Applied Text Analytics &

Natural Language Processing

with Dr. Mahdi Roozbahani
& Wafa Louhichi

Glove (Global Vectors for Word Representation)

Learning Objectives

In this course, you will learn

- GloVe: Global vectors for word representation
- Word similarity through the word vectors

GloVe Model

- GloVe stands for global vectors where global refers to global statistics of corpus and vectors are representation of words
- GloVe uses statistics of word occurrences in a corpus as the primary source of information
- GloVe model combines two widely adopted approaches for training word vectors:
    - Global matrix factorization
    - Window-based methods

Co-Occurrence Matrix

- For a corpus of vocabulary V of size d, the co-occurrence matrix is a symmetrical matrix of size $d \times d$
- X_{ij}: number of times word j occurs in the context of the word i after defining a window size
- $X_i = \sum_k X_{ik}$: summation over all the words which occur in the context of the word i
- $Pij = \frac{X_{ij}}{X_i}$: P is co-occurrence probability where P_{ij} is the probability of word j occurring in the context of word i

Example corpus: "it was the best of times, it was the worst of times" with a context window =2

Co-occurrence Matrix

	it	was	the	best	of	times	worst
it	0	2	2	0	1	1	0
was	2	0	2	1	0	1	1
the	2	2	0	1	2	0	1
best	0	1	1	0	1	1	0
of	1	0	2	1	0	2	1
times	1	1	0	1	2	0	1
worst	0	1	1	0	1	1	0

Co-Occurrence Matrix

$$X_{i=0,j=1} = 2$$
$$X_{i=0} = 6$$
$$P_{i=0,j=1} = \frac{2}{6} = 0.33$$

Example corpus: “it was the best of times, it was the worst of times” with a context window =2

Co-occurrence Matrix

	it	was	the	best	of	times	worst
it	0	2	2	0	1	1	0
was	2	0	2	1	0	1	1
the	2	2	0	1	2	0	1
best	0	1	1	0	1	1	0
of	1	0	2	1	0	2	1
times	1	1	0	1	2	0	1
worst	0	1	1	0	1	1	0

GloVe Cost Function

- GloVe suggests finding the relationship between two words in terms of probability rather than occurrence counts.
- GloVe looks to find vectors w_i and w_j such as $w_i^T w_j = \log\left(P_{ij}\right) = \log(\frac{X_{ij}}{X_i})$
- $\log(X_i)$ is independent of word j and can be represented as a bias b_i
- Adding a bias term to restore the symmetry for vector w_j, we get:

$$w_i^T w_j + b_i + b_j = \log(X_{ij})$$

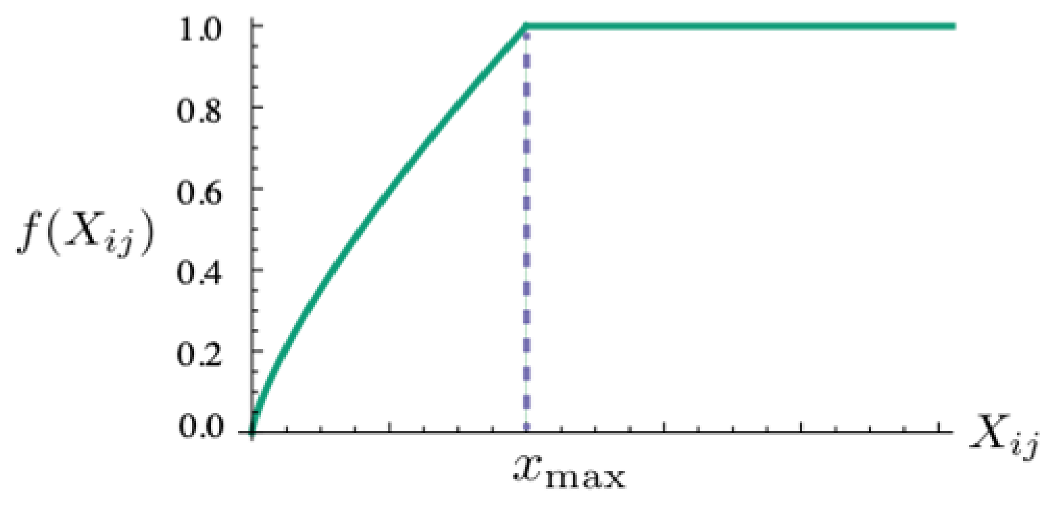
GloVe Cost Function

- A weighted least squares is used as a cost function for GloVe model:

