NATURAL LANGUAGE PROCESSING

LECTURE 3: WORDEMBEDDING

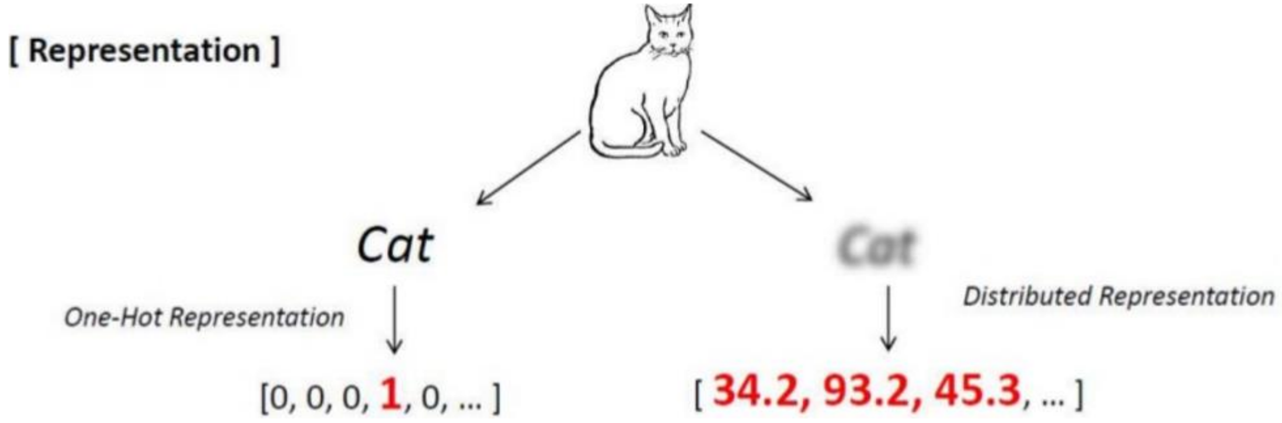
goorm KAIST AI Graduate School of AI DAVIAN

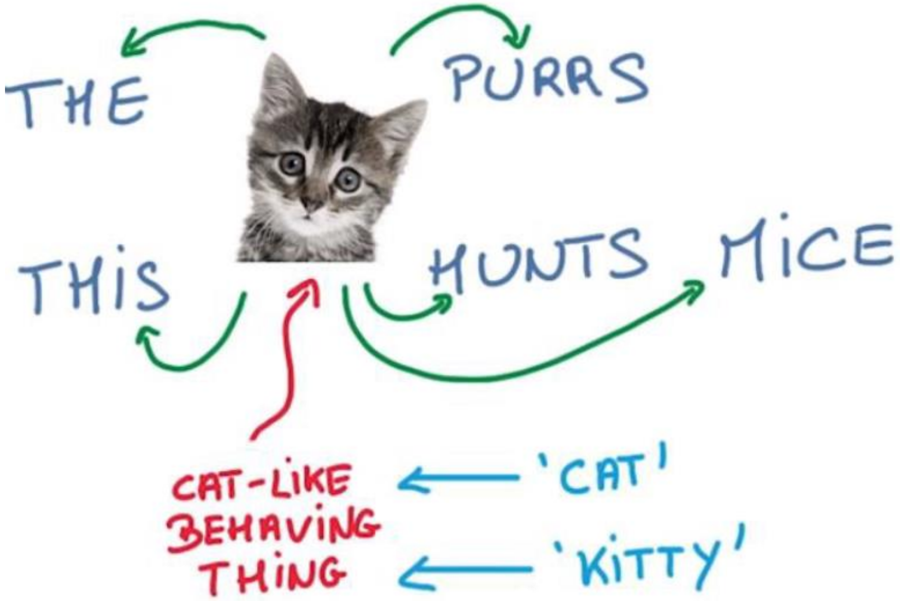
What is Word Embedding?

- Express a word as a vector
- 'cat' and 'kitty' are similar words, so they have similar vector representations □ short distance
- 'hamburger' is not similar with 'cat' or 'kitty’, so they have different vector representations □ far distance

Word2Vec

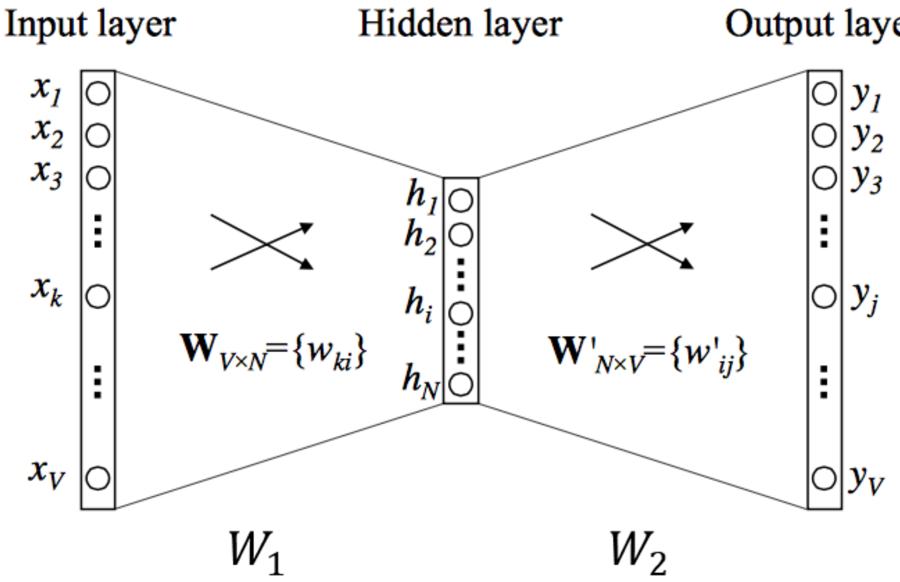
- An algorithm for training vector representation of a word from context words (adjacent words)
  - Assumption: words in similar context will have similar meanings.
- e.g.)
  - The **cat** purrs.
  - The **cat** hunts mice.



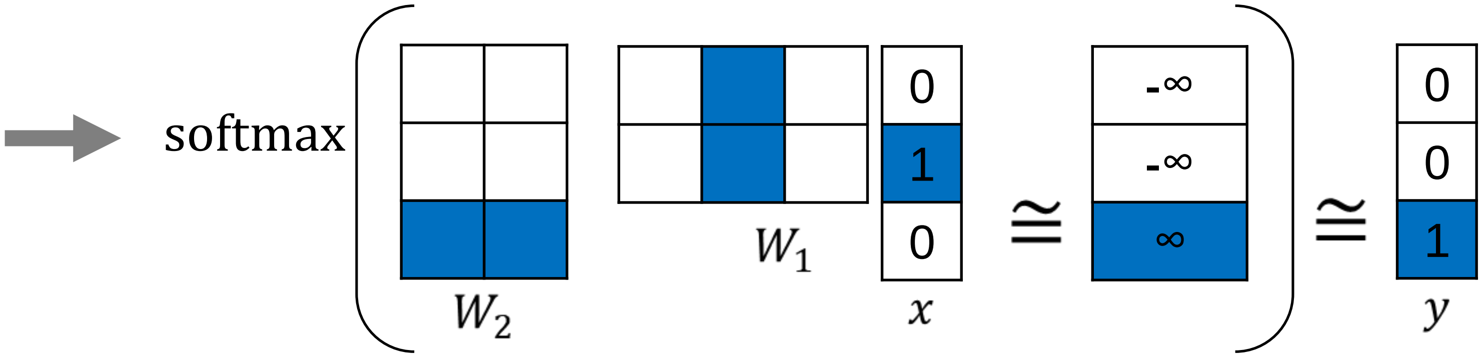


Distributed Representations of Words and Phrases and their Compositionality, NeurIPS'13

How Word2Vec Algorithm Works



$$y = \text{softmax}(W_2 W_1 x)$$



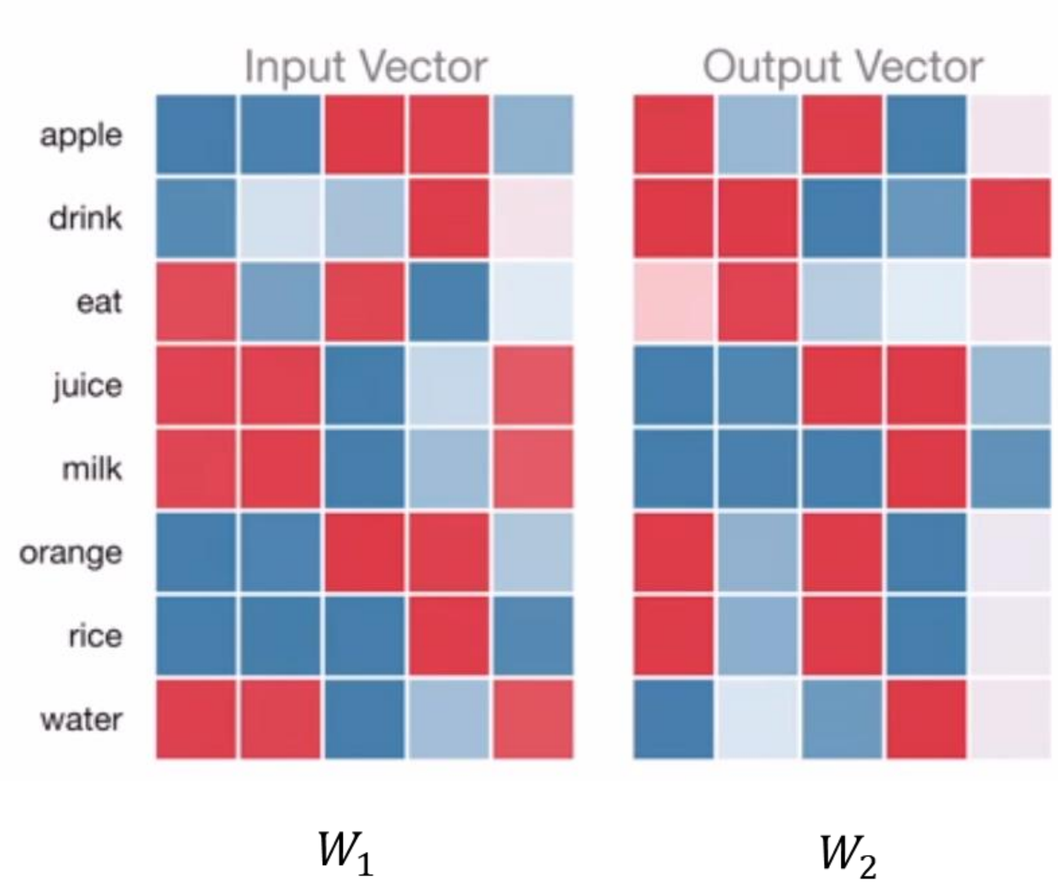
- Sentence : "I study math."
- Vocabulary: {"I", "study" "math"}
- Input: "study" [0, 1, 0]
- Output: "math" [0, 0, 1]
- Columns of W_1 and rows of W_2 represent each word.
- E.g., 'study' vector : 2nd column in W_1, 'math' vector : 3rd row in W_2.
- The 'study' vector in W_1 and the 'math' vector in W_2 should have a high inner-product value.

