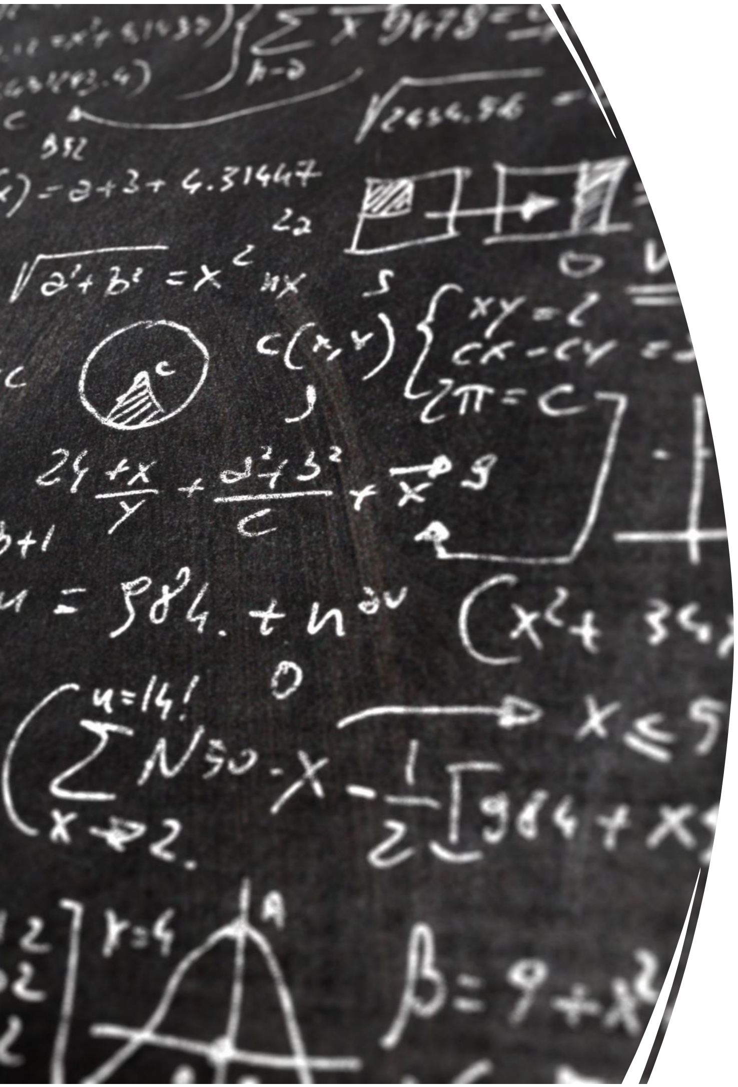


CS 383 – Computational Text Analysis

Lecture 5
CTA overview, Word Representations

Adam Poliak

02/01/2023

Slides adapted from Dan Jurafsky, Jure Lescovik

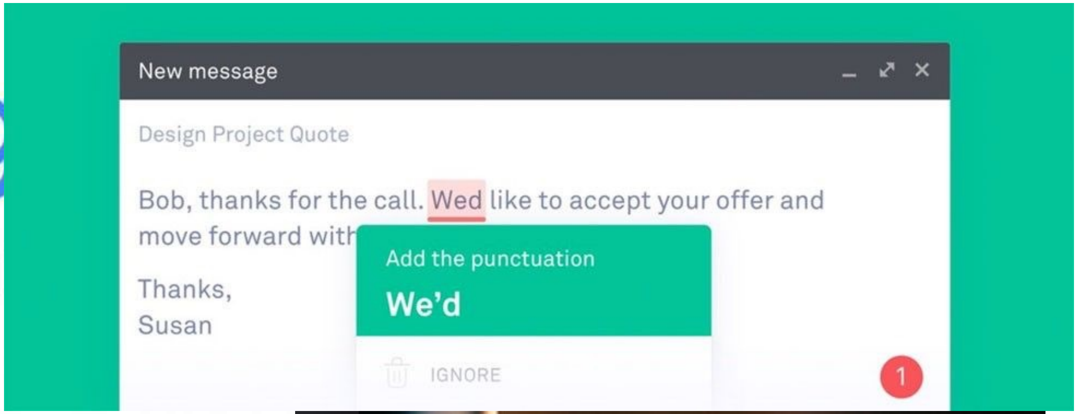
Announcements

- Office Hours:
  - This week: Thursday 3:30-4:30pm
- HW02 released last night, due Wednesday 02/08

Outline

- ML vs CTA vs CL (according to Adam)
- Recap
- Word Representations

Natural Language Processing



Natural Language Processing

HLT

Natural Language Processing

Natural Language Processing

- Goal of ML: algorithmic contribution

Natural Language Processing

- Goal of ML: algorithmic contribution

- Goal of CL: understand humans & language

Natural Language Processing

AI

HLT

Text as data

Natural Language Processing

- Goal of AI:
- 

AI

HLT

Text as data

Natural Language Processing

AI

- Goal of AI:
-

HLT

Goal of Text as Data:

