Combining Contextual Words and Knowledge Graph Embeddings

Software project, work update 3

IDMC, University of Lorraine

27/10/2020

OUTLINES

1. Dimensionality Reduction Using PCA.

2. Autoencoded Meta-Embedding

Dimensionality Reduction Using PCA

Relation Prediction

Original Models (Dimension)	MRR	MAP@k=1
Contextual Embeddings (2048)	0.554	0.554
KG Embeddings (800)	0.817	0.663
Concatenation (2848)	0.738	0.533

Reduced Dimension Models (Dimension = 400)	MRR	MAP@k=1
Contextual Embeddings	0.656	0.433
KG Embeddings	0.750	0.557
Concatenation	0.708	0.481

Entity Typing

Original Models (Dimension)	MAP@k=10	Precision @ k (mean ± sd)
Contextual Embeddings (1024)	0.631	0.449 ± 0.271
KG Embeddings (400)	0.825	0.528 ± 0.269
Concatenation (1424)	0.828	0.527 ± 0.268

Reduced Dimension Models (Dimension = 200)	MAP@k=10	Precision @ k (mean ± sd)
Contextual Embeddings	0.190	0.208 ± 0.249
KG Embeddings	0.673	0.476 ± 0.261
Concatenation	0.297	0.306 ± 0.277

Relation Prediction with Concatenated Model

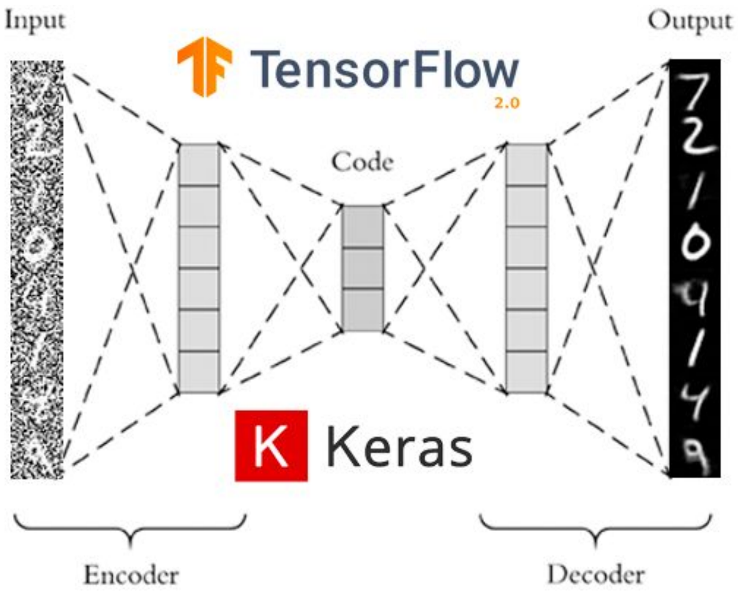
Dimension	MRR	MAP@k=1
400	0.708	0.481
800	0.716	0.493
1200	0.710	0.476
2848	0.738	0.533

Auto-Encoder Meta-Embedding:
Potential solution for the concatenating model

What is autoencoder?

★ Model learning from unlabelled data, but used for supervised learning tasks.

★ Via a non-linear transformation, the intermediate representation used by the autoencoder captures the essential information about the (potentially noisy) input so it can be accurately reconstructed.



Autoencoded Meta-embedding

- ❖ **Meta embedding learning** is a process of producing a single (meta) word embedding from a given set of pre-trained input (source) word embeddings.
- ❖ Autoencoder setting is closely related to the meta-embedding learning where we must reconstruct the information contained in individual source embeddings using a single meta-embedding.
- ❖ Typical autoencoder learning have just a single input, whereas meta-embedding learning reconstruct multiple source embeddings.
- ❖ Combined the two method in **Autoencoded Meta-embedding (AEME):** proved to have good performance compared to other method

Three approaches to autoencoded meta-embedding

❖ Three AEMEs methods :

