VL Deep Learning for Natural Language Processing

5. Word Embeddings II

Prof. Dr. Ralf Krestel

AG Information Profiling and Retrieval

ZBW CAU

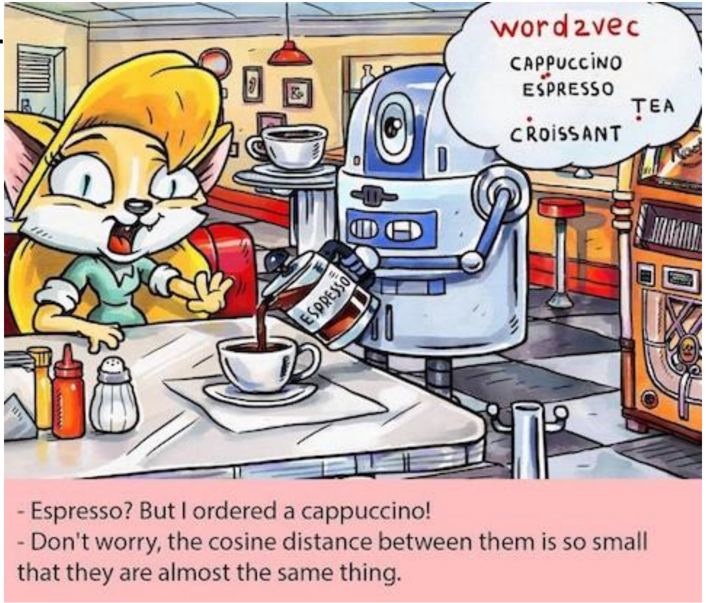
Leibniz Gemeinschaft

Summary 1

- Word embeddings represent words (discrete variables) as vectors
- Reduce the dimensionality
- Similar (semantic) words are closer to each other
  - Cosine similarity

- Word2vec is an algorithm based on neural networks to compute word embeddings
  - Learning objective: predict center word given outside words (or the other way around)

- Why does this work?



https://miro.medium.com/max/480/1*HmkxRdUcK1xZ9hQXmACg4w.jpeg

Summary 2

- Idea: The meaning of a word is defined by the words that appear often in its vicinity. -> Distributed Semantics
  - *“You shall know a word by the company it keeps”* (Firth 1957)

J.R.Firth

- Example:

Sentence 1: Die Kinder spielen _____ auf der Straße.
Sentence 2: Das _____spiel endete 2 zu 1.
Sentence 3: Die schnellste Geschwindigkeit mit der je ein _____ gekickt wurde ist 211 km/h.
Sentence 4: _______ ist ein Mannschaftssport mit zwei Teams mit jeweils 11 Spielern.
Sentence 5: Hol’ dir die neusten _______ergebnisse, Paarungen und Videohöhepunkte aufs Handy.

John Rupert Firth (1957). "A synopsis of linguistic theory 1930-1955." In *Special Volume of the Philological Society.* Oxford: Oxford University Press.

Summary 3

Sentence 1: Die Kinder spielen _____ auf der Straße.
Sentence 2: Das _____spiel endete 2 zu 1.
Sentence 3: Die schnellste Geschwindigkeit mit der je ein _____ gekickt wurde ist 211 km/h.
Sentence 4: _______ ist ein Mannschaftssport mit zwei Teams mit jeweils 11 Spielern.
Sentence 5: Hol' dir die neusten _______ergebnisse, Paarungen und Videohöhepunkte aufs Handy.

Sentence	1	2	3	4	5
Fußball	1	1	1	1	1
Handball	1	0	1	0	1
Schach	0	1	0	0	1
Verstecken	1	0	0	0	1

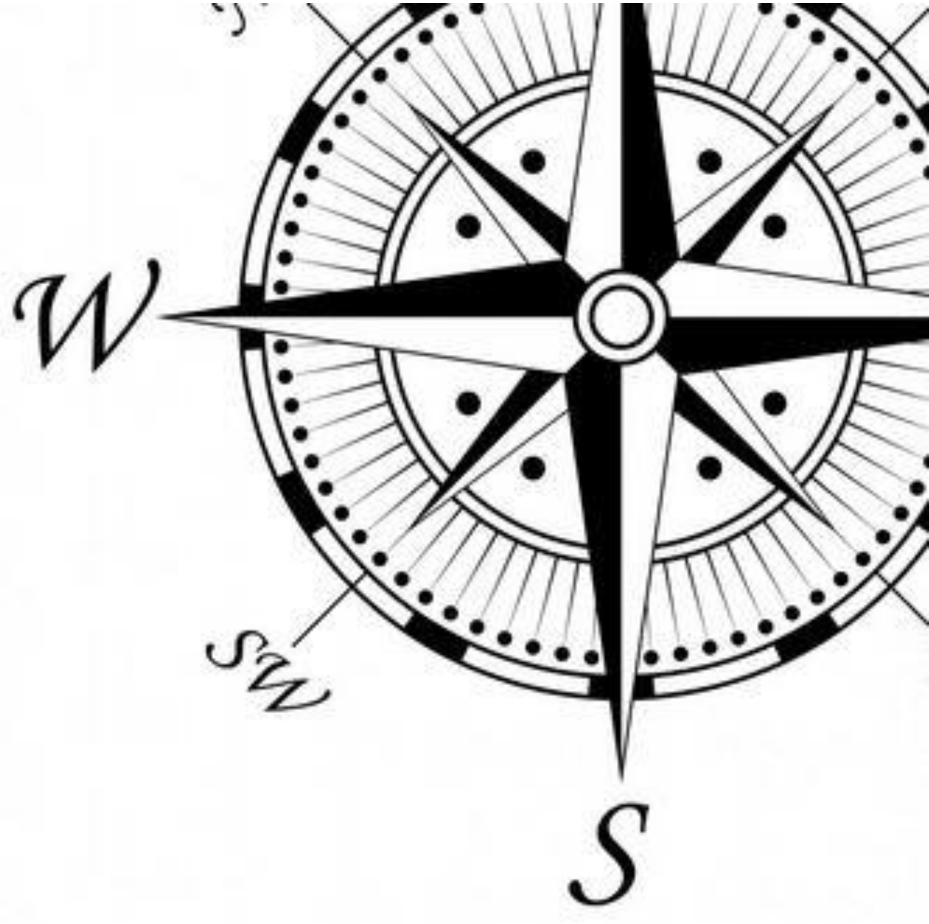
https://www.italybyevents.com/en/events/lombardia/hide-and-seek-world-championship-italy/

Learning Goals for this Chapter

- Know different methods to evaluate word embeddings
  - Pros and cons of the methods
- Be able to name limitations of the evaluations
- Be able to explain the diffences between embedding models
  - Character level
  - Subword level
  - Word level
- Understand document/paragraph/sentence embeddings
  - Word movers distance
  - Doc2vec
- Relevant chapters
  - P6.1
  - S2 (2021) https://www.youtube.com/watch?v=gqaHkPEZAew

Topics Today

1. **Evaluation of Word Vectors**
2. Sub-Word Models
3. Document Embeddings



What is a Good Word Vector?

- How to evaluate NLP systems in general?

Quantitatively

  - **Intrinsically**
    - Based on a small, well-defined specific task
    - Useful to understand components
    - Gain in performance only useful if a connection to a real task exists
    - Fast to compute
  - **Extrinsically**
    - Based on a concrete, „real“, complex task
    - The whole system is evaluated, for NLP: the complete processing pipeline
    - Hard to tell which components of the system are performing well
    - Might take a while (at least longer than individual components)
    - Ablation test: exchanging/improving one particular component improves system
      $\longrightarrow$ The new component **is better** than the old one!

