DaSciM's Weekly Group Meeting

Paper review

Antoine Tixier

October 14, 2016

Agenda

- ✓ Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change (*ACL 2016*). From Leskovec's team at Stanford.

- ✓ Corpus-independent Generic Keyphrase Extraction using Word Embedding Vectors (*WDSM 2015 workshop on deep learning*). From the University of Western Australia.

- ✓ A Graph Degeneracy-based Approach to Keyword Extraction (*our EMNLP 2016 paper*)

Diachronic Word Embeddings (ACL 2016)

Diachronic: temporal evolution. **Synchronic**: fixed point in time.

Goal: Use **historical** word embeddings to study the evolution of word meaning over time (i.e., semantic change).

Ways to quantify **semantic change**:

- Pairwise cosine similarity time series (i.e., linguistic "shifts"):

$$s^{(t)}(w_i, w_j) = \text{cos-sim}(w_i^{(t)}, w_j^{(t)}) \quad (1)$$

  between *two words* w_i and w_j over a time-period
- Semantic displacement:

$$\Delta^{(t)}(w_i) = \text{cos-dist}(w_i^{(t)}, w_i^{(t+1)}) \quad (2)$$

  for the *same word* over a time-period (i.e., "rate" of semantic change)

*$dist = 1 - sim$

Diachronic Word Embeddings (ACL 2016)

Learning word embeddings is a **stochastic** process (random initialization, negative sampling, etc.) $\Rightarrow$ word-word distances are *invariant* from training to training, but the axes are *recycled*.

$\Rightarrow$ to enable comparison between embeddings trained at different times (i.e., on different corpora), we need to perform **vector alignment**. This can be done via **orthogonal Procrustes**:

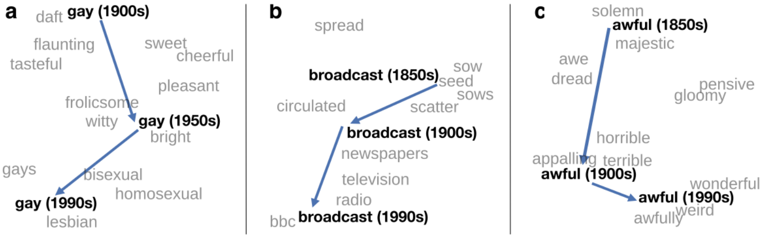
$$R^{(t)} = \arg \min_{Q^\top Q = I} \|QW^{(t)} - W^{(t+1)}\|_F, \quad (3)$$

$W^{(t)} \in \mathbb{R}^{d \times |\mathcal{V}|}$ is the embedding space at time t, $\| \cdot \|_F$ is the Frobenius norm.

The solution $R^{(t)} \in \mathbb{R}^{d \times d}$ (orthogonal) corresponds to the best rotational alignment between $W^{(t)}$ and $W^{(t+1)}$. It is a mapping/ transfer matrix.

Diachronic Word Embeddings (ACL 2016)

Qualitative results:



- *gay* shifted from meaning "cheerful" to referring to homosexuality
- in the early 20th century *broadcast* referred to "casting out seeds", it now means "transmitting signals"
- *Awful* underwent a process of pejoration, as it shifted from meaning "full of awe" to meaning "terrible"

Diachronic Word Embeddings (ACL 2016)

Quantitative results:

the **rate of semantic change** obeys the following power-law relation:

$$\Delta(w_i) \propto f(w_i)^{\alpha} \times d(w_i)^{\beta} \tag{4}$$

where f is the frequency of w_i, d its polysemy, $\alpha < 0$, and $\beta > 0$.

$\Rightarrow$ **frequent** words change at slower rates while **polysemous** words change faster

Recall that:

$$\Delta^{(t)}(w_i) = \text{cos-dist}(w_i^{(t)}, w_i^{(t+1)}) \tag{5}$$

Polysemy is roughly equivalent to the **contextual diversity** of a word, measured in terms of clustering coefficient within a co-occurrence network. High: always, great, quite... Low: retrieval, thirties, mom...

GoW combined with Word Embeddings

How to combine **local** co-occurrence statistics with **global** exterior knowledge to get better edge weights in graphs-of-words?

⇒ Word **attraction force** (adaptation of Newton's law of universal gravitation):

$$f(w_i, w_j) = \frac{freq(w_i) \times freq(w_j)}{d^2} \tag{6}$$

Where $freq(w)$ is the count of word w in document D and d is the euclidean distance between w_i and w_j in an embedding space.

$\times$

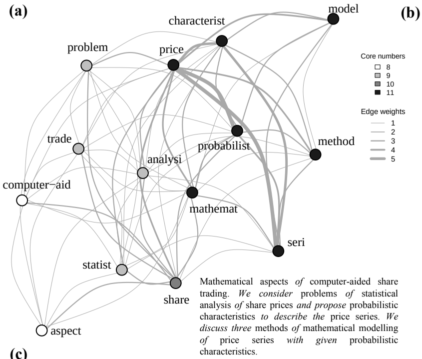
⇒ **Dice coefficient** (phraseness likelihood):

$$Dice(w_i, w_j) = \frac{2freq(w_i, w_j)}{freq(w_i) + freq(w_j)} \tag{7}$$

Where $freq(w_i, w_j)$ is the co-occurrence count of w_i and w_j in D.