Distributed Representations of Words and Phrases and their Compositionality, NeurIPS'13

How Word2Vec Algorithm Works

- A vector representation of ‘eat’ in W_1 has similar pattern with vectors of ‘apple’, ‘orange’, and ‘rice’ in W_2.

- When the input is 'eat', the model can predict 'apple', 'orange’, or 'rice’ for output, because the vectors have high inner product values.

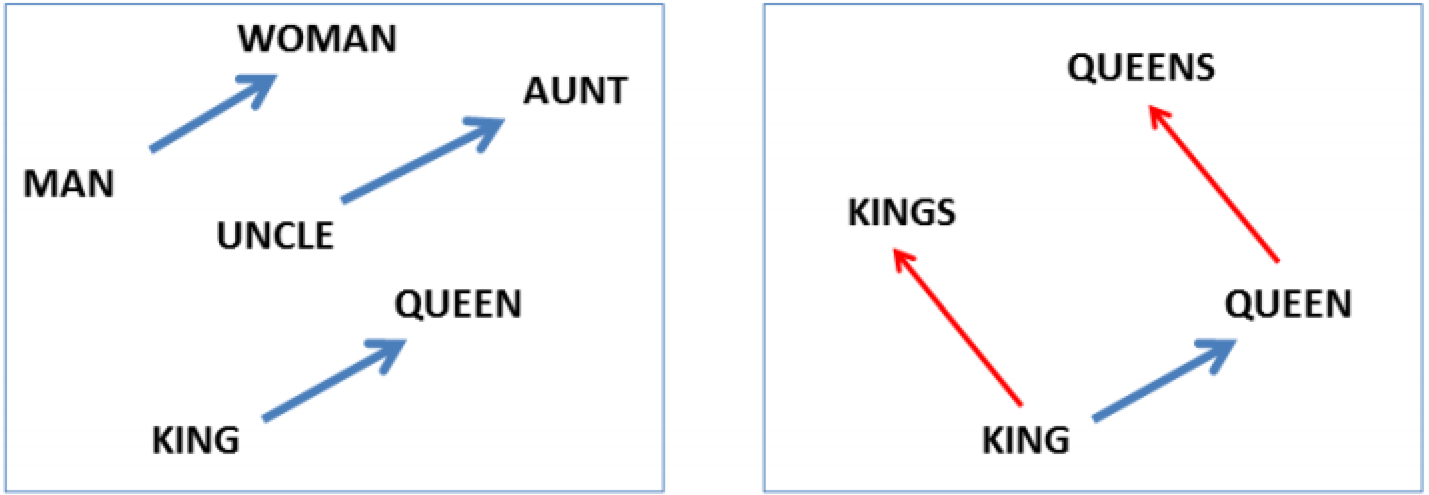


https://ronxin.github.io/wevi/

Distributed Representations of Words and Phrases and their Compositionality, NeurIPS’13

Property of Word2Vec

- The word vector, or the relationship between vector points in space, represents the relationship between the words.
- The same relationship is represented as the same vectors.



(Mikolov et al., NAACL HLT, 2013)

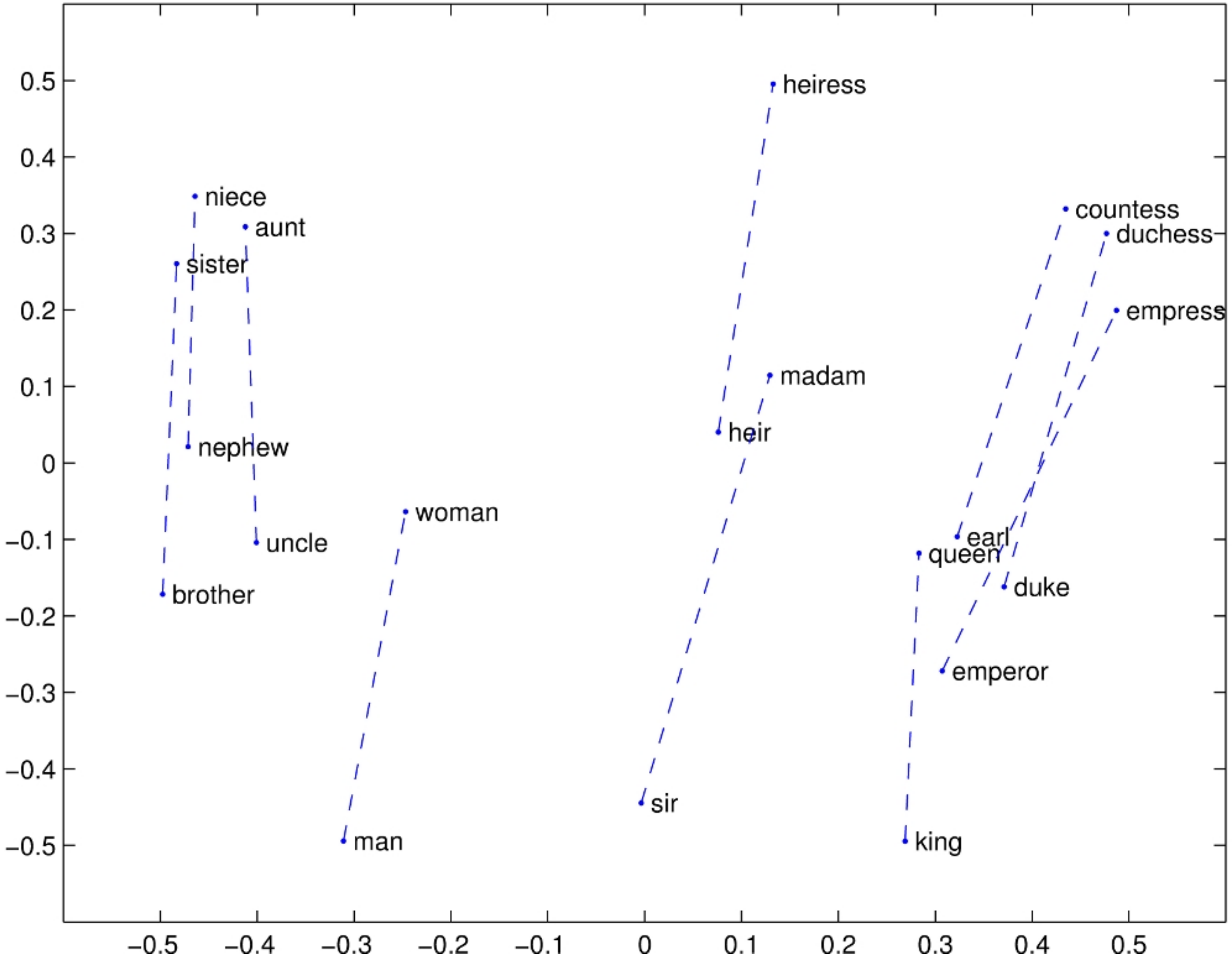
- e.g.,

vec[queen] – vec[king] = vec[woman] – vec[man]

Distributed Representations of Words and Phrases and their Compositionality, NeurIPS'13

Property of Word2Vec - Linear Substructure

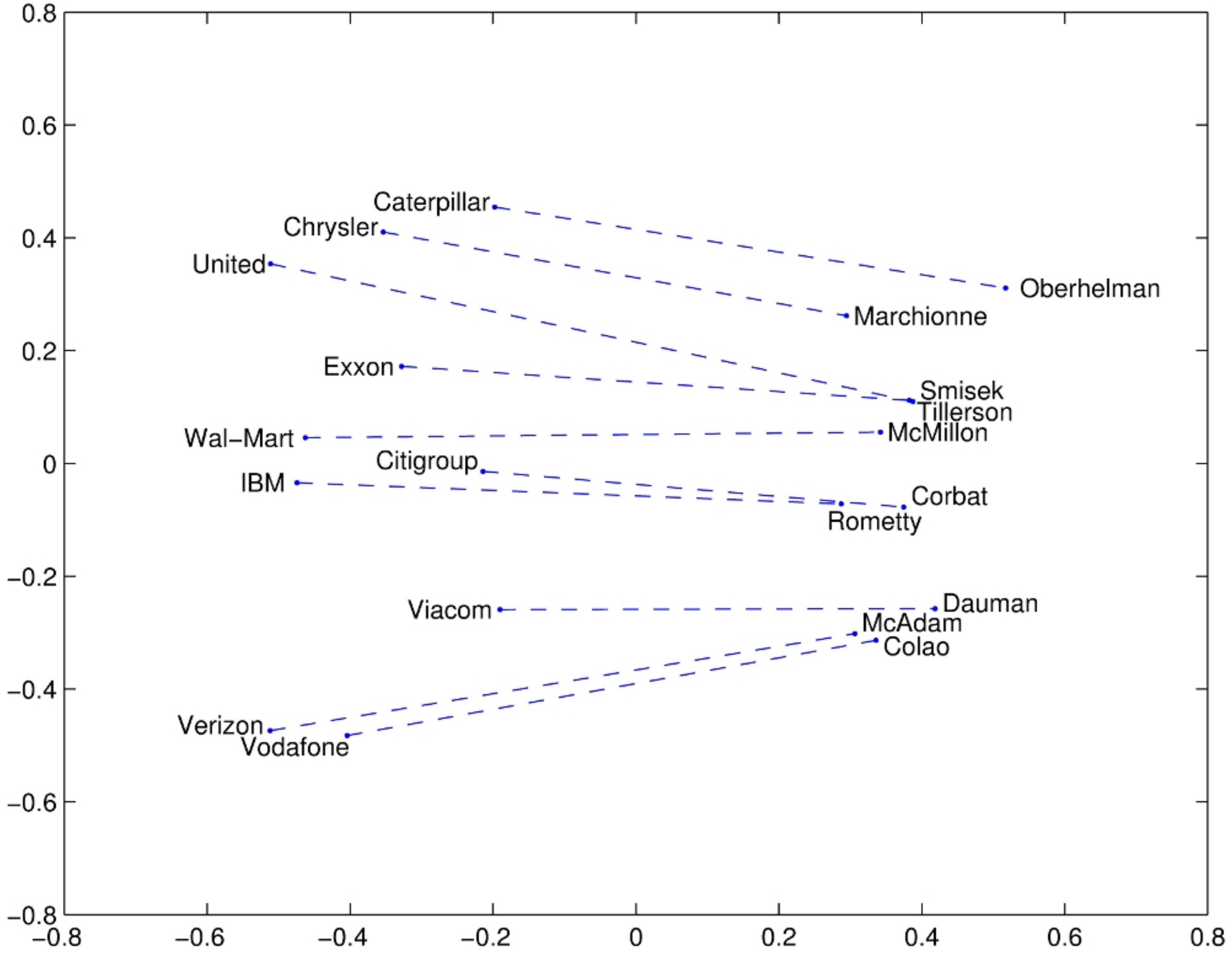
man - woman



Distributed Representations of Words and Phrases and their Compositionality, NeurIPS'13

