

Understanding Sentiment and Intent: Transfer Learning for Natural Language

Knox Data | April 18, 2019 | Nikhil Deshmukh

Do computers understand language?

How do we make words and sentences
“machine readable” ?

a.k.a. Natural Language Processing

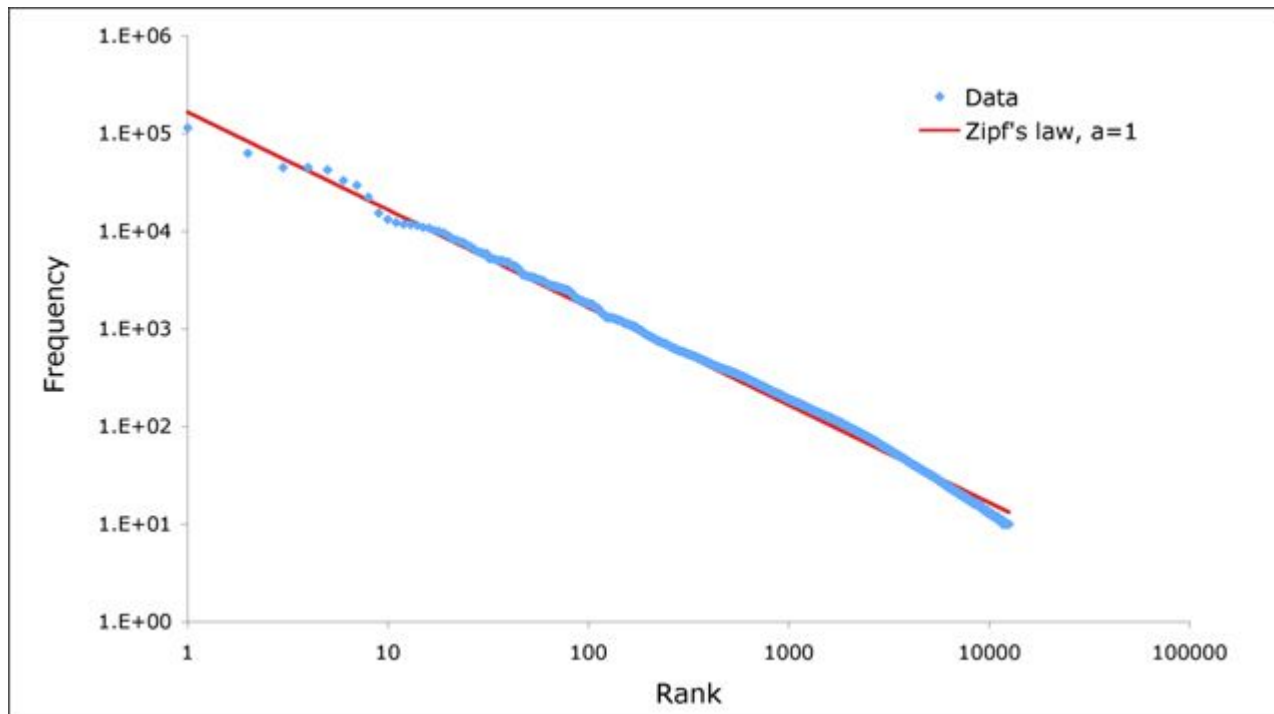
Sentiment: positive vs. negative

Language has statistical properties - Zipf's Law

Frequency proportional
to inverse of rank

$$f = 1/r$$

Rank	Word
1	the
2	be
3	to
4	of
5	and
6	a
7	in
8	that
9	have
10	I



But don't take my word for it...

The Bag of Words Representation

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!



it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
...	...

Doesn't take into account the structure

Word Embeddings

How to reduce dimensionality?