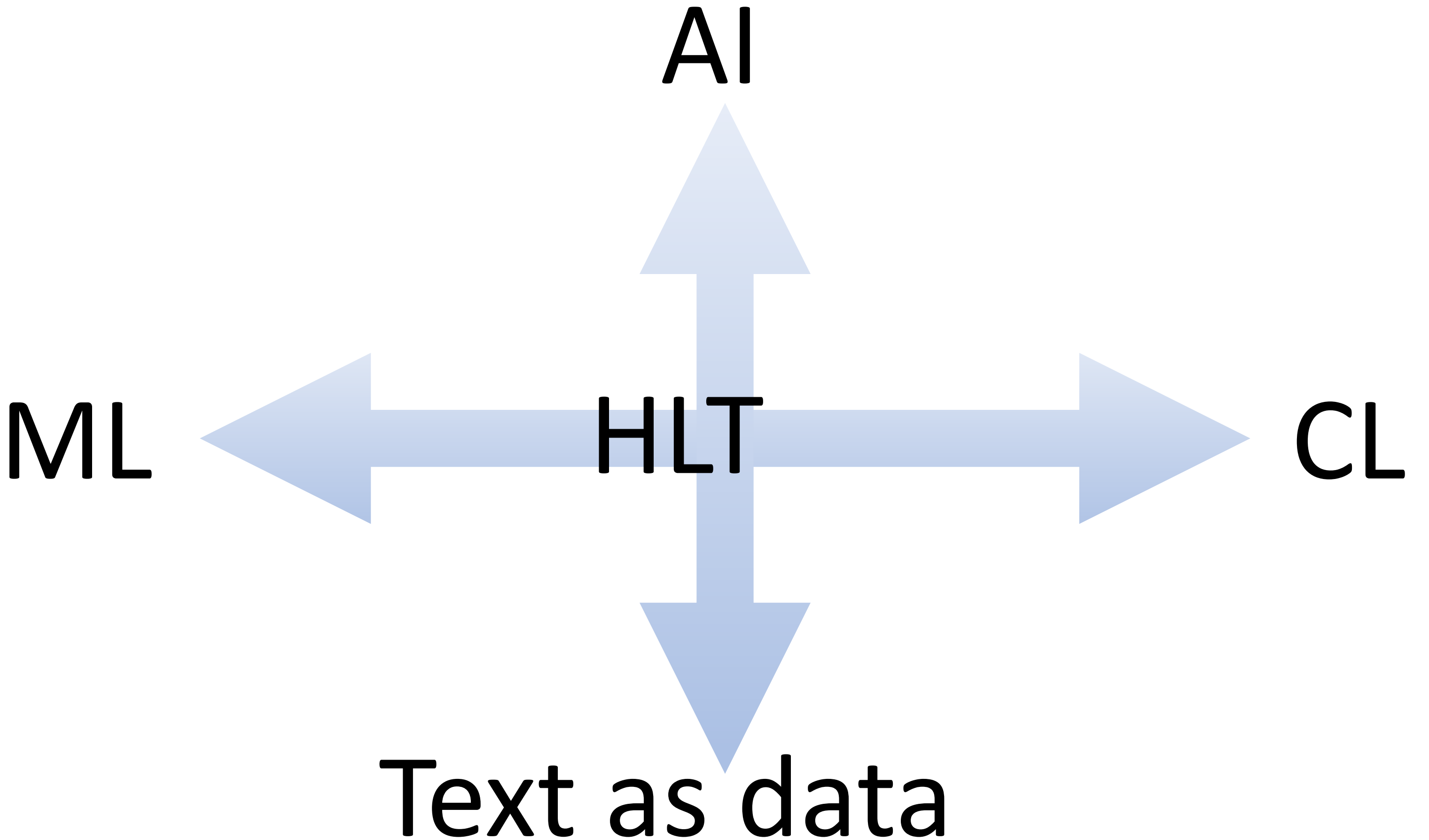
How we do things with words: Analyzing text as social and cultural data

Dong Nguyen[1,2], Maria Liakata[1,3], Simon DeDeo[4], Jacob Eisenstein[5], David Mimno[6], Rebekah Tromble[1,7], and Jane Winters[8]