Property of Word2Vec - Linear Substructure

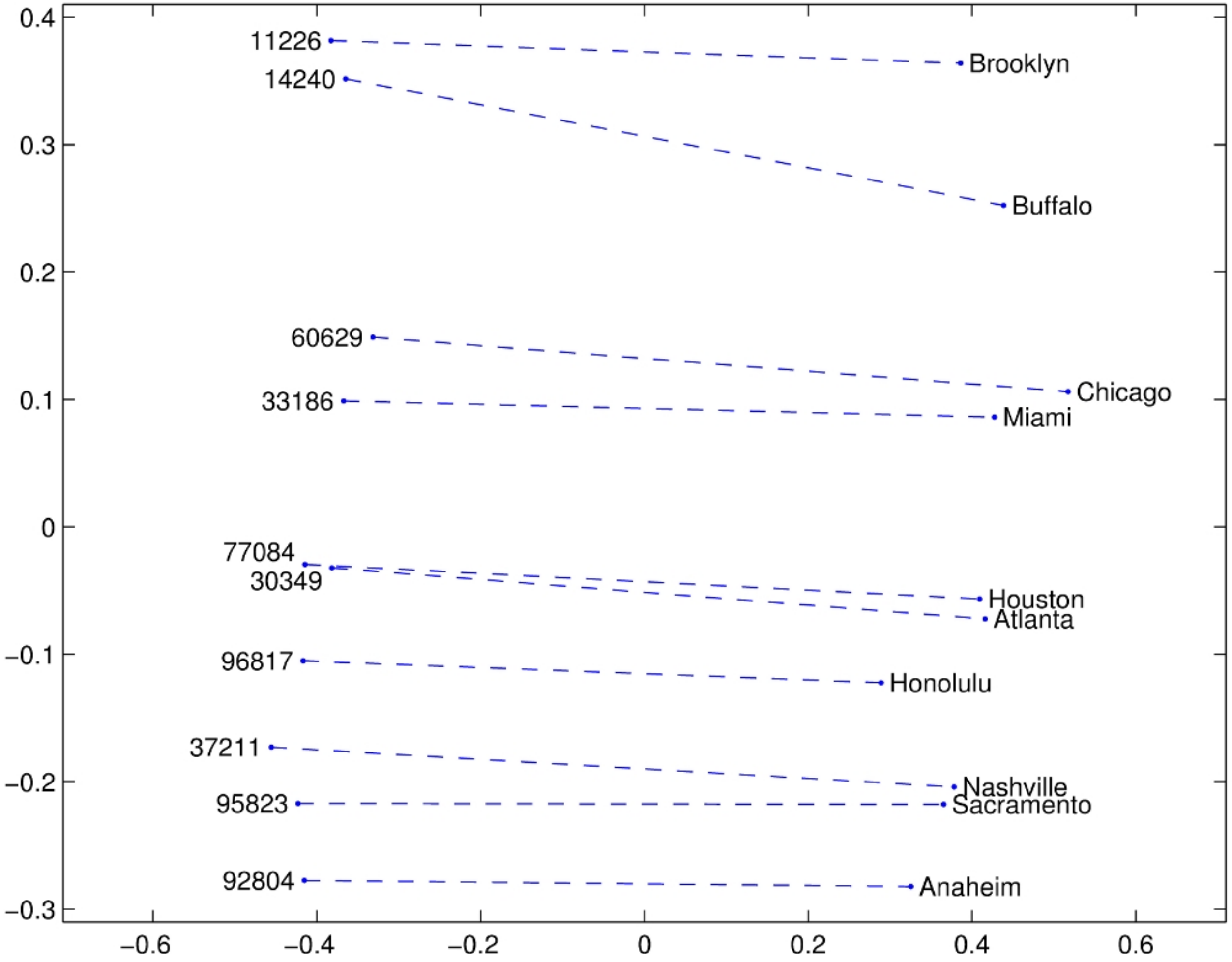
company - ceo



Distributed Representations of Words and Phrases and their Compositionality, NeurIPS'13

Property of Word2Vec - Linear Substructure

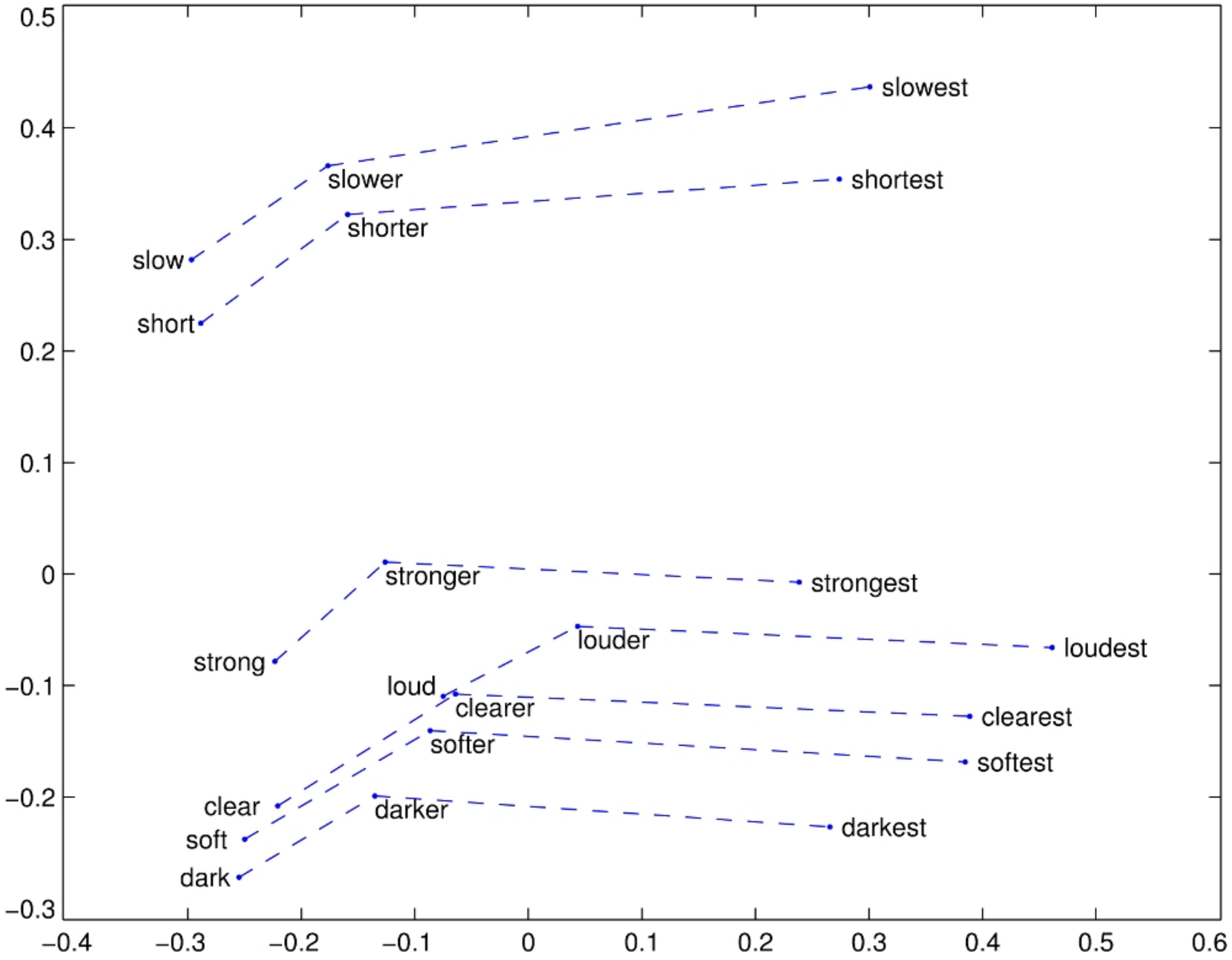
city - zip code



Distributed Representations of Words and Phrases and their Compositionality, NeurIPS’13