	Masculine	Royalty	Young
Man	1	0	0
Woman	-1	0	0
Boy	1	0	1
Girl	-1	0	1
Prince	1	1	1
Queen	-1	1	0
Monarch	0	1	0

Learn a “dense” vector

Context: surrounding words

WHAT IF MIKE WAS SHORT FOR MICYCLE

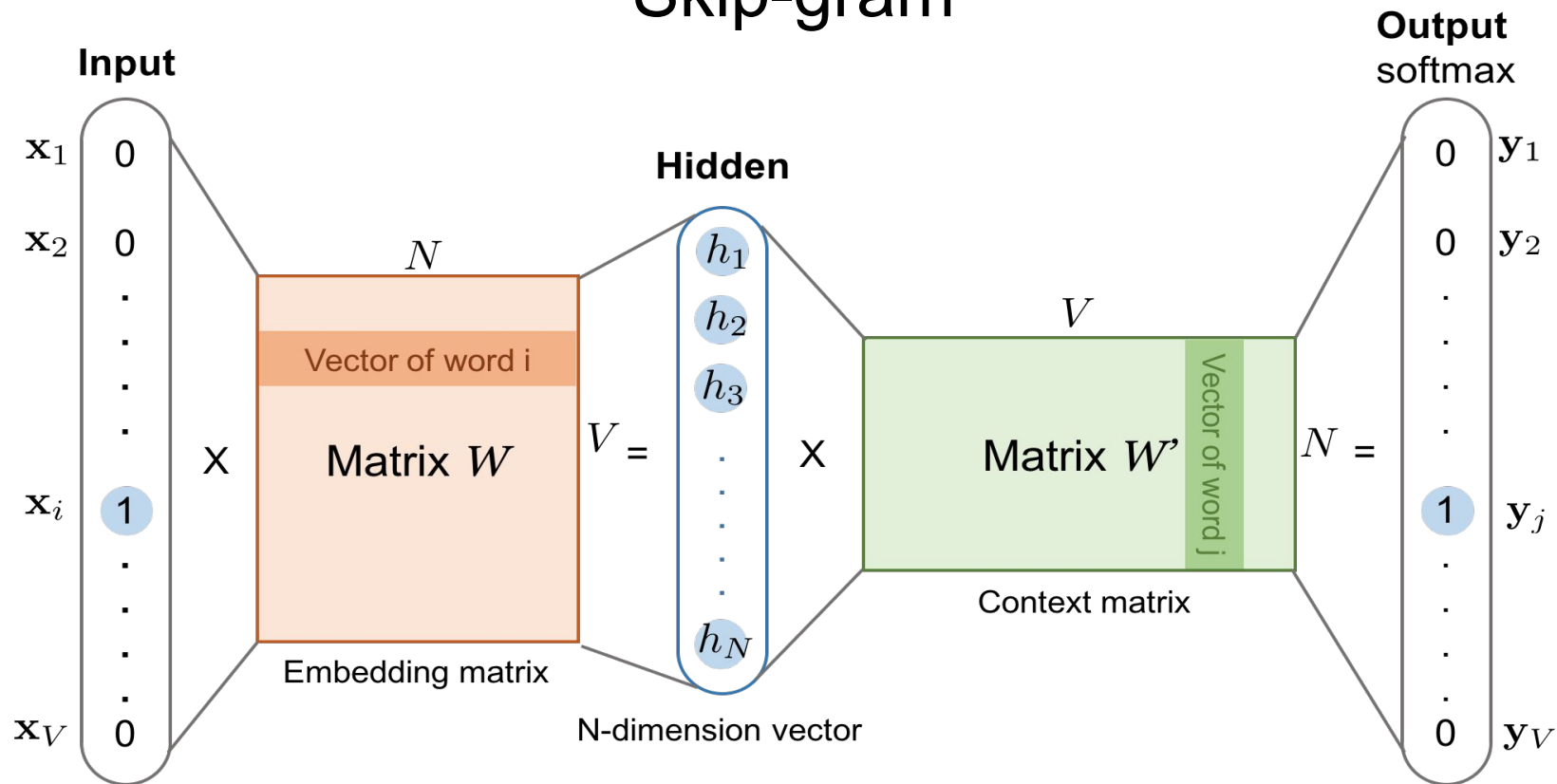
WHAT IF MIKE WAS SHORT FOR MICYCLE

WHAT IF MIKE WAS SHORT FOR MICYCLE

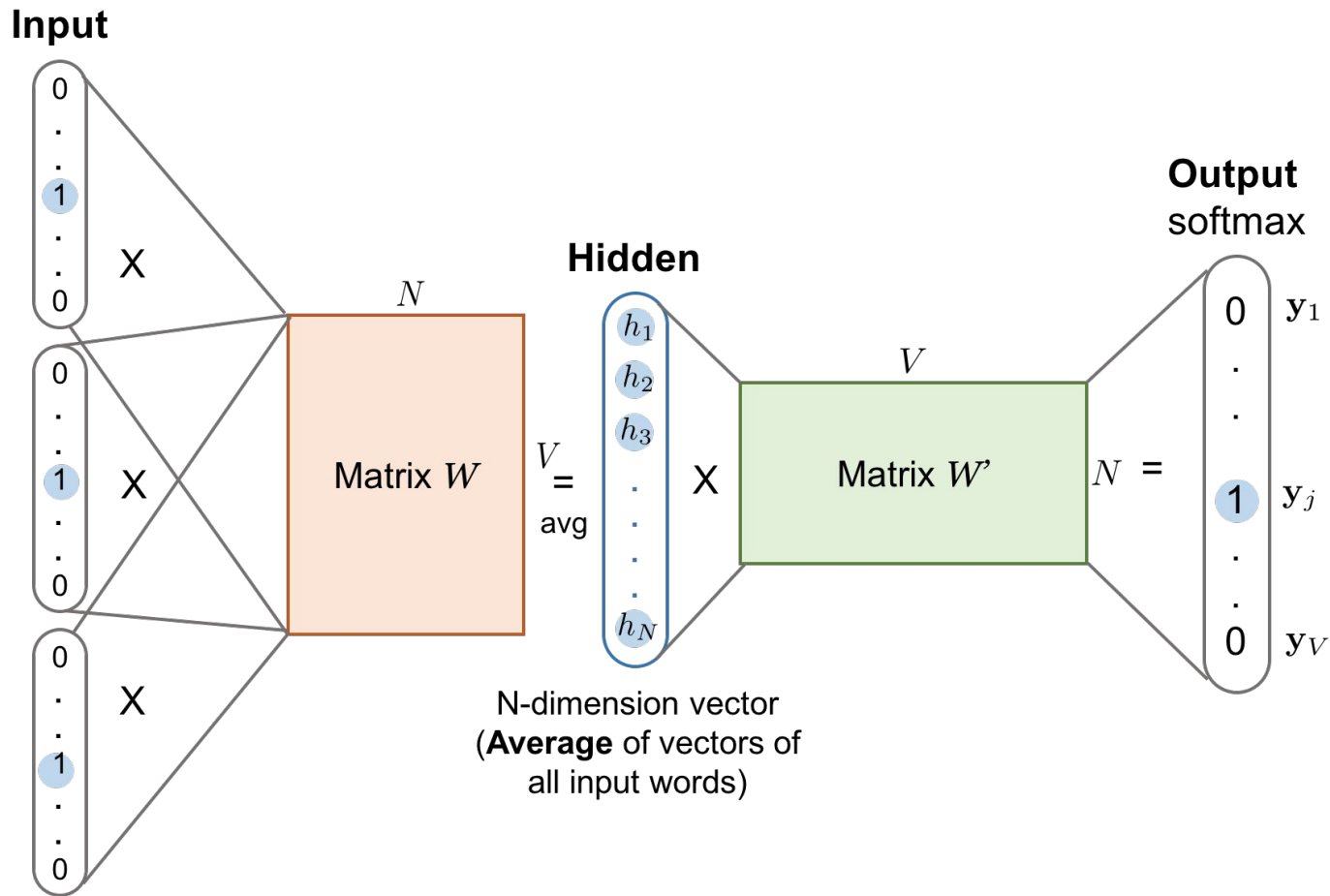
“The man who passes the sentence should swing the sword.” – Ned Stark

