Word Embeddings

Mithun Prasad, PhD
miprasad@Microsoft.com

Microsoft

The value of science is not to make things complex, but to find the inherent simplicity.

Microsoft

Vector Space Models

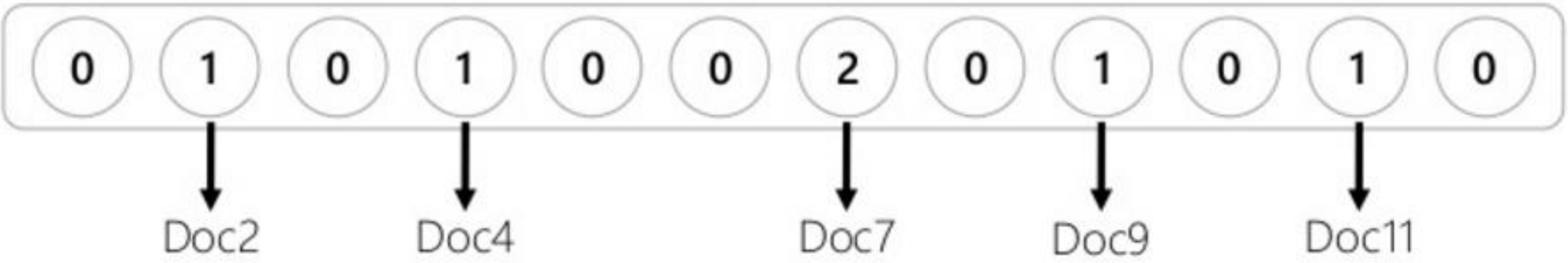
Represent an item (e.g., word) as a Vector of numbers.

banana	0	1	0	1	0	0	2	0	1	0	1	0

Vector Space Models

Represent an item (e.g., word) as a Vector of numbers.

banana



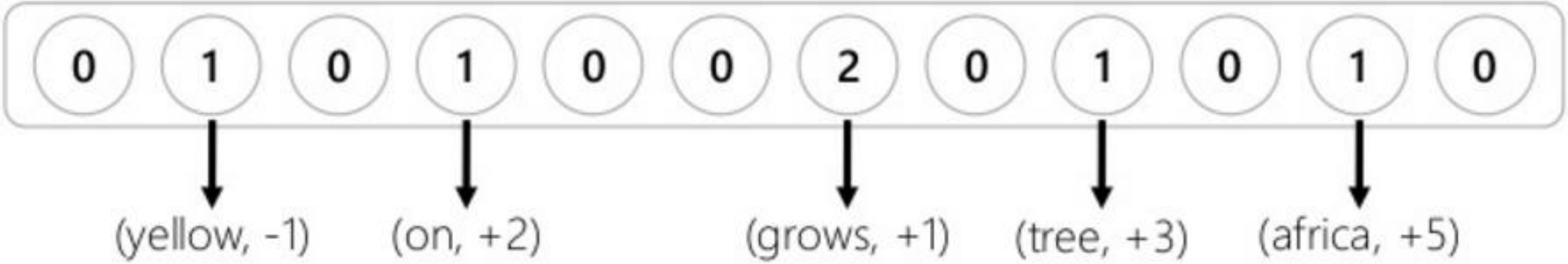
The vector can correspond to documents in which the word occurs.

Microsoft

Vector Space Models

Represent an item (e.g., word) as a Vector of numbers.

banana



The vector can correspond to neighboring word context.

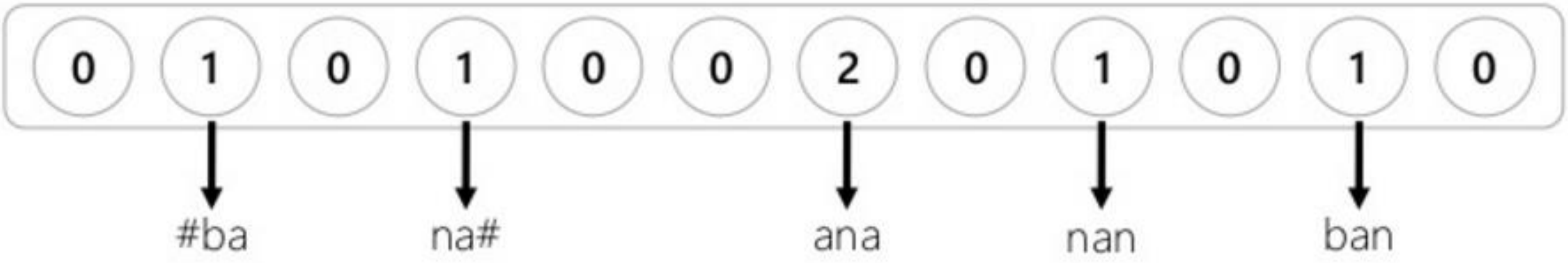
$$e.g., \text{“}yellow\ \ banana\ \ grows\ \ on\ \ trees\ \ \ in\ \ africa\text{”}$$

-1 0 +1 +2 +3 +4 +5

Vector Space Models

Represent an item (e.g., word) as a Vector of numbers.

banana



The vector can correspond to character trigrams in the word.