Property of Word2Vec - Linear Substructure

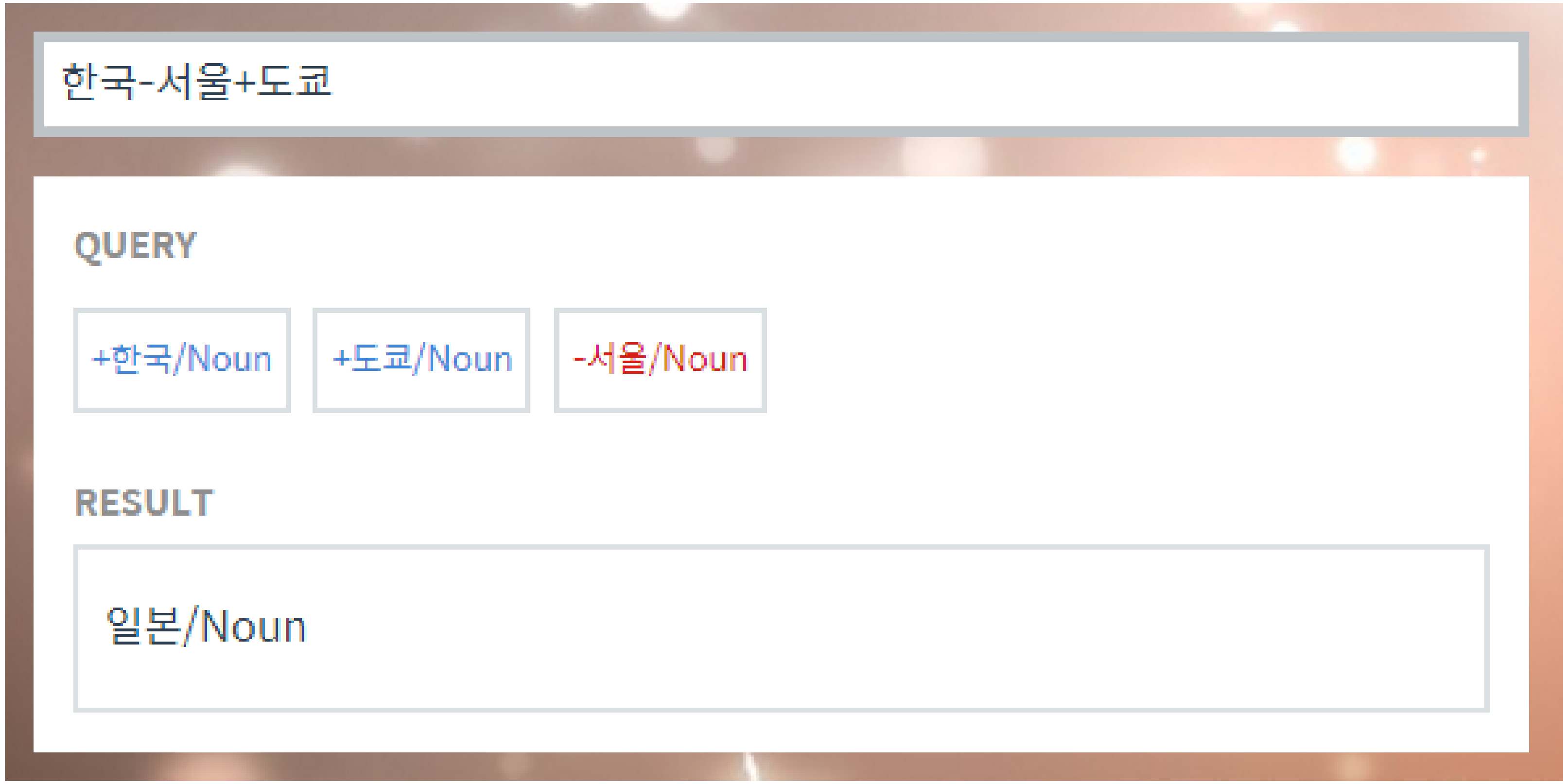
comparative - superlative



Distributed Representations of Words and Phrases and their Compositionality, NeurIPS'13

Property of Word2Vec – Analogy Reasoning

- Korean Word2Vec : http://w.elnn.kr/search/



한국-서울+도쿄

QUERY

+한국/Noun +도쿄/Noun -서울/Noun

RESULT

일본/Noun

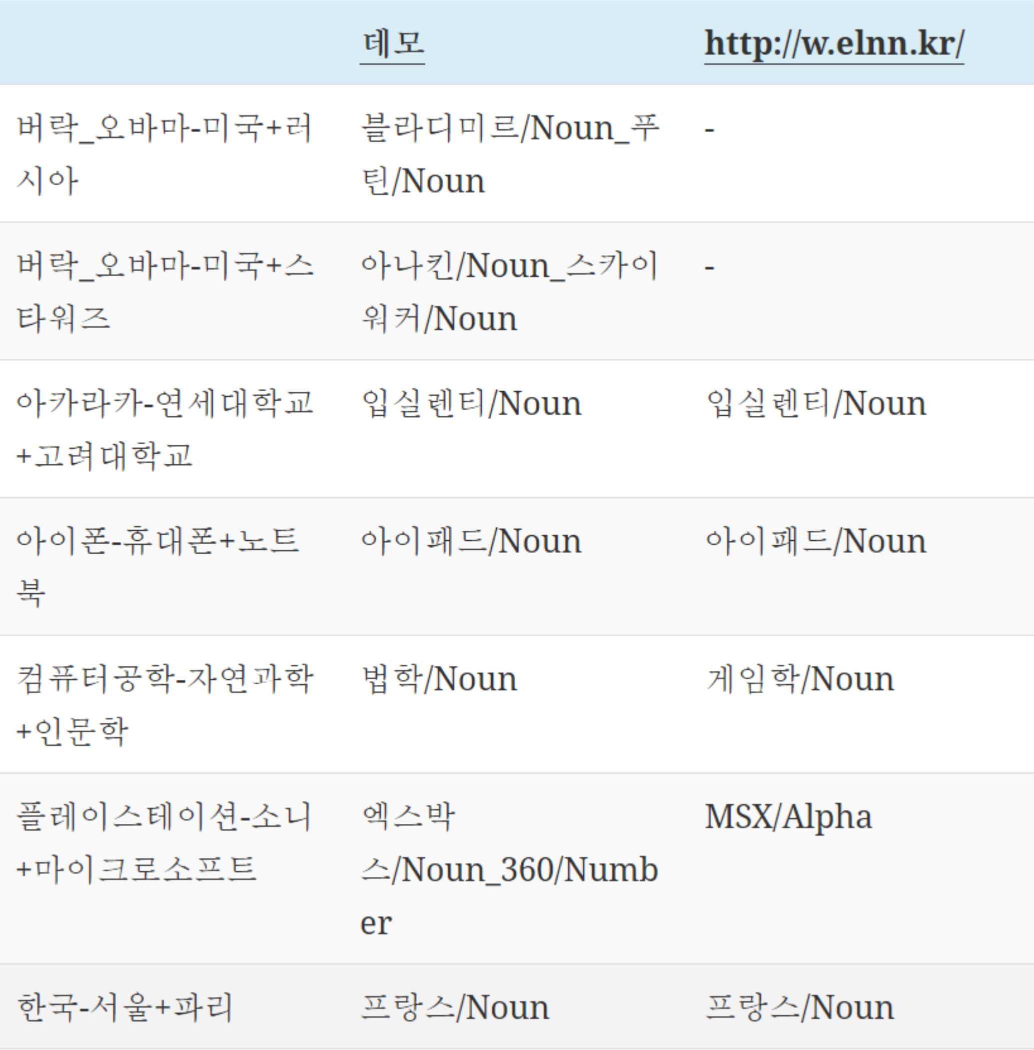
Application of Word2Vec

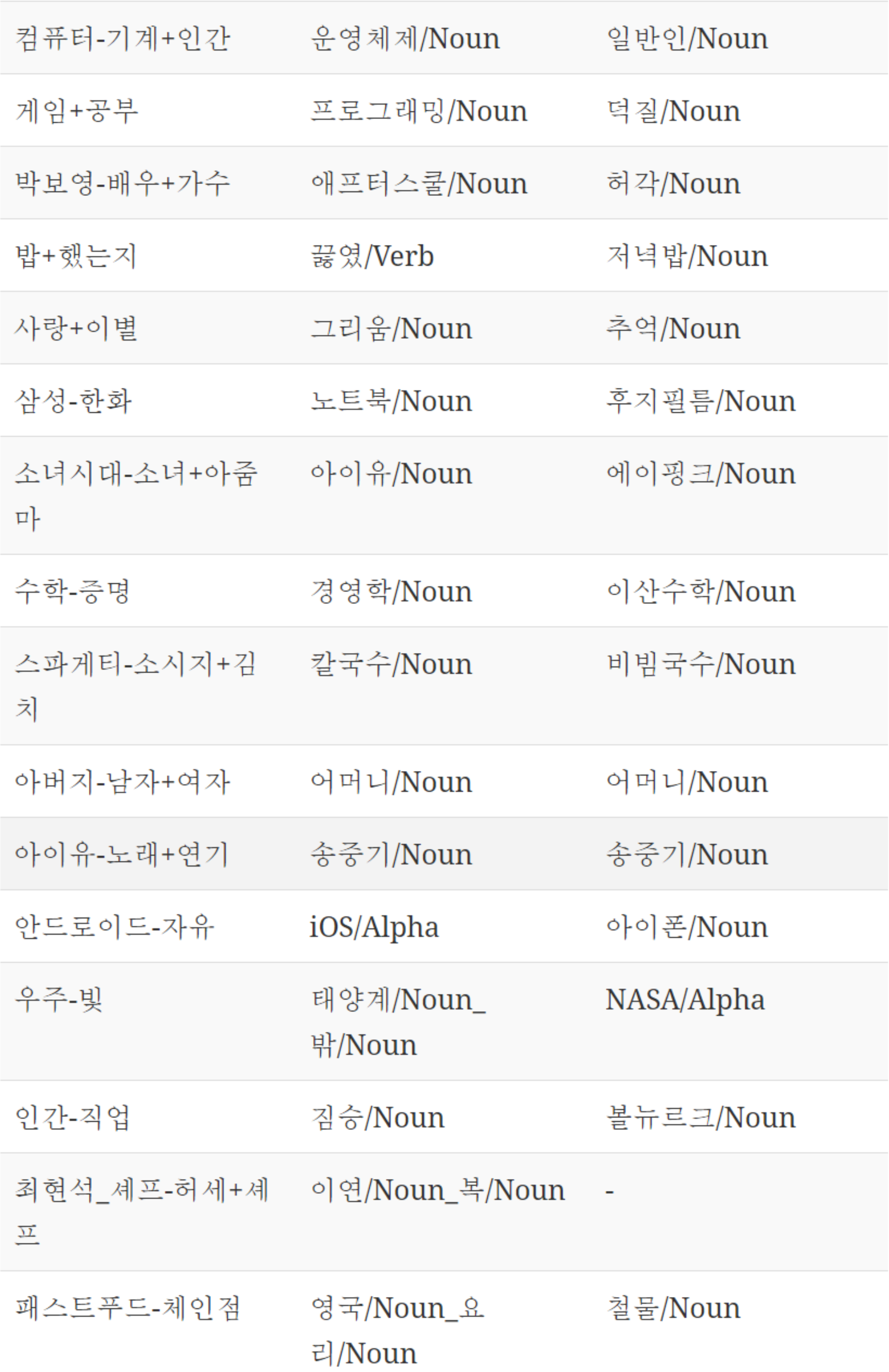
Word2Vec improves performances in most areas of NLP.