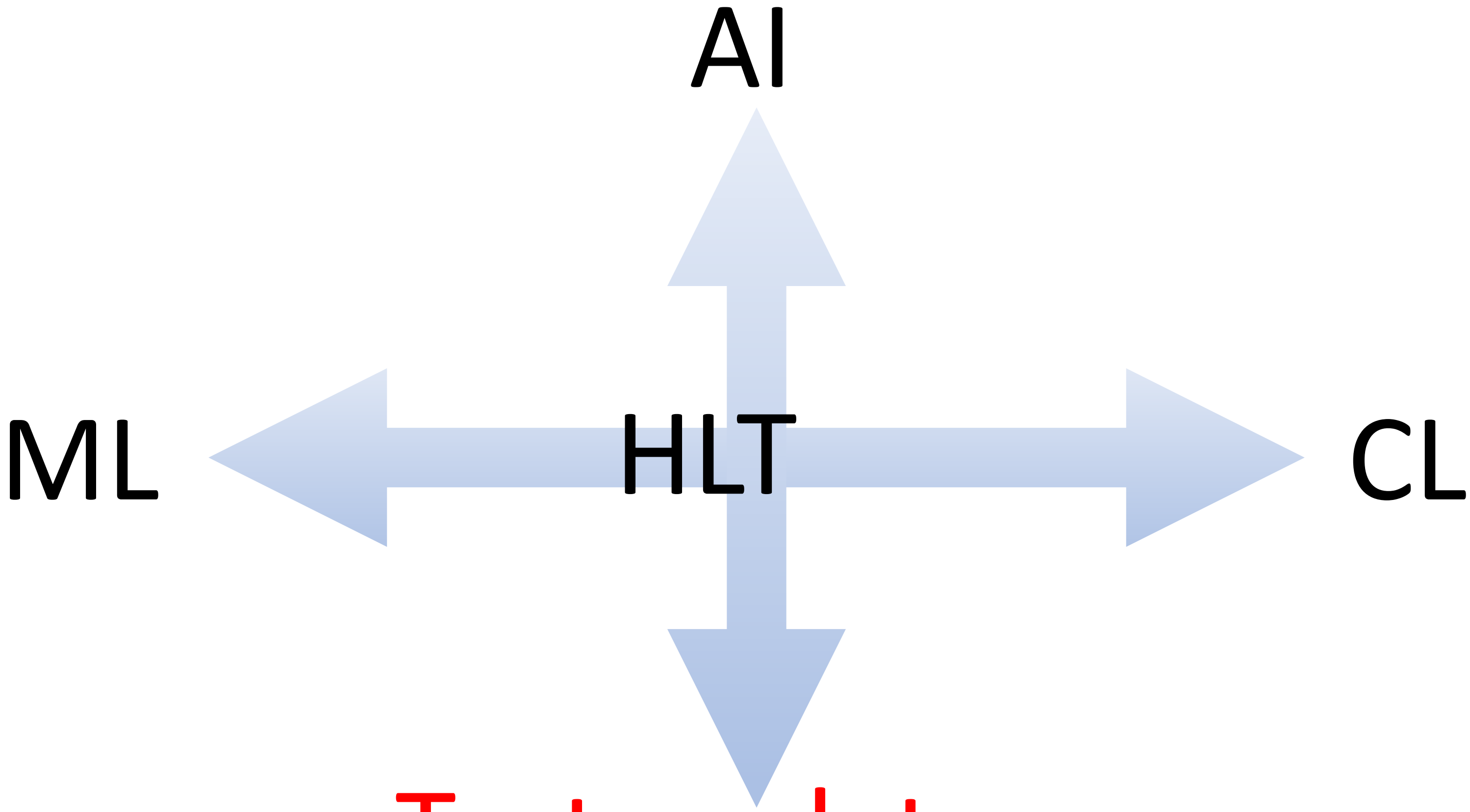
[1]Alan Turing Institute (UK), [2]Utrecht University (NL), [3]University of Warwick (UK), [4]Carnegie Mellon University (USA), [5]Georgia Institute of Technology (USA), [6]Cornell University (USA), [7]Leiden University (NL), [8]University of London (UK)

Text as data

Natural Language Processing



Natural Language Processing



Text as data

What is Computational Text Analysis?

Computational Text Analysis

practice

- *"Data science is the study of extracting value from data"* –

large scale textual

Jeannette Wing

Adam Poliak

Outline

- ML vs CTA vs CL (according to Adam)
- Recap
- Word Representations

Recap so far

The first class was all about counting words

2nd 3rd classes about the power of counting words.

By counting words we can ___ ___

learn about language
generate language
categorize language
represent documents as vectors

4th class: reducing dimensions of count matrices

Why Reduce Dimensions?

Discover hidden correlations/topics

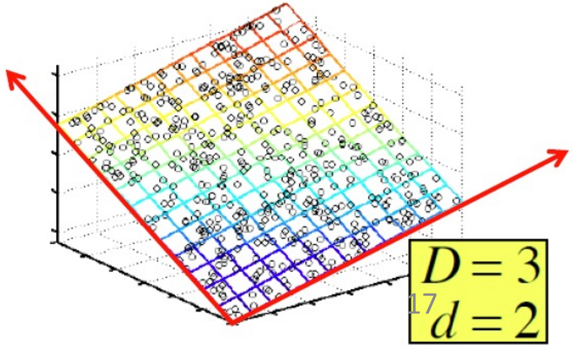
- Words that occur commonly together

Remove redundant and noisy features

- Not all words are useful

Interpretation and visualization

Easier storage and processing of the data



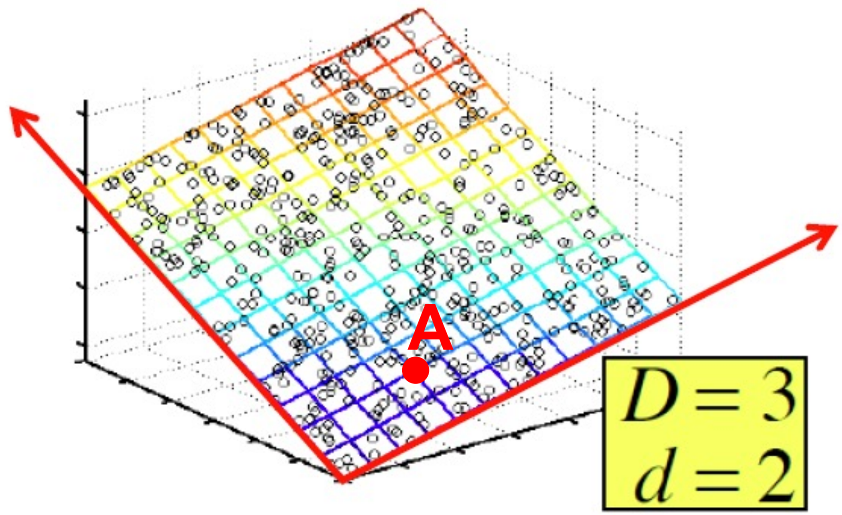
Rank of a Matrix

- **Q:** What is **rank** of a matrix **A**?
- **A:** Number of **linearly independent** columns of **A**
- **For example:**
  - Matrix **A =** $\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ has rank **r=2**
    - **Why?** The first two rows are linearly independent, so the rank is at least 2, but all three rows are linearly dependent (the first is equal to the sum of the second and third) so the rank must be less than 3.
- **Why do we care about low rank?**
  - We can write **A** as two "basis" vectors: [1 2 1] [-2 -3 1]
  - And new coordinates of : [1 0] [0 1] [1 -1]

