Grasping the Finer Point: A Supervised Similarity Network for Metaphor Detection

Marek Rei, Luana Bulat, Douwe Kiela, Ekaterina Shutova

UNIVERSITY OF CAMBRIDGE

Metaphors in Language

Metaphors arise when one concept is viewed in terms of the properties of the other. (Shutova & Teufel, 2010)

1. How can I *kill* a process?
2. "My heart with pleasure *fills*, and *dances* with the daffodils."
3. He *attacked* every weak point in my argument.
4. The ambassador will start *mending* foreign policy.

They can help explain a concept better, make the reader visualise a concept, and add richness and colour to communication.

Metaphor Detection

Helps NLP systems understand the real meaning behind the words.

Allows generation systems to generate rich and nuanced natural text.

- Semantic roles (Gedigian et al., 2006)
- Concreteness (Turney et al., 2011)
- Imageability (Strzalkowski et al., 2013)
- WordNet supersenses (Tsvetkov et al., 2014)
- Sparse distributional features (Shutova et al., 2010)
- Dense neural word embeddings (Bracewell et al., 2014)
- Visual vectors (Shutova et al., 2016)
- Attribute-based vectors (Bulat et al., 2017)

We propose a neural architecture that is specially designed for metaphor detection.

Cosine Similarity

Shutova et al. (2016) showed that the cosine similarity between neural embeddings of the two words in a phrase is indicative of its metaphoricity.

$$s = cos(x_c, x_p)$$

For example

$$cos(colourful, personality)$$

would have lower similarity than

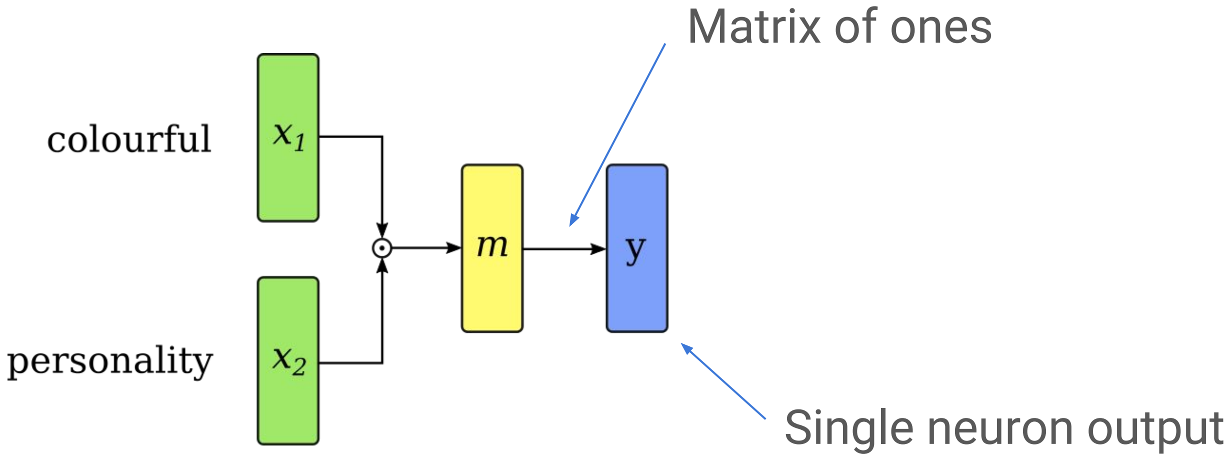
$$cos(nice, personality)$$

Cosine Similarity

If the input vectors x1 and x2 are normalised to unit length, the cosine similarity between them is equal to their dot product:

$$cos(x_1, x_2) \propto \sum_i x_{1,i} x_{2,i}$$

We can formulate that as a small neural network:

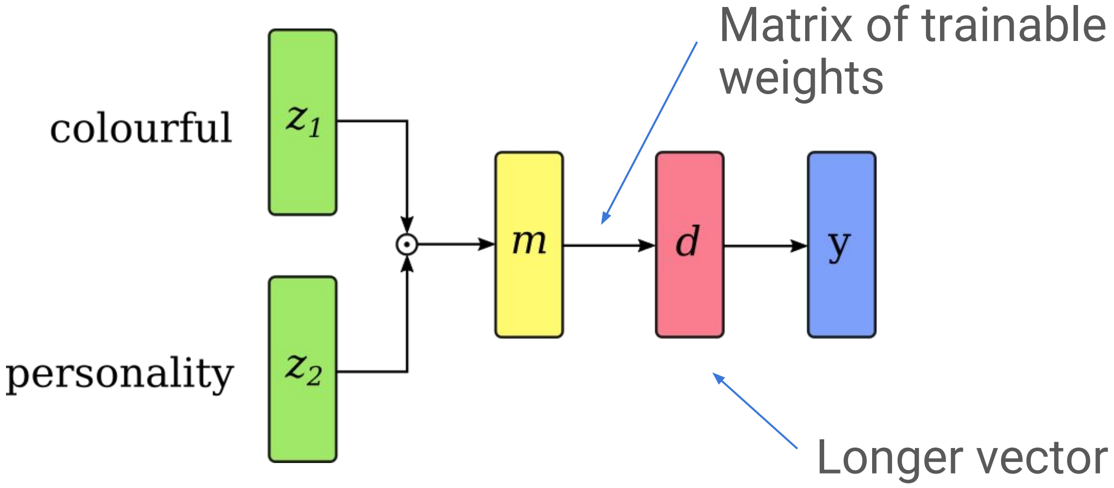


Weighted Cosine

We can instead create a version where vector m is passed through another layer, with weights that are optimised during training.

$$m_i = z_{1,i} z_{2,i}$$

$$d = \gamma(W_d m)$$



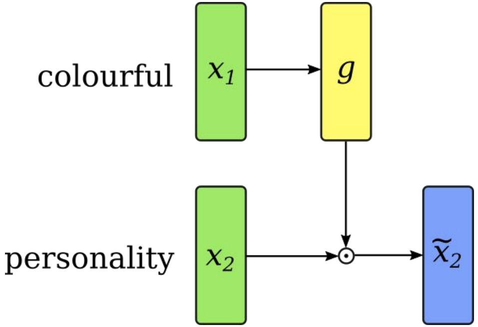
Word Representation Gating

Example: “healthy balance”

The source domain properties of the adjective "healthy” are projected onto the target domain noun “balance”, resulting in the interaction of the two domains in the interpretation of the metaphor.

$$g = \sigma(W_g x_1)$$

$$\widetilde{x}_2 = x_2 \odot g$$



Gating the noun vector, based on the adjective vector.

Vector Space Mapping

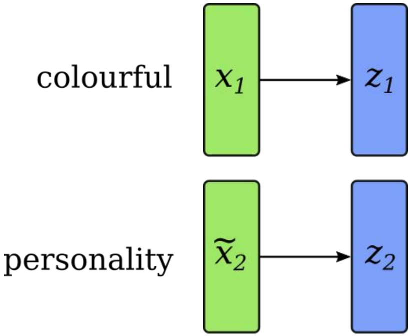
The original method uses basic pre-trained skip-gram vectors.

Let's add a transformation that maps them into a space that is specific for metaphor detection.

Importantly, we will use separate mappings for adjectives and nouns.