Sliding window (size = 5)	Target word	Context
[The man who]	the	man, who
[The man who passes]	man	the, who, passes
[The man who passes the]	who	the, man, passes, the
[man who passes the sentence]	passes	man, who, the, sentence
...
[sentence should swing the sword]	swing	sentence, should, the, sword
[should swing the sword]	the	should, swing, sword
[swing the sword]	sword	swing, the

Skip-gram

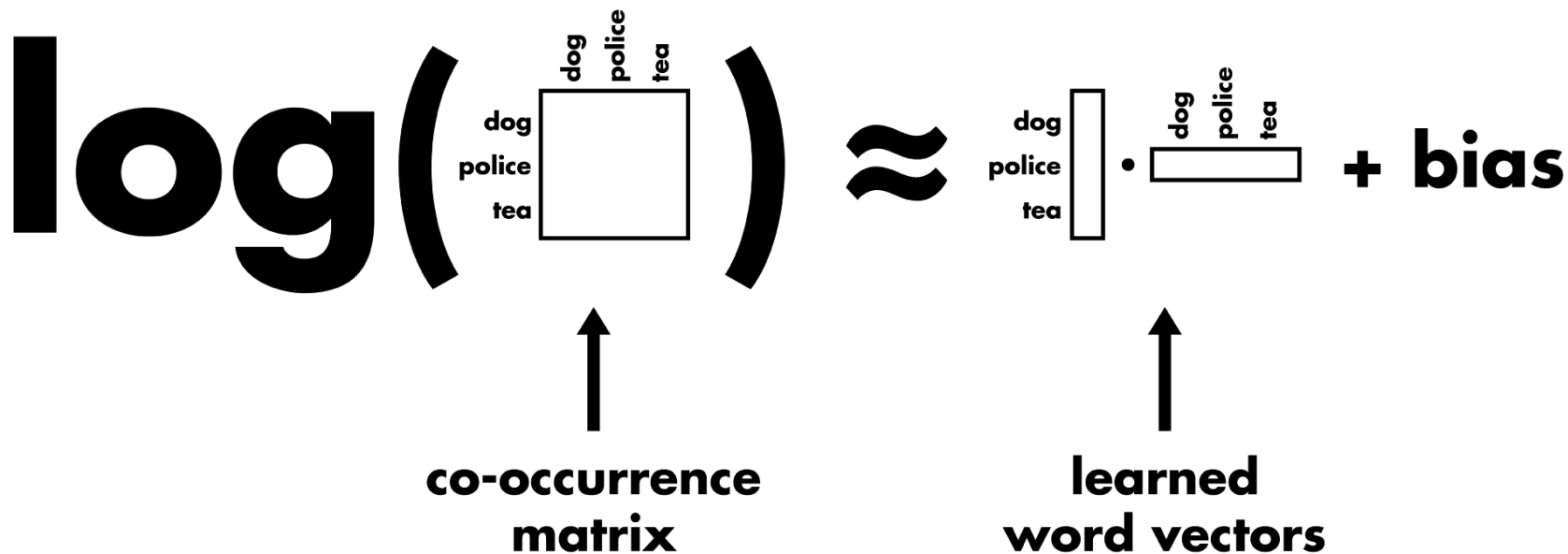


Continuous Bag-of-Words



$$\log(\text{co-occurrence matrix}) \approx \text{learned word vectors} + \text{bias}$$