Rank is "Dimensionality"

- **Cloud of points 3D space:**
  - Think of point positions as a matrix:

1 row per point:
$$\begin{array}{c} \\ \end{array}\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}\begin{array}{l} \textbf{A} \\ \textbf{B} \\ \textbf{C} \end{array}$$



- **We can rewrite coordinates more efficiently!**
  - Old basis vectors: [1 0 0] [0 1 0] [0 0 1]
  - **New basis vectors: [1 2 1] [-2 -3 1]**
  - Then **A** has new coordinates: [1 0]. **B**: [0 1], **C**: [1 1]
    - **Notice: We reduced the number of coordinates!**

SVD - Definition

$$\mathbf{A}_{[m \times n]} = \mathbf{U}_{[m \times r]} \, \Sigma_{[r \times r]} \, (\mathbf{V}_{[n \times r]})^T$$

- **A**: **Input data matrix**
  - *m* x *n* matrix (e.g., *m* documents, *n* terms)
- **U**: **Left singular vectors**
  - *m* x *k* matrix (*m* documents, *r* concepts) **r << n**
- **Σ**: **Singular values**
  - *r* x *r* diagonal matrix (strength of each 'concept')
    (*r* : rank of the matrix **A**)
- **V**: **Right singular vectors**
  - *n* x *r* matrix (*n* terms, *r* concepts)

Outline

- ML vs CTA vs CL
- Recap
- Word Representations
  - One hot
  - Co-occurrence matrix (& SVD)
  - Embeddings

Word Representations

Document-Term Matrix

DMT:

- Rows represent a document
- Columns represent a word
- Values represent some feature of word w_i in document d_j

	w_1	w_2	w_3	w_4	…	…	…	…	w_v
d_1									
d_1									
…									
d_n									10

Document-Term Matrix

We represent each word in our vocabulary as …

an index in our matrix

	w_1	w_2	w_3	w_4	…	…	…	…	w_v
d_1									
d_1									
…									
d_n									

One Hot Vector

- Unique vector for each word

- n-1 elements in vector are 0

- One element in vector is 1

One hot vector example

a pioneer of computer science for work combining statistics and linguistics, and an advocate for women in the field

One hot vector example

a pioneer of computer science for work combining statistics and linguistics, and an advocate for women in the field

a	?	…	?	…	?	…	?
pioneer	?	…	?	…	?	…	?
science	?	…	?	…	?	…	?
…	?	…	?	…	?	…	?
advocate	?	…	?	…	?	…	?

One hot vector example

a pioneer of computer science for work combining statistics and linguistics, and an advocate for women in the field

a	1	…	0	…	0	…	0
pioneer	?	…	?	…	?	…	?
science	?	…	?	…	?	…	?
…	?	…	?	…	?	…	?
advocate	?	…	?	…	?	…	?

One hot vector example

a pioneer of computer science for work combining statistics and linguistics, and an advocate for women in the field

a	1	…	0	…	0	…	0
pioneer	0	…	1	…	0	…	0
science	?	…	?	…	?	…	?
…	?	…	?	…	?	…	?
advocate	?	…	?	…	?	…	?

One hot vector example

a pioneer of computer science for work combining statistics and linguistics, and an advocate for women in the field

a	1	…	0	…	0	…	0
pioneer	0	…	1	…	0	…	0
science	0	…	0	…	1	…	0
…	?	…	?	…	?	…	?
advocate	?	…	?	…	?	…	?

One hot vector example

a pioneer of computer science for work combining statistics and linguistics, and an advocate for women in the field

a	1	…	0	…	0	…	0
pioneer	0	…	1	…	0	…	0
science	0	…	0	…	1	…	0
…	0	…	0	…	0	…	1
advocate	0	…	0	…	0	…	1

One hot vector example

a pioneer of computer science for work combining statistics and linguistics, and an advocate for women in the field

	a	…	pioneer	…	science	…	advocate
a	1	…	0	…	0	…	0
pioneer	0	…	1	…	0	…	0
science	0	…	0	…	1	…	0
…	0	…	0	…	0	…	1
advocate	0	…	0	…	0	…	1

Issues with one-hot vector

- Sparse
  - Lots of 0's
- Very big
  - As big as vocabulary
- Doesn't capture any meaning of the word
  - DTM actually captures some aspects of the documents' meaning
  - We'd like the same for our word representations

How do we figure out the meaning of a new word?

Meaning from Context: Tezguino

A bottle of *tezgüino* is on the table.
Everyone likes *tezgüino*.
Tezgüino makes you drunk.
We make *tezgüino* out of corn.

Lin, ACL 1998; Nida, 1975 p.167

Meaning from Context: Tezguino

A bottle of *tezgüino* is on the table.
Everyone likes *tezgüino*.
Tezgüino makes you drunk.
We make *tezgüino* out of corn.

