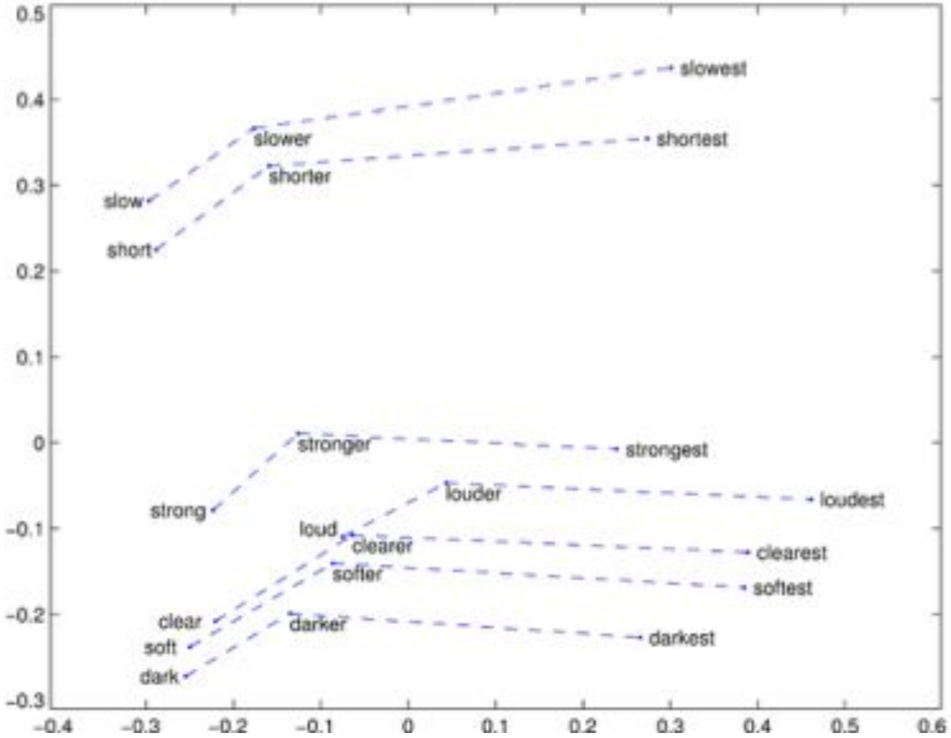
DS-GA 1011 Fall 2018 - lab 5

Word Embedding & Intrinsic Evaluation

Word Embedding

- **Representation of words**
- **Relation among vectors can denote relation among words**
- **Intrinsic Evaluation by analogies**