- Word similarity
- Machine translation
- Part-of-speech tagging and named entity recognition
- Sentiment analysis
- Clustering
- Semantic lexicon building

Property of Word2Vec – Analogy Reasoning

- More examples: http://wonjaekim.com/archives/50





	데모	http://w.elnn.kr/
버락_오바마-미국+러시아	블라디미르/Noun_푸틴/Noun	-
버락_오바마-미국+스타워즈	아나킨/Noun_스카이워커/Noun	-
아카라카-연세대학교+고려대학교	입실렌티/Noun	입실렌티/Noun
아이폰-휴대폰+노트북	아이패드/Noun	아이패드/Noun
컴퓨터공학-자연과학+인문학	법학/Noun	게임학/Noun
플레이스테이션-소니+마이크로소프트	엑스박스/Noun_360/Number	MSX/Alpha
한국-서울+파리	프랑스/Noun	프랑스/Noun
컴퓨터-기계+인간	운영체제/Noun	일반인/Noun
게임+공부	프로그래밍/Noun	덕질/Noun
박보영-배우+가수	애프터스쿨/Noun	허각/Noun
밥+했는지	끓였/Verb	저녁밥/Noun
사랑+이별	그리움/Noun	추억/Noun
삼성-한화	노트북/Noun	후지필름/Noun
소녀시대-소녀+아줌마	아이유/Noun	에이핑크/Noun
수학-증명	경영학/Noun	이산수학/Noun
스파게티-소시지+김치	칼국수/Noun	비빔국수/Noun
아버지-남자+여자	어머니/Noun	어머니/Noun
아이유-노래+연기	송중기/Noun	송중기/Noun
안드로이드-자유	iOS/Alpha	아이폰/Noun
우주-빛	태양계/Noun_밖/Noun	NASA/Alpha
인간-직업	짐승/Noun	볼뉴르크/Noun
최현석_셰프-허세+셰프	이연/Noun_복/Noun	-
패스트푸드-체인점	영국/Noun_요리/Noun	철물/Noun

Property of Word2Vec – Semantic Similarity

- Example: https://github.com/dhammack/Word2VecExample

- Word intrusion detection
  - staple hammer saw drill
  - math shopping reading science
  - rain snow sleet sun
  - eight six seven five three owe nine
  - breakfast cereal dinner lunch
  - england spain france italy greece germany portugal australia

Application of Word2Vec – Machine Translation

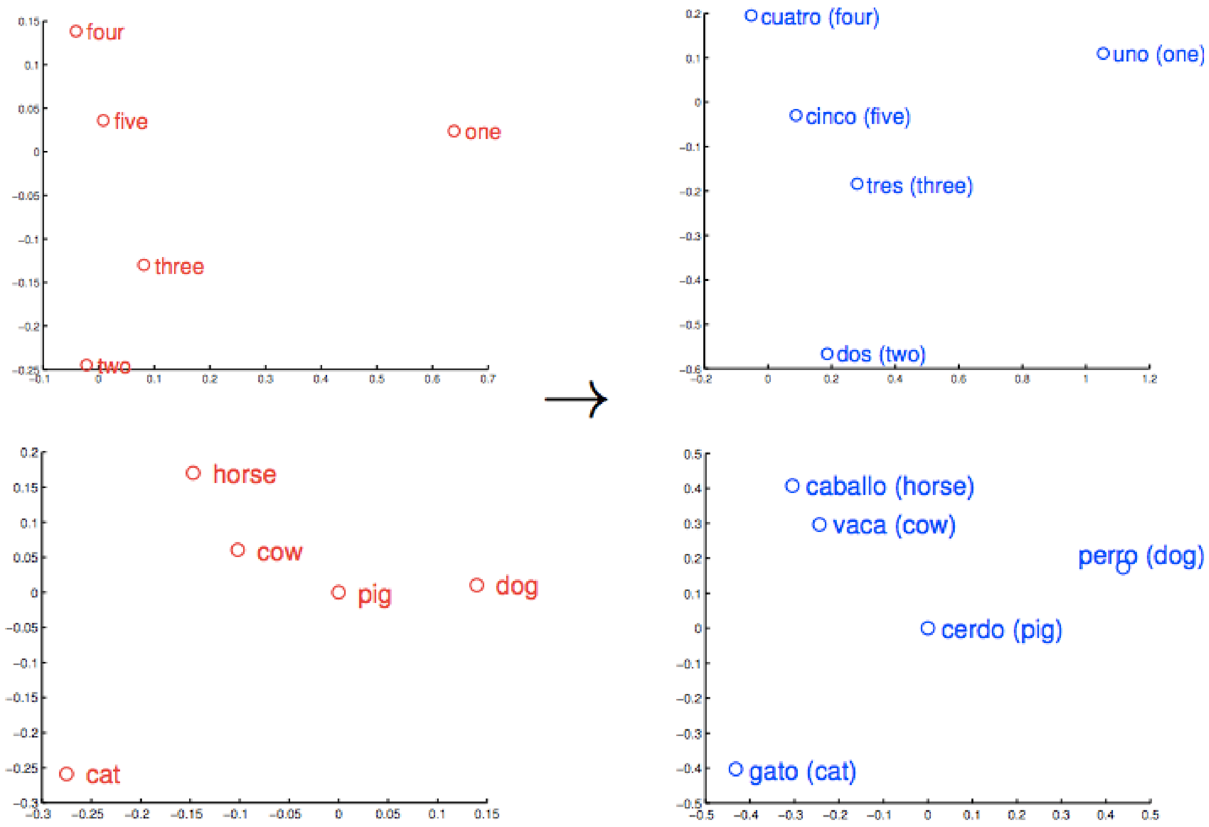
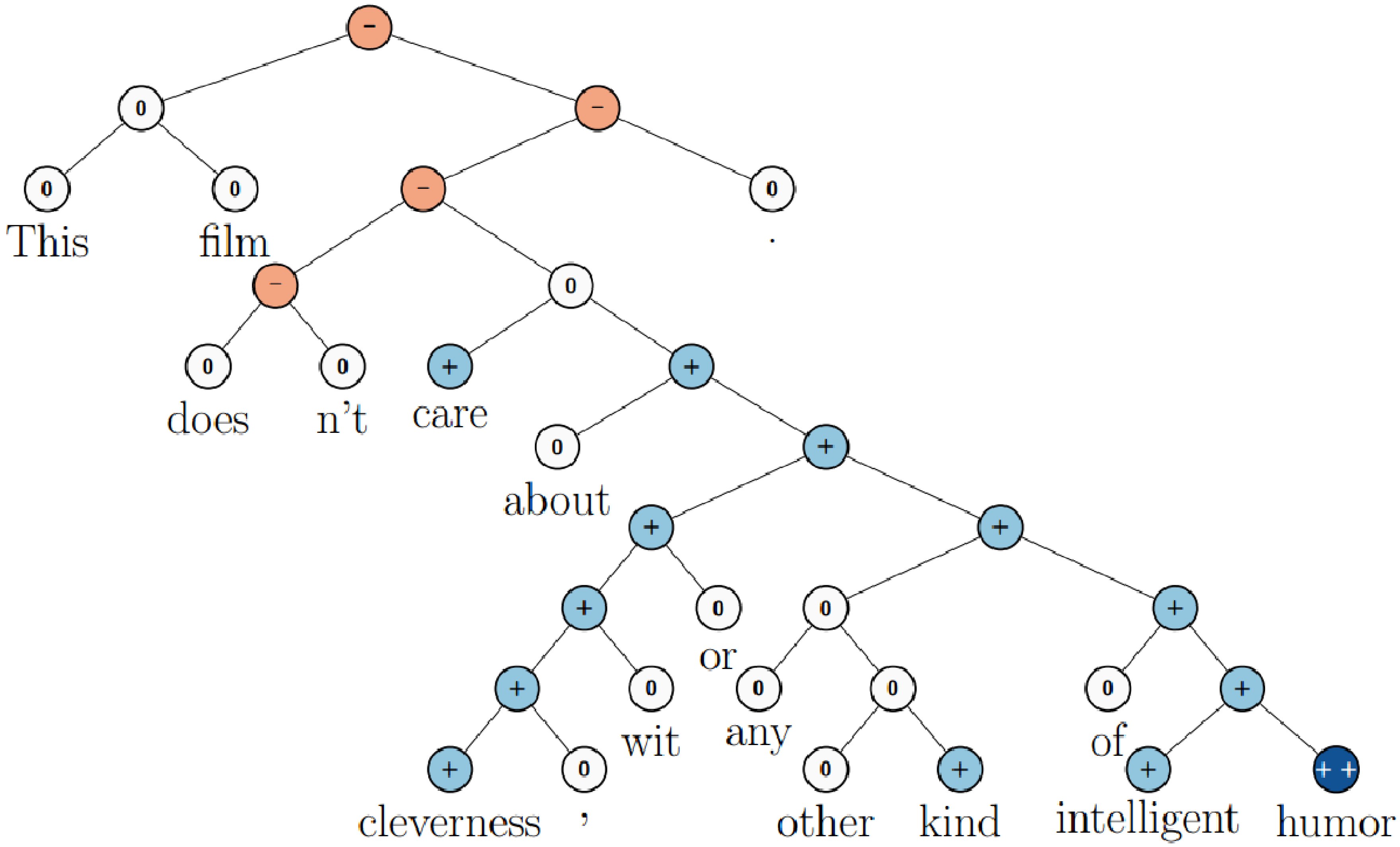
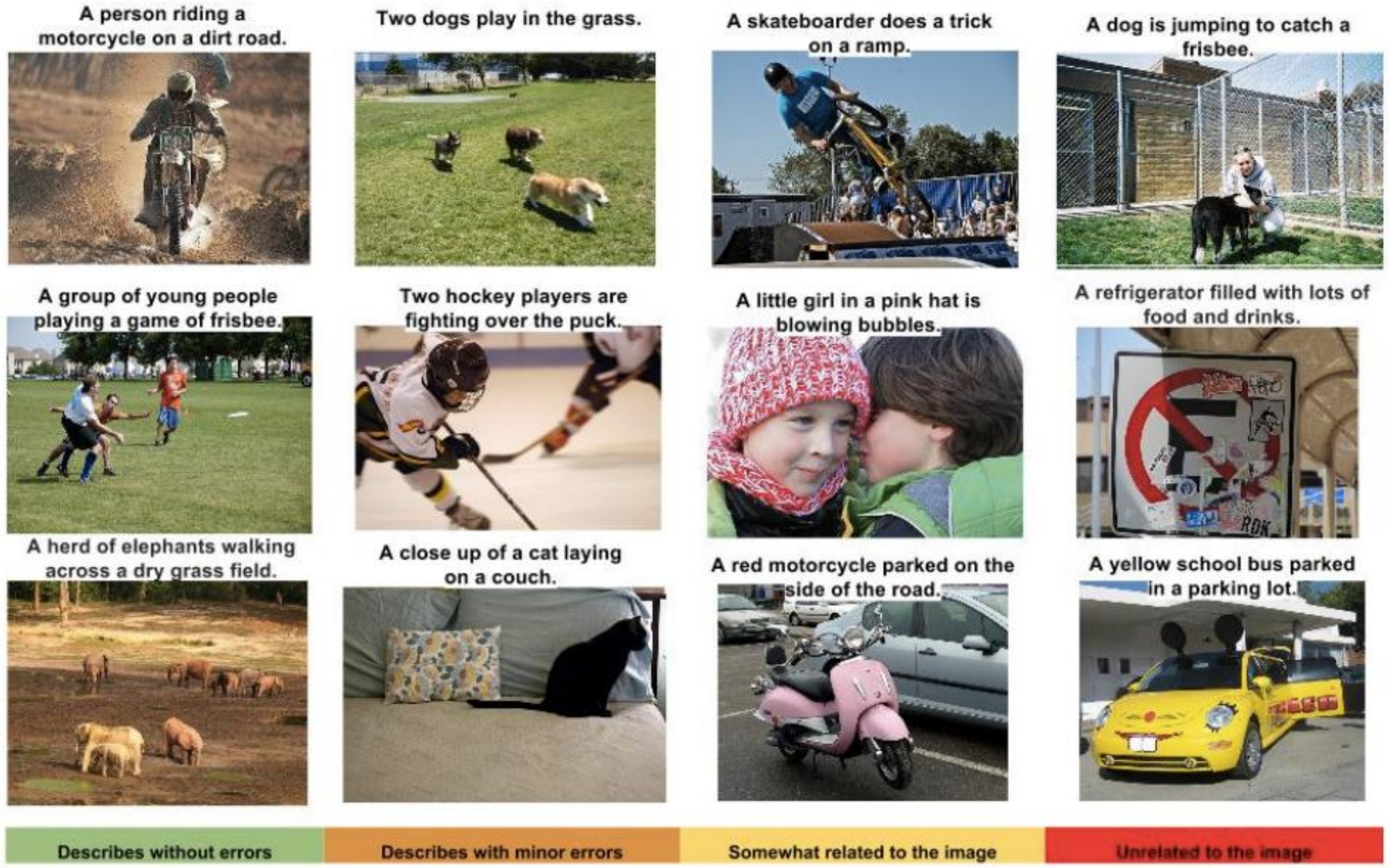