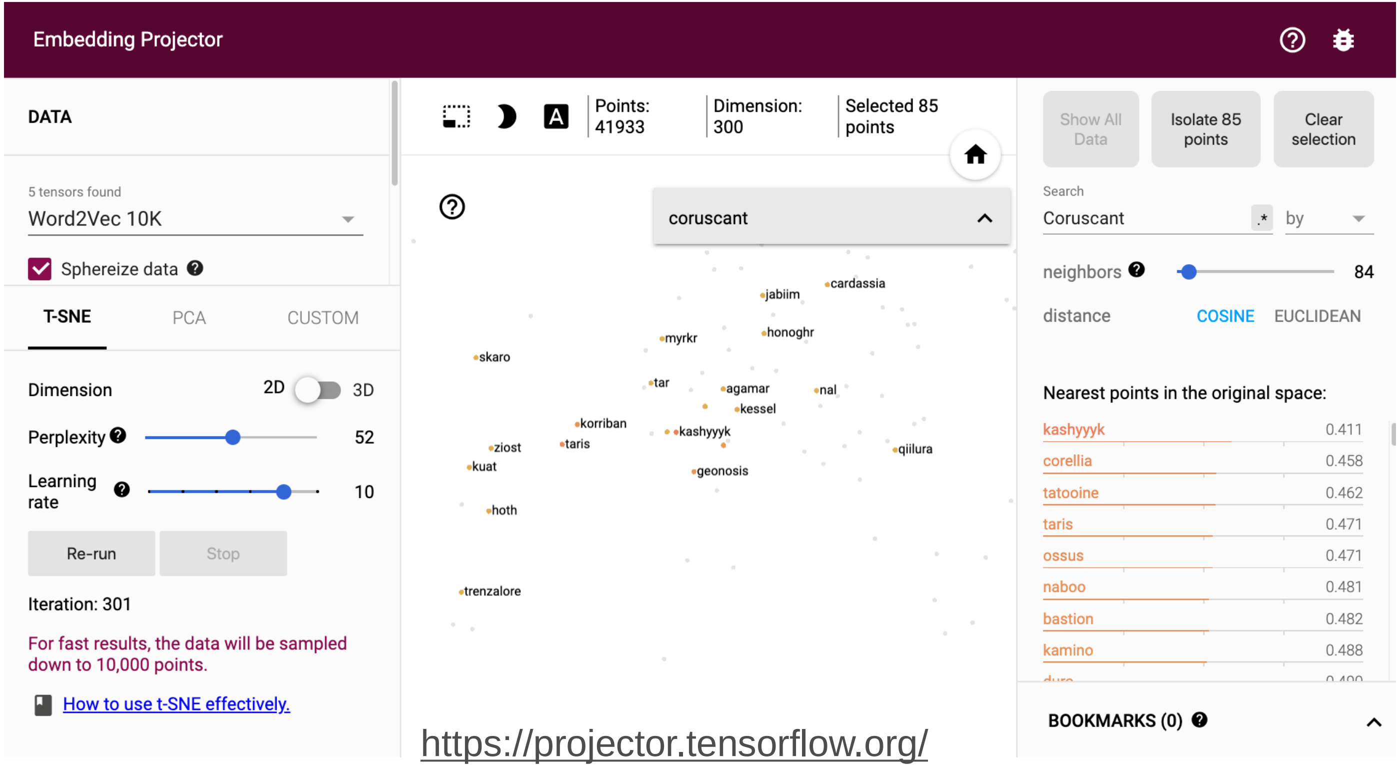
Evaluation Methods for Word Embeddings

- Intrinsically
  - Word analogy task
  - Correlation with human assessment
- Extrinsically
  - All kinds of downstream tasks
    - Classification of documents
    - Classification of words
    - Clustering of documents/words
    - …and many more

- Qualitatively (anecdotal evidence)
  - Nearest neighbors

In contrast: quantitatively (empirical evidence)

Interactive Demo



https://projector.tensorflow.org/

Links to Word Embedding Visualizations

- https://projector.tensorflow.org/?config=https://gist.githubusercontent.com/julian-risch/9d6d125b7b5e49eb9b1ffacfd6de922a/raw/22b39854e2f5acb31dac1d586871d3fee7a4d0ca/1-10k-projector_config.json
- https://projector.tensorflow.org/?config=https://gist.githubusercontent.com/julian-risch/0e4bc9ac0d5fdae61639faae8eddf23e/raw/640cf0266eee2d468b0294f8616d98da183b9881/2-projector_config.json
- https://projector.tensorflow.org/?config=https://gist.githubusercontent.com/julian-risch/15ceed73ed4beb0e84849804bd08f1d2/raw/efa486f76b49207f3111dfa318f9ea2fc7ae111e/3-projector_config.json

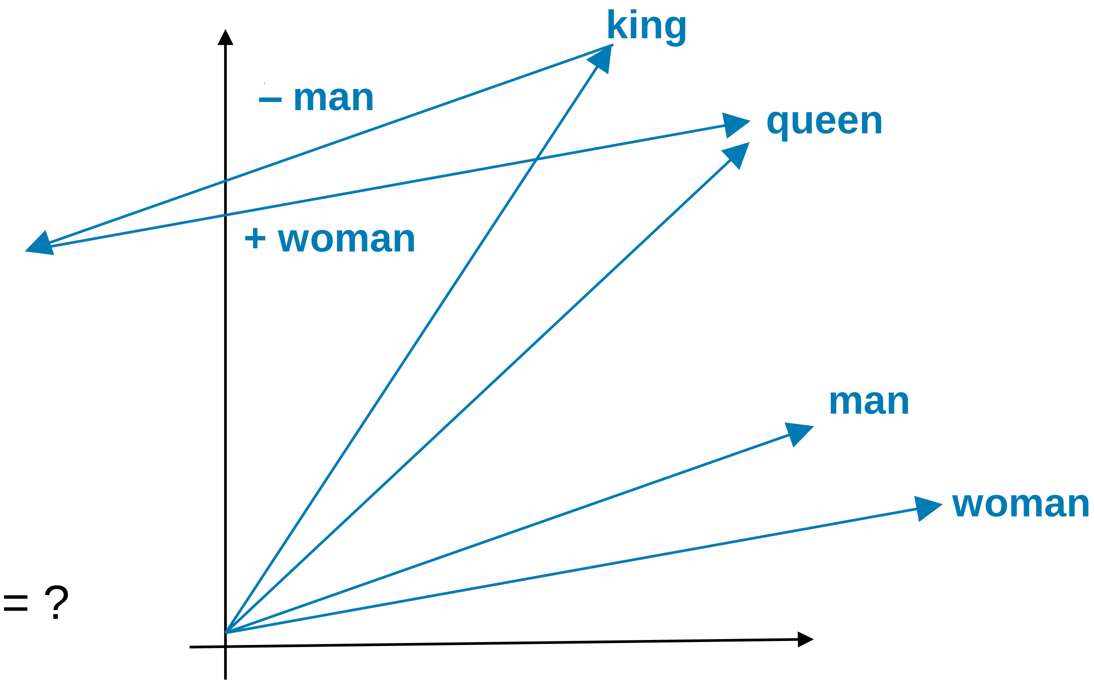
Arithmetics in Vector Space

- Word embeddings model semantic similarity or relatedness

$$House = \begin{bmatrix} 0.44 \\ 0.24 \\ 0.11 \\ 0.22 \\ 0.13 \\ 0.45 \\ 0.16 \\ 0.39 \\ 0.15 \end{bmatrix} \qquad Home = \begin{bmatrix} 0.46 \\ 0.22 \\ 0.16 \\ 0.12 \\ 0.41 \\ 0.39 \\ 0.18 \\ 0.38 \\ 0.10 \end{bmatrix}$$

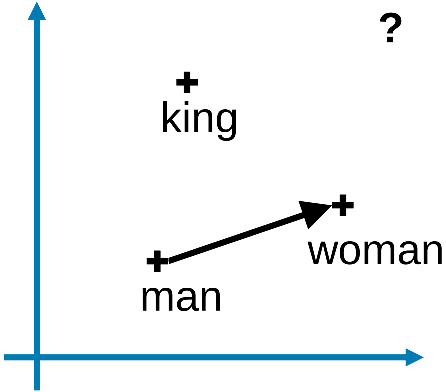
- But they can do more…

King – man + woman = ?



Word Analogy Task

- Word vector analogies:
  - $a : b \;::\; c : ?$
    - man:woman :: king:?
  - $d = arg\max_i \frac{(x_b - x_a + x_c)^T x_i}{\|x_b - x_a + x_c\| \|x_i\|}$



Word2Vec -> Nerd2Vec

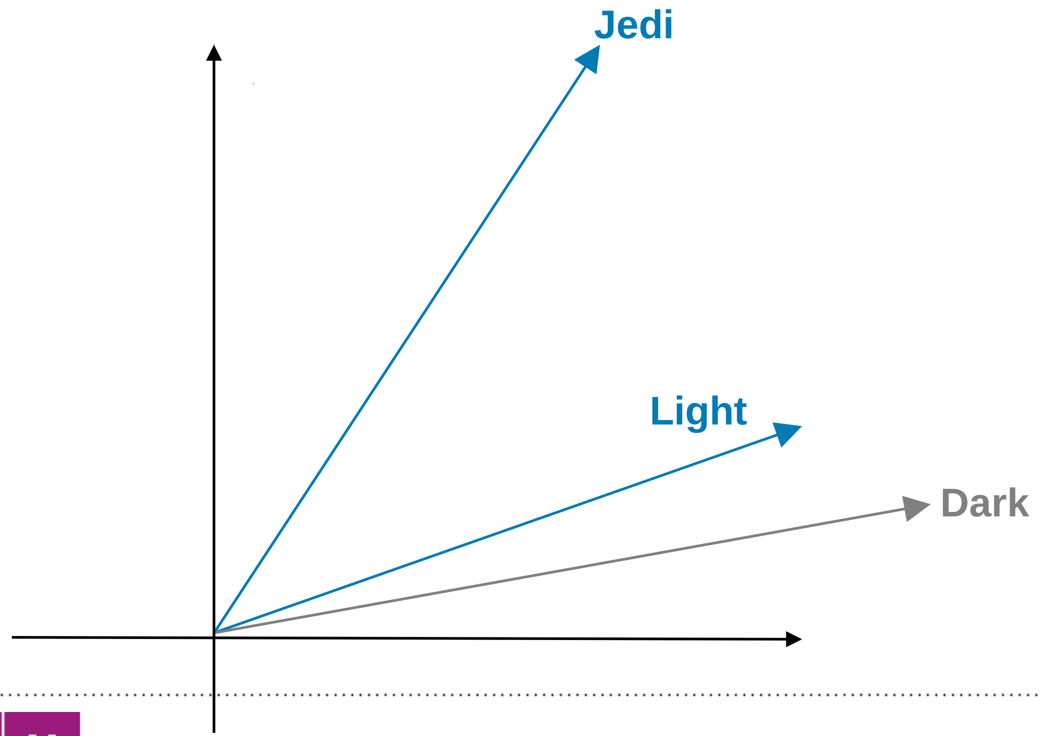
- Word2Vec
  - Assign a real-valued vector representation to each word
  - Learn the vectors on large corpora
  - Words that appear in similar context shall have similar vectors

