A Regression Model of Adjective-Noun Compositionality in Distributional Semantics

Emiliano Guevara

Sanne Brinkhorst

December 7, 2012

Overview

1. Introduction
2. Model
3. Experimental setup
4. Evaluation
5. Conclusion

Compositionality of Meaning

Three models are begin compared:

- Vector addition
  $v1_i + v2_i = v3_i$
- Point-wise mulitplication
  $v1_i \times v2_i = v3_i$
- Partial Least Squares Regression (PLSR)

PLSR

- Multivariabte regression technique
- Very suitable for problems high dimensionality and limited data
- Robust against overtraining
- Some dimensions contribute more than others

PLSR

- Function mapping of two vectors
- Each dimension is important
- We use adjective and noun as predictors and the observed pair as dependent variable

Dataset

- British National Corpus
- 1380 Adjective-Noun pairs with frequency >400 e.g. *nice_house*
- Vector space with 40,000 most frequent tokens
- Resulting in a 40,000 $\times$ 500 matrix
- Training set of 1000 pairs, test set of 380 pairs
- $v1 =$ *nice*, $v2 =$ *house* and $v3 =$ *nice_house*

Adjective-Noun pairs

Euclidian distance between predicted vectors and observed vectors

	Min.	1st Q.	Median	Mean	3rd Q.	Max.
ADD	0.877	1.402	1.483	1.485	1.570	1.814
MUL	0.973	0.998	1.002	1.002	1.005	1.019
PLSR	0.624	0.805	0.856	0.866	0.919	1.135

Table 1: Summary of distance values between the 380 observed A-N pairs and the predictions from each model (ADD=additive, MUL=multiplicative, PLSR=Partial Least Squares Regression).

Adjective-Noun pairs

For each A-N pair in the test set the 10 nearest neighbours in the whole subspace are determined

	1	2	3	4	5	6	7	8	9	10	Tot.
ADD	0	0	0	0	0	0	0	0	0	0	0
MUL	0	1	0	2	1	0	0	0	0	0	4
PLSR	94	51	24	18	10	7	7	5	2	1	219

Table 2: Nearest predicted neighbours and their positions in the top-10 list.

Predicting Neighbours

- For each A-N pair the 10 nearest neighbours are determined
- This is very restricting!
- For each A-N pair the nearest prediction neighbours are determined as well
- For each prediction you check if this shares a top-10 neighbour with the gold standard

Predicting Neighbours

	Shared Neigh.	Predicted Neigh.	Total
ADD	48	577	625
MUL	0	37	37
PLSR	0	263	263
Not shared:			6,675

Table 3: Shared neighbours with respect to the gold standard and shared predicted neighbours.

Predicting Neighbours

	ADD	MUL	PLSR	OBS
ADD	2,144 (56%)	–	–	–
MUL	59 (1%)	3,800 (100%)	998 (26%)	1,555 (40%)
PLSR	1,472 (38%)	–	2,802 (73%)	2,190 (57%)
OBS	125 (3%)	–	–	55 (1%)

Table 4: Origins of neighbours in each models' top-10 list of neighbours extracted from the full space composed of observations and predictions ($380 \times 4 = 1,440$ items) (ADD=additive, MUL=multiplicative, PLSR=Partial Least Squares Regression, OBS=observed vectors) .

Conclusion

- This method is not tried before in this field
- For combining adjectives and nouns this gives good results
- For the second evaluation the results are mixed, the additive model works better here.

Future research

- Develop better evaluation methods
- Extending experiments to other techniques
- *Any semantic relation instantiated by any syntactic structure could be learned if sufficient data is provided*

Bibliography

- Guevara (2010) *A Regression Model of Adjective-Noun Compositionality in Distributional Semantics*, GEMS Workshop.
- Baroni & Zamparelli (2010) *Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space*, EMNLP
- Mitchell & Lapata (2008) *Vector-based Models of Semantic Composition*, ACL