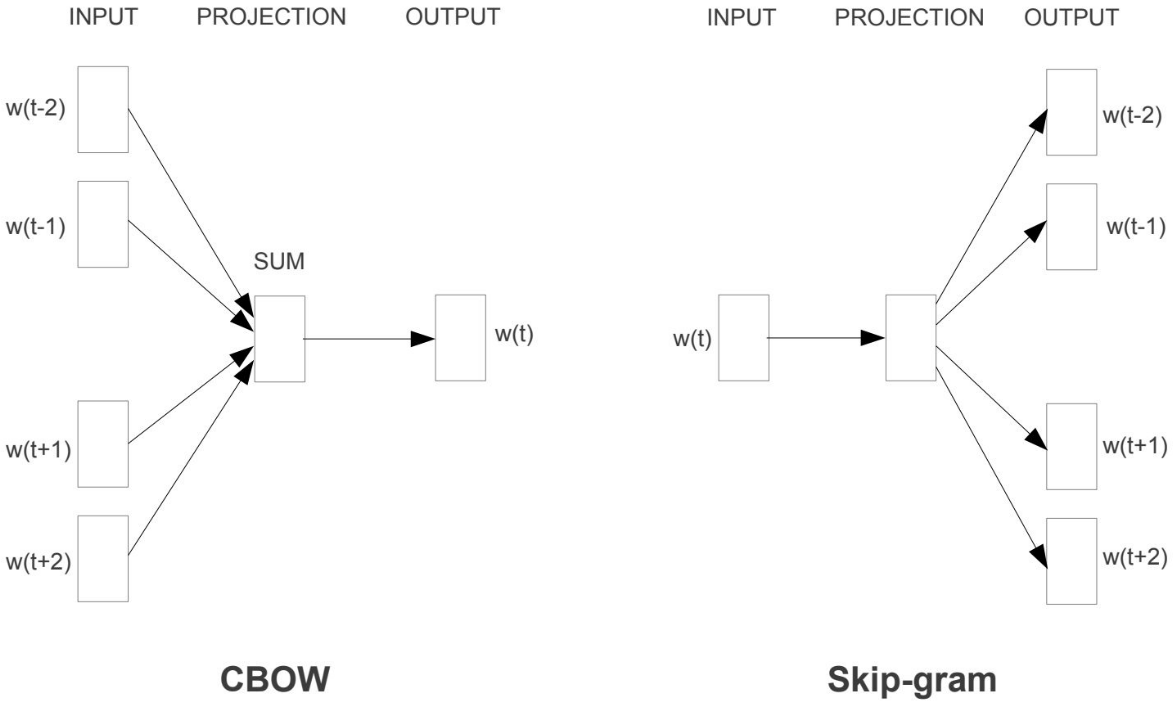


Training Word-Vectors

“You shall know a word by the company it keeps”

Firth, John R., 1957. *Modes of meaning.* Oxford: Oxford University Press

Training Word-Vectors



CBOW

Skip-gram

Pre-trained word embeddings

- **GloVE vectors**

$$J(\theta) = \frac{1}{2} \sum_{i,j=1}^{W} f(P_{ij})(u_i^T v_j - \log P_{ij})^2$$

- **FastText**
  - CBOW on subword n-grams

Cosine Similarity

- **Measure the angle between two vectors**

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2}$$

- **Range: [-1, 1]**
  - -1 when two vectors point to exactly opposite direction
  - 1 when two vectors point to the same direction

Intrinsic Word Vector Evaluation

- **A : B :: C : D**
- **Given A, B, and C, find best match for D**
- **Examples:**

  Athens to Greece = Berlin to ?

  King to Men = Queen to ?

  Dark to Darker = Soft to ?

  Flock to Bird = School to ?