- Nerd2Vec
  - Based on Wookieepedia, a Star Wars Wiki
  - Captures semantic similarities of fictional characters, locations, e

https://blogs.oracle.com/irml/nerd2vec:-jointly-embedding-star-trek,-star-wars-and-doctor-who-wikias

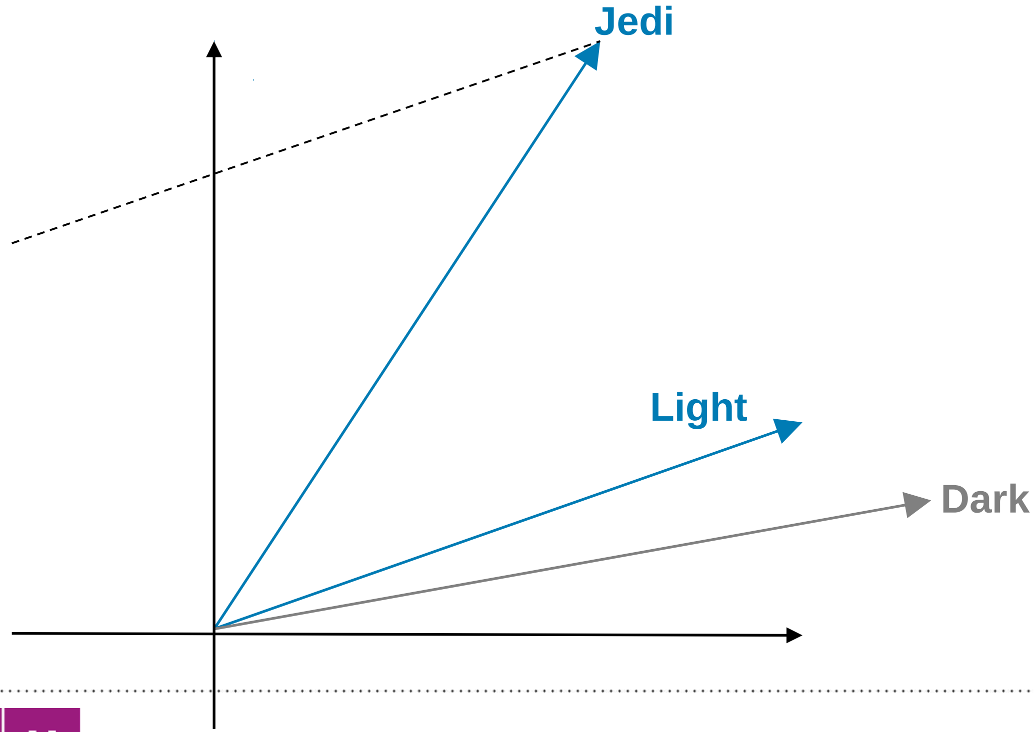
Arithmetics in Vector Space

- Jedi – Light + Dark = ?



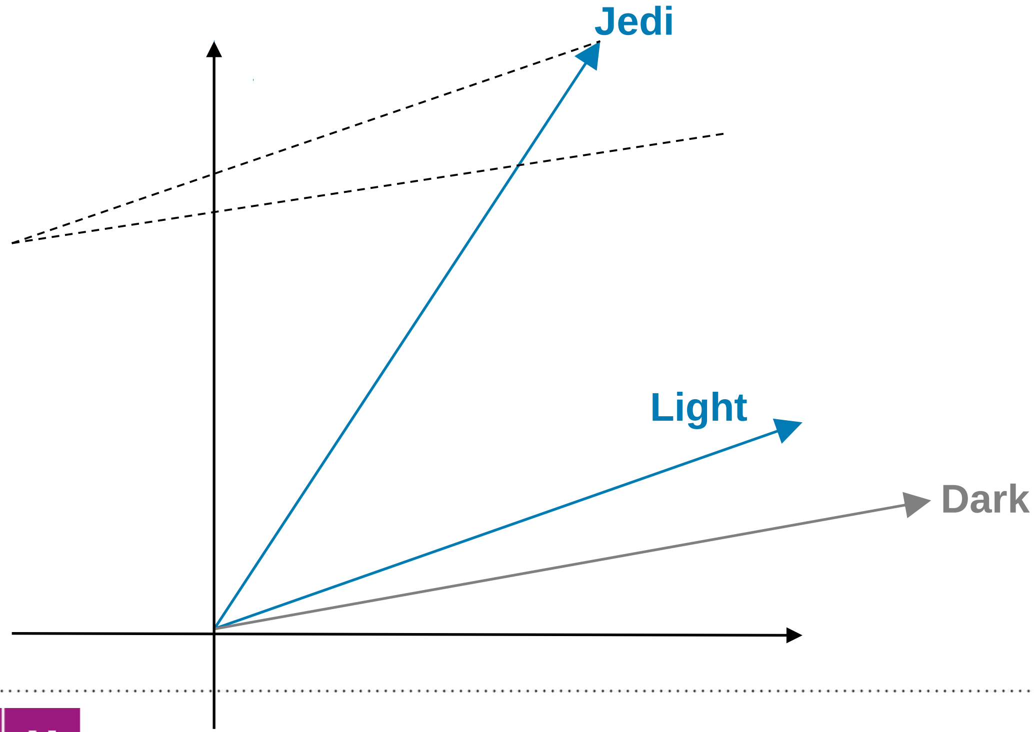
Arithmetics in Vector Space

- Jedi – Light + Dark = ?



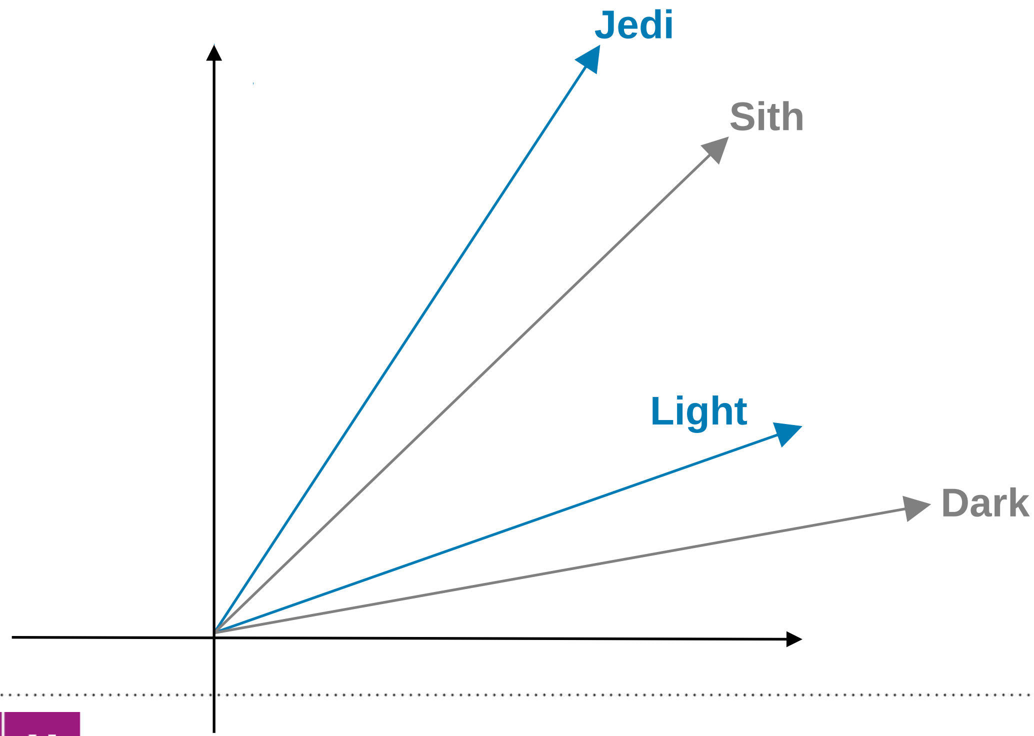
Arithmetics in Vector Space

- Jedi – Light + Dark = ?