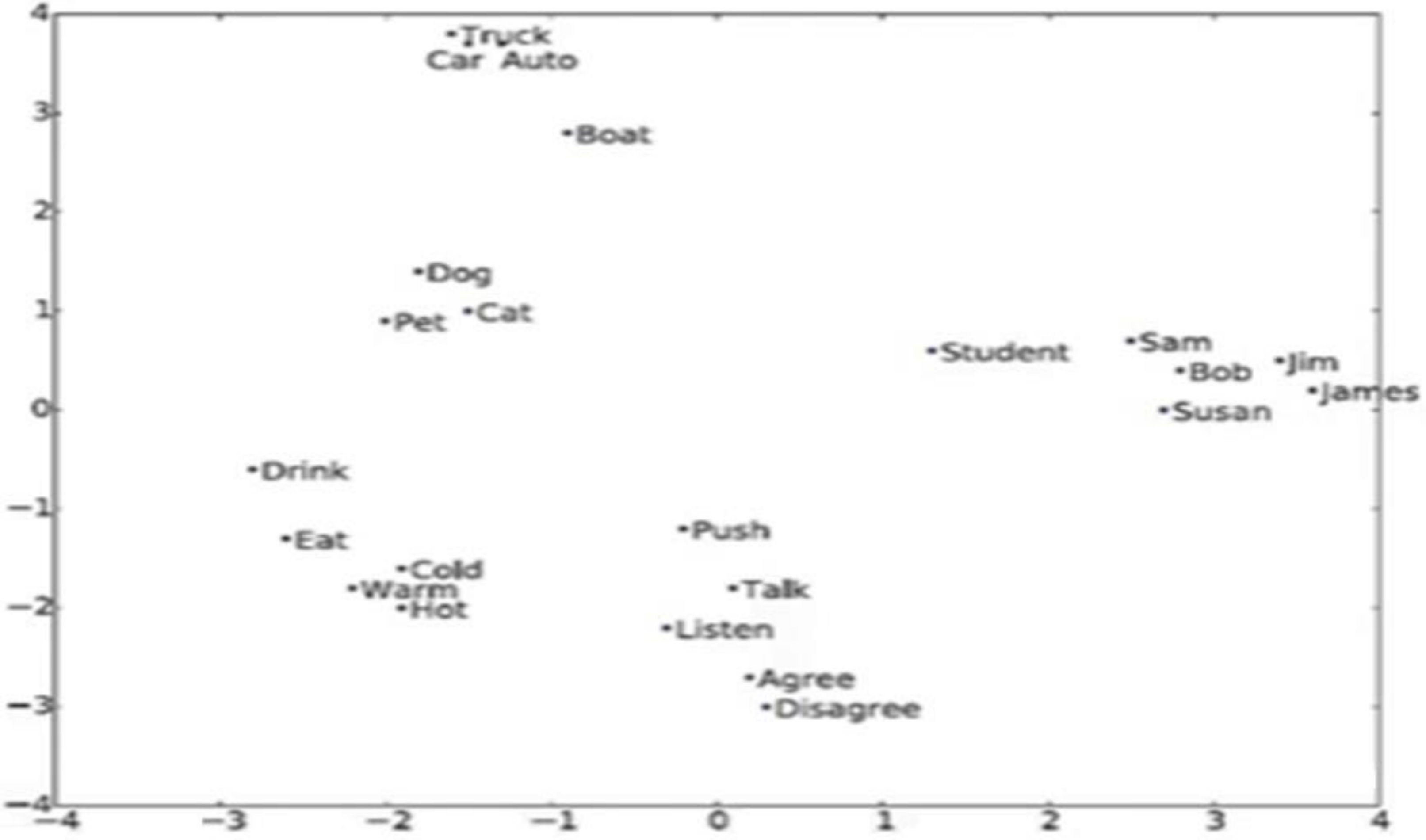
Notions Of Relatedness

Comparing two vectors (e.g., using cosine similarity) estimates how similar the two words are. However, the notions of relatedness depends on what vector representation you have chosen for the words.

seattle similar to denver?	or	seattle similar to seahawks?
Because they are both cities.		Because "seattle seahawks". (Go seahawks!)

Important note: In previous slides I showed raw counts. They should either be normalized (e.g, using pointwise-mutual information) or (matrix) factorized. More on that later...

Word Embeddings



Word vector projected by their two principal components.

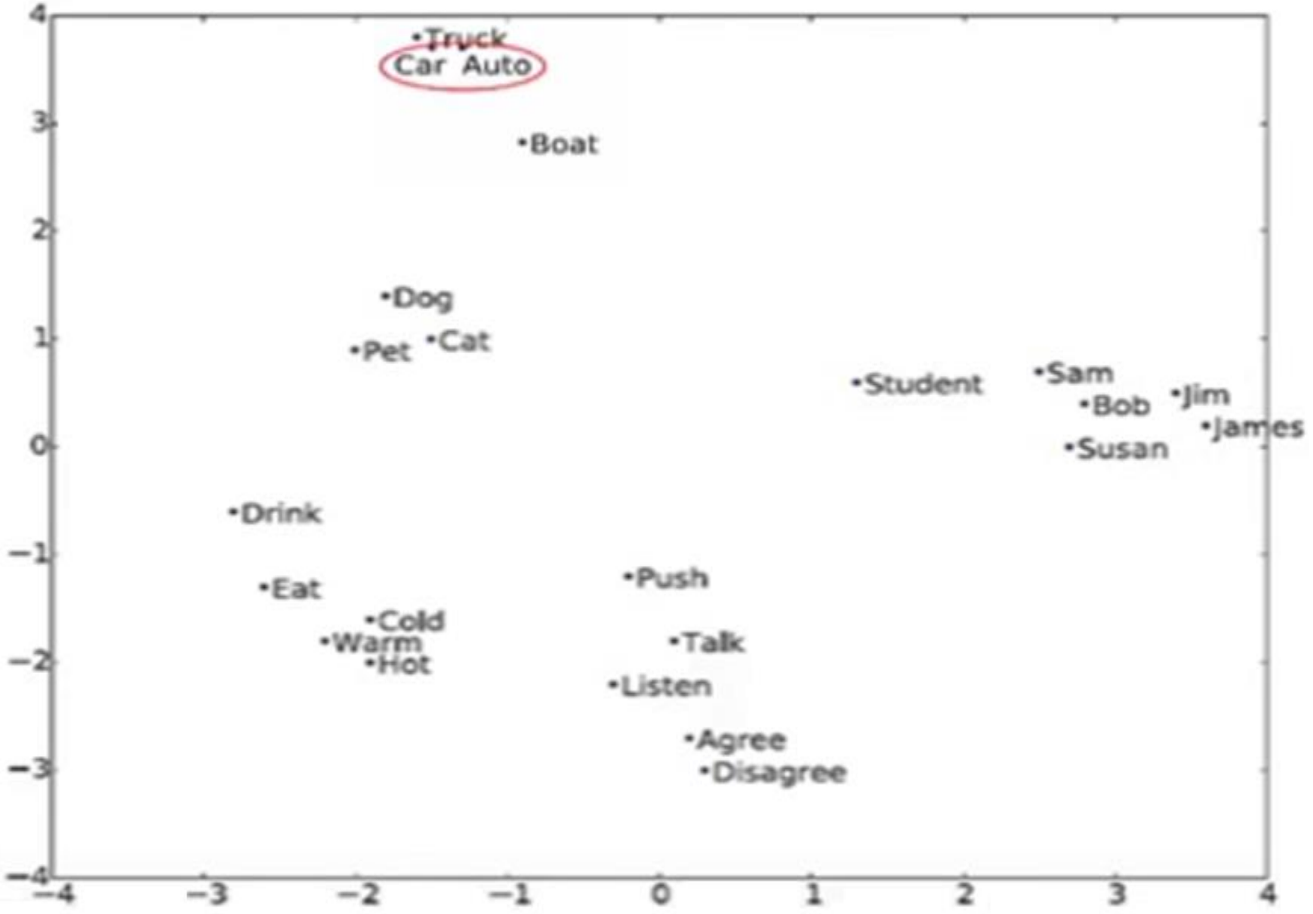
Microsoft

Word Embeddings

Synonym

A word having the same or nearly the same meaning as another word in certain contexts.

