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

Pegah Alipoormolabashi
Sharif University of Technology
Islamic Republic of Iran
palipoor@ce.sharif.edu

Sabine Schulte im Walde
Institute for Natural Language Processing
University of Stuttgart, Germany
schulte@ims.uni-stuttgart.de

Abstract

Predicting the degree of compositionality of noun compounds is a crucial ingredient for lexicography and NLP applications, to know whether the compound should be treated as a whole, or through its constituents. Computational approaches for an automatic prediction typically represent compounds and their constituents within a vector space to have a numeric relatedness measure for the words. This paper provides a systematic evaluation of using different vector-space reduction variants for the prediction. We demonstrate that Word2vec and nouns-only dimensionality reductions are the most successful and stable vector space reduction variants for our task.[1]

1 Introduction

The semantic relations between compounds and their constituents do not follow a strict rule. Compare, for example, the English noun compounds *snowball* –a ball consisting of snow, where clearly both constituents *snow* and *ball* contribute to the meaning of the compound– and *butterfly* –where the semantic contribution of the modifier noun *butter* is not obvious without knowing about the etymology of the compound. Computational approaches to predict the degree of compositionality typically represent compounds and their constituents within a vector space, and then compare the compound vectors with the constituent vectors as a proxy to the compounds' degree of compositionality (Reddy et al., 2011b; Reddy et al., 2011a; Salehi and Cook, 2013; Schulte im Walde et al., 2013; Salehi et al., 2014; Schulte im Walde et al., 2016; Cordeiro et al., 2019). Previous works have explored variants of vector space models in different ways. Our contribution in this paper was to provide a systematic evaluation of vector-space reductions across kinds, i.e., exploring part-of-speech-based reduction, Principal Components Analysis using Singular Value Decomposition, and Word2vec embeddings. As the gold standard, we used the dataset of English noun compounds created by Reddy et al. (2011b). This dataset contains a random list of English noun compounds annotated by compositionality ratings on the semantic contribution of the modifier to the compound meaning, the semantic contribution of the head noun to the compound meaning, and the compositionality of the compound as a whole phrase. Table 1 shows an example set of the gold data.

Compound	Word1	Word2	Phrase
climate change	4.90±0.30	4.83±0.38	4.97±0.18
polo shirt	1.73±1.41	5.00±0.00	3.37±1.38
search engine	4.62±0.96	2.25±1.70	3.32±1.16

Table 1: Examples of compounds and judgements on their compositionality (mean value and standard deviation, based on 30 annotators) from Reddy et al. (2011b).

[1] In accordance with the multiple-submission policy of WiNLP 2020 this work has already been published in (Alipoor and Schulte im Walde, 2020)

Vector Space Variant	Prediction Function	Correlation Coefficient
All	ADD/COMB	0.630
VV	ADD	0.581
NN	ADD/MULT	0.658
Word2vec	**COMB**	**0.689**

Vector Space Variant	Prediction Function	Correlation Coefficient
NN	ADD/MULT	0.658
NN-1000	COMB	0.483
NN-10000	ADD	0.638
NN-20000	ADD	0.661
NN-30000	**MULT**	**0.663**
NN-40000	MULT	0.659

Vector Space Variant	Prediction Function	Correlation Coefficient
All	**ADD/COMB**	**0.630**
All-PCA-100	ADD	0.527
All-PCA-200	ADD	0.577
All-PCA-500	ADD	0.584
All-PCA-1000	MULT	0.574
All-PCA-2000	COMB	0.609
All-PCA-5000	ADD/COMB	0.616

Vector Space Variant	Prediction Function	Correlation Coefficient
NN	**ADD/MULT**	**0.658**
NN-PCA-100	ADD	0.620
NN-PCA-200	COMB	0.595
NN-PCA-500	MULT	0.631
NN-PCA-1000	COMB	0.640
NN-PCA-2000	MULT	0.657
NN-PCA-5000	MULT	0.654

Table 2: Best results for each vector space variant

2 Experiment

We experimented with several vector space variants for representing the compounds and the constituents. All of these vector space variants were created based on the ENCOW16[2] corpus with a window size of 10. We also applied the *TreeTagger* for part-of-speech (pos) tagging and lemmatisation (Schmid, 1994). The vector space variants are listed below.