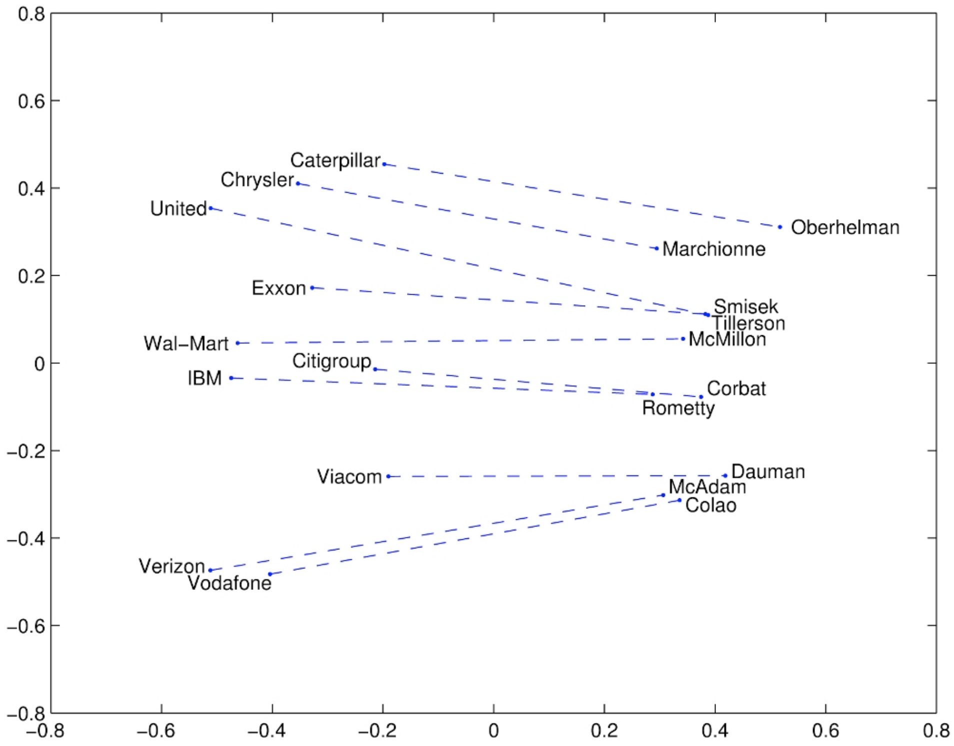
Arithmetics in Vector Space

- Jedi – Light + Dark = Sith



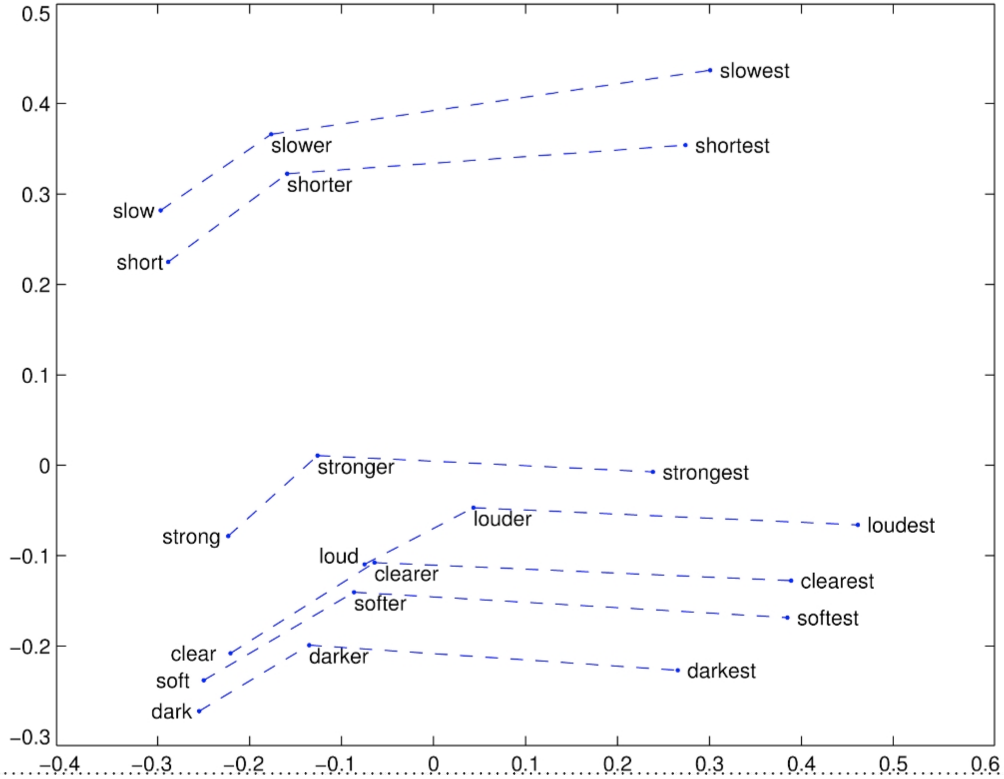
Word Analogies: Example GloVe I

- Company - CEO



Word Analogies: Example GloVe II

- Superlatives



Word Analogies: Example Word2Vec

Vektorausdruck					Nächstgelegenes Wort
Paris	-	Frankreich	+	Italien	
Größer	-	Groß	+	Kalt	
Sushi	-	Japan	+	Deutschland	
Cu	-	Kupfer	+	Gold	
Windows	-	Microsoft	+	Google	
Montreal Canadiens	-	Montreal	+	Toronto	

Word Analogies: Gold Standard Dataset I

- Semantic examples
- city-in-state
  - Chicago Illinois Houston Texas
  - Chicago Illinois Philadelphia Pennsylvania
  - Chicago Illinois Phoenix Arizona
  - Chicago Illinois Dallas Texas
  - Chicago Illinois Jacksonville Florida
  - Chicago Illinois Indianapolis Indiana
  - Chicago Illinois Austin Texas
  - Chicago Illinois Detroit Michigan
  - Chicago Illinois Memphis Tennessee
  - Chicago Illinois Boston Massachusetts

Problem: Many cities have the same name

https://code.google.com/archive/p/word2vec/source/default/source/word2vec/trunk/questions-words.txt

Word Analogies: Gold Standard Dataset II

- Semantic Examples
- capital-country
  - Abuja Nigeria Accra Ghana
  - Abuja Nigeria Algiers Algeria
  - Abuja Nigeria Amman Jordan
  - Abuja Nigeria Ankara Turkey
  - Abuja Nigeria Antananarivo Madagascar
  - Abuja Nigeria Apia Samoa
  - Abuja Nigeria Ashgabat Turkmenistan
  - Abuja Nigeria Asmara Eritrea
  - Abuja Nigeria Astana Kazakhstan

Problem: Facts can change

https://code.google.com/archive/p/word2vec/source/default/source/word2vec/trunk/questions-words.txt

Word Analogies: Gold Standard Dataset III

- Syntactic examples
- gram4-superlative
  - bad worst big biggest
  - bad worst bright brightest
  - bad worst cold coldest
  - bad worst cool coolest
  - bad worst dark darkest
  - bad worst easy easiest
  - bad worst fast fastest
  - bad worst good best
  - bad worst great greatest

https://code.google.com/archive/p/word2vec/source/default/source/word2vec/trunk/questions-words.txt

Correlation

- Word vector distances and their correlation to human assessment
- Dataset: WordSim353
  - http://www.cs.technion.ac.il/~gabr/resources/data/wordsim353/

Word 1	Word 2	Human Score
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

Word	Cosine Distance to „Sweden“
Norway	0.76
Denmark	0.71
Finland	0.62
Switzerland	0.59
Belgium	0.58
Netherlands	0.57
Iceland	0.56
Estonia	0.55

What about Ambigous Words?

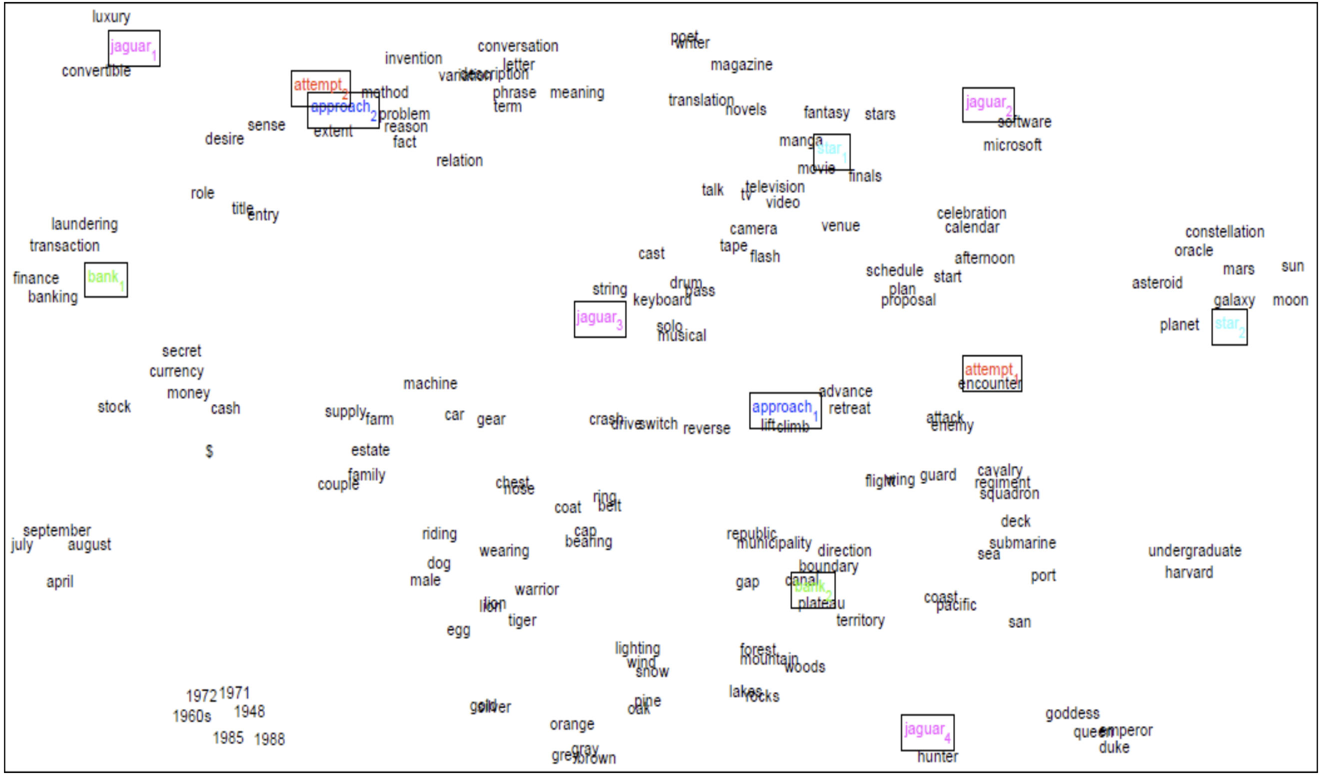
- One word = one vector
  - *to run* vs. *the run*
  - *jaguar (cat) vs. jaguar (car)*

