Distributed Semantic Models for Word Vectors

Hyperspace Analogue to Language (HAL)

Ramaseshan Ramachandran

METHODS TO CREATE WORD VECTORS

- Brown clustering - statistical algorithms for assigning words to classes based on the frequency of their co-occurrence with other words
- Hyperspace Analogue to Language - HAL
- Correlated Occurrence Analogue to Lexical Semantics - COALS
- Latent Semantic Analysis or Latent Semantic Indexing
- Global Vectors - GloVe
- Neural networks using skip grams and CBOW
  - CBOW - uses the surrounding words to predict the center word
  - Skip-gram - uses the center word to predict the surrounding words

Hyperspace Analogue To Language[1] (HAL[1])

[1] K. Lund and C. Burgess. Producing high-dimensional semantic spaces from lexical co-occurrence. Be- havior Research Methods, Instruments, & Computers, 28(2):203208, 1996.

HIGHLIGHTS OF THE HAL PAPER

Motivation

Human semantic memories are presumably constructed through experience with the world; as concepts are encountered, information about their meanings is accumulated.

Two problems of constructing semantic spaces manually

- Find a set of axes that define a concept
- Determine where each word should fall on each axis

This is a tedious and error-prone process; HAL instead learns the space automatically from co-occurrence statistics

HAL METHODOLOGY

1. A **weighted window** representing a span of words(n-grams) is moved across the corpus in one-word increments.

   Example
   Small corpus: A weighted window representing a span of words(n-grams) is moved across the corpus in one-word increments.

   n-grams
   ('By', 'moving', 'this') ('moving', 'this', 'window') ('this', 'window', 'over') ('window', 'over', 'the') ('over', 'the', 'source') ('the', 'source', 'corpus') $(\cdots)$

2. Capture the co-occurrence values of the words within it at every window movement to form a co-occurrence matrix.
3. Each cell of the matrix represents the summed co-occurrence counts for a single word pair (w_t, w_n)
4. The accumulation of values is direction sensitive
   The count of sequence "$w_1 w_2$" and count for the sequence "$w_2 w_1$" are different
5. For every word, there is both a row and a column containing relevant co-occurrence values, each one representing its concept axis

HAL SCANNING

Corpus: the horse raced past the barn fell

Left2Right Scanning

the	horse	raced	past	the	barn	fell
K	5	4	3	2	1	0

the	horse	raced	past	the	barn	fell
	K	5	4	3	2	1

Right2Left Scanning

fell	barn	the	past	raced	horse	the
K	5	4	3	2	1	0

fell	barn	the	past	raced	horse	the
	K	5	4	3	2	1

Incidence Matrix

	the	horse	raced	past	barn	fell
the	2	3				
horse	5					
raced	4	5				
past	3	4				
barn	1	2				
fell		1				

	the	horse	raced	past	barn	fell
the						
horse						
raced						
past						
barn						
fell	4	1	2	3	5	

Co-occurrence matrix using HAL

HAL ALGORITHM

Require a big corpus >5 GB for a reasonable similarity measures

1. Preprocess to limit the vocabulary size
2. Perform two scans using a ramping window of size 11 - first $\longrightarrow$ direction and later in the $\longleftarrow$ direction
3. Use the first word as the key word and the rest as context words
4. Use the last word as the key and rest as the context words, during the $\longleftarrow$ scanning
5. The nearest neighbor of the key gets the weight 10 and the 10th word gets the weight 1
6. Construct an incidence matrix using the co-occurrence values
7. For every word, concatenate its row vector and its column vector to obtain the full word vector
8. The number of elements in each word vector is therefore $2|V|$

HAL EXPERIMENT

160 million words from Usenet news groups

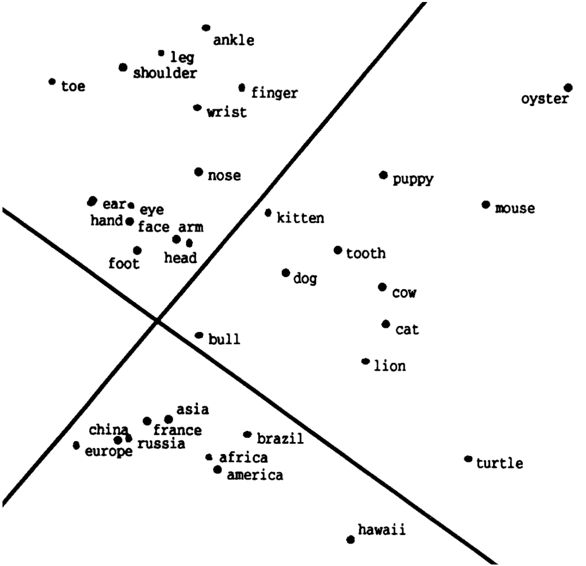
Window size $= 10$

- Vocabulary - Words with a frequency > 50
- Zipf's law is used to eliminate most common and rare words
- Minkowski distance is used for computing word similarities
  $d(x,y) = (\sum_i |x_i - y_i|^r)^{1/r}$
- The word vectors produce high dimensional semantic space - associative
- This is an unsupervised analysis of text
- Demonstrated sizable correlation between vector similarity and basic cognitive effects

RESULTS

Target	n1	n2	n3	n4	n5
jugs	juice	butter	vinegar	bottles	cans
leningrad	rome	iran	dresdan	azerbaijan	tibet
lipstick	lace	pink	cream	purple	soft
triumph	beauty	prime	grand	former	rolling
cardboard	plastic	rubber	glass	thin	tiny
monopoly	threat	huge	moral	gun	large

HAL WORD VECTORS - SIMILARITY CHART



CONCLUDING REMARKS - HAL

- HAL acquires contextual understanding of words by using the moving window and weighting co-occurrence distance.
- Using a large corpus, the co-occurrence matrix carries the history of this contextual experience
- The semantic vectors are representations that are essentially measures of context

IMPACT OF FREQUENCY MEASURE ON SIMILARITY

- Even if t_1 and t_2 are unrelated, if $p(t_1) \approx p(t_2)$, then their vectors will contain elements with similar magnitudes.

  $\implies$ any similarity measure

  For example, words a, an, the co-occur with many words in the vocabulary
- Conversely if they are related but $p(t_1) \ll p(t_2)$ then their vectors will contain elements with widely differing magnitudes, simply due to their differing co-occurrence probability.

In general, similar relative frequency does not imply semantic similarity. Hence we require normalized measures to build word vectors.

SUMMARY

- A count-based word embedding model
- Captures word meanings through the unsupervised analysis of text
- Produces word vectors that are semantic (similar words) and associative in nature
- Acquires word meanings as a function of keeping track of how words are used in context
- Carries the history of the contextual experience by using a moving window and weighting of co-occurring words based on the distance
- Exploits the regularities of language such that conceptual generalisations can be captured in a data matrix

HAL Demo

REFERENCES I

[1] Kevin Lund and Curt Burgess. "Producing high-dimensional semantic spaces from lexical co-occurrence". en. In: *Behavior Research Methods, Instruments, & Computers* 28.2 (1996), pp. 203–208. ISSN: 0743-3808, 1532-5970. DOI: 10.3758/BF03204766. URL: http://link.springer.com/article/10.3758/BF03204766 (visited on 09/09/2015).