The diagram illustrates the relationship between a co-occurrence matrix and learned word vectors. On the left, a 3x3 co-occurrence matrix is shown with words 'dog', 'police', and 'tea' as both rows and columns. An arrow points from the label 'co-occurrence matrix' to this matrix. In the center is an approximation symbol \approx . On the right, the equation is shown as the product of a 3x1 column vector (learned word vectors) and a 1x3 row vector (bias), plus a bias term. An arrow points from the label 'learned word vectors' to the column vector. The words 'dog', 'police', and 'tea' are used for both vectors.



```
model.most_similar('king', topn=10)
```

(word, similarity with 'king')

('kings', 0.897245)

('baratheon', 0.809675)

('son', 0.763614)

('robert', 0.708522)

('lords', 0.698684)

('joffrey', 0.696455)

('prince', 0.695699)

('brother', 0.685239)

('aerys', 0.684527)

('stannis', 0.682932)

```
model.most_similar('queen', topn=10)
```

(word, similarity with 'queen')

('cersei', 0.942618)

('joffrey', 0.933756)

('margaery', 0.931099)

('sister', 0.928902)

('prince', 0.927364)

('uncle', 0.922507)

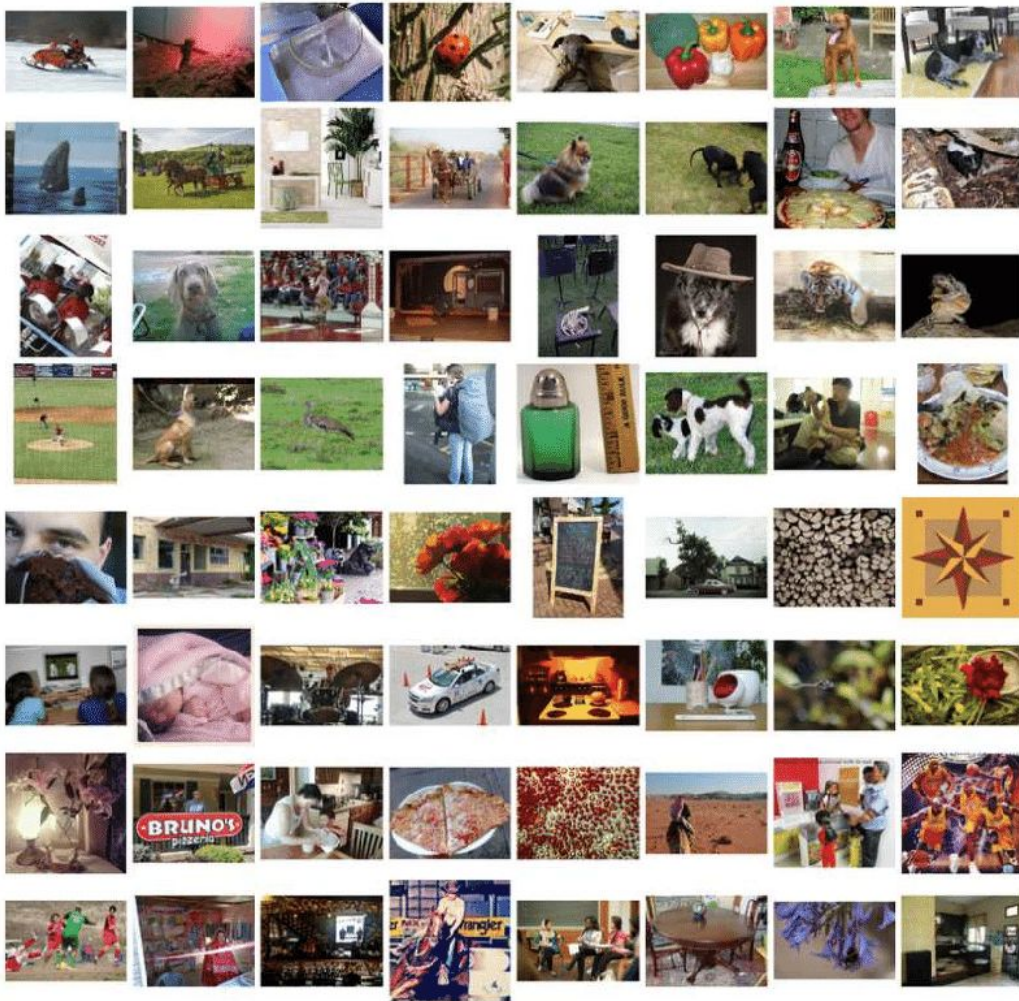
('varys', 0.918421)