- **ALL** The whole co-occurrence matrix of the words as the baseline.
- **POS** Subsets of the co-occurrence matrix with only context dimensions of specific parts-of-speech (specifying on nouns/NN vs. verbs/VV).
- **PCA** All and nouns-only matrices after Principle Component Analysis reduction.
- **WORD2VEC** Standard Word2vec embedding (Mikolov et al., 2013) with 300-dimensional vectors.
- **NN-K** Noun-only matrix reduced to contain only neighbour nouns within the k most frequent nouns.

We used **cosine** as a similarity measure between compounds and constituents, assuming that the stronger the distributional similarity (i.e., the higher the cosine values), the stronger the semantic relatedness and therefore the degree of compositionality. Next to assessing the individual contributions of compound–modifier and compound–head relatedness, we applied the same functions as in (Reddy et al., 2011b) to combine the compound–constituent cosine scores for predicting the degree of compositionality of the compounds. Given that each component within the functions might provide a different weight to the overall prediction, we applied a linear regression model to find the corresponding coefficients, and here we report the best results from three-fold cross-validation with human judgments. The vector space predictions were evaluated against the mean human ratings on the degree of compositionality, using the Spearman Rank-Order Correlation Coefficient ρ (Siegel and Castellan, 1988). Table 2 shows the best results among prediction functions using each vector-space variant.

- **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

	All			NN			All-PCA-5000			NN-PCA-2000			Word2vec		
	high	mid	low	high	mid	low	high	mid	low	high	mid	low	high	mid	low
Compound Frequency Range	0.409	**0.470**	0.268	0.372	**0.606**	0.330	0.396	**0.484**	0.196	0.447	**0.616**	0.262	**0.678**	0.585	0.352
Modifier Productivity Range	0.594	**0.619**	0.394	0.543	0.543	**0.653**	**0.600**	0.554	0.414	0.543	0.625	**0.636**	**0.631**	0.571	0.584
Head Productivity Range	0.555	**0.784**	0.245	0.648	**0.840**	0.337	0.559	**0.746**	0.227	0.677	**0.827**	0.316	0.701	**0.801**	0.474
Compound Compositionality Range	0.240	0.256	**0.525**	0.293	0.290	**0.469**	**0.409**	0.196	0.196	0.375	**0.614**	0.259	**0.631**	0.571	0.584
Modifier Compositionality Range	0.498	**0.560**	0.434	0.620	**0.662**	0.343	0.535	**0.601**	0.434	0.649	**0.681**	0.363	**0.650**	0.585	0.417
Head Compositionality Range	**0.777**	0.476	0.361	**0.752**	0.589	0.496	**0.779**	0.519	0.253	**0.753**	0.585	0.468	**0.761**	0.592	0.288

Table 3: Best results for vector space variants across compound and constituent properties.

In order to zoom into specific strengths of individual vector space variants, we applied the variants to subsets of our compound targets according to the targets' compositionality, compound frequency, modifier productivity, and head productivity. For each of these conditions, we created three disjunctive subsets of the 90 compound targets with 30 targets each. We observed that training the regression on the whole set and testing it on the subsets has the same results as training it on the subsets. The subsets contain the strongest, weakest and in-between targets as based on the respective condition, e.g., regarding the compound frequency condition we distinguish between high-frequency, mid-frequency and low-frequency compounds. The empirical information relies on a refinement of the Reddy et al. dataset by Schulte im Walde et al. (2016) using ENCOW14 (the predecessor of ENCOW16 used in this study). The best results are shown in table 3.