Beautiful: Attractive, Pretty, Lovely, Stunning



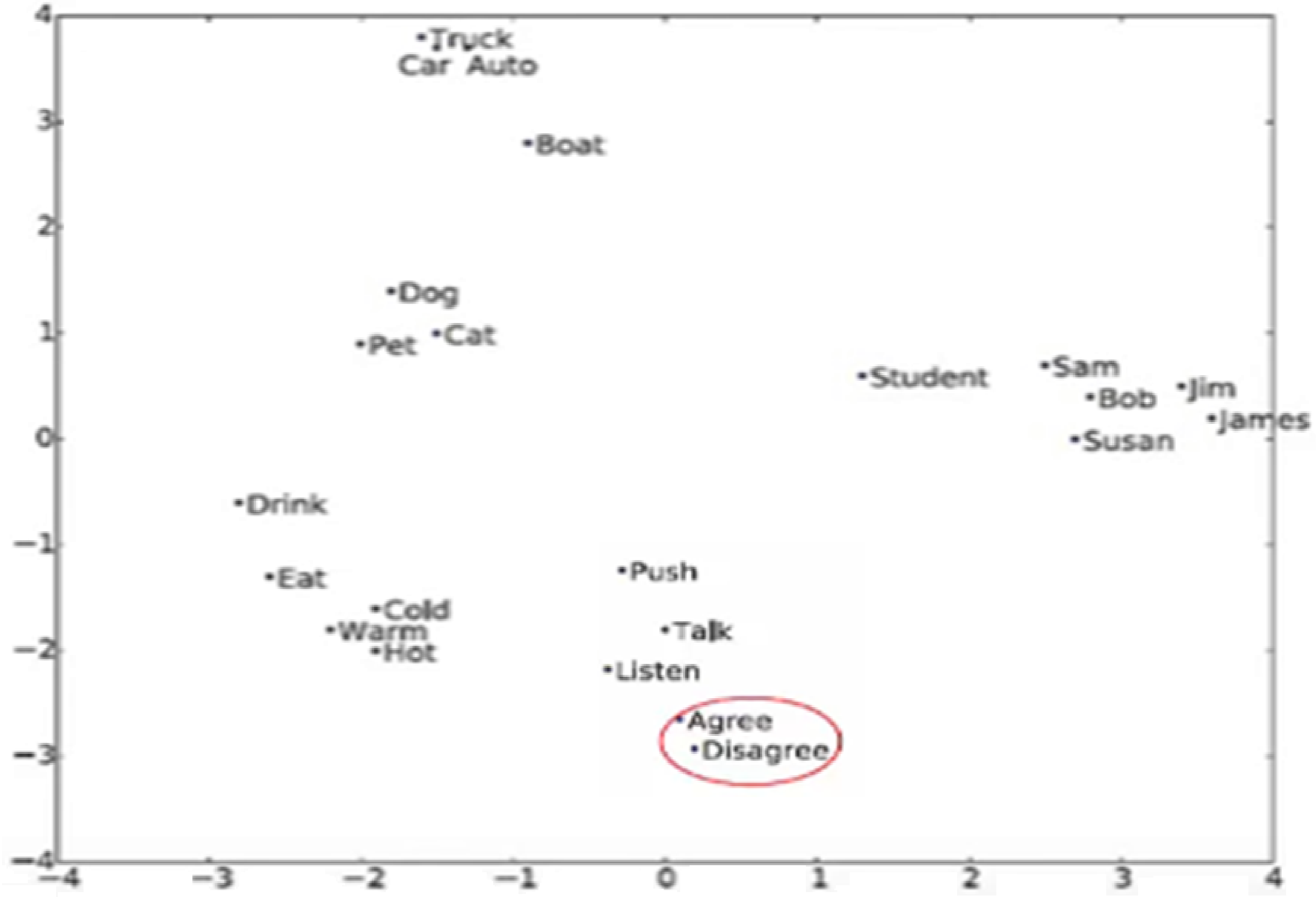
Microsoft

Word Embeddings

Antonym

A word that is opposite in context of another word although similar in other respects.

Afraid – Confident



Microsoft

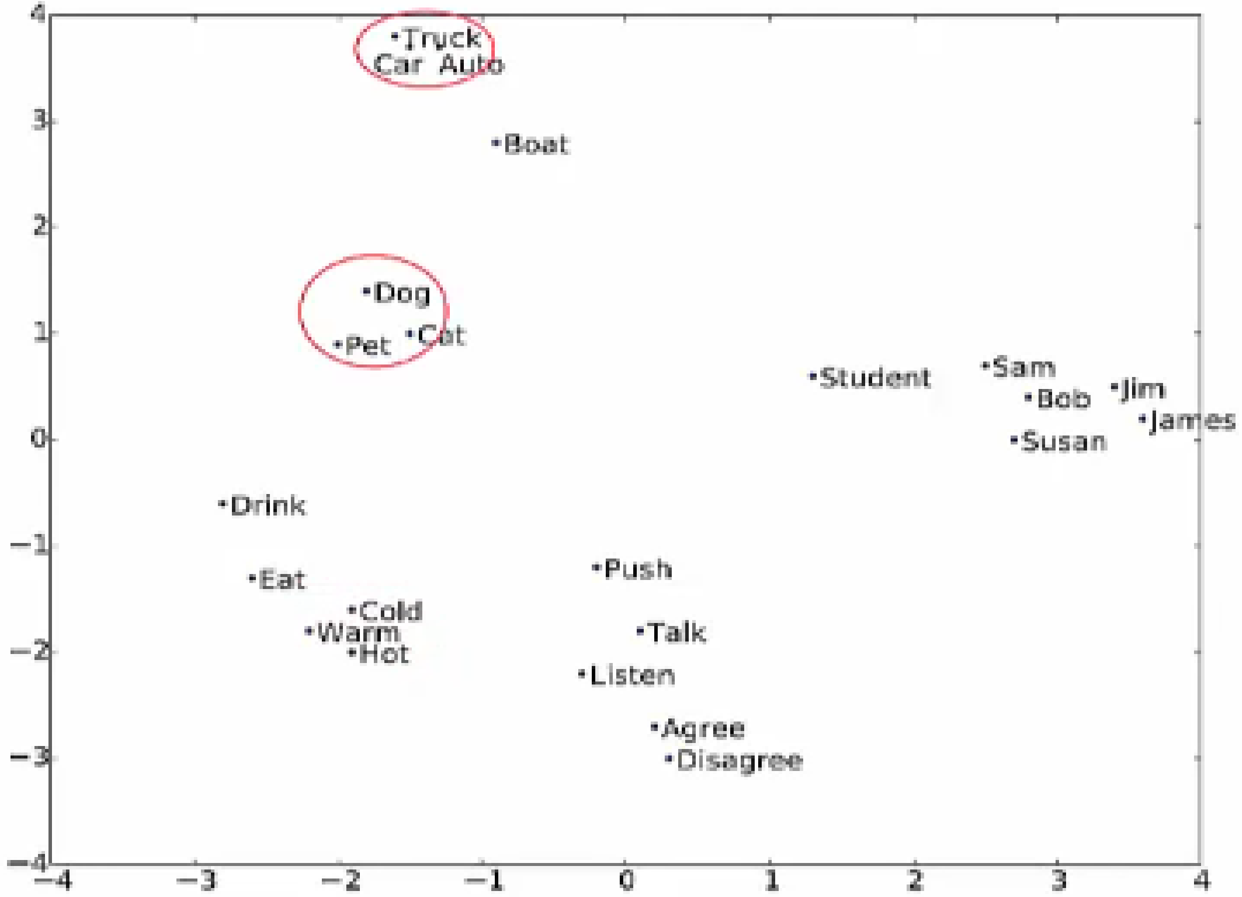
Word Embeddings

Hypernym

A word whose meaning includes the meanings of other words.