('ned', 0.917492)

('melisandre', 0.915403)

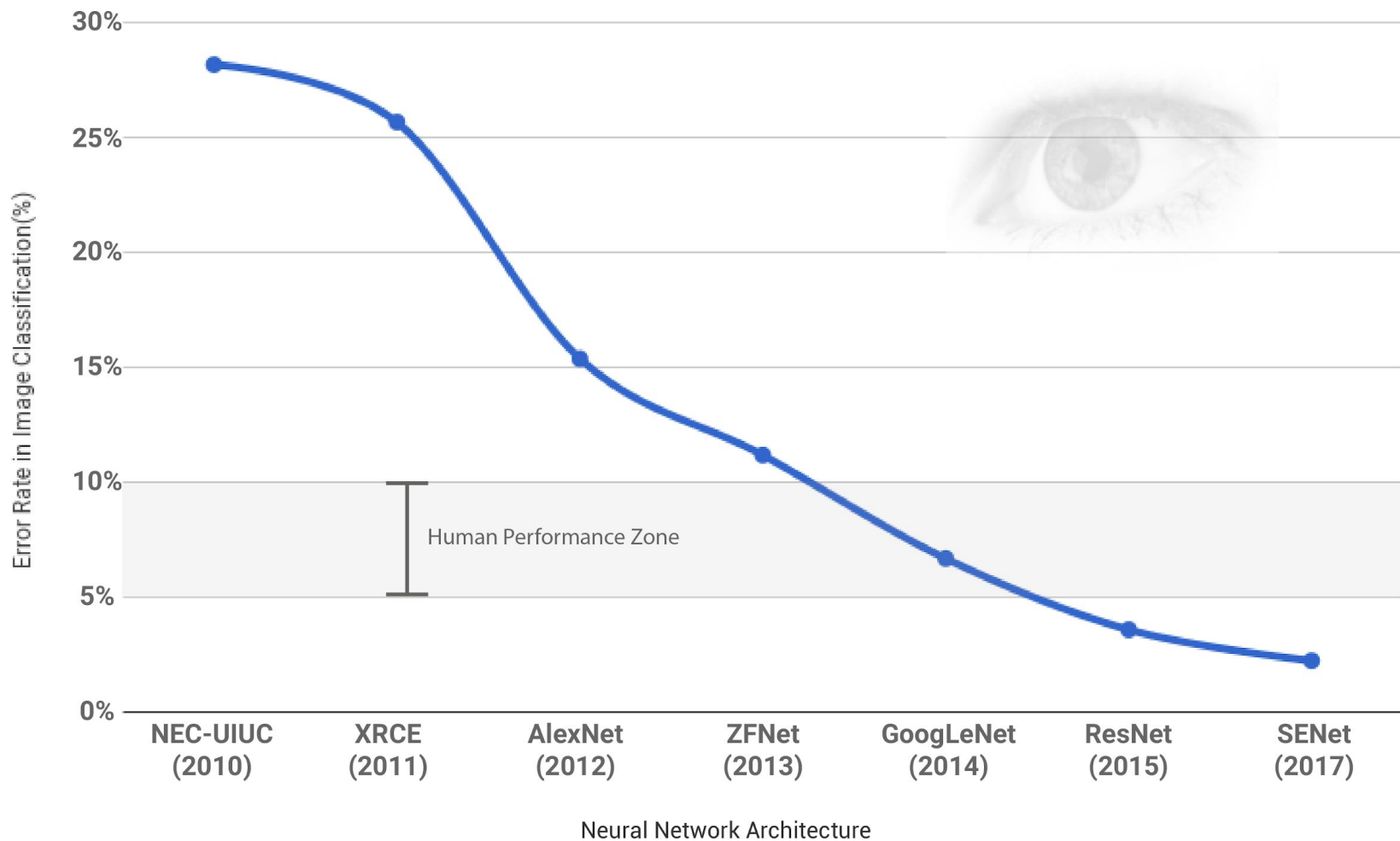
('robb', 0.915272)