Our EMNLP 2016 paper: motivation

(a)

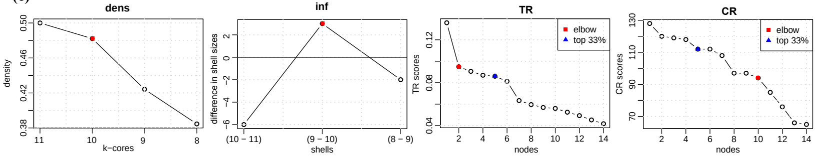


(b)

Core numbers				TR scores				CR scores			
	P	R	F1		P	R	F1		P	R	F1
MAIN	0.86	0.55	0.67	ELB	1	0.18	0.31	ELB	0.90	0.82	0.86
INF	**0.83**	**0.91**	**0.87**								
DENS	0.88	0.64	0.74	PER	1	0.45	0.63	PER	1	0.45	0.63

Core numbers		TR scores		CR scores	
mathemat	11	**price**	.1359	**mathemat**	128
price	11	**share** *ELB*	.0948	**price**	120
probabilist	11	**probabilist**	.0906	**analysi**	119
characterist	11	**characterist**	.0870	**share**	118
seri	11	**seri** *PER*	.0860	**probabilist** *PER*	112
method	11	**mathemat**	.0812	**characterist**	112
model *MAIN*	11	**analysi**	.0633	**statist**	108
share *DENS*	10	**statist**	.0595	**trade**	97
trade	9	method	.0569	problem	97
problem	9	problem	.0560	**seri** *ELB*	94
statist	9	**trade**	.0525	method	85
analysi *INF*	9	**model**	.0493	**computer-aid**	76
aspect	8	**computer-aid**	.0453	**model**	66
computer-aid	8	aspect	.0417	aspect	65

(c)

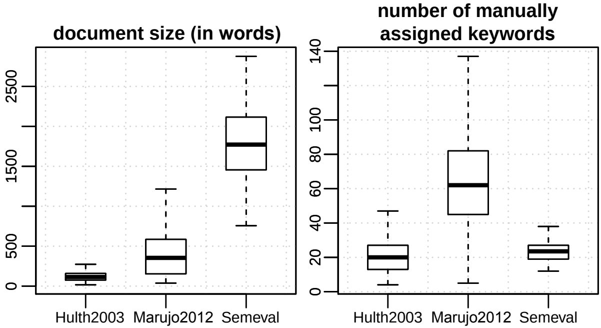


Our EMNLP 2016 paper: experiments

Hulth2003: 500 abstracts drawn from the Inspec database of physics and engineering papers

Marujo2012: 450 web news stories covering 10 different topics from art and culture to business, sport, and technology

Semeval: 100 scientific papers from the ACM Digital Library



Our EMNLP 2016 paper: results

	precision	recall	F1-score
dens	**48.79**	**72.78**	**56.09***
inf	48.96	72.19	55.98*
CRP	61.53	38.73	45.75
CRE	65.33	37.90	44.11
main	51.95	54.99	50.49
TRP	65.43	41.37	48.79
TRE	71.34	36.44	45.77

Hulth2003, K-truss, $W = 11$.

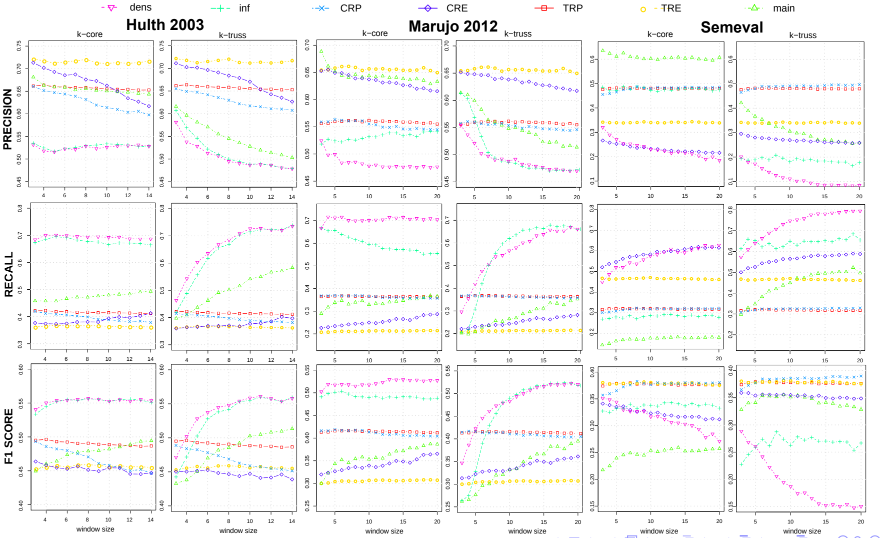
	precision	recall	F1-score
dens	**47.62**	**71.46**	**52.94***
inf	53.88	57.54	49.10*
CRP	54.88	36.01	40.75
CRE	63.17	25.77	34.41
main	64.05	34.02	36.44
TRP	55.96	36.48	41.44
TRE	65.50	21.32	30.68

Marujo2012, k-core, $W = 13$.

	precision	recall	F1-score
dens	8.44	79.45	15.06
inf	17.70	65.53	26.68
CRP	**49.67**	**32.88**	**38.98***
CRE	25.82	58.80	34.86
main	25.73	49.61	32.83
TRP	47.93	31.74	37.64
TRE	33.87	46.08	37.55

Semeval, K-truss, $W = 20$.

Our EMNLP 2016 paper: impact of window size



References

- Hamilton, W. L., Leskovec, J., Jurafsky, D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. arXiv preprint arXiv:1605.09096.

- Wang, C., Mahadevan, S. (2008, July). Manifold alignment using procrustes analysis. In Proceedings of the 25th international conference on Machine learning (pp. 1120-1127). ACM.

- Wang, R., Liu, W., McDonald, C. (2014, November). Corpus-independent Generic Keyphrase Extraction Using Word Embedding Vectors. In Software Engineering Research Conference (p. 39).

- Tixier, A. J. P., Malliaros, F. D., Vazirgiannis, M. A Graph Degeneracy-based Approach to Keyword Extraction.