Lin, ACL 1998; Nida, 1975 p.167

Distributional Hypothesis

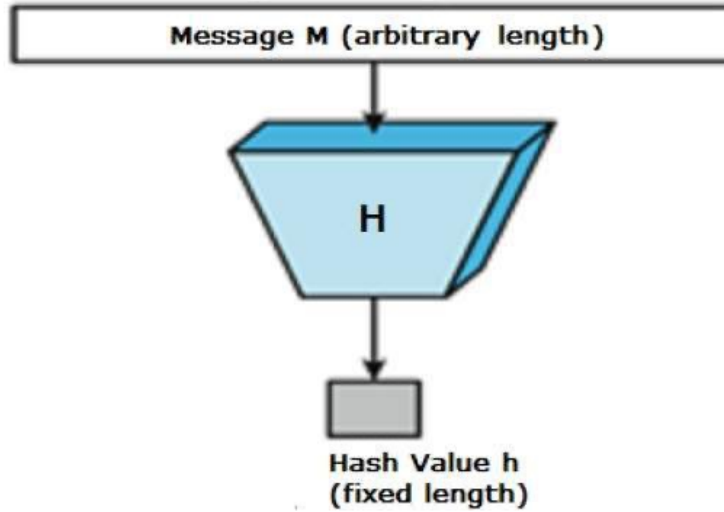
words with similar contexts share similar meanings

(Harris, 1954)

you shall know a word by the company it keeps

(Firth 1957)

Meaning from Context: *Hash*



Meaning from Context: *Hash*

about to get my hands on some top shelf ***hash*** *but I have no idea what the* ***hash*** *price is in my area. There is no one that sells* ***hash*** *in my area actually.*

Co-occurrence matrix

about to get my hands on some top shelf ***hash*** *but I have no idea what the* ***hash*** *price is in my area. There is no one that sells* ***hash*** *in my area actually.*

	on	hands	hash	price	actually	area	my
on							
hands							
hash							
price							
actually							
area							
my							

v x v matrix

Co-occurrence matrix

*about to get my hands on some top shelf **hash** but I have no idea what the **hash** price is in my area. There is no one that sells **hash** in my area actually.*

Window size of 2

	on	hands	hash	price	actually	area	my
on							
hands							
hash							
price							
actually							
area							
my						????	

Co-occurrence matrix

*about to get my hands on some top shelf **hash** but I have no idea what the **hash** price is in my area. There is no one that sells **hash** in my area actually.*

Window size of 2

	on	hands	hash	price	actually	area	my
on							
hands							
hash							
price							
actually							
area							
my						????	

Co-occurrence matrix

*about to get my hands on some top shelf **hash** but I have no idea what the **hash** price is in my area. There is no one that sells **hash** in my area actually.*

Window size of 2

	on	hands	hash	price	actually	area	my
on							
hands							
hash							
price							
actually							
area							
my						2	

Co-occurrence matrix

about to get my hands on some top shelf ***hash*** *but I have no idea what the* ***hash*** *price is in my area. There is no one that sells* ***hash*** *in my area actually.*

Window size of 2

	on	hands	hash	price	actually	area	my
on							
hands							
hash							
price							
actually							
area							2
my						2	

Co-occurrence matrix

about to get my hands on some top shelf ***hash*** *but I have no idea what the* ***hash*** *price is in my area. There is no one that sells* ***hash*** *in my area actually.*

Window size of 2

	on	hands	hash	price	actually	area	my
on							
hands							
hash							
price			????				
actually							
area							2
my						2	

Co-occurrence matrix

about to get my hands on some top shelf ***hash*** *but I have no idea what the* ***hash*** *price is in my area. There is no one that sells* ***hash*** *in my area actually.*

Window size of 2

	on	hands	hash	price	actually	area	my
on							
hands							
hash							
price			1				
actually							
area							2
my						2	

Co-occurrence matrix

about to get my hands on some top shelf ***hash*** *but*
I have no idea what the ***hash*** *price is in my area.*
There is no one that sells ***hash*** *in my area actually.*

Window size of 2

	on	hands	hash	price	actually	area	my
on	0	1	0	0	0	0	0
hands	1	0	0	0	0	0	1
hash	0	0	0	1	0	0	1
price	0	0	1	0	0	0	0
actually	0	0	0	0	0	1	0
area	0	0	0	0	1	0	2
my	0	1	1	0	1	2	0

Issues with co-occurrence matrix

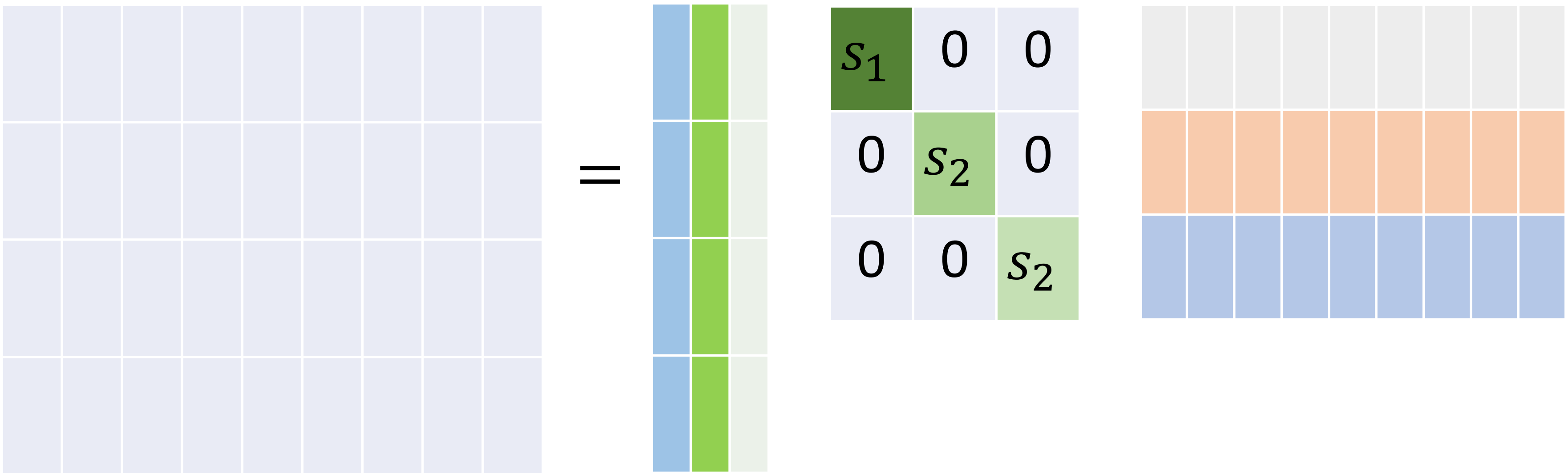
- Large dimensions

- Still sparse
  - Not as much as one-hot but still sparse

- Is meaning captured?

