EMNLP 2019

Rotate King to get Queen:
Word Relationships as Orthogonal Transformations in Embedding Space

Kawin Ethayarajh - Stanford University

Summary

Word relationships are generally thought of as simple geometric translations or complex non-linear transformations in embedding space. We find that there are parsimonious representations of relationships between these two extremes: orthogonal and linear maps.

For example, where $\overrightarrow{king} + (\overrightarrow{woman} - \overrightarrow{man}) \approx \overrightarrow{queen}$, we can find an orthogonal map R such that $R(\overrightarrow{king}) \approx \overrightarrow{queen}$ and $R(\overrightarrow{man}) \approx \overrightarrow{woman}$. Using orthogonal maps for analogical reasoning is almost as accurate as using geometric translations, and using linear maps is more accurate than both.

Put simply, where $\overrightarrow{b}$ is a translation vector such that

👑 $+ \overrightarrow{b} \approx$ 👸 and 👨 $+ \overrightarrow{b} \approx$ 👩

we can find an orthogonal or linear map R such that

$R($ 👑 $) \approx$ 👸 and $R($ 👨 $) \approx$ 👩

Representing Word Relationships

Given a set of word pairs S, how can we find an orthogonal or linear map f such that $\forall\ (x, y) \in S : f(\overrightarrow{x}) \approx \overrightarrow{y}$?

1. Calculate the mean translation $\overrightarrow{b} = \frac{1}{|S|} \sum_{(x,y) \in S} \overrightarrow{y} - \overrightarrow{x}$.
2. Uniformly randomly sample n words from the vocabulary and stack their embeddings to get the source matrix X.
3. Add $\overrightarrow{b}$ to each sampled word to get the target matrix Y.
4. Use the closed-form solution to the orthogonal Procrustes problem to find the orthogonal matrix that most closely maps X to Y.
5. Use the closed-form solution to ordinary least squares (OLS) to find the linear map that best maps X to Y.

Implications

Evaluating Embeddings: Solving analogies arithmetically should not be treated as a proxy for embedding quality.

Model Architecture: Individual attention heads in Transformers have the capacity to represent semantic relationships, as has been observed with syntactic ones.

Debiasing Embeddings: Traditional methods of debiasing word vectors do not remove social biases in orthogonal and linear maps, only geometric translations.

Evaluating Representations of Word Relationships

See the table below for the accuracy on our word analogy task when relationships are represented as orthogonal, linear, and translative functions. The highest accuracy for each category is in bold.

As seen in the last row, orthogonal maps are almost as accurate as translations for analogical reasoning (0.761 vs. 0.782). Linear maps are more accurate than both (0.798).

Analogy Category	Accuracy		
	Orthogonal	Linear	Translative
capital-common-countries	**0.957**	**0.957**	**0.957**
capital-world	0.922	**0.966**	**0.966**
currency	0.300	**0.467**	0.267
city-in-state	0.529	0.897	**0.926**
family	**0.913**	**0.913**	**0.913**
gram1-adjective-to-adverb	0.438	**0.500**	**0.500**
gram2-opposite	**0.621**	0.586	0.517
gram3-comparative	0.865	0.865	**0.892**
gram4-superlative	**0.912**	0.882	**0.912**
gram5-present-participle	0.909	**0.939**	0.848
gram6-nationality-adjective	0.902	0.902	**0.927**
gram7-past-tense	0.600	**0.650**	0.625
gram8-plural	0.892	**0.919**	0.892
gram9-plural-verbs	**0.900**	0.733	0.800
Avg	0.761	**0.798**	0.782

The accuracy plateaus for $n \geq 250$, where n is the number of sampled words used to create X, suggesting that a robust transformation can be learned with little data.