1. Decoupled Autoencoded Meta-Embedding (DAEME)

2. Concatenated Auto-encoded Meta-Embedding (CAEME)

3. Averaged Auto-encoded Meta-Embedding (AAEME)

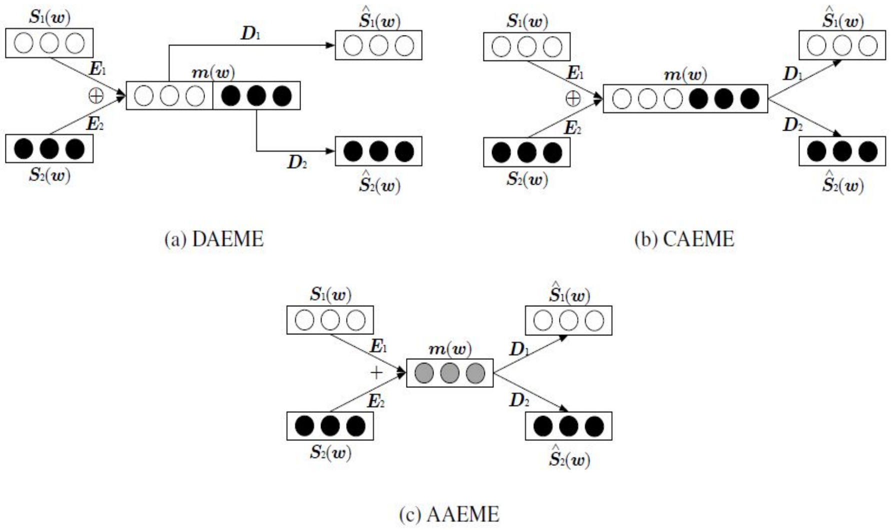


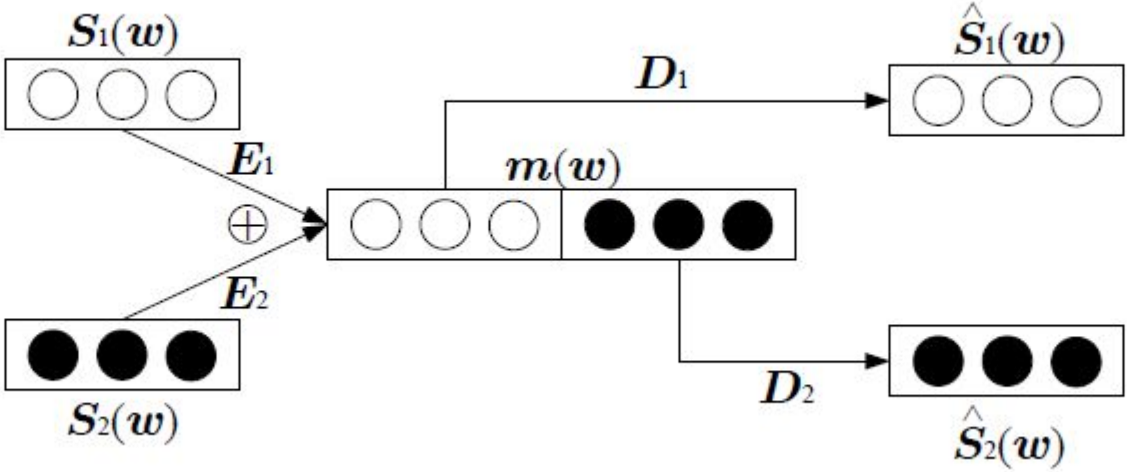
Figure 1: The architectures of proposed AEMEs. Rectangles represent word vectors, circles represent a single dimension of a word vector, and filled circles represent source of word embeddings. In (c), greyed circles in $m(w)$ indicates mixing of vectors from the two source embeddings.

Decoupled Autoencoder meta-Embedding (DAEME):

$$\boldsymbol{m}(w) = E_1(\boldsymbol{s}_1(w)) \oplus E_2(\boldsymbol{s}_2(w))$$

- Include concatenation in the mechanism, good for comparison with previous results.
- Source Embeddings : Contextual word embeddings(S1), and KG word embeddings(S2).
- Two encoders and two decoders will be used E1,E2,D1,D2.

Decoupled Autoencoder meta-Embedding (DAEME):



Shortcomings

- Source embeddings differ significantly, non-obvious to reconcile.
- Trained on different corpora or knowledge bases, results in different type of embeddings.
- Resources used to train source embeddings may not published, therefore, difficult to train all source embeddings with similar data.

References :

1. Bollegala, Danushka, and Cong Bao. "Learning word meta-embeddings by autoencoding." Proceedings of the 27th International Conference on Computational Linguistics. 2018.
2. James O'Neill and Danushka Bollegala: Semi-Supervised Multi-Task Word Embeddings arXiv, 2018. [arXiv]

Thank you! Questions?