Figure 1: Distributed word vector representations of numbers and animals in English (left) and Spanish (right). The five vectors in each language were projected down to two dimensions using PCA, and then manually rotated to accentuate their similarity. It can be seen that these concepts have similar geometric arrangements in both spaces, suggesting that it is possible to learn an accurate linear mapping from one space to another. This is the key idea behind our method of translation.

Application of Word2Vec – Sentiment Analysis



Application of Word2Vec – Image Captioning



A person riding a motorcycle on a dirt road.

Two dogs play in the grass.

A skateboarder does a trick on a ramp.

A dog is jumping to catch a frisbee.

A group of young people playing a game of frisbee.

Two hockey players are fighting over the puck.

A little girl in a pink hat is blowing bubbles.

A refrigerator filled with lots of food and drinks.

A herd of elephants walking across a dry grass field.

A close up of a cat laying on a couch.

A red motorcycle parked on the side of the road.

A yellow school bus parked in a parking lot.

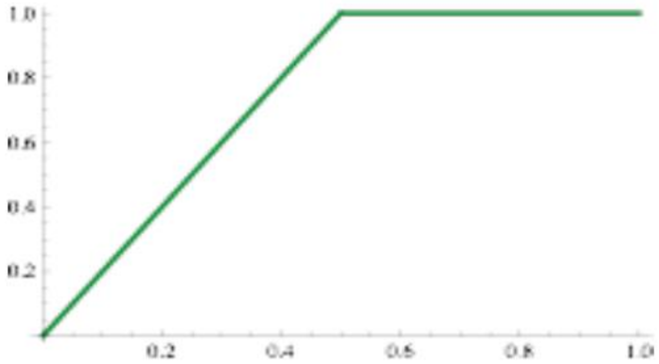
Figure 5. A selection of evaluation results, grouped by human rating.

GloVe: Another Word Embedding Model

GloVe: Global Vectors for Word Representation

- Rather than going through each pair of an input and an output words, it first computes the co-occurrence matrix, to avoid training on identical word pairs repetitively.
- Afterwards, it performs matrix decomposition on this co-occurrent matrix.

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



- Fast training
- Works well even with a small corpus

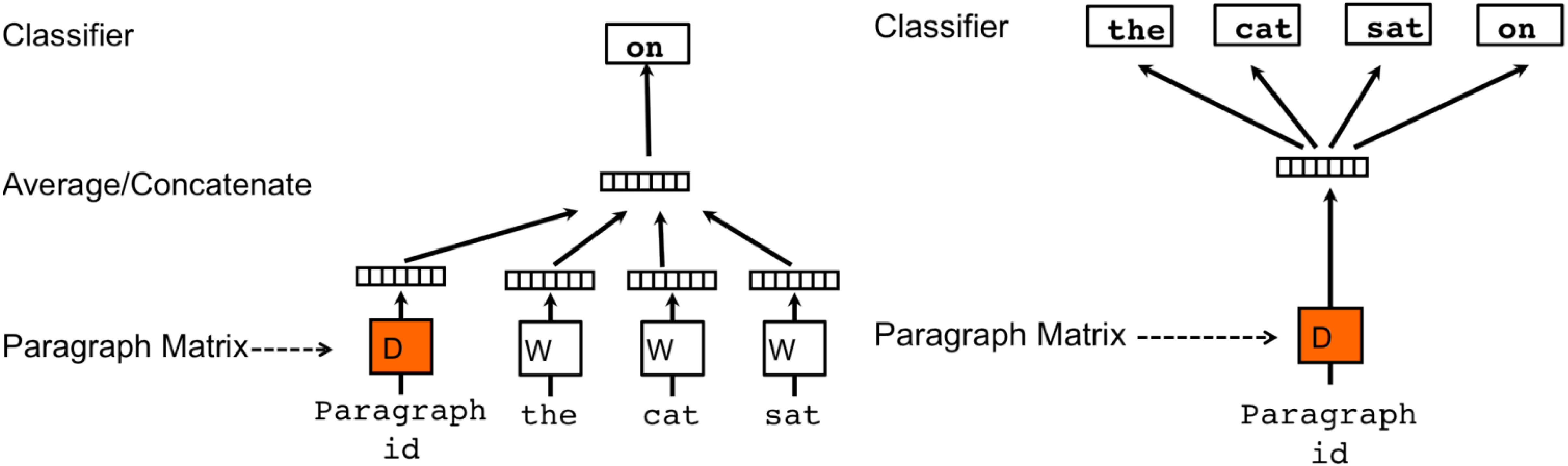
GloVe: Global Vectors for Word Representation, EMNLP'14

Word Embedding Methods

- Distributed vector representations
  - Vector representation in the form of nonzero values across multiple dimensions, as opposed to conventional one-hot vector.
  - Euclidean distance, inner product, and cosine similarity of two different word vectors encode their semantic similarity
- Today's topics
  - Word2Vec
  - GloVe
  - **Doc2Vec (adaptation of Word2Vec for a document vector)**

Doc2Vec (Paragraph2Vec)

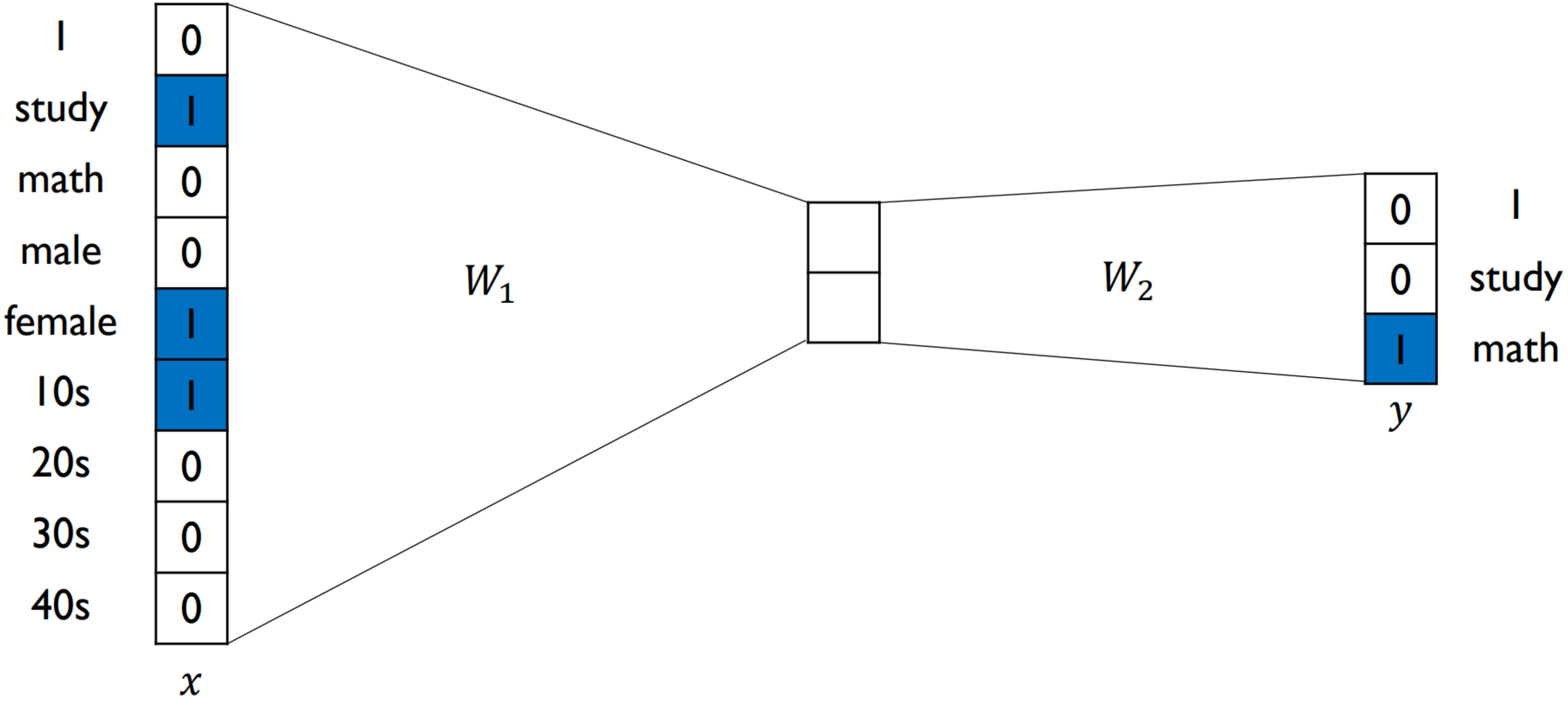
- Idea: Represent a document (or paragraph) vector as a word
- Properties and Advantages
  - The words in the same paragraphs and documents would have high similarity
  - A document can be embedded to the same space as a word vector.