$$z_1 = tanh(W_{z_1} x_1)$$

$$z_2 = tanh(W_{z_2} \widetilde{x}_2)$$

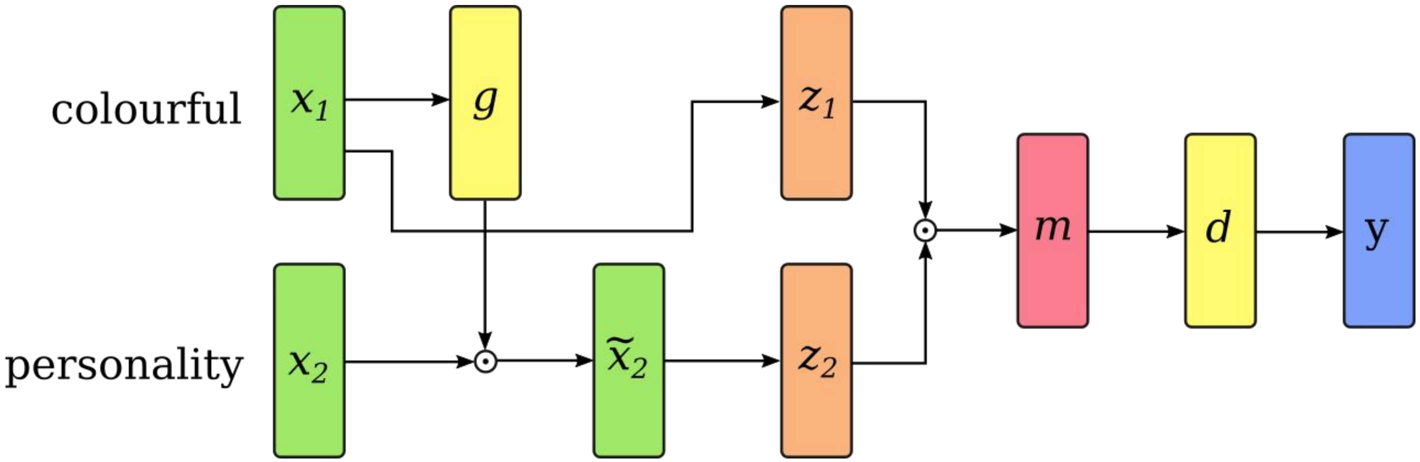


The weight matrices are optimised during training, while the pre-trained embeddings are kept fixed.

Supervised Similarity Network

The final network architecture, using:

1. Word gating
2. Representation mapping
3. Vector combination based on weighted cosine



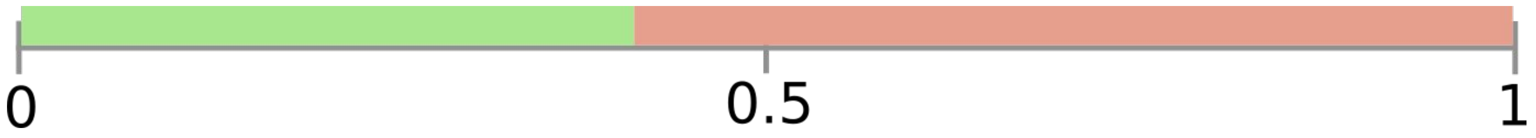
Optimisation

Output between 0 and 1 with sigmoid activation

$$y = \sigma(W_y d)$$

Minimising squared distance with margin 0.1

$$E = \sum_k q_k \qquad q_k = \begin{cases} (\tilde{y} - y)^2 & \text{if } |\tilde{y} - y| > 0.4 \\ 0, & \text{otherwise} \end{cases}$$



Word Representations

Skip-gram embeddings

Skip-gram with negative sampling (Mikolov et al., 2013).

100-dimensional

Trained on English Wikipedia for 3 epochs.

Easy to obtain and generate.

Attribute vectors

Each dimension is a property for the word (Bulat et al., 2017).

2526-dimensional

Trained on McRae norms, predicted for missing words.

Give the best results in previous work.

Datasets

TSV - Tsvetkov et al. (2014)

Adjective–noun pairs annotated for metaphoricity by 5 annotators.

1768 pairs for training, 200 for testing.

Metaphorical	Literal
deep understanding	cold weather
empty promise	dry skin
green energy	empty can
healthy balance	gold coin

MOH - Mohammad et al. (2016)

Verb-subject and verb-object pairs annotated for metaphoricity by 10 annotators.

647 pairs, using 10-fold cross-validation

Metaphorical	Literal
absorb cost	accommodate guest
attack problem	attack village
breathe life	deflate mattress
design excuse	digest milk

Results: Tsvetkov dataset

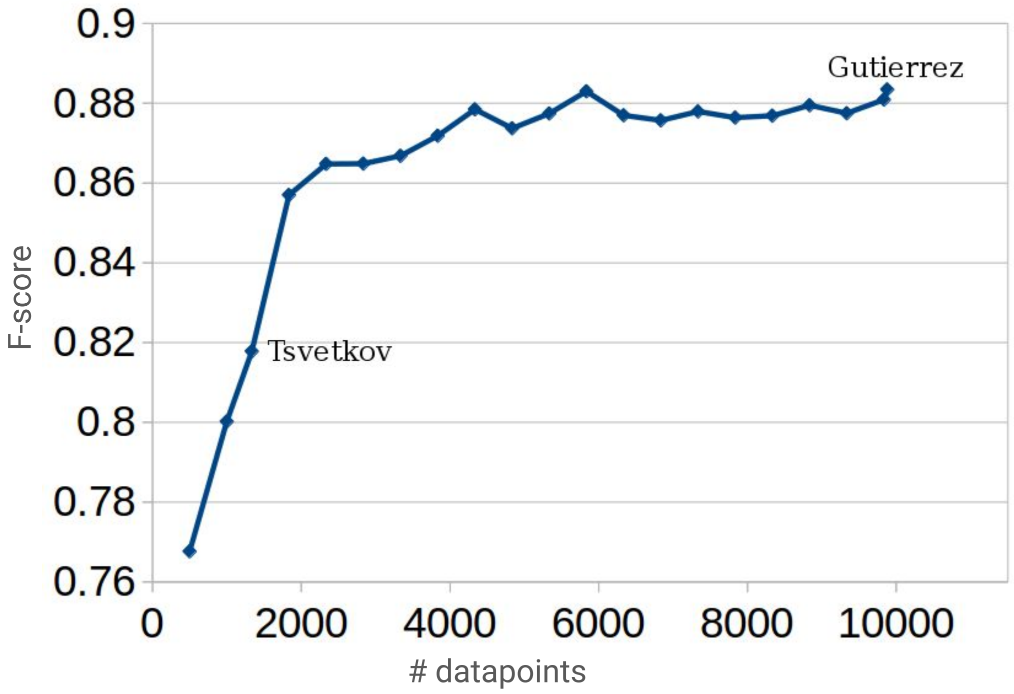
	F1
Tsvetkov et al. (2014)	**85**
Shutova et al. (2016) linguistic	76
Shutova et al. (2016) multimodal	79
Bulat et al. (2017)	77
FFN skip-gram	74.4
FFN attribute	74.5
SSN skip-gram	80.1
SSN attribute	80.6
SSN fusion	**81.1**