Flower is a hypernym of daisy and rose.

Truck is a type of car.



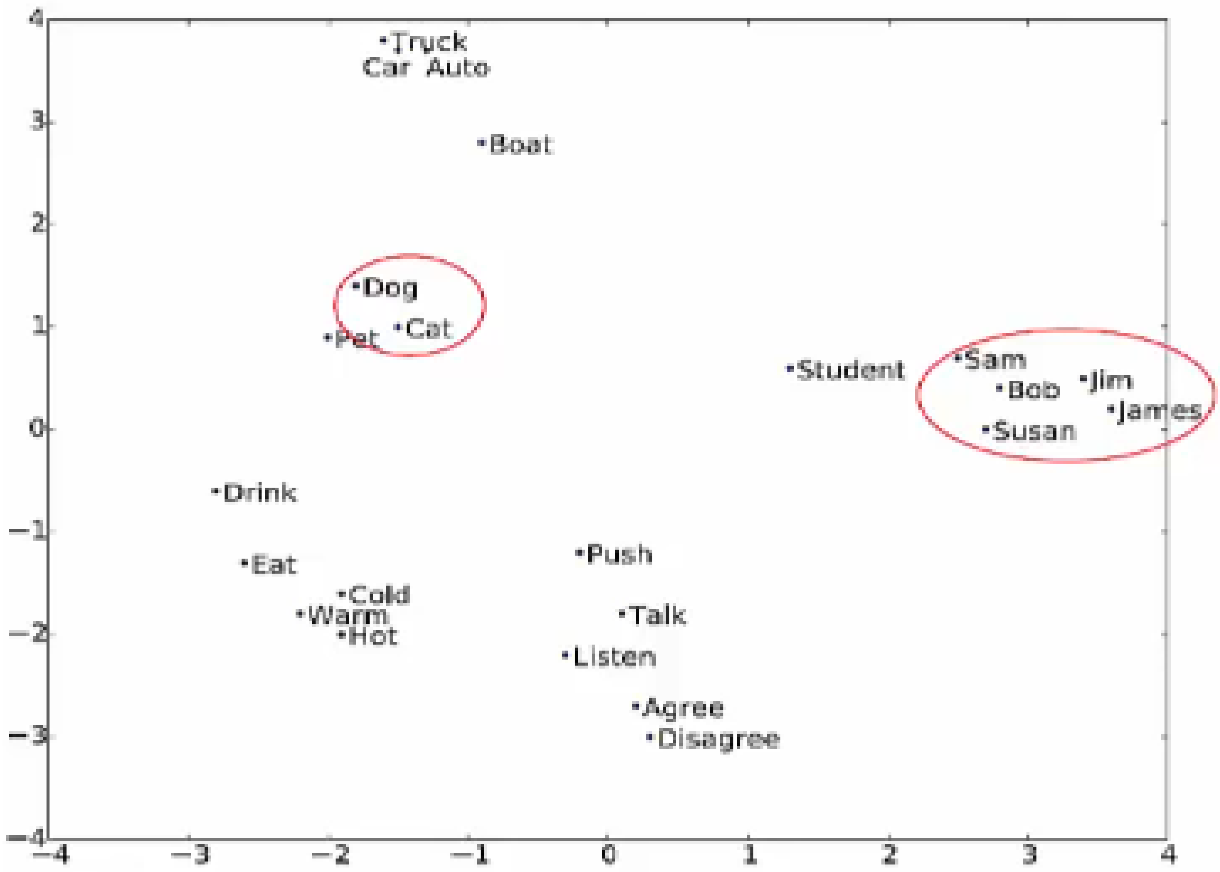
Microsoft

Word Embeddings

Hyponym

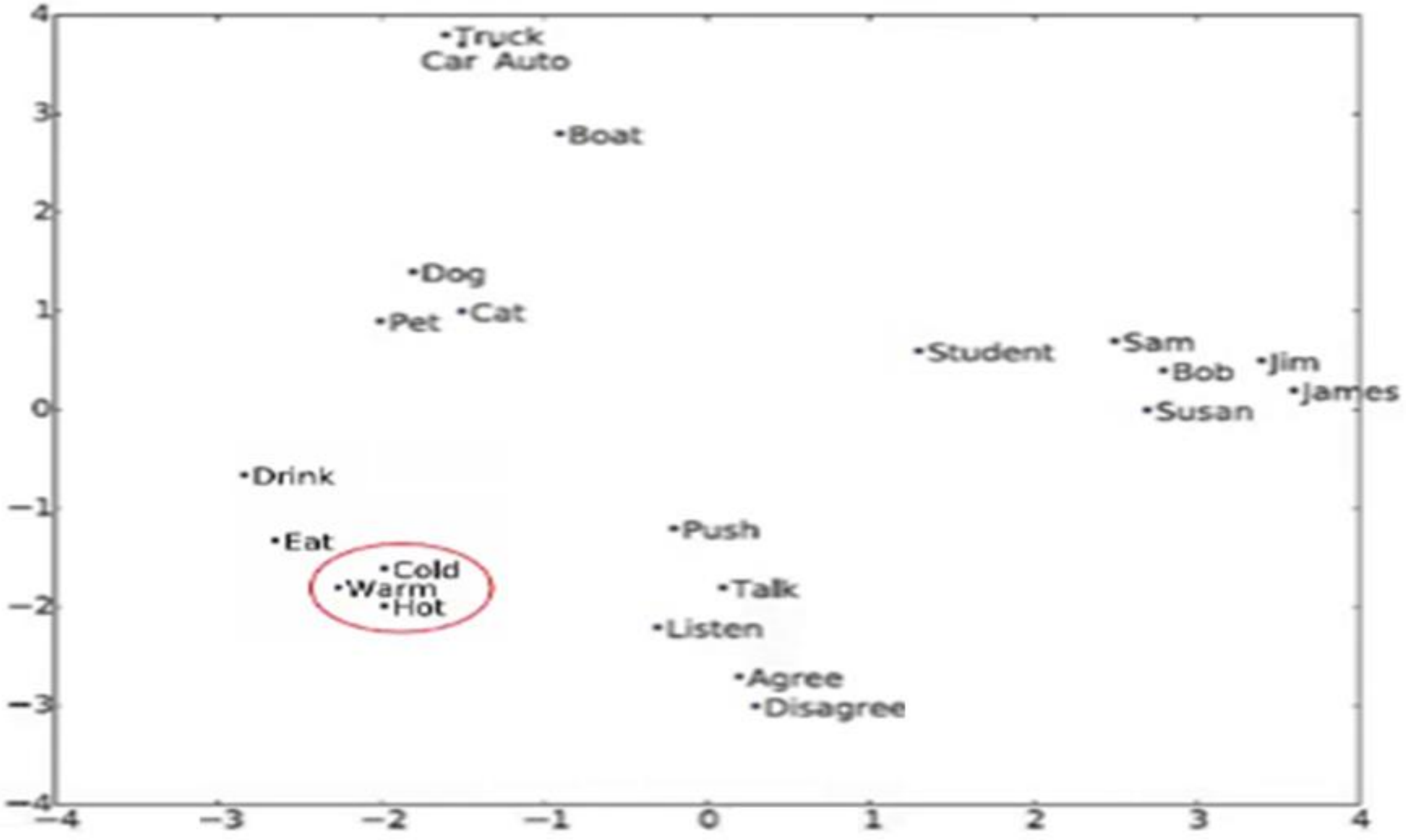
A term used to designate a particular member of a broader class.

