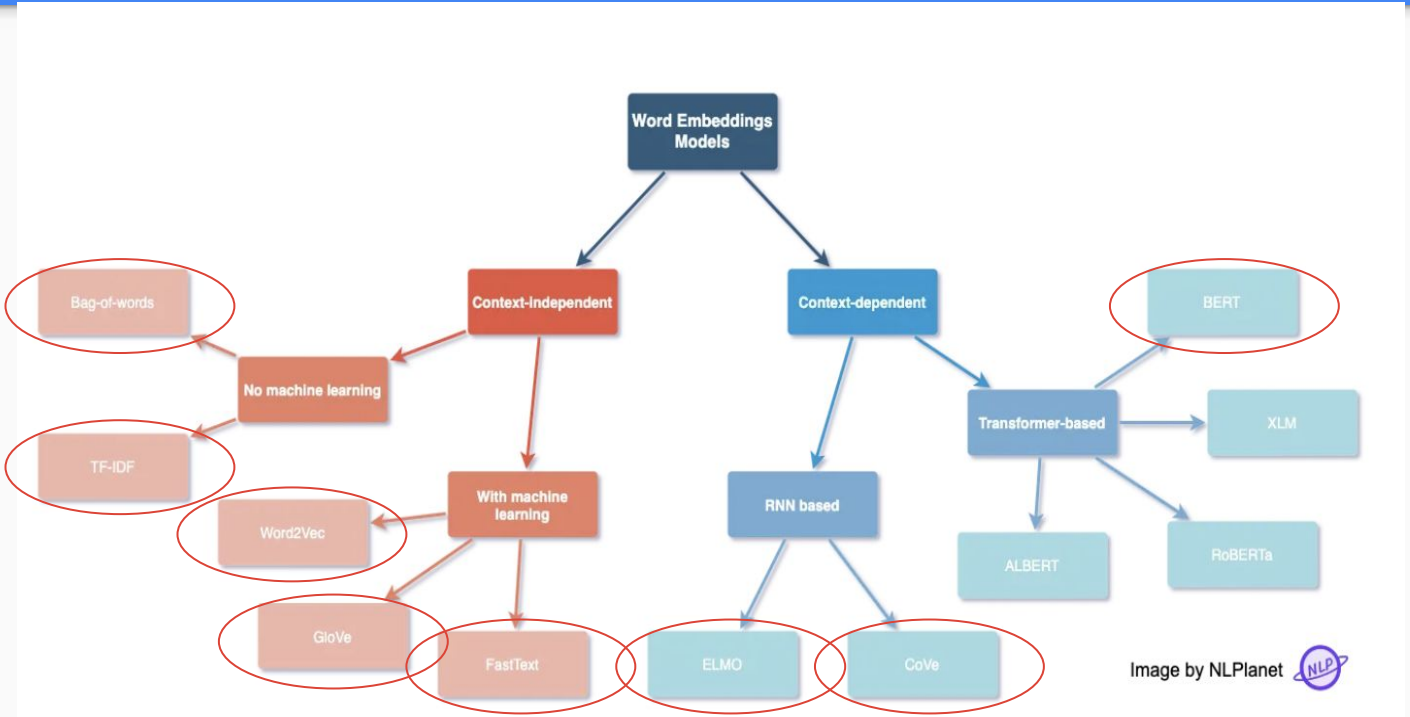
Dow Meeting 04/21

Andrew Simon, Meenal Rawlani, Shijie Zhang

Contents

- Finished Gantt chart
- Updated model list
- Used bag of words and tf-idf on imdb dataset
- Extrinsic vs. Intrinsic methods
- Embedding & meta embedding
- Updated Roadmap for future meetings

Updated model list



Updated model list

Model	Bag-of-words	TF-IDF	Word2Vec	GloVe	FastText	ELMO	CoVe	BERT
Type	No machine learning	No machine learning	Machine learning	Machine learning	Machine learning	RNN based	RNN based	Transformer based
Context dependent	No	No	No	No	No	Yes	Yes	Yes

Extrinsic vs. Intrinsic Methods

- Extrinsic methods evaluate performance on downstream tasks, while intrinsic methods attempt to determine performance from just the embeddings alone.

Extrinsic vs. Intrinsic Methods

Extrinsic Methods:

- Accuracy Score
- F1 Score
- Recall
- AUC & ROC curves

Intrinsic Methods (taken from Wang et. al):

- Similarity Tasks
- Outlier Detection
- Analogy Tasks
- QVEC

Ensemble Embedding vs. Meta Embedding

- Similar to ensemble embedding, however attaches weights to each embedding method
  - Experimentally found to have greater accuracy than ensemble embedding.

$$\mathbf{w}_j^{DME} = \sum_{i=1}^{n} \alpha_{i,j} \mathbf{w'}_{i,j}$$

Ensemble Embedding vs. Meta Embedding

- Ensemble embedding uses the concatenated ensemble of word embeddings as a single hyper parameter in the downstream task
  - Experimentally shown to increase accuracy over any one embedding method
  - Computationally expensive since it requires all embeddings methods in the ensemble to be trained on the same dataset

$$\mathbf{w}_j^{CAT} = [\mathbf{w}_{1,j}, \mathbf{w}_{2,j}, \ldots, \mathbf{w}_{n,j}].$$

Plans for the future

- Complete model training
- Start to use extrinsic and intrinsic methods to evaluate model performance
- Start to incorporate meta embedding

References

- Task-Optimized Word Embeddings for Text Classification
- ETNLP: a visual-aided systematic approach to select pre-trained embeddings for a downstream task
- Dynamic Meta-Embeddings for Improved Sentence Representations
- Evaluating Word Embedding Models: Methods and Experimental Results
- 11 word-embedding you need to know