$$J = \sum_{ij} f(X_{ij})\left(w_i^T w_j + b_i + b_j - \log(X_{ij})\right)^2$$

the function $f(x)$ is defined as: $f(x) = \left(\frac{x}{x_{\text{max}}}\right)^a$ if $x < x_{\text{max}}$ else, 1

Weighting function f with α = 3/4.



Source: https://nlp.stanford.edu/pubs/glove.pdf

GloVe Word Vectors

- The model is trained in batches of the training sample with optimizer to minimize the cost function and hence generate word and context vectors for each word
- Each word in the corpus is thus represented by a dense vector of a fixed size length
- The word vectors obtained by GloVe showcase the meaning was captured in these vector representations through similarity as well as linear structure
- Using Euclidean distance or cosine similarity between word vectors represents the linguistics or semantic similarity of the corresponding words. For example, here are the closest words to the target word summer:

```
model.most_similar('summer')
```

```
[('winter', 0.8896949291229248),
 ('spring', 0.8580389022827148),
 ('autumn', 0.7742397785186768),
 ('weekend', 0.7385302782058716),
 ('year', 0.7348464131355286),
 ('days', 0.725011944770813),
 ('beginning', 0.7218300104141235),
 ('during', 0.7205086946487427),
 ('season', 0.7031365633010864),
 ('day', 0.7015056610107422)]
```

GloVe Word Vectors

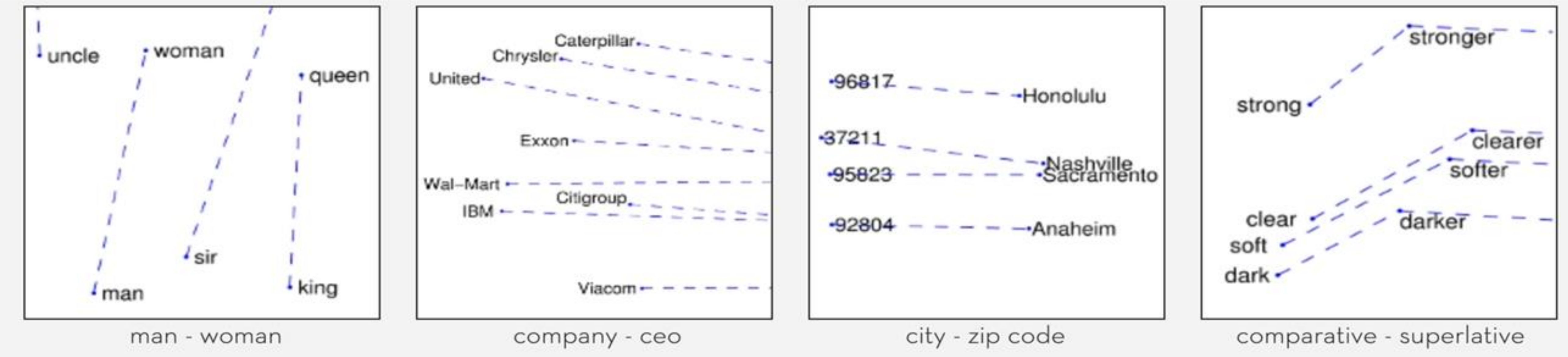
- The word vectors by GloVe conserve linear substructures. Vector differences captures as much as possible the meaning specified by the two words. For example, the underlying concept that differentiates man and woman, meaning gender, may be equivalently specified by other word pairs such as king and queen; $w_{man} - w_{woman} = w_{king} - w_{queen}$

```
1 analogy('man', 'king', 'woman')
```

'queen'

```
1 analogy('england', 'english', 'italy')
```

'italian'



Source: https://nlp.stanford.edu/projects/glove/

Summary

- We learned about dense vector representations using GloVe
- We learned about the interesting results that show how GloVe vector representations capture the word semantics