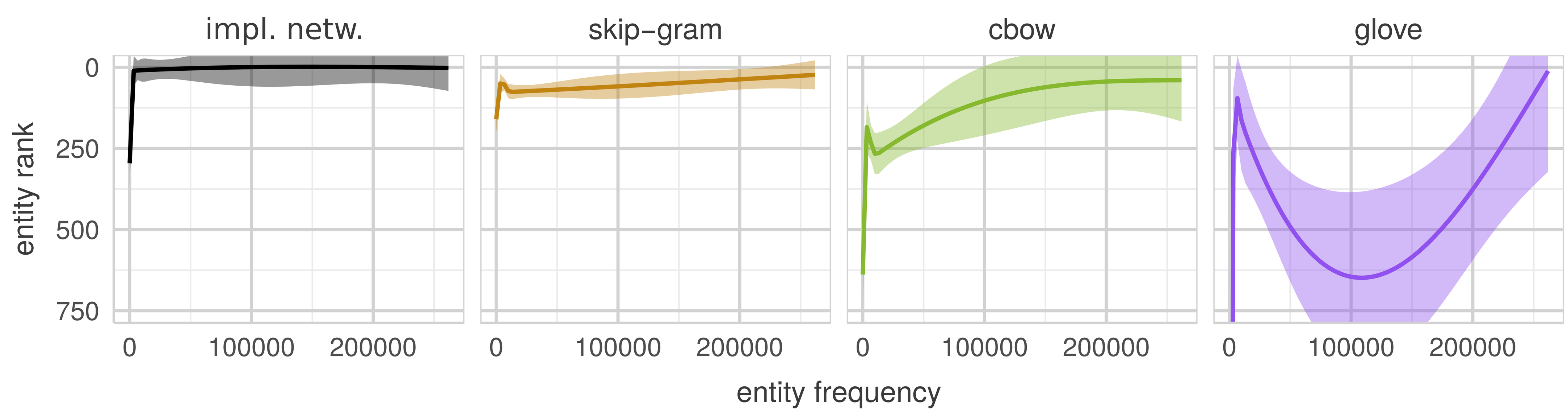




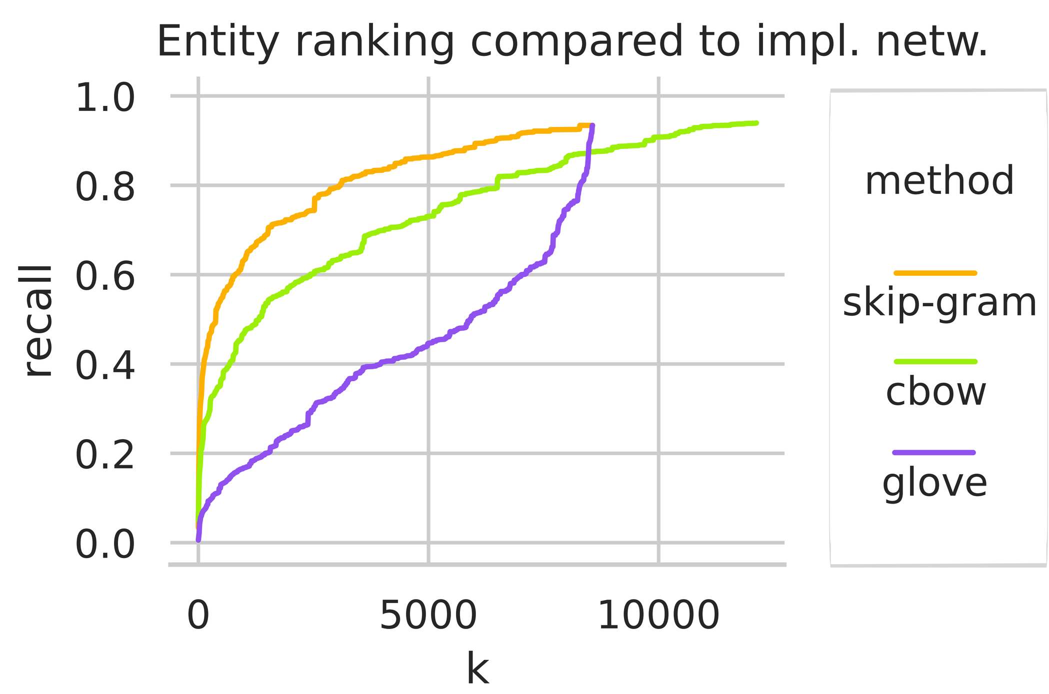
Conclusions:

- SUM performs best.
- Higher embedding dimensions lead to better performance.
- An extended window size of 21 words yields the best results for any method.

Evaluation: The Importance of Entity Frequencies



Open Research Questions



- GloVe models different relations: How can GloVe be used to detect new entity associations?
- How to create ensemble methods?
- Are context-sensitive embeddings, such as ELMo / BERT, a suitable replacement for implicit networks?

References

[1] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. **Distributed Representations of Words and Phrases and their Compositionality**. 2013, *NeurIPS'13*

[2] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. **Glove: Global Vectors for Word Representation.** 2014, *EMNLP'14*

[3] Andreas Spitz and Michael Gertz. **Terms over LOAD: Leveraging Named Entities for Cross-Document Extraction and Summarization of Events**. 2016, *SIGIR'16*

Contact Information

Gloria Feher

gloria.e.feher@gmail.com

`http://dbs.ifi.uni-heidelberg.de/`

This work was presented at the 42nd International Conference on Research and Development in Information Retrieval (SIGIR'19), July 21-25, 2019, Paris, France.