Distributed Representations of Sentences and Documents, ICML'14

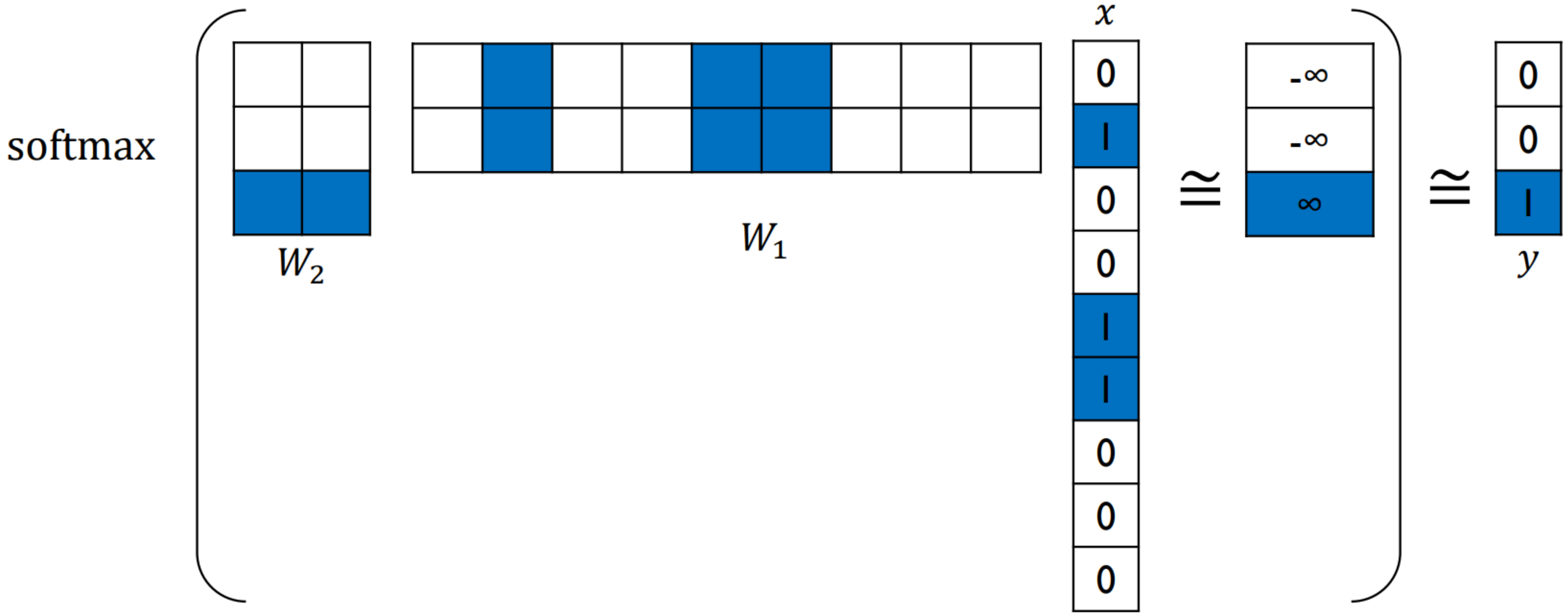
Doc2Vec (Paragraph2Vec)

- Doc2Vec can encode any other types of attributes associated with text data.
  - You can use multi-hot vector to represent associated attributes.
  - ((study, female, 10s), math)



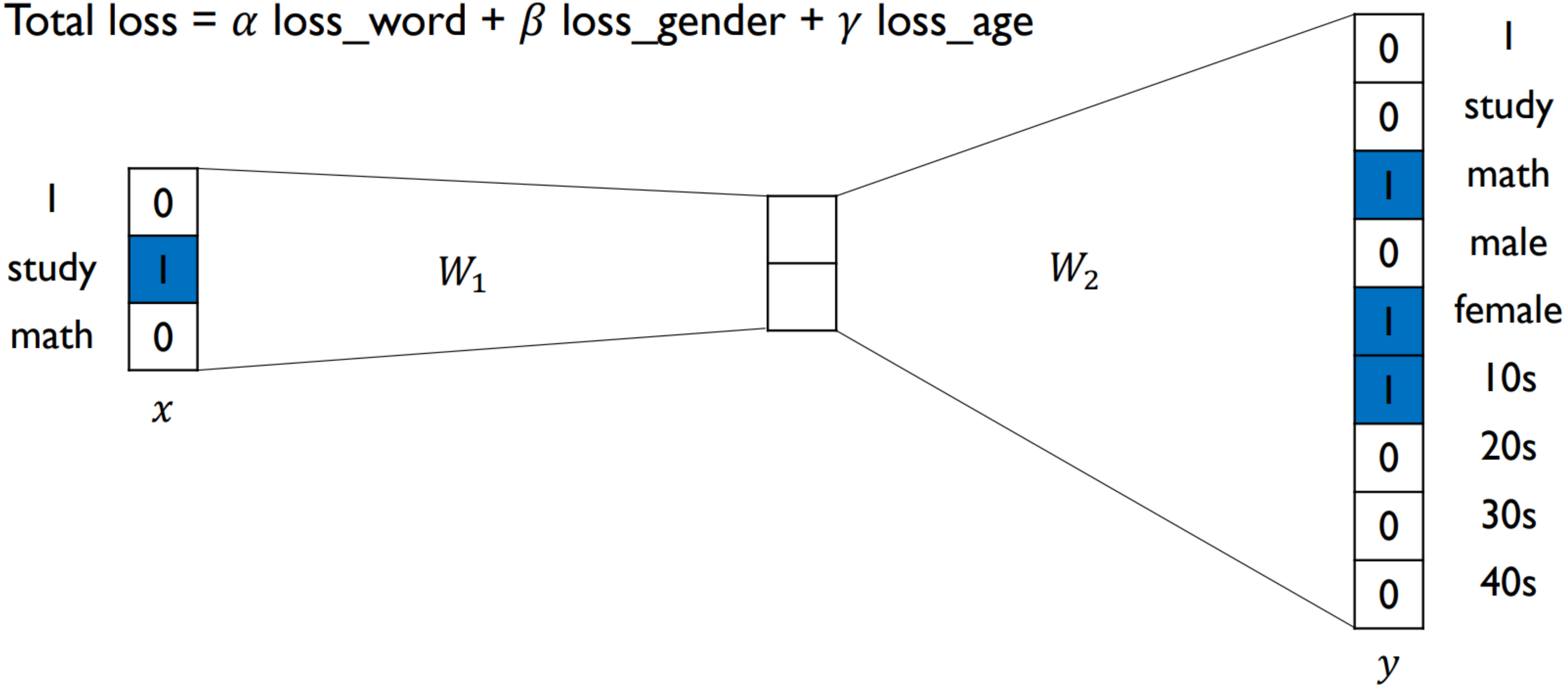
Doc2Vec (Paragraph2Vec)

- Doc2Vec can encode any other types of attributes associated with text data.
  - ((study, female, 10s), math)
  - The sum of inner product values between 'study', 'female', and '10s' vector in W_1 and the 'math' vector in W_2 should be high.



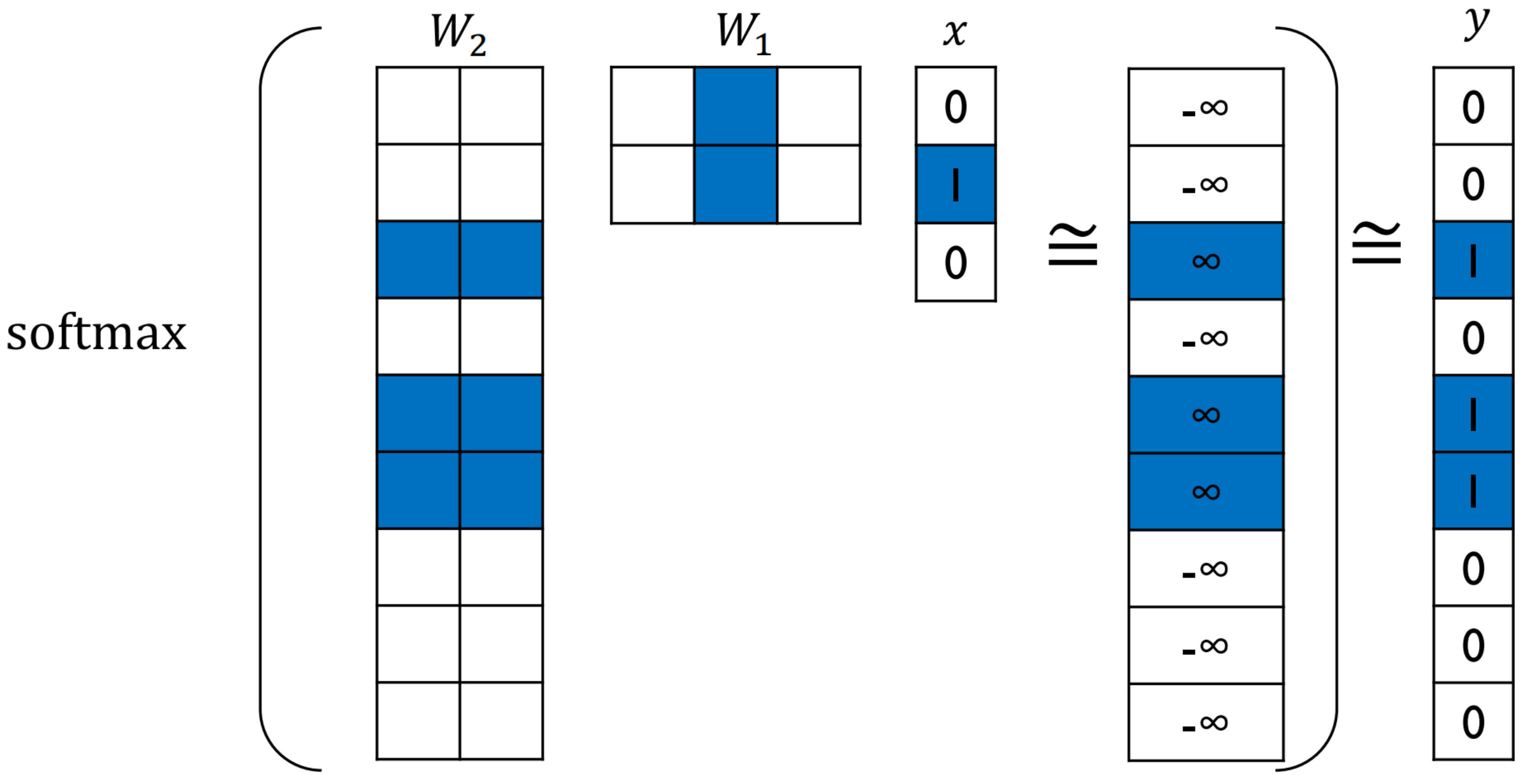
Doc2Vec (Paragraph2Vec)

