Word Vectors - IV

Pawan Goyal

CSE, IIT Kharagpur

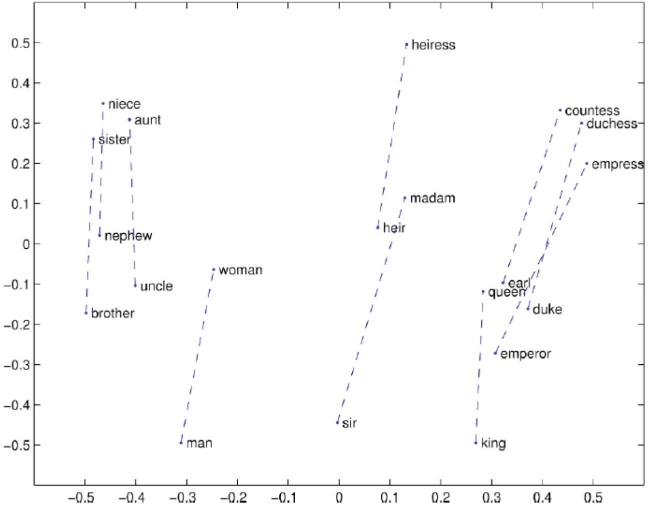
February 3rd, 2017

Questions from yesterday

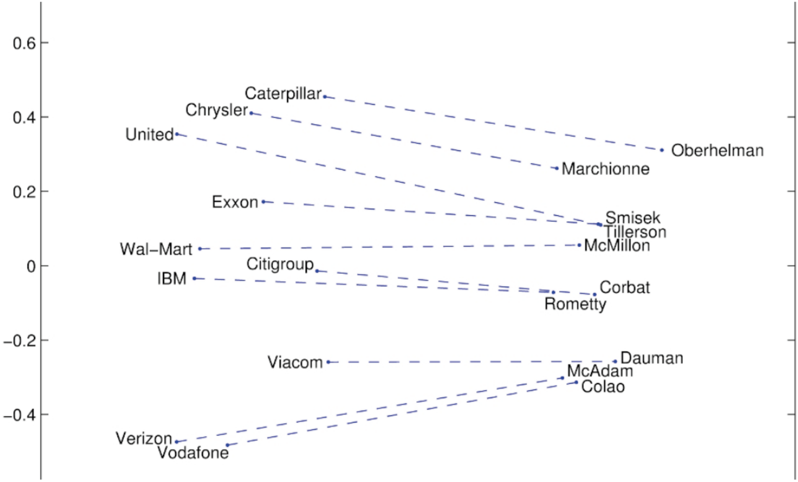
On Board

- CBOW / Skip-Gram: Form of the objective / error function for gradient descent?
- $f(P_{ij})$ in Glove: Why should it give low weight in the regression to low co-occurrences?
- How is Glove faster than CBOW / Skip-Gram?

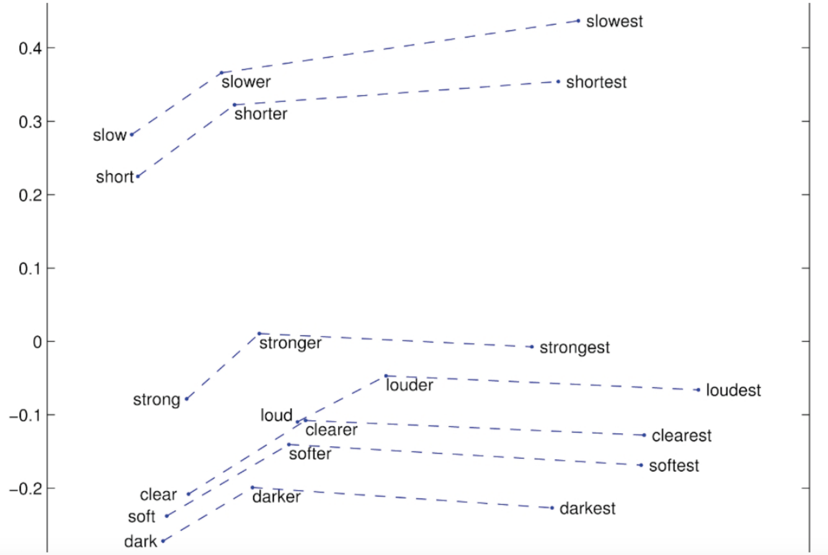
Glove Visualisations



Glove Visualisations



Glove Visualisations



Visualization using t-SNE

t-distributed Stochastic Neighbor Embedding

Objective: Given a set of N high-dimensional objects $x_1, \ldots, x_N$, learn a $d-$dimensional map $y_1, \ldots, y_N$ (with $y_i \in R^d$) such that the new similarities q_{ij} reflect the old similarities p_{ij} as closely as possible, i.e., the KL-divergence between the two similarity distributions is minimized

$$KL(P||Q) = \sum_{i \neq j} p_{ij} log \frac{p_{ij}}{q_{ij}}$$

Intrinsic Evaluation

- Word vector distances and their correlation with human judgements
- Example dataset: WordSim353

