Variants of Vector Space Reductions for Predicting the Compositionality of English Noun Compounds

Pegah Alipoormolabashi, Sabine Schulte im Walde

Observation

The relationship between noun compounds and their constituents' meanings is not always the same.

- Snow-ball: A **ball** made of **snow**
- Butter-fly: Something that **flies**; Not clearly related to butter

It's crucial for NLP applications that we know the interaction between compounds and constituents' properties.

Literature

There's been research on automatic prediction of the **degree of compositionality** of compounds.

Automatic prediction uses properties of compounds and constituents and/or the compounds' similarity to their constituents.

Degree of Compositionality: A measure of relatedness between a compound's meaning and it's constituents' meanings

Measure of Similarity: Usually words are represented with vectors in a vector space and their similarity is calculated as a function of vectors.

Our Contribution

We evaluated the role of vector-space reductions on the prediction of the compositionality degree of English noun compounds.

Zooming on Compounds and Constituents' Properties

- We split the words into categories based on their value of:
  - Compound frequency
  - Head productivity
  - Modifier productivity
  - Compound compositionality
  - Head compositionality
  - Modifier compositionality
- We then evaluated the predictions on each subset

Setting

- **Datasets**:
  - Text data for making word vectors: ENCOW16 (Schäfer and Bildhauer (2012))
    - English COrpora from the Web
  - Gold standard data for compounds' compositionality degree:
    - Provided by Reddy, et al (2011)
    - A list of English noun compounds and their constituents
    - Human judgement on the compositionality degree of the compounds and also properties of constituents
- **Training word vectors**
  - All trained with a window size of 10
- **POS parser**
  - The TreeTagger by Schmid, 1994 is used for POS tagging and lemmatization
- **Measure**
  - We used **cosine** as a measure of similarity between word vectors.
  - We used the Spearman Rank-Order Correlation Coefficient (Siegel and Castellan, 1988) to compare the predicted results with human judgement.

Main Results

- **Word2vec** performs generally better than the other vector-space variants.
- The nouns matrix outperforms the verbs matrix and the whole matrix.
- Performing **PCA reduction doesn't improve** the results.
- Reducing the nouns matrix to the k most frequent nouns leads to better results for some values of k. It gets better with increasing k, but reaches to a maximum around k=25000-30000.

Properties' Results

- All variants but Word2vec perform **better on mid-frequent compounds** and the prediction on that subset is better than on average.
- Modifier productivity doesn't seem to affect prediction results, but the results are **better for compounds with mid-productive heads**.
- Results are significantly **better for compounds with high-compositional heads**.

Vector Space Variations

- **ALL :** All context words
- **POS Matrices**
  - **VV:** All context verbs
  - **NN:** All context nouns
  - **NN-k:** k most frequent corpus nouns
- **PCA reductions**
  - **ALL-PCA-k:** PCA with k dimensions on ALL
  - **NN-PCA-k:** PCA with k dimensions on NN
- **Word2vec:** 300d Word2vec vectors

Prediction Functions

- **WORD1:** Use only the compound–modifier cosine score
- **WORD2:** Use only the compound–head cosine score
- **ADD:** Add the compound–modifier and compound–head cosine scores
- **MULT:** Multiply the compound–modifier and compound–head cosine scores
- **COMB:** Add the compound–modifier, compound–head and the multiplication of both cosine scores

Conclusion

- **Word2vec with 300 dimensions is the winner**, both in best performance and stability over different prediction functions.
- The second best results were obtained when using a **large subset of context nouns.**
- While ADD,MULT and COMB are better prediction functions overall, while zooming on subsets of words using just head-compound or modifier-compound similarity can be enough.