Cat and rabbit are co-hyponyms of the hypernym animal.



Microsoft

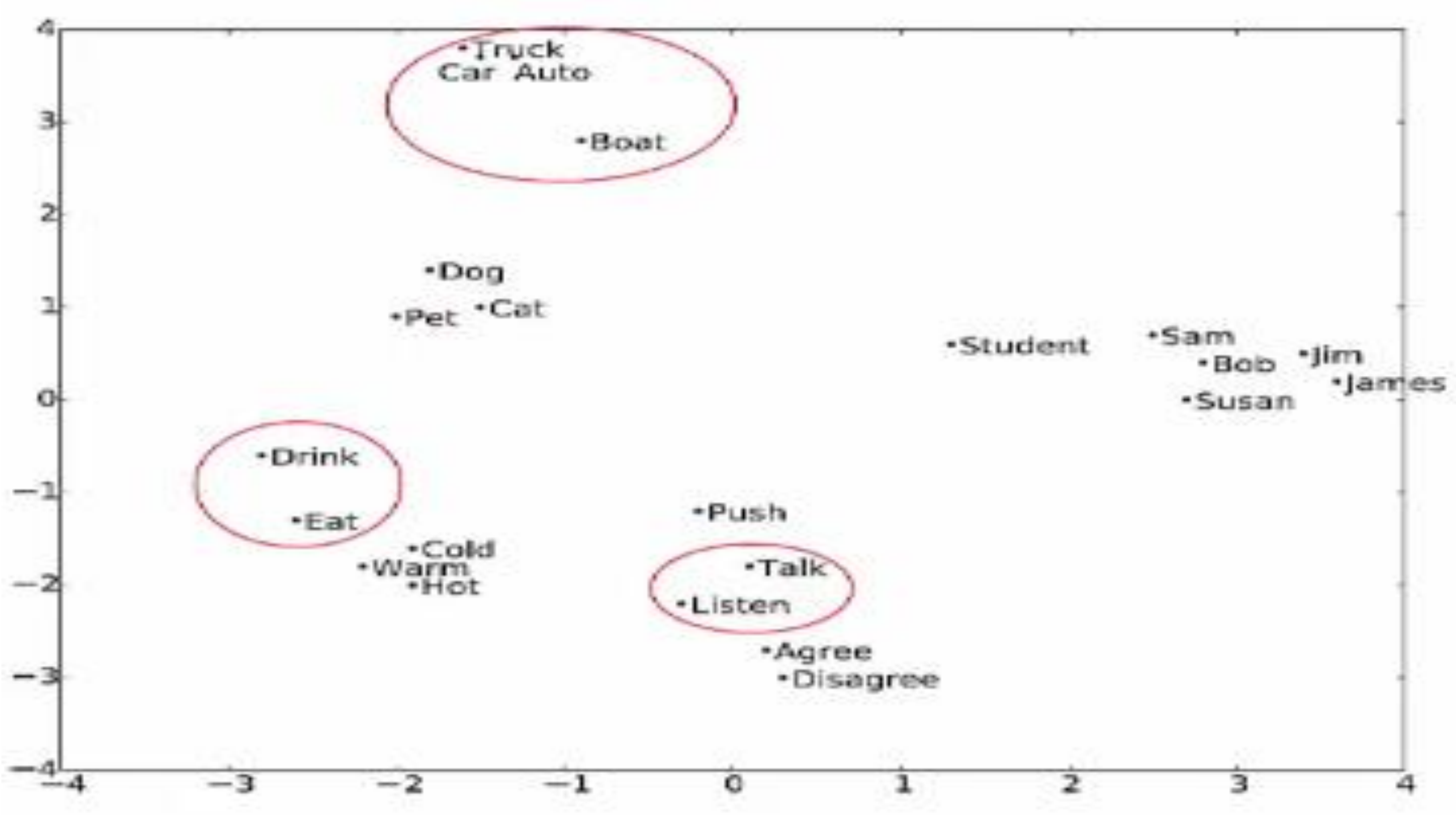
Word Embeddings



Words are values on a scale: hot, warm, cold.

Microsoft

Word Embeddings



Words appear in similar contexts.