3 Summary of Results

This study provided a systematic evaluation of vector-space reductions across kinds, for the task of predicting the compositionality degree of noun compounds. Our vector-space variant experiments identified Word2vec with 300 dimensions as the clear winner. Similarly good and stable predictions have been achieved when using a large subset of context nouns, even without any further PCA reduction. As a result, we suggest using these vectors in NLP applications. The baseline with using all context dimensions is worse in comparison to all optimised reduced conditions other than running PCA on the whole matrix. Therefore, next to identifying a clear winner (Word2vec) we can induce from our results that using only the most frequent noun dimensions is a reasonable alternative.
Regarding the prediction functions, ADD, MULT and COMB (with only marginal differences between them in most cases) generally outperformed WORD1 and WORD2. So combining the relatedness information for compound–modifier and compound–head pairs is better for the prediction of the compounds' degree of compositionality than relying on just one or the other.
These results vary strongly across subsets representing different ranges of compositionality, frequency and productivity. We observed that predictions for mid-frequency compounds are better than on average and predictions for low-frequency compounds are particularly bad. For head productivity ranges we observed very high prediction results for mid-productivity and very low prediction results for low-productivity subsets. Finally, regarding compositionality the predictions were better for overall mid- and high-compositional in comparison to low-compositional compounds; and generally good for compounds with mid-compositional compound-modifier relatedness and compounds with strongly compositional compound-head relatedness.

[2] `http://corporafromtheweb.org/encow16/`

Acknowledgements

We thank Anna Hätty for her help on training the Word2vec and Gabriella Lapesa for guiding through the SVD reductions.

References

Pegah Alipoor and Sabine Schulte im Walde. 2020. Variants of vector space reductions for predicting the compositionality of english noun compounds. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 4379–4387, Marseille, France, May. European Language Resources Association.

Silvio Cordeiro, Aline Villavicencio, Marco Idiart, and Carlos Ramisch. 2019. Unsupervised Compositionality Prediction of Nominal Compounds. *Computational Linguistics*, 45(1):1–57.

Tomas Mikolov, Wen tau Yih, and Geoffrey Zweig. 2013. Linguistic Regularities in Continuous Space Word Representations. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 746–751, Atlanta, GA, USA.

Siva Reddy, Ioannis P. Klapaftis, Diana McCarthy, and Suresh Manandhar. 2011a. Dynamic and Static Prototype Vectors for Semantic Composition. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 705–713, Chiang Mai, Thailand.

Siva Reddy, Diana McCarthy, and Suresh Manandhar. 2011b. An Empirical Study on Compositionality in Compound Nouns. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 210–218, Chiang Mai, Thailand.

Bahar Salehi and Paul Cook. 2013. Predicting the Compositionality of Multiword Expressions Using Translations in Multiple Languages. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics*, pages 266–275, Atlanta, GA, USA.

Bahar Salehi, Paul Cook, and Timothy Baldwin. 2014. Using Distributional Similarity of Multi-way Translations to Predict Multiword Expression Compositionality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 472–481, Gothenburg, Sweden.

Helmut Schmid. 1994. Probabilistic Part-of-Speech Tagging using Decision Trees. In *Proceedings of the 1st International Conference on New Methods in Language Processing*.

Sabine Schulte im Walde, Stefan Müller, and Stephen Roller. 2013. Exploring Vector Space Models to Predict the Compositionality of German Noun-Noun Compounds. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics*, pages 255–265, Atlanta, GA, USA.

Sabine Schulte im Walde, Anna Hätty, and Stefan Bott. 2016. The Role of Modifier and Head Properties in Predicting the Compositionality of English and German Noun-Noun Compounds: A Vector-Space Perspective. In *Proceedings of the 5th Joint Conference on Lexical and Computational Semantics*, pages 148–158, Berlin, Germany.

Sidney Siegel and N. John Castellan. 1988. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, Boston, MA, USA.