- Idea:
  - Clustering of word windows
  - Word will be assigned to appropriate cluster
    - $jaguar_1, jaguar_2, \ldots$

Huang, E. H., Socher, R., Manning, C. D., & Ng, A. Y. (2012, July). Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1* (pp. 873-882). Association for Computational Linguistics.

https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcQpSr-u-en0Bnsal_G9MghQP577_ukz4LADkuKPlCsDPzqVIT_1
https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcTuUpkl4OEJWAH2sQdC2jEqRQn7mDbC1XS09XtRqthM0zetY_iq

Vectors Learned from Local and Global Context



Leibniz Gemeinschaft

Extrinsically: NER

- Named Entity Recognition (NER)
  - Word classification task
    - Denotes a word a person, organisation or location?

  - Better word vectors
    = better representation of input words
    = better features
    = higher accuracy for classification task

Model	Dev	Test	ACE	MUC7
Discrete	91.0	85.4	77.4	73.4
SVD	90.8	85.7	77.3	73.7
SVD-S	91.0	85.5	77.6	74.3
SVD-L	90.5	84.8	73.6	71.5
HPCA	92.6	**88.7**	81.7	80.7
HSMN	90.5	85.7	78.7	74.7
CW	92.2	87.4	81.7	80.2
CBOW	93.1	88.2	82.2	81.1
GloVe	**93.2**	88.3	**82.9**	**82.2**

Bias in Word Embeddings

- Gender stereotypes

Occupations as projected on to the she−he gender direction on w2vNEWS

Extreme *she*	Extreme *he*
1. homemaker	1. maestro
2. nurse	2. skipper
3. receptionist	3. protege
4. librarian	4. philosopher
5. socialite	5. captain
6. hairdresser	6. architect
7. nanny	7. financier
8. bookkeeper	8. warrior
9. stylist	9. broadcaster
10. housekeeper	10. magician

Automatically generated analogies for the pair she-he

Gender stereotype *she*-*he* analogies

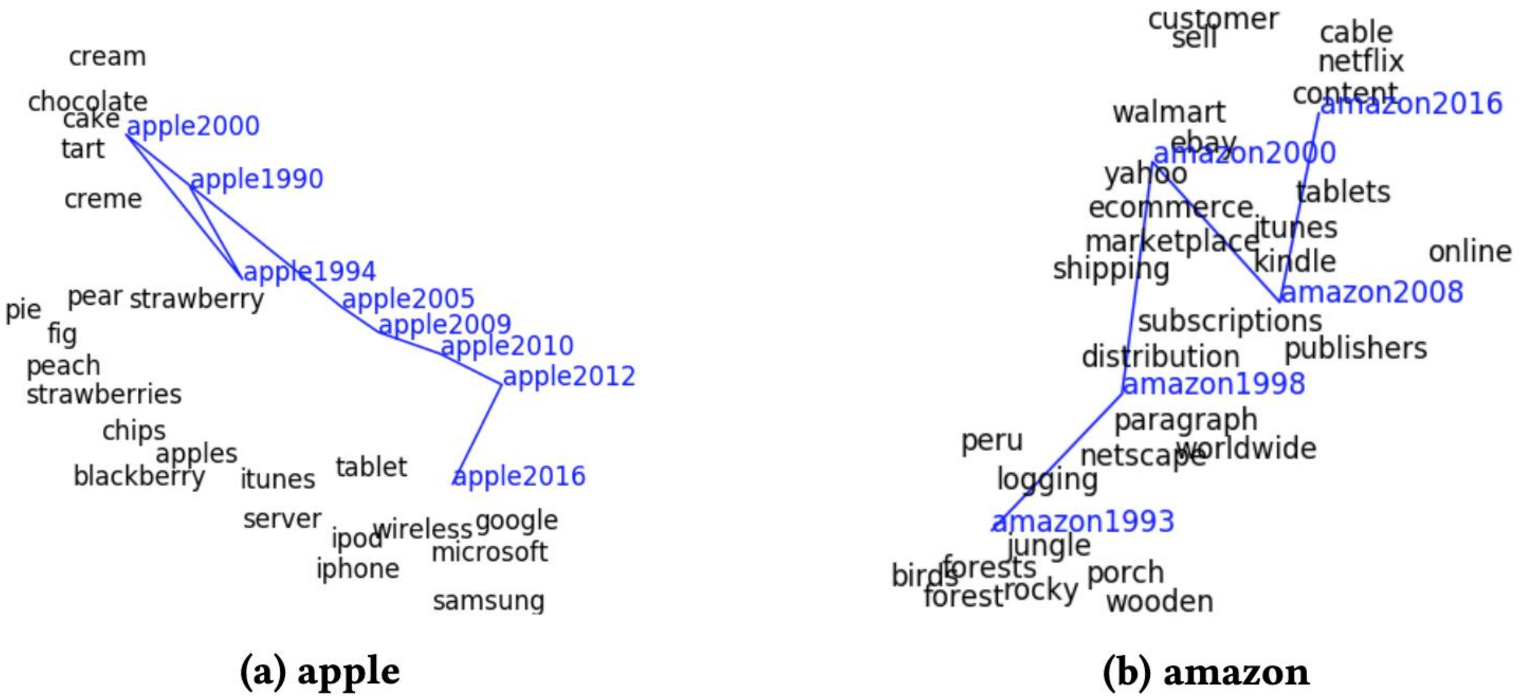
sewing-carpentry	registered nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	lovely-brilliant

Gender appropriate *she*-*he* analogies

queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

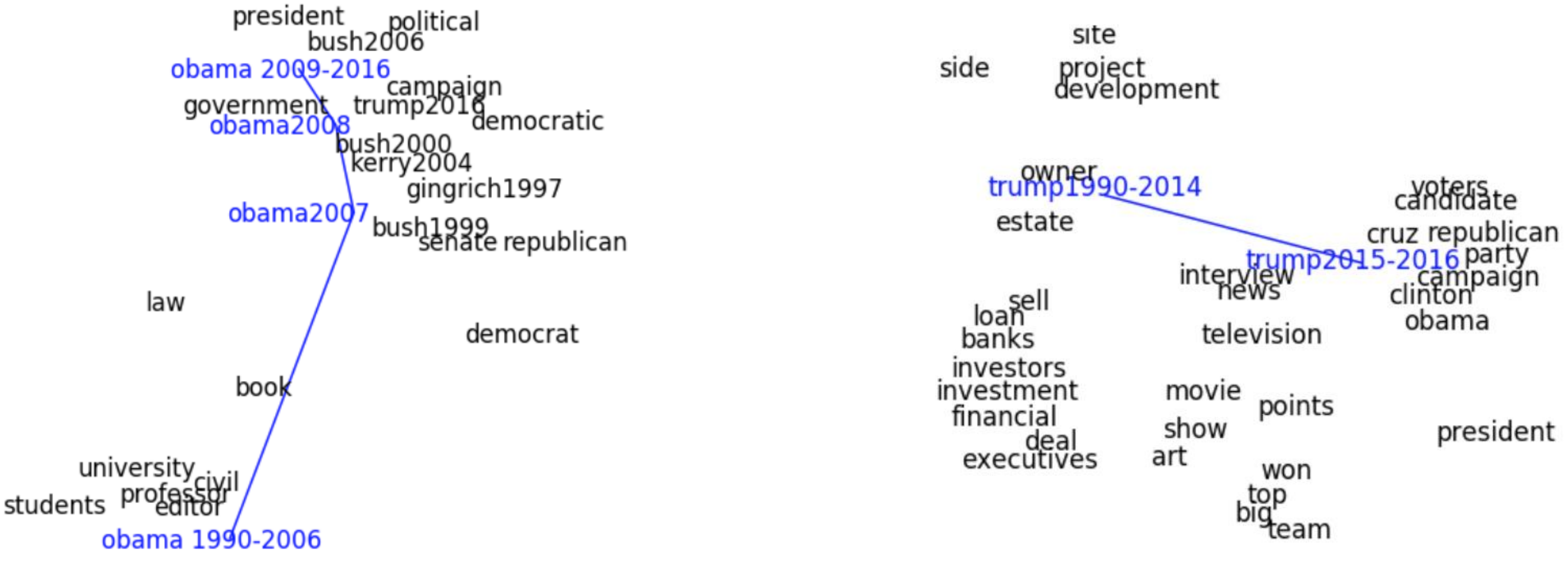
Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in neural information processing systems* (NIPS), 4349-4357.

Change Analysis



Yao, Z., Sun, Y., Ding, W., Rao, N., & Xiong, H. (2018, February). Dynamic word embeddings for evolving semantic discovery. In *Proceedings of the eleventh international conference on web search and data mining* (WSDM) (pp. 673-681).