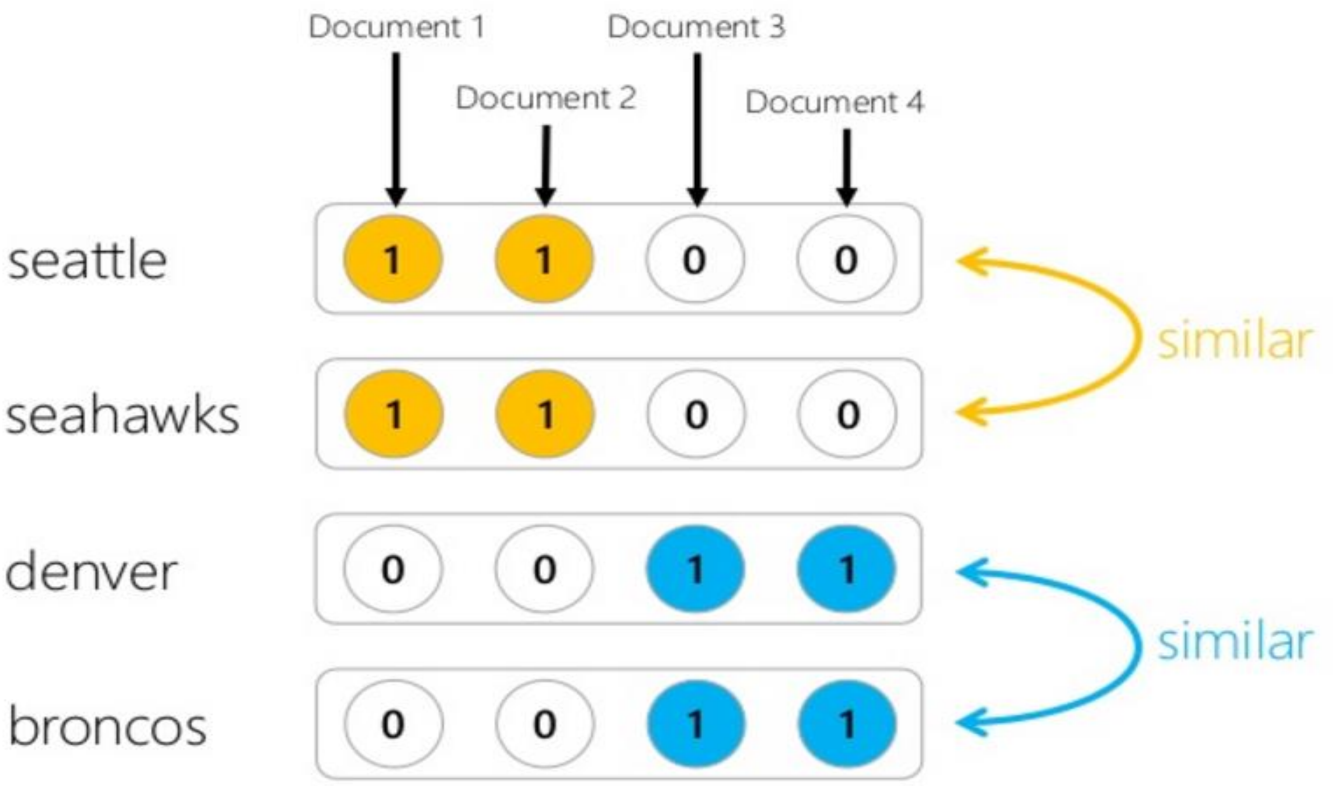
Microsoft

Let's Consider The Following Example…

We have four (tiny) documents,

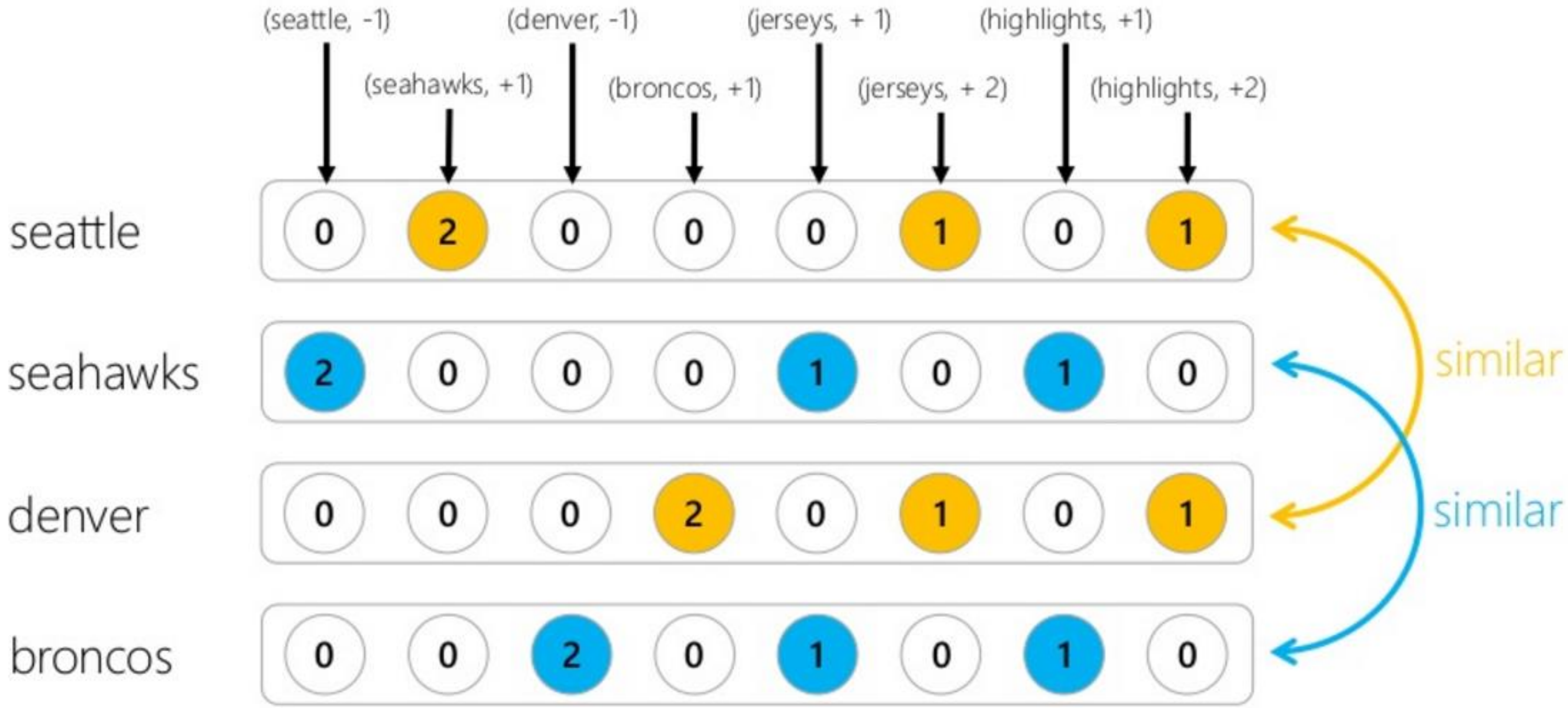
Document 1 : “seattle seahawks jerseys”
Document 2 : “seattle seahawks highlights”
Document 3 : “denver broncos jerseys”
Document 4 : “denver broncos highlights”

If We Use Document Occurrence Vectors...



In the rest of this talk, we refer to this notion of relatedness as Topical similarity.

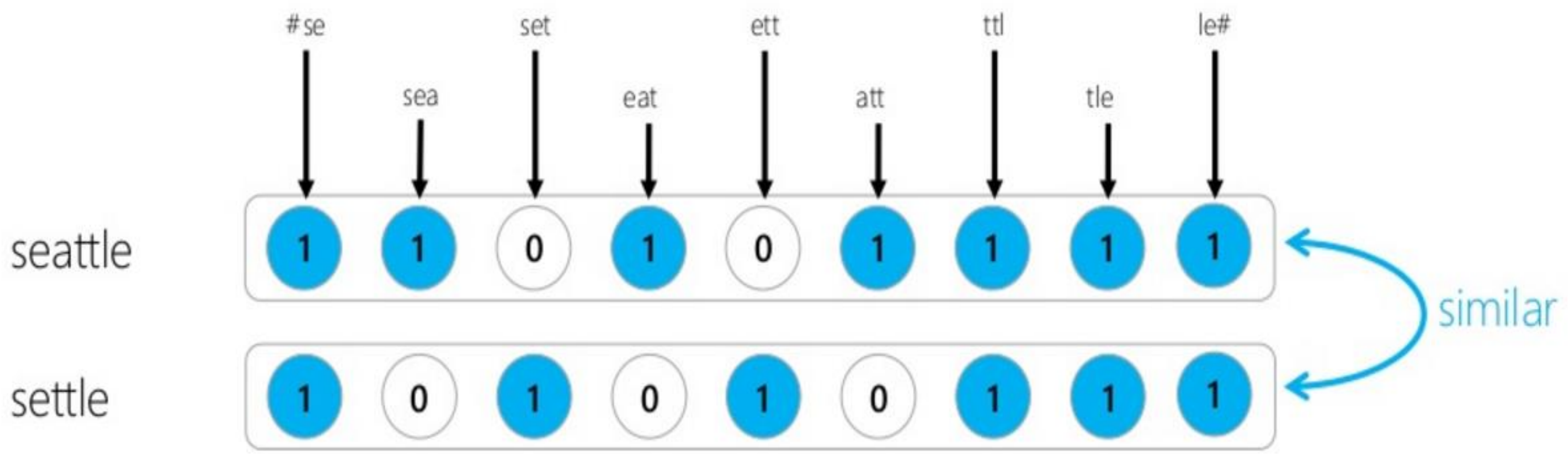
If We Use Word Context Vectors…



In the rest of this talk, we refer to this notion of relatedness as Topical (by-type) similarity.

