CS565: Intelligent Systems and Interfaces

Lecture: Vector Semantics

30th Mar, 2016

Semester: Jan – May 2016

Ashish Anand

IIT Guwahati

- Reference
  - Speech and Language Processing, Jurfsky and Martin : Chapter 19, draft version, 24th Aug, 2015 [http://web.stanford.edu/~jurafsky/slp3/19.pdf]
  - Slides: [http://web.stanford.edu/~jurafsky/slp3/slides/19.pdf]

  - **REMARK**: Skip the topics not discussed in the class. But you have to go through the topics which were left for self-reading.

Vector Semantics – What is it?

There is no lecture on Friday, instead we will have project presentation.

There is no class on Friday, instead we will have project presentation.

“class” is similar to “lecture”.

Vector-Semantics: vector representation of words in vector-space but maintaining their similarity.

Why we need such representation?

- Can we not do with Thesaurus?
  - Answers lie in questions like –
    - Do we have exhaustive and updated thesaurus ?
    - Do we have thesaurus for all languages? [In other words, can we develop it in language independent manner?]

Context determines meaning of word

A bottle of *tesguino* is on the table.

Everybody likes *tesguino*.

Tesguino makes you drunk.

We make *tesguino* out of corn.

Context determines word similarity

- Harris (1954)
  - "*Oculist and eye-doctor … occur in almost the same environments*"
  - Generalize it: *"If A and B have almost identical environments … we say that they are synonyms"*

- Firth (1957)
  - "*You shall know a word by the company it keeps!*"

Meaning of word is determined by the distribution of words around it.

Broad categories of vector space models

- Long and Sparse vector representation
  - Co-occurrence matrix based methods (term-doc, term-term matrices based on MI, tf-idf etc.)
- Short and Dense vector representation
  - Dimensionality reduction techniques such as Singular value decomposition (Latent Semantic Analysis) on co-occurrence matrix
  - Neural language inspired models (skip-grams, CBOW)
- Other Methods
  - Clustering methods: Brown Clusters [Collins lecture]
  - Hybrid methods: GloVe

Co-occurrence Matrix

Building block of vector space models

Term Document Matrix: Document Vector

- Each cell: count of word *w* in a document *d*:
  - Each document is a count vector in $\mathbb{N}^V$: a column below

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	1	8	15
soldier	2	2	12	36
fool	37	58	1	5
clown	6	117	0	0

Term-Document Matrix: Document Vector

- Initially defined as vector representation for documents.
- Each document is being represented in |V|- dimensional vector space.
- Notion: Similar documents tend to use similar words.
- Document vectors used in document clustering and several other Information Retrieval (IR) tasks.