Word 1	Word 2	Human (mean)
tiger	cat	7.35
tiger	tiger	10.00
book	paper	7.46
computer	internet	7.58
plane	car	5.77
professor	doctor	6.62
stock	phone	1.62
stock	CD	1.31
stock	jaguar	0.92

Hyperparameters

Skip-gram: using evaluation on analogy testing

- **Dimensions:** 300 dimensions work the best
- **Window size:** 8 words around each center word works well.
- **More training time and data helps!!**

Handling Polysemy

Problem with word vectors

Multiple senses of a given word get the same representation!!

Handling Polysemy

Problem with word vectors

Multiple senses of a given word get the same representation!!

Huang et al., “Improving Word Representations via Global Context and Multiple Word Prototypes”, ACL 2012.

Basic Idea

Cluster words windows around words, retrain with each word assigned to multiple different clusters, e.g., bank$_1$, bank$_2$ etc.

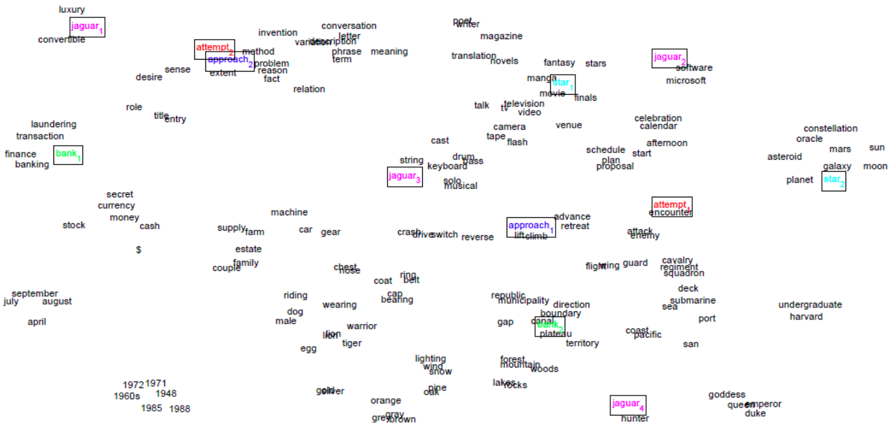
Multiple Word Prototypes: Nearest Neighbors

Center Word	Nearest Neighbors
bank_1	corporation, insurance, company
bank_2	shore, coast, direction
star_1	movie, film, radio
star_2	galaxy, planet, moon
cell_1	telephone, smart, phone
cell_2	pathology, molecular, physiology
left_1	close, leave, live
left_2	top, round, right

Code and dataset:

http://www.socher.org/index.php/Main/ImprovingWordRepresentationsViaGlobalC

Multiple Word Prototypes: Visualization



Cross-lingual applications: Word Embeddings

Cross-lingual information retrieval task

Query is in one language and documents in another language

Cross-lingual applications: Word Embeddings

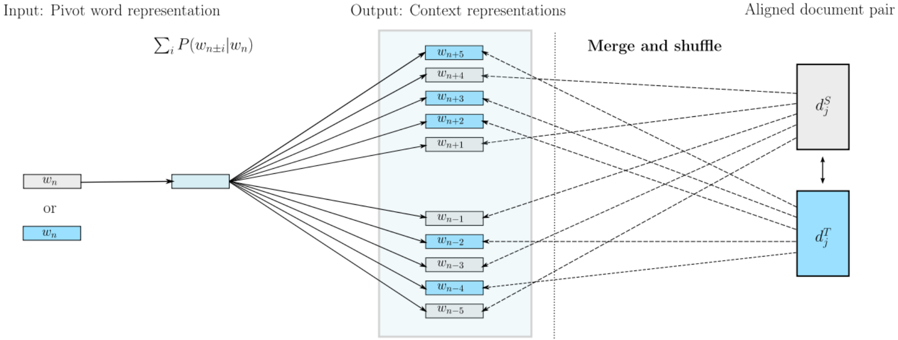
Cross-lingual information retrieval task

Query is in one language and documents in another language

Obtaining common representations for words in multiple languages

So that you can visualize words in multiple languages in the same space

Bilingual Word Embeddings: Merge and Shuffle



Vulić, Ivan, and Marie-Francine Moens. "Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings." SIGIR 2015.

Other Approaches

- The previous approach requires a comparable corpora. What if you do not have such corpora?
- Suppose you have a dictionary to start with, what would be an approach?