Results: Mohammad dataset

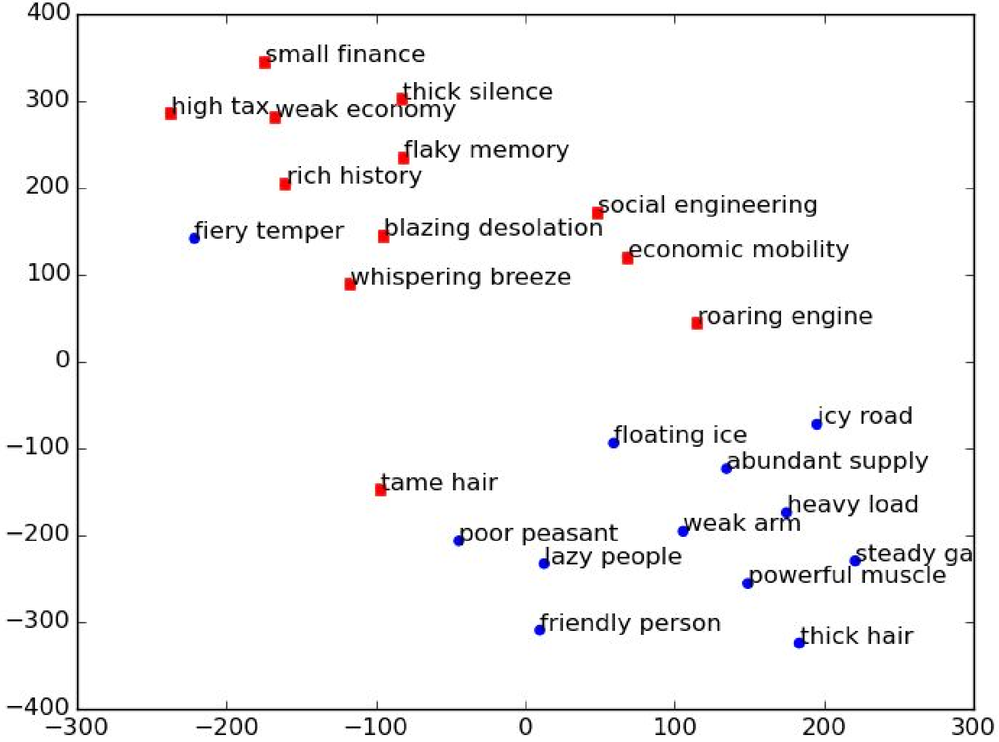
		F1
Shutova et al. (2016)	linguistic	71
Shutova et al. (2016)	multimodal	**75**
FFN skip-gram		70.5
FFN attribute		68.3
SSN skip-gram		**74.2**
SSN attribute		68.8
SSN fusion		69.9

The Effects of Training Data

Adding in 8,592 adjective-noun pairs by Gutierrez et al. (2016)



Vector Space Analysis



SSN (layer m)

Conclusion

- We introduced the first deep learning architecture designed to capture metaphorical composition and evaluated it on a metaphor identification task.
- The network outperforms a metaphor-agnostic baseline and previous corpus-driven approaches.
- Using more training data we can also outperform hand-crafted approaches.
- The framework could potentially be useful for other word pair classification tasks.

Thank you!