- Solution:
  - Dimensionality Reduction to the rescue

Singular Value Decomposition
Document Term Matrix



M = U S V

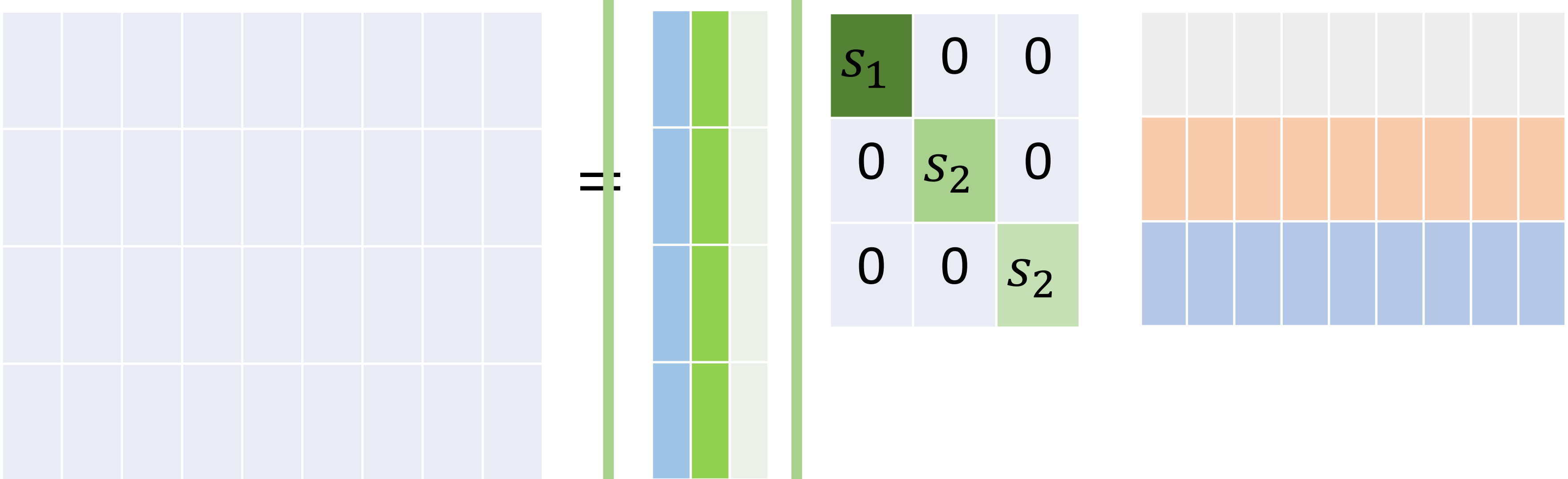
n x v   n x k   k x k   k x v

Singular Value Decomposition
Co-occurrence matrix

M

v x v

Singular Value Decomposition
Co-occurrence matrix



M	=	U	S	V
v x v		v x k	k x k	k x v

Word Embeddings

Initialize random vectors

This is a look-up table where each row indicates the list of numbers for a word

Update word embeddings by reading a corpus

Example

Posted by u/SaltyPositive 1 year ago

0

Ziip Disposable Device

Where are all the ziip device posts at?!I recently bought the ziip refilled disposable device and I'm so so unsure on what to make of it, because there is NO hit, but the cloud is dense upon exhaling, but I don't feel a rush and I'm not sure how hard you have to pull(????) it really doesn't feel like I'm pulling at anything at all. I'm posting here because I bought this pod for 7 cad as a substitute for the Juul ones but don't know if I just got a faulty device? Any other similar experiences?