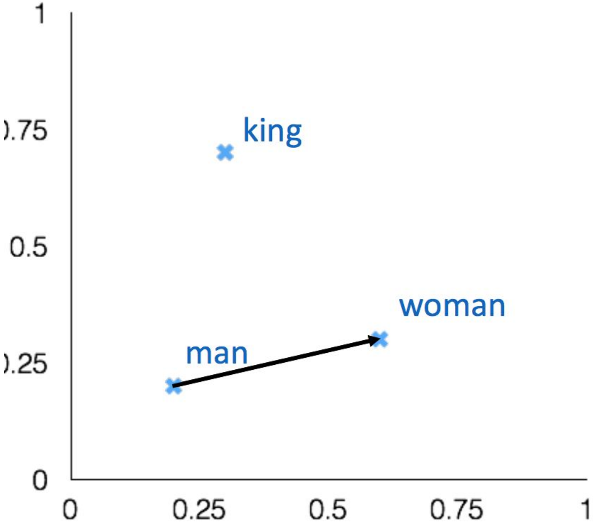
Word Vector Analogy

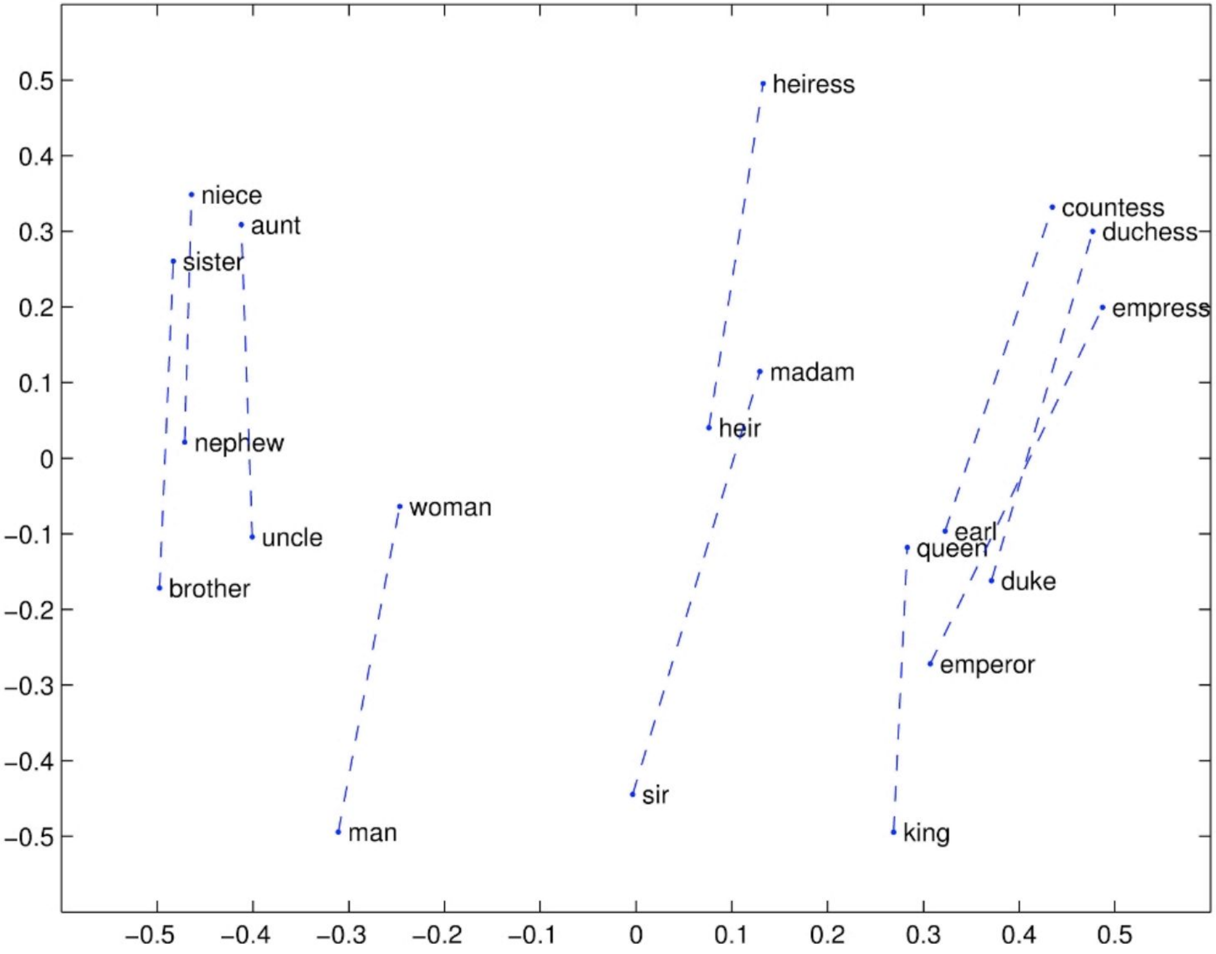
a:b :: c:?

man:woman :: king:?

$$d = \arg\max_i \frac{(x_b - x_a + x_c)^T x_i}{||x_b - x_a + x_c||}$$

Evaluate word vectors based on how well their cosine distance after addition captures intuitive semantic and syntactic analogy questions.





0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
−0.5
−0.5
−0.4
−0.3
−0.2
−0.1
0
0.1
0.2
0.3
0.4
0.5
niece
aunt
sister
nephew
uncle
brother
woman
man
heiress
madam
heir
sir
countess
duchess
empress
earl
queen
duke
emperor
king

Reading

- **GloVE**: Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation.
- **FastText**: Armand Joulin, Edouard Grave, Piotr Bojanowski, Tomas Mikolov. 2016. Bag of Tricks for Efficient Text Classification
- **Skip-gram/CBOW**: Mikolov et al. 203. Efficient Estimation of Word Representations in Vector Space
- **Negative Sampling/Hierarchical Softmax**: Mikolov et al. Distributed Representations of Words and Phrases and their Compositionality