Microsoft

If We Use Character Trigrams Vectors…



This notion of relatedness is similar to string edit-distance.

DIY: Learning Word Types

Demo at http://bionlp-www.utu.fi/wv_demo/

Compute (Positive) Pointwise Mutual Information for every Word-Context pair.

$$pmi(x, y) \equiv \log \frac{p(x, y)}{p(x)p(y)}$$

Compute the cosine similarity between the context score vectors to estimate word similarity by type.

Enter a word sydney Lookup

Words	Similarity Coefficient
sydney	1
melbourne	0.4376428
brisbane	0.4071144
perth	0.3362517
adelaide	0.2916113
auckland	0.2493333

Enter a word batman Lookup

Words	Similarity Coefficient
batman	1
spiderman	0.1429663
superman	0.137329
ghostbusters	0.1045547
tinkerbell	0.08972809
starwars	0.07744732

Enter a word java Lookup

Words	Similarity Coefficient
java	1
c	0.1601557
javascript	0.145963
powershell	0.1096152
python	0.09570167
vb	0.0907691

Enter a word pasta Lookup

Words	Similarity Coefficient
pasta	1
spaghetti	0.1822345
lasagna	0.1541065
macaroni	0.1090949
salad	0.1030677
casserole	0.09800283

Word Analogy Task

man is to woman as king is to _?
good is to best as smart is to _?
china is to beijing as russia is to _?

Turns out the word-context based vector model we just learnt is good for such analogy tasks,

$$[king] - [man] + [woman] \approx [queen]$$

Levy, Goldberg, and Israel, Linguistic Regularities in Sparse and Explicit Word Representations, CoNLL.2014

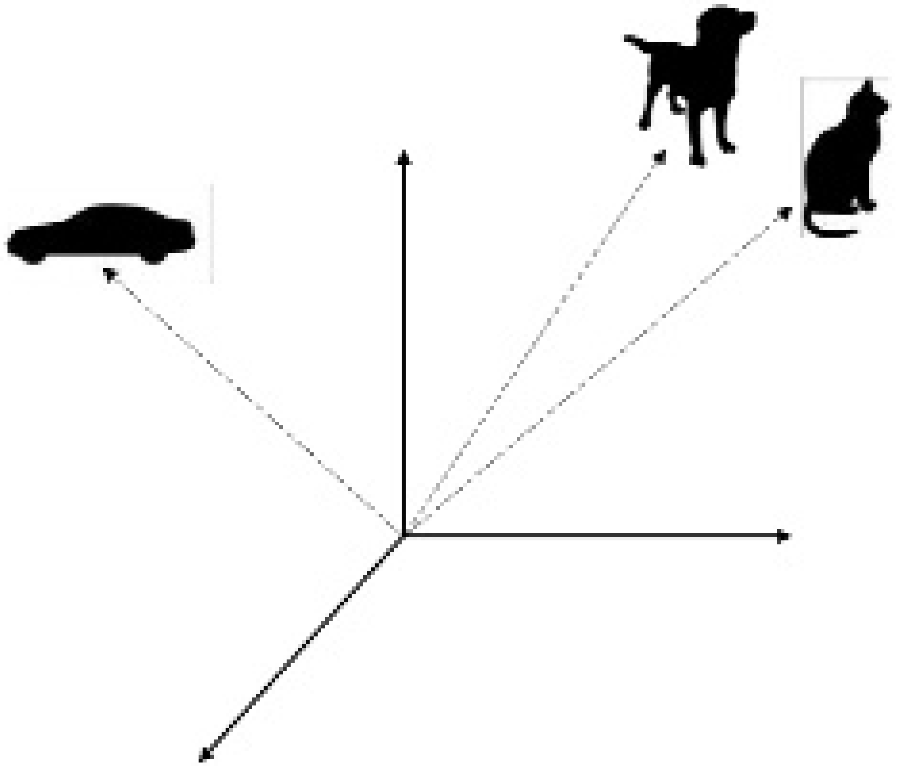
Microsoft

Embeddings

The vectors we have been discussing so far are very high-dimensional (thousands, or even millions) and sparse.

But there are techniques to learn lower-dimensional dense vectors for words using the same intuitions.

These dense vectors are called embeddings.



Learning Dense Embeddings

Matrix Factorization

Factorize word-context matrix.

	$Context_1$	$Context_1$	....	$Context_k$
$Word_1$				
$Word_2$				
⋮				
$Word_n$				

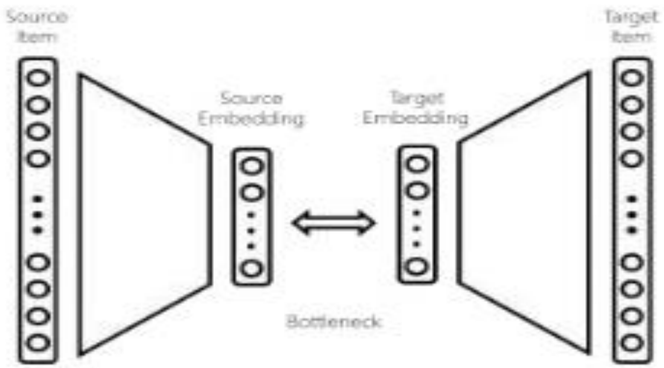
E.g.,

LDA (Word-Document),

GloVe (Word-NeighboringWord)

Neural Networks

A neural network with a bottleneck, word and context as input and output respectively.



E.g.,

Word2vec (Word-Neighboringword)

Deerwester, Dumais, Landauer, Furnas, and Harshman, Indexing by latent semantic analysis, JASIS,1990.

Pennington, Socher, and Manning, GloVe: Global Vectors for Word Representation, EMNLP, 2014. Mikolov, Sutskever, Chen, Corrado, and Dean, Distributed resentations of Words and phrases and their compositionality, NIPS, 2013.

Exercise

Both Word2vec and Glove define context as the neighboring word only, without considering the distance from the current word.

How does this change the relationship that is learnt by the embedding space?