2 Comments Share Save Hide Report 50% Upvoted

Posted by u/SaltyPositive 1 year ago

0

Ziip Disposable Device

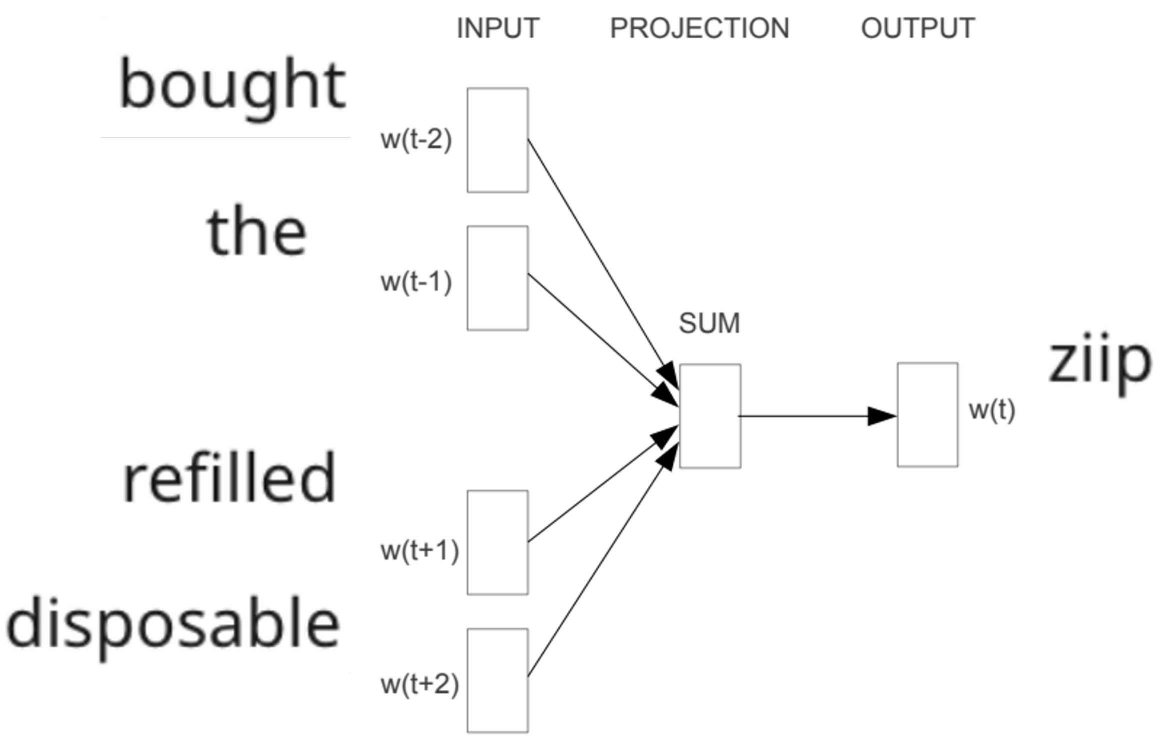
Where are all the ziip device posts at?!I recently bought the ? refilled disposable device and I'm so so unsure on what to make of it, because there is NO hit, but the cloud is dense upon exhaling, but I don't feel a rush and I'm not sure how hard you have to pull(????) it really doesn't feel like I'm pulling at anything at all. I'm posting here because I bought this pod for 7 cad as a substitute for the Juul ones but don't know if I just got a faulty device? Any other similar experiences?

2 Comments Share Save Hide Report 50% Upvoted

Continuous Bag of Words (CBOW)

(Mikolov et al. 2013)

- Predict a word given its context



Posted by u/SaltyPositive 1 year ago

Ziip Disposable Device

Where are all the ziip device posts at?!I recently bought the ziip refilled disposable device and I'm so so unsure on what to make of it, because there is NO hit, but the cloud is dense upon exhaling, but I don't feel a rush and I'm not sure how hard you have to pull(????) it really doesn't feel like I'm pulling at anything at all. I'm posting here because I bought this pod for 7 cad as a substitute for the Juul ones but don't know if I just got a faulty device? Any other similar experiences?

2 Comments Share Save Hide Report 50% Upvoted

0

Posted by u/SaltyPositive 1 year ago

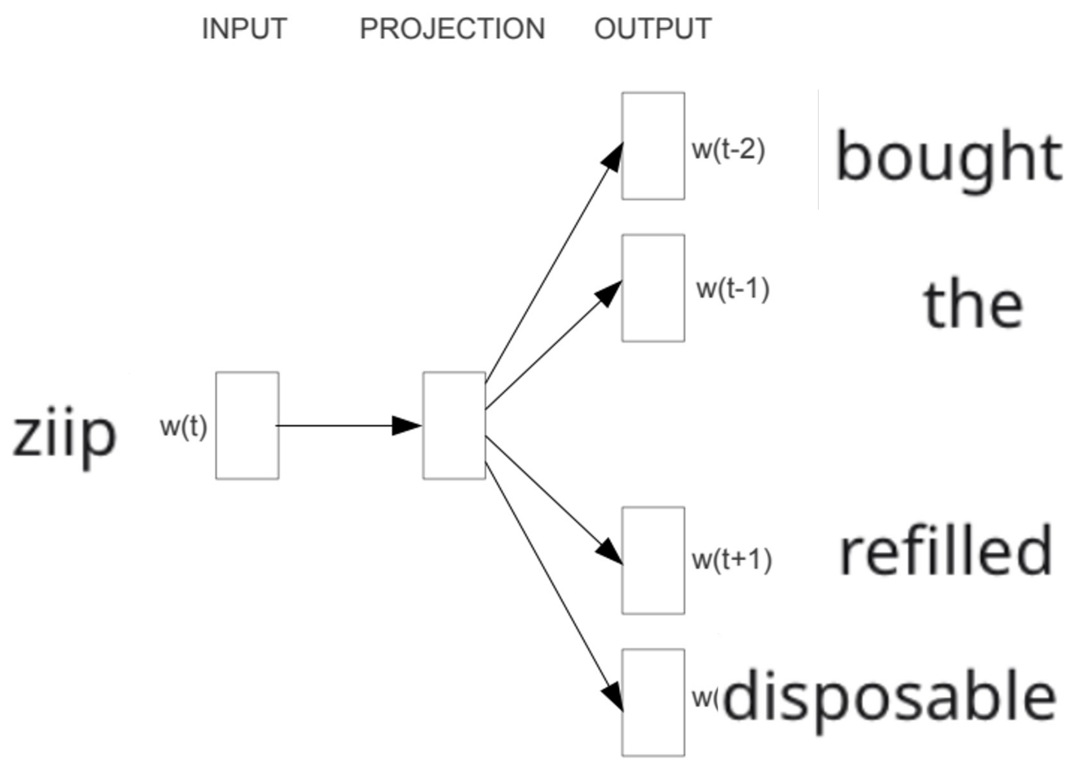
Ziip Disposable Device

Where are all the ziip device posts at?!I recently ? ziip ?
device and I'm so so unsure on what to make of it, because there is NO hit, but the cloud is dense upon exhaling, but I don't feel a rush and I'm not sure how hard you have to pull(????) it really doesn't feel like I'm pulling at anything at all. I'm posting here because I bought this pod for 7 cad as a substitute for the Juul ones but don't know if I just got a faulty device? Any other similar experiences?