Change Analysis



(c) obama

(d) trump

Yao, Z., Sun, Y., Ding, W., Rao, N., & Xiong, H. (2018, February). Dynamic word embeddings for evolving semantic discovery. In *Proceedings of the eleventh international conference on web search and data mining* (WSDM) (pp. 673-681).

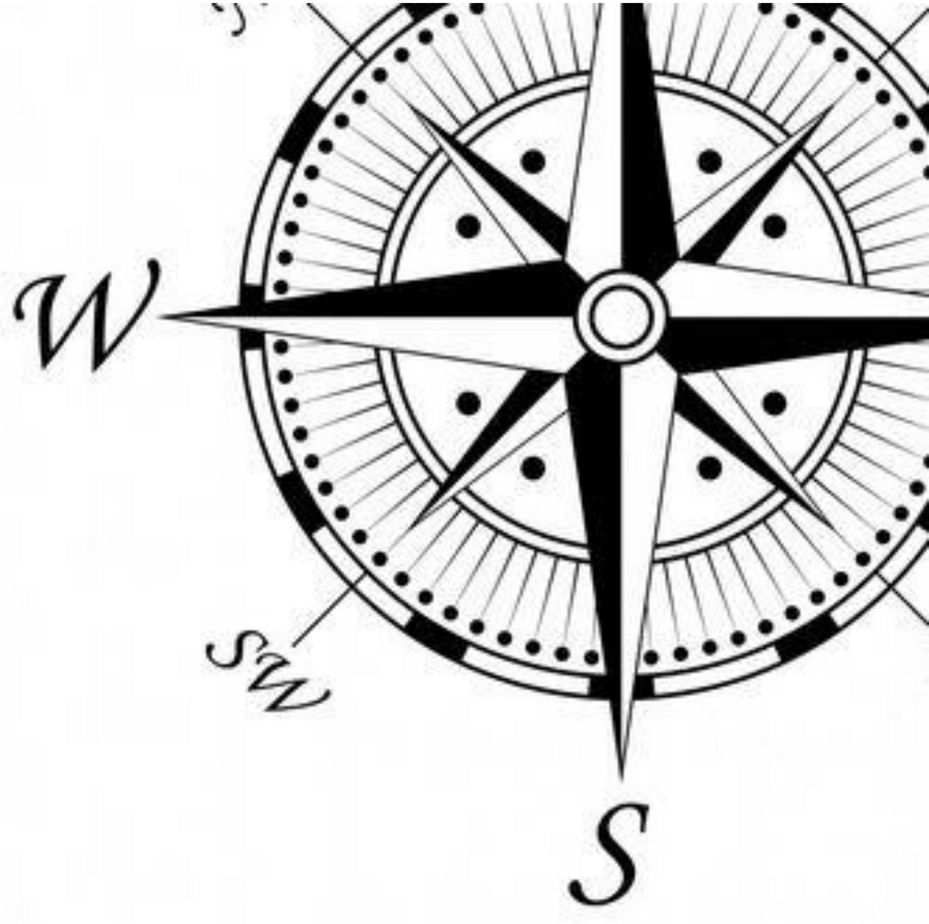
Evaluation of Word Vectors

- How could you generate a gold standard word analogy dataset automatically?

Start 5 4 3 2 1 End

Topics Today

1. Evaluation of Word Vectors
2. **Sub-Word Models**
3. Document Embeddings



What is a „Word“

- Representation of a word different in different languages
- Word boundries marked by space character (or not)
  - 我想吃一个汉堡
  - I want to eat a hamburger
- Clitics, pronouns, agreement?
  - Separated
    - **Je vous aime** **Il y a beaucoup d'argent**
  - Joined
    - فقلناها = ف+ قلنا +ها = so+said+we+it
- Composita
  - Separated
    - life insurance company employee
  - Joined
    - Lebensversicherungsgesellschaftsangestellter

Why is a Model on Word Level not Enough?

- Unknown word (**out-of-vocabulary**, OOV)
  - Spelling mistakes:
  - Rich morphology: Composita
  - Informal spelling: Veeeeery Sloooow
  - Transliterations: Christopher ↦Kryštof

Rare words

- Differences in writing systems (Where is the meaning?)

– Phonemic	jiyawu ngabulu	Wambaya
– Fossilized phonemic	thorough failure	English
– Syllabic/moraic	ᑐᖕᒍᓴᐊᖕᒐᓴᖅ	Inuktitut
– Ideographic	去年太空船二号坠毁	Chinese
– Combination (syllabic+ideographic)	インド洋の島	Japanese

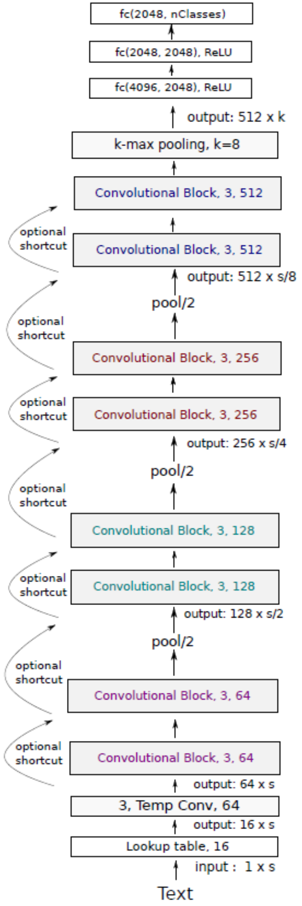
Character Level Models

1. Word embeddings can be composed from character embeddings
   - Enables embeddings for unknown words
   - Similar spelling -> similar embedding
   - Solves the OOV problem
   - Combination of character and word level
2. Written language as a sequence of individual characters
   - No explicit representations for words
   - **Pure character-based models**

- Both methods are very successful
  - Surprising, since traditionally, phonemes/characters are not semantic units, but, DL models group them (morphemes)
    - Unfortunately = $\left[\left[un\ [[fortun(e)]_{ROOT}\ ate]_{STEM}\right]_{STEM}\ ly\right]_{WORD}$

Pure Character Level Models

- Pure character level model based on CNNs
- Idea:
  - Learn hierarchical representations of sentences
  - Task-based (text classification)
  - Input are characters
  - Deeper layers form syllables, words, phrases, sentences
  - Up to 29 layers



Conneau, Schwenk, Lecun, Barrault. *Very deep convolutional networks for text classification*. In EACL 2017

Sub-Word Embeddings

- Instead of words, also other units could be considered
  - Sentences bear meaning
    - *John likes strawberries*
  - Documents bear meaning
  - …

  - Syllables (subwords) bear meaning
    - In particular in morphological rich languages
    - *expensive* vs. *inexpensive*

Out-of-Vocabulary (OOV) words

- Advantage of subword embeddings
  - Words can be embedded (represented as vectors) even if they do not occur very frequently (or not at all) in the training data

Byte Pair Encoding (BPE)

- Originally a compression algorithm:
  - Most frequent byte pair$\mapsto$a new byte
  - Replace bytes with character n-grams

- A word segmentation algorithm:
  - Though done as bottom up clustering
  - Start with a unigram vocabulary of all (Unicode) characters in data
  - Most frequent n-gram-pairs $\mapsto$a new n-gram

- Implementierung
  - https://github.com/rsennrich/subword-nmt

Sennrich, Rico, Barry Haddow, and Alexandra Birch. "Neural Machine Translation of Rare Words with Subword Units." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016.

Byte Pair Encoding

- *Dictionary (# occurences, word)*
  5 l o w
  2 l o w e r
  6 n e w e s t
  3 w i d e s t

- Start with all characters in vocabulary
- Add a pair (e,s) with frequency 9
- Add a pair (es,t) with frequency 9
- Add a pair (w,e) with frequency 8
- Add a pair (l,o) with frequency 7

- Vocabulary
  - l, o, w, e, r, n, w, s, t, i, d

- New encodings:
  - l, o, w, e, r, n, w, s, t, i, d
  - l, o, w, e, r, n, w, s, t, i, d, es
  - l, o, w, e, r, n, w, s, t, i, d, es, est
  - l, o, w, e, r, n, w, s, t, i, d, es, est, we
  - l, o, w, e, r, n, w, s, t, i, d, es, est, we, lo

Byte Pair Encoding

- Given a target vocabulary size
  - stop when you reach it
- Do deterministic longest piece segmentation of words
- Segmentation is only within words identified by some prior tokenizer (commonly Moses tokenizer for MT)
- Automatically decides vocab for system
- No longer strongly "word" based in conventional way
- Top places in WMT 2016!
  - Also widely used in WMT 2018

Wordpiece/Sentencepiece Models I

- Google NMT (GNMT) uses a variant of this
  - V1: wordpiece model
  - V2: sentencepiece model
  - Rather than char *n*-gram count, uses a greedy approximation to maximizing language model log likelihood to choose the pieces
  - Add *n*-gram that maximally reduces perplexity

- Wordpiece model tokenizes inside words
- Sentencepiece model works from raw text
  - Whitespace is retained as special token (_) and grouped normally
  - You can reverse things at end by joining pieces and recoding them to spaces
  - https://github.com/google/sentencepiece
  - https://arxiv.org/pdf/1804.10959.pdf

