Dynamic Meta-Embeddings for Improved Sentence Representations

by Facebook

Receipt to win DA/ML competition

1. Take number of models.
2. Ensemble.
3. Profit!!!

Dynamic meta-embeddings

1. Take number of models.
2. Ensemble.
3. Create neural-network to rule for specific task.
4. Profit!!!

Pros

- Coverage - one of the main problems with NLP systems is dealing with out-of vocabulary words: our method increases lexical coverage by allowing systems to take the union over different embeddings.
- Multi-domain - standard word embeddings are often trained on a single domain, such as Wikipedia or newswire. With our method, embeddings from different domains can be combined, optionally while taking into account contextual information.
- Multi-modality - multi-modal information has proven useful in many tasks, yet the question of multi-modal fusion remains an open problem. Our method offers a straightforward solution for combining information from different modalities.

Pros (part 2)

- Evaluation - while it is often unclear how to evaluate word embedding performance, our method allows for inspecting the weights that networks assign to different embeddings, providing a direct, task-specific, evaluation method for word embeddings.
- Interpretability and Linguistic Analysis - different word embeddings work well on different tasks. This is well-known in the field, but knowing why this happens is less well-understood. Our method sheds light on which embeddings are preferred in which linguistic contexts, for different tasks, and allows us to speculate as to why that is the case.

Contras

1. Learning performance: relative accuracy improvement for 2-3% asks for 3-4 times more time on learning process.
2. Inference performance: from 3-4 times more time for inference. Most of NLP applications are time-sensitive.
3. Implementation sensitivity.

Sources:

1. Dynamic Meta-Embeddings for Improved Sentence Representations
2. Enriching Word Vectors with Subword Information
3. From Frequency to Meaning: Vector Space Models of Semantics
4. Long Short-Term Memory-Networks for Machine Reading
5. About that BiLSTM approach for supervised learning of sentence repres.