Exercises on Vector Semantics and Meaning Composition

Quynh Do

LIIR lab, Department of Computer Science, KU Leuven

22^{nd} November 2016

Vector Semantics - Exercise 1

The Pointwise Mutual Information (PMI) is a measure of how often two pointwise mutual information events x and y occur, compared with what we would expect if they were independent:

$I(x, y) = log_2 \frac{P(x,y)}{P(x)P(y)}$

Apply this intuition to co-occurrence vectors by defining the PMI association between a target word w and a context word c as:

$I(w, c) = log_2 \frac{P(w,c)}{P(w)P(c)}$

Vector Semantics - Exercise 1

	aardvark	...	computer	data	pinch	result	sugar	...
apricot	0	...	0	0	1	0	1	
pineapple	0	...	0	0	1	0	1	
digital	0	...	2	1	0	1	0	
information	0	...	1	6	0	4	0	

Figure: Co-occurrence counts

	p(w,context)					p(w)
	computer	data	pinch	result	sugar	p(w)
apricot	0	0	0.05	0	0.05	0.11
pineapple	0	0	0.05	0	0.05	0.11
digital	0.11	0.05	0	0.05	0	0.21
information	0.05	.32	0	0.21	0	0.58
p(context)	0.16	0.37	0.11	0.26	0.11	

Figure: Replacing co-occurrence counts with joint probabilities, showing the marginals around the outside.

Vector Semantics - Exercise 1

Table: PMI vectors

	computer	data	pinch	result	sugar
apricot	-inf	-inf	2.25	-inf	2.25
pineapple	-inf	-inf	2.25	-inf	2.25
digital	1.66	-0.56	-inf	-0.07	-inf
information	-0.8	0.57	-inf	0.47	-inf

Vector Semantics - Exercise 1

Advantage of PMI over raw frequency:
One problem is that raw frequency is very skewed and not very discriminative. If we want to know what kinds of contexts are shared by apricot and pineapple but not by digital and information, we're not going to get good discrimination from words like the, it, or they, which occur frequently with all sorts of words and aren't informative about any particular word.

Vector Semantics - Exercise 1

Problem of PMI:
PMI values range from negative to positive infinity. But negative PMI values (which imply things are co-occurring less often than we would expect by chance) tend to be unreliable unless our corpora are enormous.
To distinguish whether two words whose individual probability is each 10^{-6} occur together more often than chance, we would need to be certain that the probability of the two occurring together is significantly different than 10^{-12}, and this kind of granularity would require an enormous corpus. Furthermore it's not clear whether it's even possible to evaluate such scores of 'unrelatedness' with human judgments.

Vector Semantics - Exercise 1

It is more common to use Positive PMI (PPMI) which replaces all negative PMI values with zero (Church and Hanks 1989, Dagan et al. 1993, Niwa and Nitta 1994)

Table: PPMI vectors

	computer	data	pinch	result	sugar
apricot	0	0	2.25	0	2.25
pineapple	0	0	2.25	0	2.25
digital	1.66	0	0	0	0
information	0	0.57	0	0.47	0

Vector Semantics - Exercise 2

$$\text{cosine}(\vec{v},\vec{w}) = \frac{\vec{v}\cdot\vec{w}}{|\vec{v}||\vec{w}|} = \frac{\sum_{i=1}^{N} v_i w_i}{\sqrt{\sum_{i=1}^{N} v_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}}$$

Figure: Cosine similarity

Vector Semantics - Exercise 2

- $apricot = [2, 0, 0]$
- $digital = [0, 1, 2]$
- $information = [1, 6, 1]$
- $cosine(apricot, information) = \frac{2*1+0*6+0*1}{\sqrt{2^2+0^2+0^2}\sqrt{1^2+6^2+1^2}} = 0.16$
- $cosine(digital, information) = \frac{0*1+1*6+2*1}{\sqrt{0^2+1^2+2^2}\sqrt{1^2+6^2+1^2}} = 0.58$

Since $0.58 > 0.16$, then digital is closer to information than apricot.

Meaning Composition - Exercise 3

Step 1: Software Preparation

- Download WEKA tool from http://www.cs.waikato.ac.nz/ml/weka/downloading.html
- You can use any programming language that you are familiar with. But we recommend Python 3. The library ”numpy” may be needed.

Meaning Composition - Exercise 3

Step 2: Read word embedding file The idea is that you create a dictionary that maps a word to a float vector.

russian 0.033725 0.070442 0.180771 0.301392 0.17898
clear -0.02888 0.359714 -0.132314 0.338405 0.297073
english 0.193309 0.10998 -0.240711 -0.433992 -0.159352

Meaning Composition - Exercise 3

Step 3: Read data file Each instance should contains four attributes:

- adjective
- noun1
- noun2
- label – positive (True) or negative (False)

Meaning Composition - Exercise 3

Step 4: Make composition representation The presentation for noun 2 is easily extracted from the word embedding dictionary. But we need composition representation for the phrase “adjective noun1”.
Idea for composition:

- concatenation
- sum
- weighted sum

Meaning Composition - Exercise 3

Step 5: For each composition method, create a csv file for the dataset: For each instance, representations of “adjective noun1” and “noun 2” are used as features for the classification problem. For example, if “noun2” = [1,2,3,4] and “adjective noun1” = [5,6] and the label is positive (True) then you should create a csv line of:
1,2,3,4,5,6,True
Note that you need also line that contains the names of features and class on top of the csv file:
fea_1,fea_2,fea_3,fea_4,fea_5,fea_6,class The csv file is used as input for WEKA tool.

Meaning Composition - Exercise 3

Step 6: Let's play with WEKA and the csv files.

Meaning Composition - Exercise 3

If you have problem with programming, try to understand the sample codes.