2 Comments Share Save Hide Report 50% Upvoted

Skip-Gram

- Predict the context around a word



Updated Word Embeddings as byproduct of training

After training the neural network, we have updated values in our look-up table

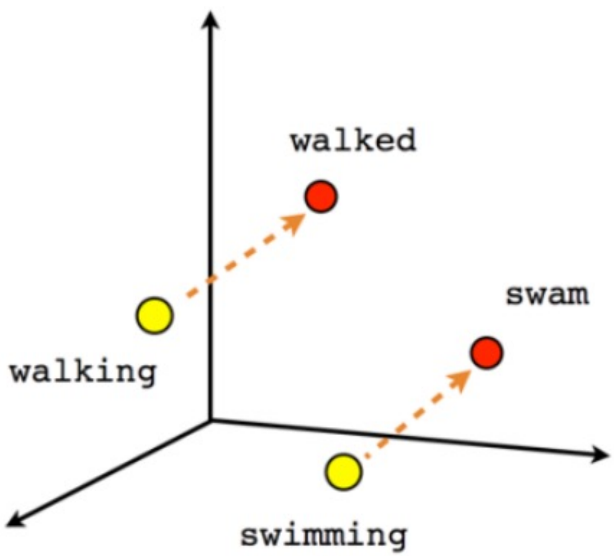
Word Embeddings

a	0.4420	…	0.167	…	0.4838	…	0.2314
pioneer	0.2401	…	0.3732	…	0.9653	…	0.6366
science	0.7532	…	0.3245	…	0.5893	…	0.7772
…	0.2032	…	0.5792	…	0.9302	…	0.4924
advocate	0.3424	…	0.2944	…	0.3923	…	0.3492

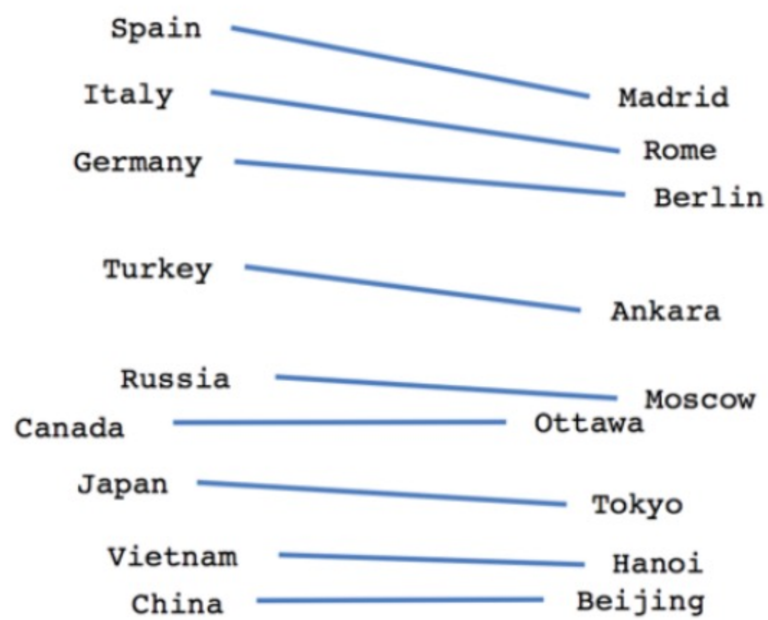
Word2vec: how to learn vectors

- Given the set of positive and negative training instances, and an initial set of embedding vectors
- The goal of learning is to adjust those word vectors such that we:
  - **Maximize** the similarity of the target word, context word pairs (w , c_{pos}) drawn from the positive data
  - **Minimize** the similarity of the (w , c_{neg}) pairs drawn from the negative data.

Word Embeddings Preserve Meaning



Verb tense



Country-Capital

Embeddings as a window onto historical semantics

Train embeddings on different decades of historical text to see meanings shift

~30 million books, 1850-1990, Google Books data

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. Proceedings of ACL.

Embeddings reflect cultural bias!

Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." In *NeurIPS*, pp. 4349-4357. 2016.

- Ask "Paris : France :: Tokyo : x"
  - x = Japan
- Ask "father : doctor :: mother : x"
  - x = nurse
- Ask "man : computer programmer :: woman : x"
  - x = homemaker

Algorithms that use embeddings as part of e.g., hiring searches for programmers, might lead to bias in hiring

Historical embedding as a tool to study cultural biases

Garg, N., Schiebinger, L., Jurafsky, D., and Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. Proceedings of the National Academy of Sciences 115(16), E3635–E3644.

- Compute a **gender or ethnic bias** for each adjective: e.g., how much closer the adjective is to "woman" synonyms than "man" synonyms, or names of particular ethnicities
  - Embeddings for **competence** adjective (*smart, wise, brilliant, resourceful, thoughtful, logical)* are biased toward men, a bias slowly decreasing 1960-1990
  - Embeddings for **dehumanizing** adjectives (barbaric, monstrous, bizarre) were biased toward Asians in the 1930s, bias decreasing over the 20th century.
- These match the results of old surveys done in the 1930s