- Doc2Vec can encode any other types of attributes associated with text data.
  - You can use multi-hot vector to represent associated attributes.
  - ((study, female, 10s), math)
  - Total loss = α loss_word + β loss_gender + γ loss_age



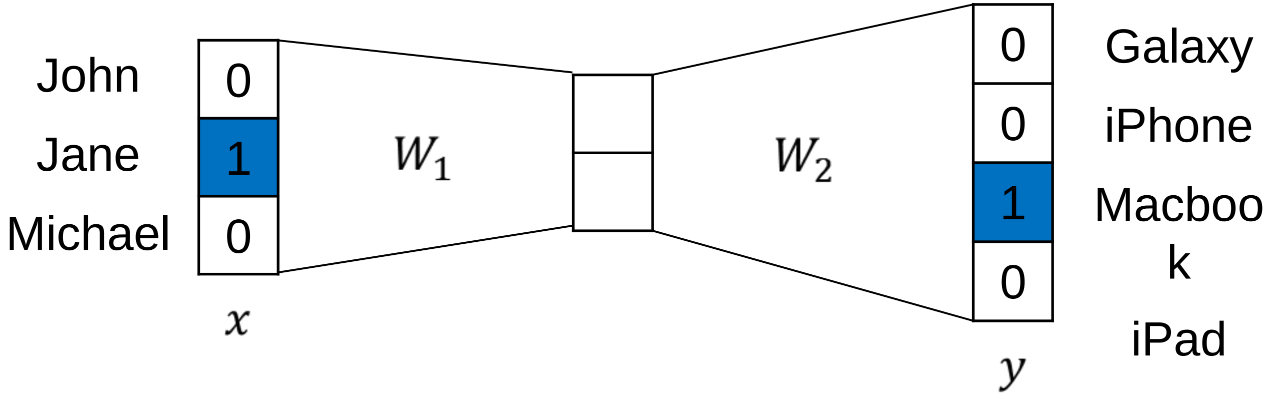
Doc2Vec (Paragraph2Vec)

- Doc2Vec can encode any other types of attributes associated with text data.
  - (study, (math, female, 10s)
  - The ‘math’, ‘female’, and ‘10s’ vector in W_2 and the ‘study’ vector in W_1 should have high inner product value.



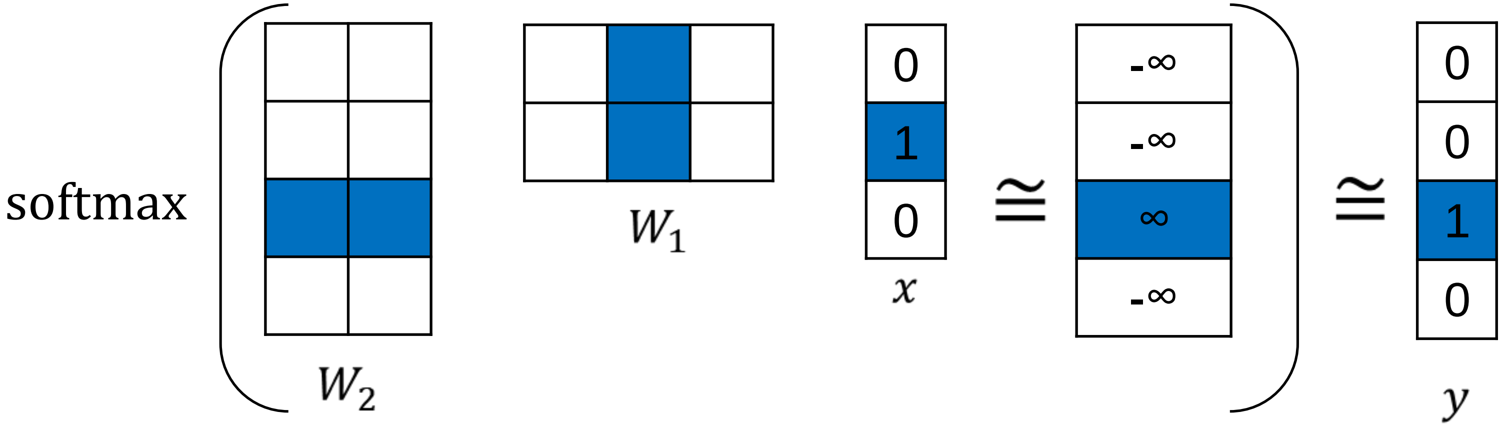
Other Applications of Word2Vec

- Word2Vec can embed different types of entities in a common vector space.
- It is similar to a collaborative filtering approach in recommender systems.
- User vocabulary: {John, Jane, Michael}
- Item vocabulary: {Galaxy, iPhone, Macbook, iPad}



Other Applications of Word2Vec

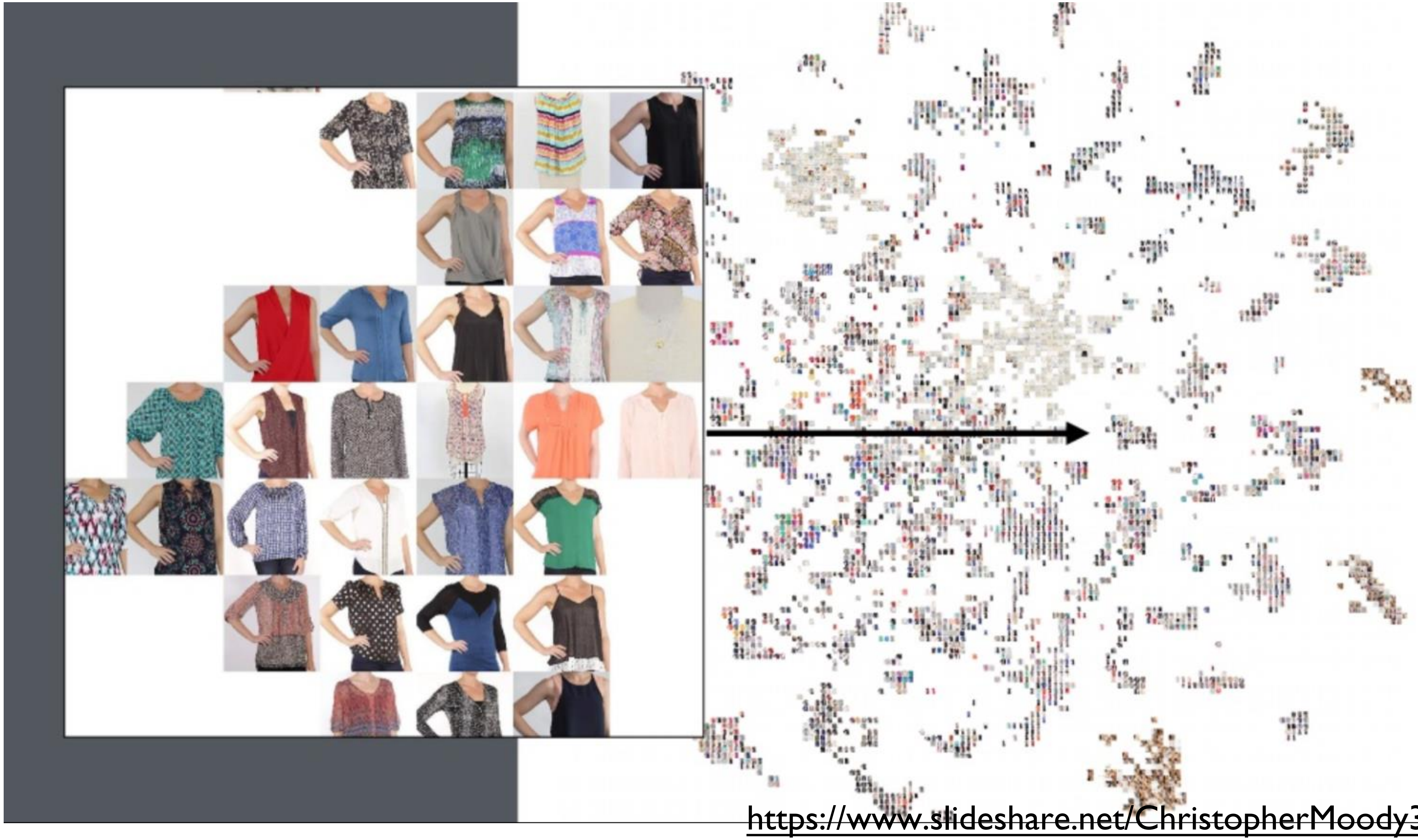
- User vocabulary: {John, Jane, Michael}
- Item vocabulary: {Galaxy, iPhone, Macbook, iPad}
- If the words 'Jane' and 'Mac' co-occur frequently, then the 'Jane' vector in W_1 and the 'Mac' vector in W_2 would have a high inner product value.



Other Applications of Word2Vec

https://www.slideshare.net/ChristopherMoody3/word2vec-lda-and-introducing-a-new-hybrid-algorithm-lda2vec

Other Applications of Word2Vec



https://www.slideshare.net/ChristopherMoody3/word2vec-lda-and-introducing-a-new-hybrid-algorithm-lda2vec

References

Codes

- Python: https://radimrehurek.com/gensim/models/word2vec.html
- C++: https://code.google.com/archive/p/word2vec/
- FastText: https://github.com/facebookresearch/fastText

Useful resources

- https://shuuki4.wordpress.com/2016/01/27/word2vec-%EA%B4%80%EB%A0%A8-%EC%9D%B4%EB%A1%A0-%EC%A0%95%EB%A6%AC/
- https://code.facebook.com/posts/1438652669495149/fair-open-sources-fasttext/
- https://ronxin.github.io/wevi/
- https://www.lucypark.kr/slides/2015-pyconkr/

Other General References

- Stanford University CS224n: Deep Learning for Natural Language Processing
- https://arxiv.org/pdf/1705.00108.pdf
- https://arxiv.org/abs/1602.02410
- https://blog.openai.com/language-unsupervised/
- https://nlp.stanford.edu/seminar/details/jdevlin.pdf