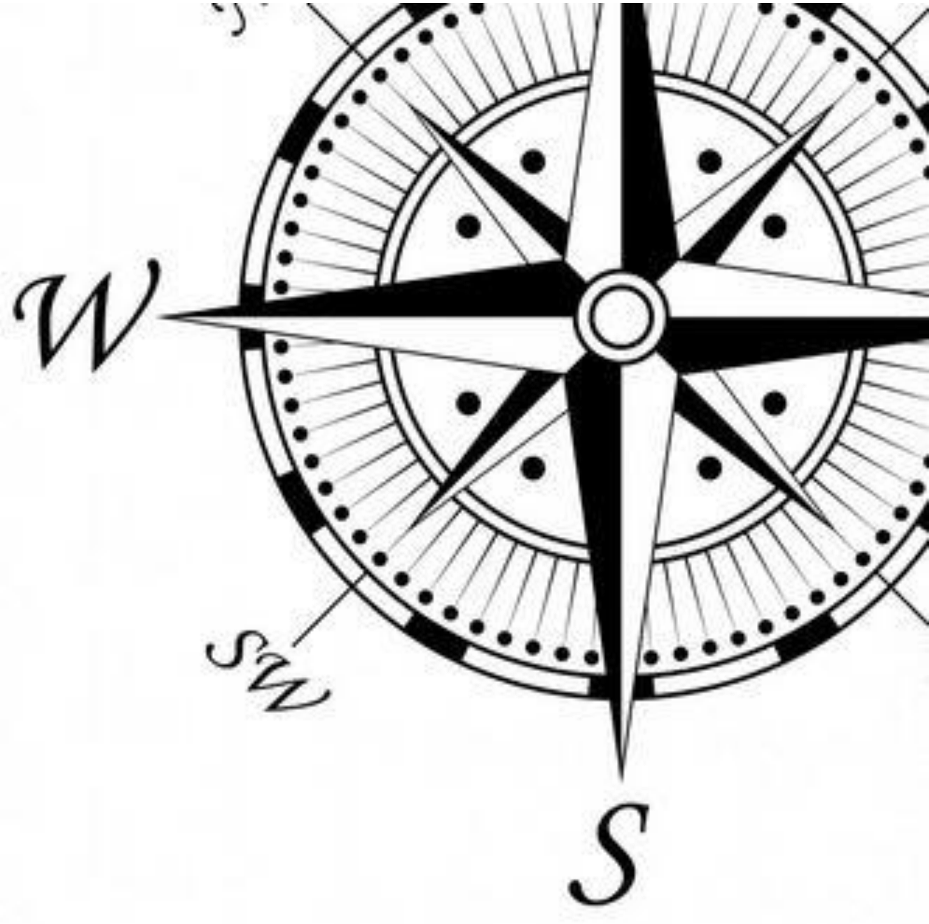
Wordpiece/Sentencepiece Models II

- BERT (later in lecture) uses a variant of the wordpiece model
  - (Relatively) common words are in the vocabulary:
    - *at, fairfax, 1910s*
  - Other words are built from wordpieces:
    - *hypatia = h ##yp ##ati ##a*
- If you’re using BERT in an otherwise word based model, you have to deal with this

https://pixy.org/src/425/4254306.png

```
from tokenizers import BertWordPieceTokenizer
from transformers import BertTokenizer
```

Topics Today



1. Evaluation of Word Vectors
2. Sub-Word Models
3. **Document Embeddings**

Representation of a Document

- Document ≈ long text, paragraph, sentence
- Many options:
  - Bag-of-words
  - TF-IDF
  - N-grams
- Problem:
  - Same content ≠ same words

Obama speaks to the media in Illinois

The President greets the press in Chicago

- Solution: methods, that capture semantics:
  - E.g. topic models
  - **Something with word embeddings**

Kusner, Matt, et al. "From word embeddings to document distances." *International conference on machine learning*. 2015.

Document Embeddings

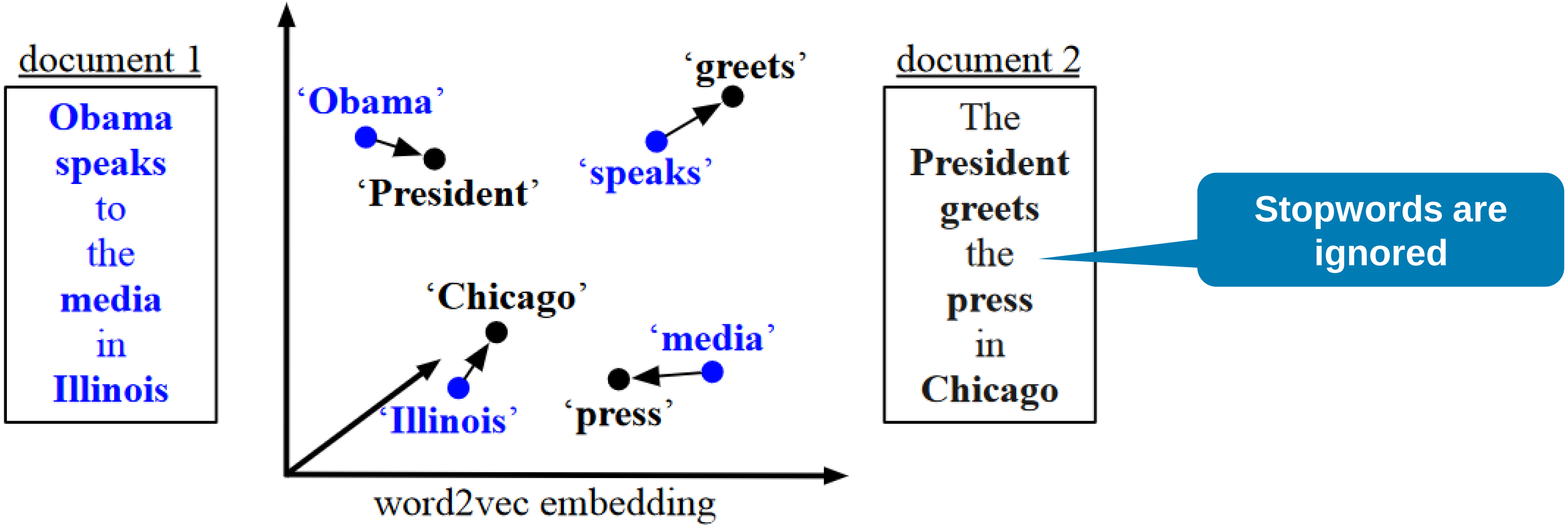
- Most simple appraoch (usually works quite well!)
  - (Weighted) **Average** of individual word embeddings

- More sophisticated:
  - **Word movers distance**
  - „Bag-of-word-embeddings“
  - Word vectors are mapped between documents

- Data-driven:
  - Direct learning of vectors for groups of words or whole documents
  - Most popular: **doc2vec**
  - Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." *International conference on machine learning*. 2014.

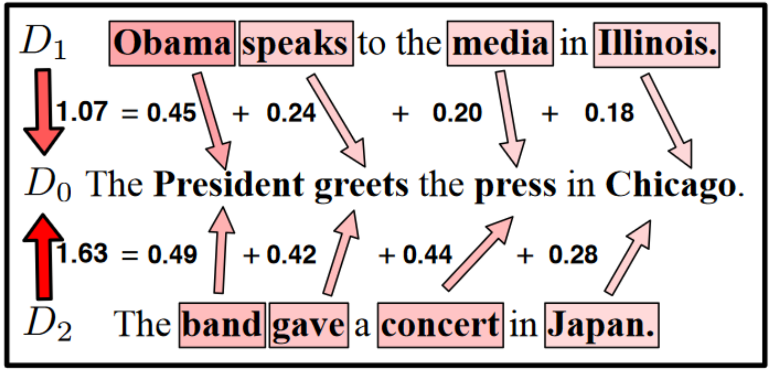
Word Movers Distance I

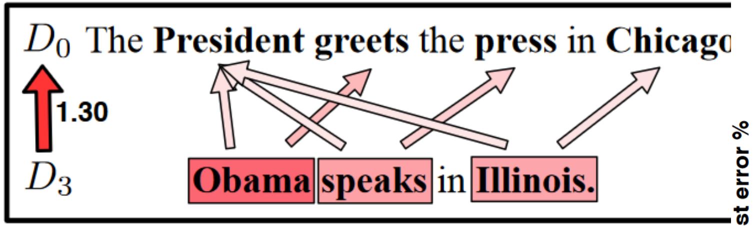
- Simlarity of documents
  - Sum of the minimal distances in the embedding space to move from one document to the other



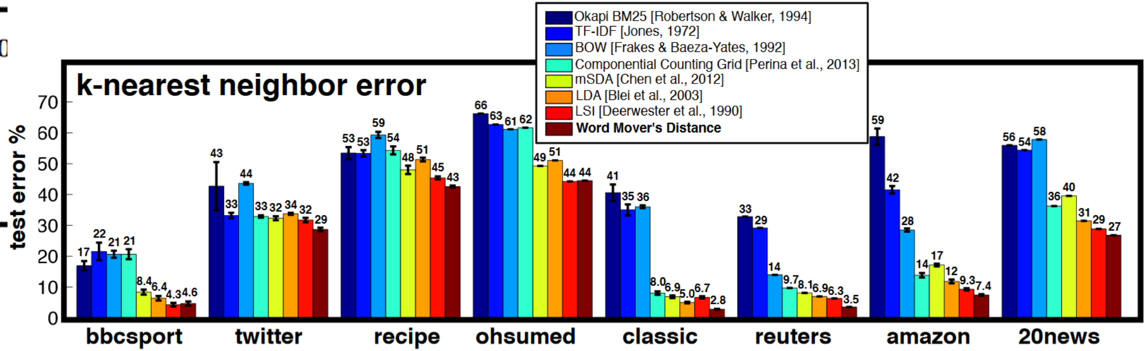
Kusner, Matt, et al. "From word embeddings to document distances." *International conference on machine learning*. 2015.