Term-Document Matrix: Document vector

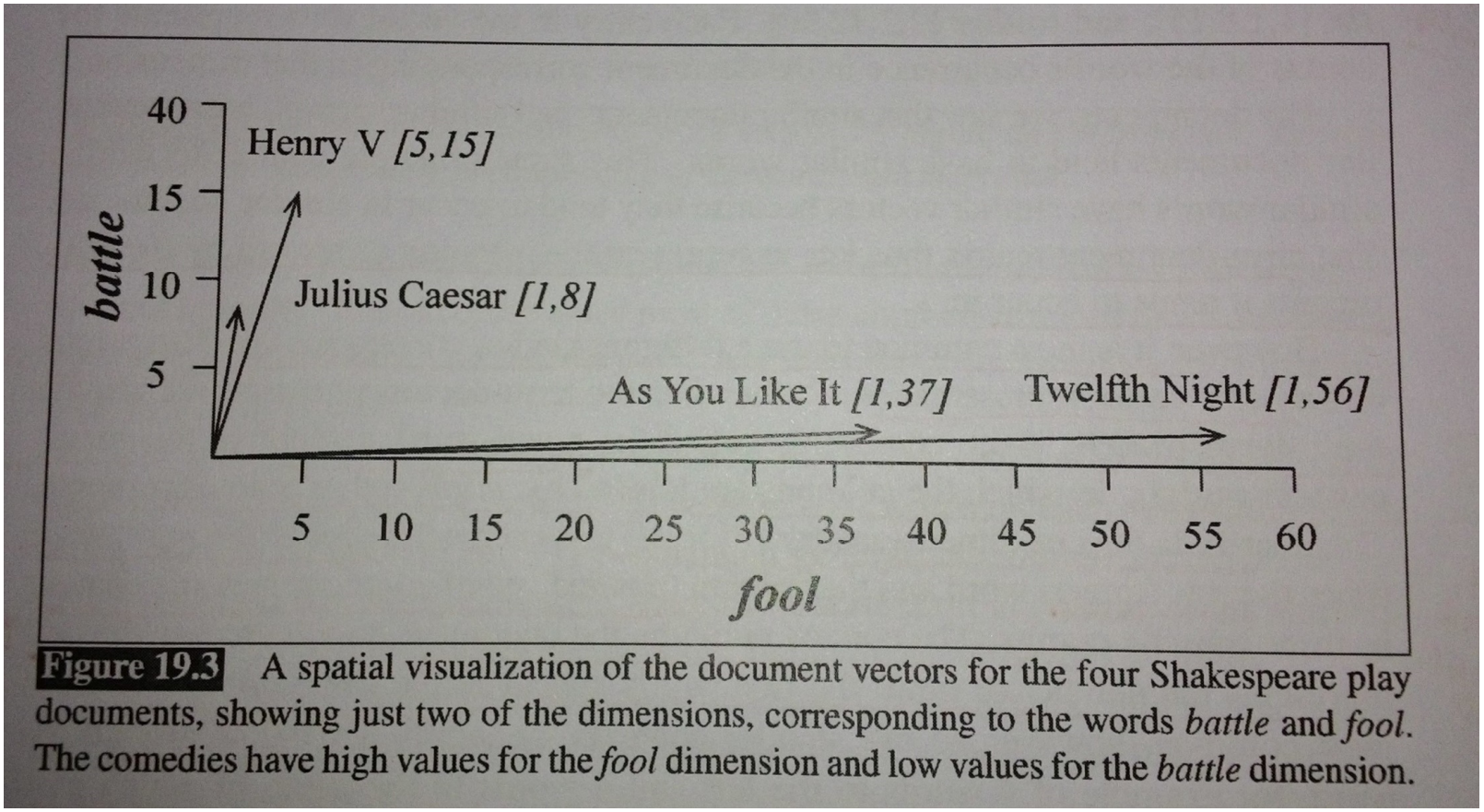


Figure 19.3 A spatial visualization of the document vectors for the four Shakespeare play documents, showing just two of the dimensions, corresponding to the words *battle* and *fool*. The comedies have high values for the *fool* dimension and low values for the *battle* dimension.

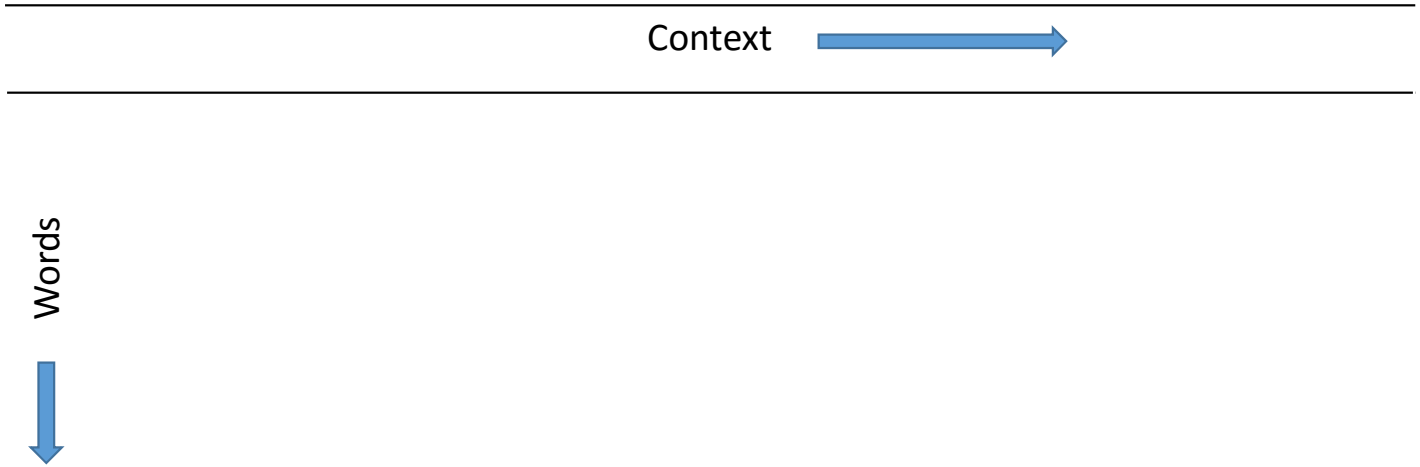
Term-Document Matrix: Word vector

- Row-vector can be used as vector representation of word
- Notion: meaning of a word can be inferred from the documents it tends to occur in.
- Two words are similar if their vectors are similar.

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	1	8	15
soldier	2	2	12	36
fool	37	58	1	5
clown	6	117	0	0

Term-Term Matrix: Word vector

- Alternate names
  - Word-word matrix
  - word-context matrix



Term-term Matrix

- Multiple ways to fill the |V| x |V| matrix
  - Each cell records number of times the *row (target)* words co-occur with the *column (context)* words.
  - ***context***: document, then "how many times the two words co-occur in the same document.
  - ***context***: window of *n* words around the word, then “number of times column words occur within *n* words either side of the row word”.

Co-occurrence takes into account two kinds of association

- Syntagmatic Association (First-order association)
  - They occur nearby each other.
  - *Drink* is first-order associate of *water*

- Paradigmatic Association (Second-order association)
  - They occur with similar words.
  - *Drink* is second-order associate of words like *sip, swallow*

Word-context matrix: An Example

sugar, a sliced lemon, a tablespoonful of **apricot** preserve or jam, a pinch each of,
their enjoyment. Cautiously she sampled her first **pineapple** and another fruit whose taste she likened
well suited to programming on the digital **computer**. In finding the optimal R-stage policy from
for the purpose of gathering data and **information** necessary for the study authorized in the

	aardvark	computer	data	pinch	result	sugar	...
apricot	0	0	0	1	0	1	
pineapple	0	0	0	1	0	1	
digital	0	2	1	0	1	0	
information	0	1	6	0	4	0	

Word-context matrix: what determines size of context ?

- Objective at hands
  - Shorter window (1-3) , more **syntactic** representation
  - Longer window (4-10), more **semantic** representation

Word-context matrix: Issue with raw count

- Raw word count or frequency is not a good measure. [Why?]
  - May not be very informative
  - Example of very frequent and common words such as "*the*" and *"of*" not having discriminative power.

- Can you think of measure which can say whether a ***context word*** is informative about the ***target word*** ?
  - Answer lies in realizing similarity with a particular topic we discussed earlier in the course.

Word-context matrix: Alternative Measures

- Positive Pointwise Mutual Information [PPMI] [DIY]
  - Definition
  - Why positive adjective?
  - What happens with rare context words?
  - Do we need smoothing methods here?

- Tf-idf (Term Frequency – Inverse Document Frequency)

- t- and Likelihood Ratio tests