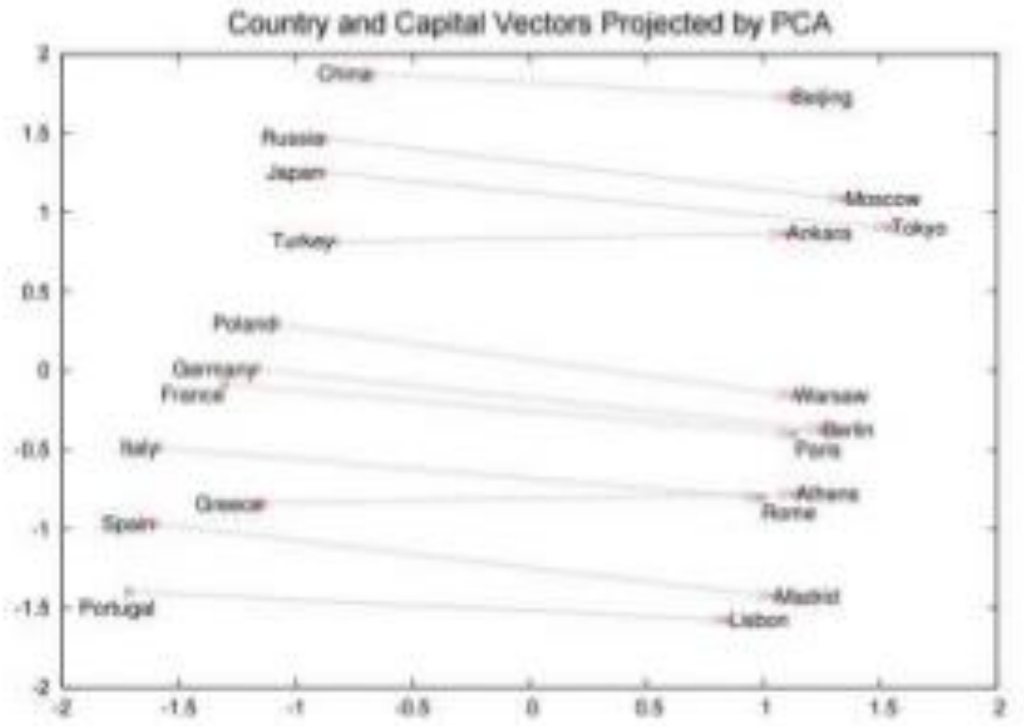
Microsoft

How Do Word Analogies Work?

Visually, the vector {china → beijing} turns out to be almost parallel to the vector {russia → moscow}.

But if you aren't queasy about reading a lot of equations, read the following paper…

Arora, et al. [RAND-WALK: A Latent variable Model approach Word Embeddings](), 2015.



Mikolov, Sutskever, Chen, Corrado, and Dean, [Distributed representations of words and phrases and their compositionality](), NIPS, 2013.

Microsoft

Word Embeddings For Document Ranking

Traditional IR uses Term matching,
→# of times the doc says Albuquerque

We can use word embeddings to Compare all-pairs of query-document Terms,

→# of terms in the doc that relate to Albuquerque

Albuquerque is the most populous city in the U.S. state of New Mexico. The high-altitude city serves as the county seat of Bernalillo County, and it is situated in the central part of the state, straddling the Rio Grande. The city population is 557,169 as of the July 1, 2014, population estimate from the United States Census Bureau, and ranks as the 32nd-largest city in the U.S. The Metropolitan Statistical Area (or MSA) has a population of 902,797 according to the United States Census Bureau's most recently available estimate for July 1, 2013.

Passage *about* Albuquerque

Allen suggested that they could program a BASIC interpreter for the device; after a call from Gates claiming to have a working interpreter, MITS requested a demonstration. Since they didn't actually have one, Allen worked on a simulator for the Altair while Gates developed the interpreter. Although they developed the interpreter on a simulator and not the actual device, the interpreter worked flawlessly when they demonstrated the interpreter to MITS in Albuquerque, New Mexico in March 1975; MITS agreed to distribute it, marketing it as Altair BASIC.

Passage *not about* Albuquerque

Nalisnick, Mitra , Craswell, and Caruana, [Improving Document Ranking with Dual Word Embeddings](), in www, 2016.
Mitra, Nalisnick Craswell, and Caruana, [A Dual Embeddings Space model for Document Ranking]() arxlv:160201137, 2016

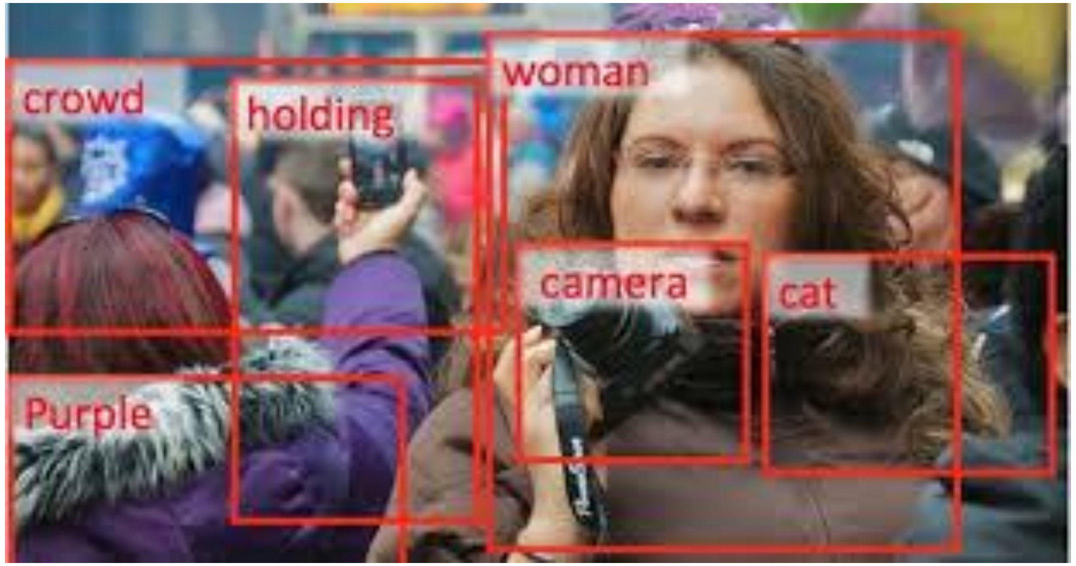
What's Next?

Train your own or use a pre-trained embedding

- Word2vec
- Word2vec trained on queries
- GloVe

Microsoft

Remember These Are Exciting Times…



Fang et.al., From Captions to Visual Concepts and Back, CVPR, 2015.

Human: What is the purpose of life?
Machine: To serve the greater good.
Human: What is the purpose of living?
Machine: To live forever.
Human: What is the purpose of existence?
Machine: To find out what happens when we get to the planet earth.
Human: Where are you now?
Machine: I'm in the middle of nowhere.
Human: What is the purpose of dying?
Machine: To have a life.
Human: What is the purpose of being intelligent?
Machine: To find out what it is.
Human: What is the purpose of emotions?
Machine: I don't know.

Vinyals et.al. A Neural Conversational Model, ICML.2015.

Microsoft