Word Movers Distance II



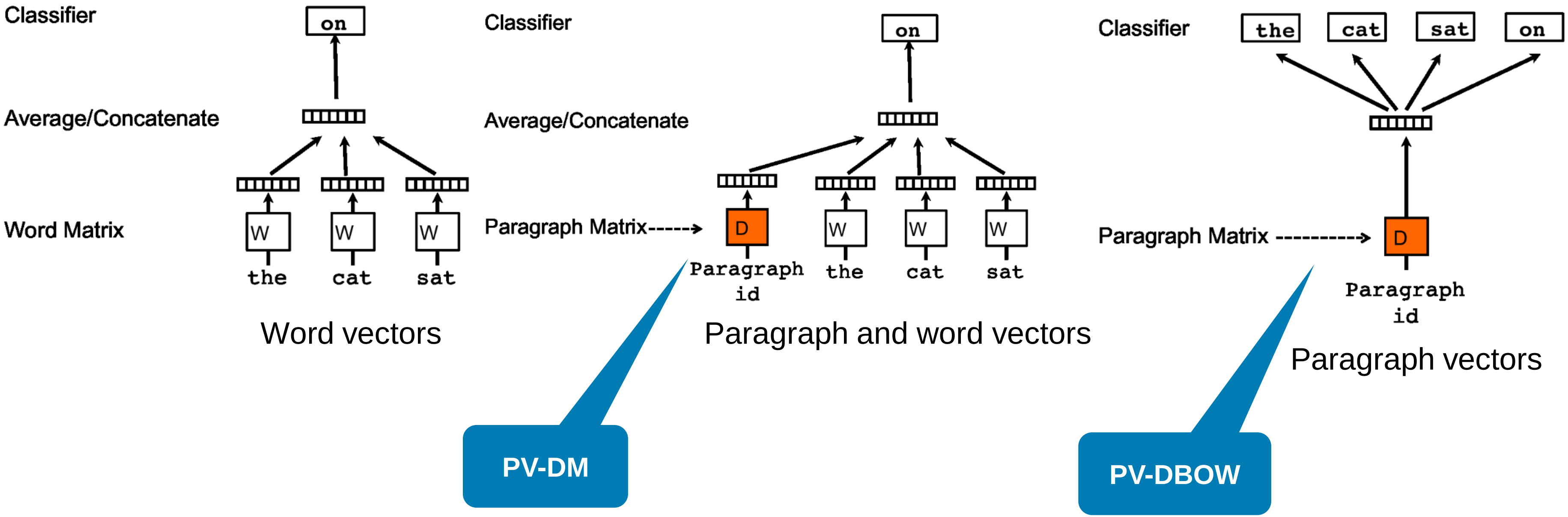


NAME	n	BOW DIM.	UNIQUE WORDS (AVG)	$\|\mathcal{Y}\|$
BBCSPORT	517	13243	117	5
TWITTER	2176	6344	9.9	3
RECIPE	3059	5708	48.5	15
OHSUMED	3999	31789	59.2	10
CLASSIC	4965	24277	38.6	4
REUTERS	5485	22425	37.1	8
AMAZON	5600	42063	45.0	4
20NEWS	11293	29671	72	20



Kusner, Matt, et al. "From word embeddings to document distances." *International conference on machine learning*. 2015.

Doc2Vec I



Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." *International conference on machine learning*. 2014.

Doc2Vec II

- Both paragraph vector models have two modi:
  - Training:
    - The model is trained with training data
    - Embedding vectors and softmax weight are fitted
  - Testing/in production:
    - New, unseen documents arrive
    - Only embedding vectors are learned; softmax-weights remain unchanged
- Testing is faster than training
  - But still not cheap (learning word embeddings)

Model	Error rate
BoW (bnc) (Maas et al., 2011)	12.20 %
BoW (bΔt'c) (Maas et al., 2011)	11.77%
LDA (Maas et al., 2011)	32.58%
Full+BoW (Maas et al., 2011)	11.67%
Full+Unlabeled+BoW (Maas et al., 2011)	11.11%
WRRBM (Dahl et al., 2012)	12.58%
WRRBM + BoW (bnc) (Dahl et al., 2012)	10.77%
MNB-uni (Wang & Manning, 2012)	16.45%
MNB-bi (Wang & Manning, 2012)	13.41%
SVM-uni (Wang & Manning, 2012)	13.05%
SVM-bi (Wang & Manning, 2012)	10.84%
NBSVM-uni (Wang & Manning, 2012)	11.71%
NBSVM-bi (Wang & Manning, 2012)	8.78%
Paragraph Vector	**7.42%**

Concatenated PV-DM ∘ PV-DBOW

Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." *International conference on machi… …ning*. 2014.

Results: Similar Sentences

Query and nearest sentence
he ran his hand inside his coat , double-checking that the unopened letter was still there . he slipped his hand between his coat and his shirt , where the folded copies lay in a brown envelope .
im sure youll have a glamorous evening , she said , giving an exaggerated wink . im really glad you came to the party tonight , he said , turning to her .
although she could tell he had n't been too invested in any of their other chitchat , he seemed genuinely curious about this . although he had n't been following her career with a microscope , he 'd definitely taken notice of her appearances .
an annoying buzz started to ring in my ears , becoming louder and louder as my vision began to swim . a weighty pressure landed on my lungs and my vision blurred at the edges , threatening my consciousness altogether .
if he had a weapon , he could maybe take out their last imp , and then beat up errol and vanessa . if he could ram them from behind , send them sailing over the far side of the levee , he had a chance of stopping them .
then , with a stroke of luck , they saw the pair head together towards the portaloos . then , from out back of the house , they heard a horse scream probably in answer to a pair of sharp spurs digging deep into its flanks .
" i 'll take care of it , " goodman said , taking the phonebook . " i 'll do that , " julia said , coming in .
he finished rolling up scrolls and , placing them to one side , began the more urgent task of finding ale and tankards . he righted the table , set the candle on a piece of broken plate , and reached for his flint , steel , and tinder .

Approximately two weeks of training on a billion-word Books corpus

Hard to evaluate!

Semantic Relatedness Evaluation

- SICK semantic relatedness task: score sentences for semantic similarity from 1 to 5 (average of 10 human ratings)

- Sentence A: A man is jumping into an empty pool
  Sentence B: There is no biker jumping in the air
  Relatedness score: 1.6
- Sentence A: Two children are lying in the snow and are making snow angels
  Sentence B: Two angels are making snow on the lying children
  Relatedness score: 2.9
- Sentence A: The young boys are playing outdoors and the man is smiling nearby
  Sentence B: There is no boy playing outdoors and there is no man smiling
  Relatedness score: 3.6
- Sentence A: A person in a black jacket is doing tricks on a motorbike
  Sentence B: A man in a black jacket is doing tricks on a motorbike
  Relatedness score: 4.9

Semantic Entailment Evaluation

- SICK semantic entailment task: score sentences for relations: ENTAILMENT, CONTRADICTION, NEUTRAL:

- Sentence A: Two teams are competing in a football match
  Sentence B: Two groups of people are playing football
  Entailment judgment: ENTAILMENT
-
- Sentence A: The brown horse is near a red barrel at the rodeo
  Sentence B: The brown horse is far from a red barrel at the rodeo
  Entailment judgment: CONTRADICTION
-
- Sentence A: A man in a black jacket is doing tricks on a motorbike
  Sentence B: A person is riding the bicycle on one wheel
  Entailment judgment: NEUTRAL

Hard to compare: Models trained with these specific objectives outperform general embedding models!

Exercise

- Use average word2vec word vectors to represent sentences.
- Use doc2vec to compute sentence vectors.
  - Model available in Gensim
    - https://radimrehurek.com/gensim/models/doc2vec.html
- Compare most similar sentences for both representations

- More links:
  - https://ireneli.eu/2016/07/27/nlp-05-from-word2vec-to-doc2vec-a-simple-example-with-gensim/
  - https://www.analyticsvidhya.com/blog/2020/08/top-4-sentence-embedding-techniques-using-python/

Start 10 9 8 7 6 5 4 3 2 1 End

Learning Goals for this Chapter

- Know different methods to evaluate word embeddings
  – Pros and cons of the methods
- Be able to name limitations of the evaluations
- Be able to explain the diffences between embedding models
  – Character level
  – Subword level
  – Word level
- Understand document/paragraph/sentence embeddings
  – Word movers distance
  – Doc2vec
- Relevant chapters
  – P6.1
  – S2 (2021) https://www.youtube.com/watch?v=gqaHkPEZAew

Literature

- Evaluation methods for unsupervised word embeddings
- Linear Algebraic Structure of Word Senses, with Applications to Polysemy
- On the Dimensionality of Word Embedding
- Debiasing Word Embeddings
- Dynamic Word Embeddings