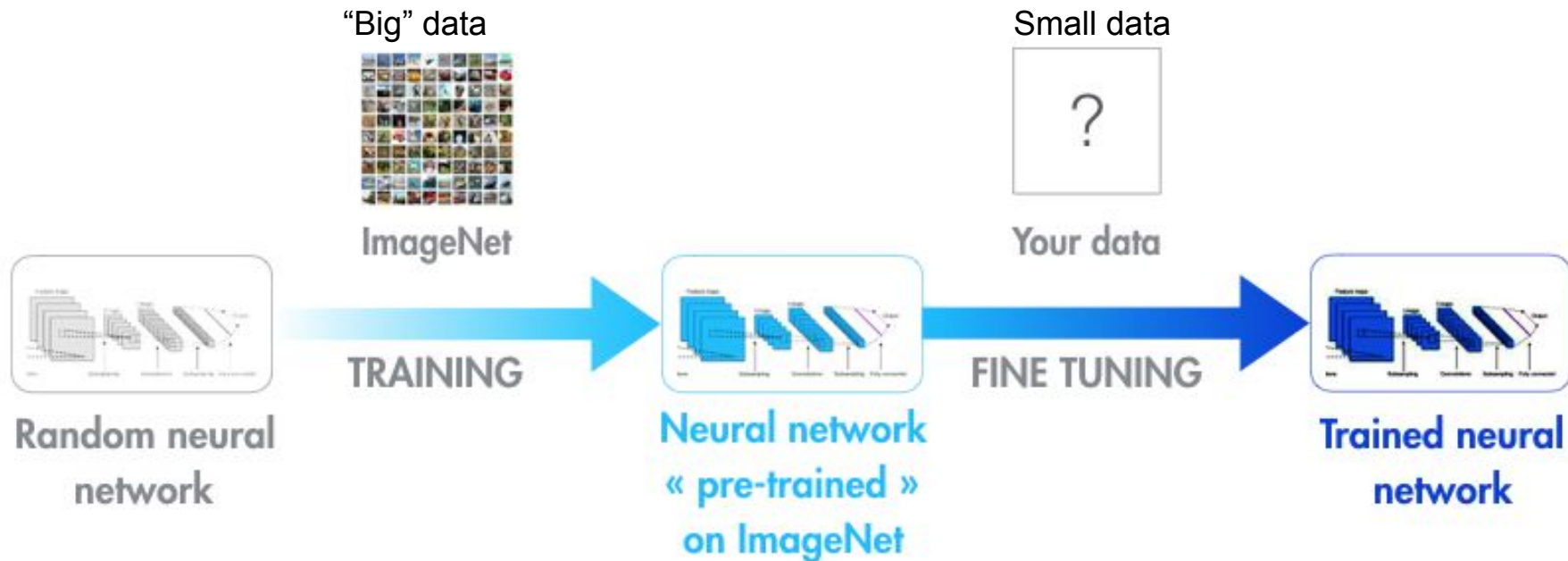
Transfer learning



Imagenet Database

- 14 million images
- 20,000 categories
- Large Scale Visual Recognition Challenge

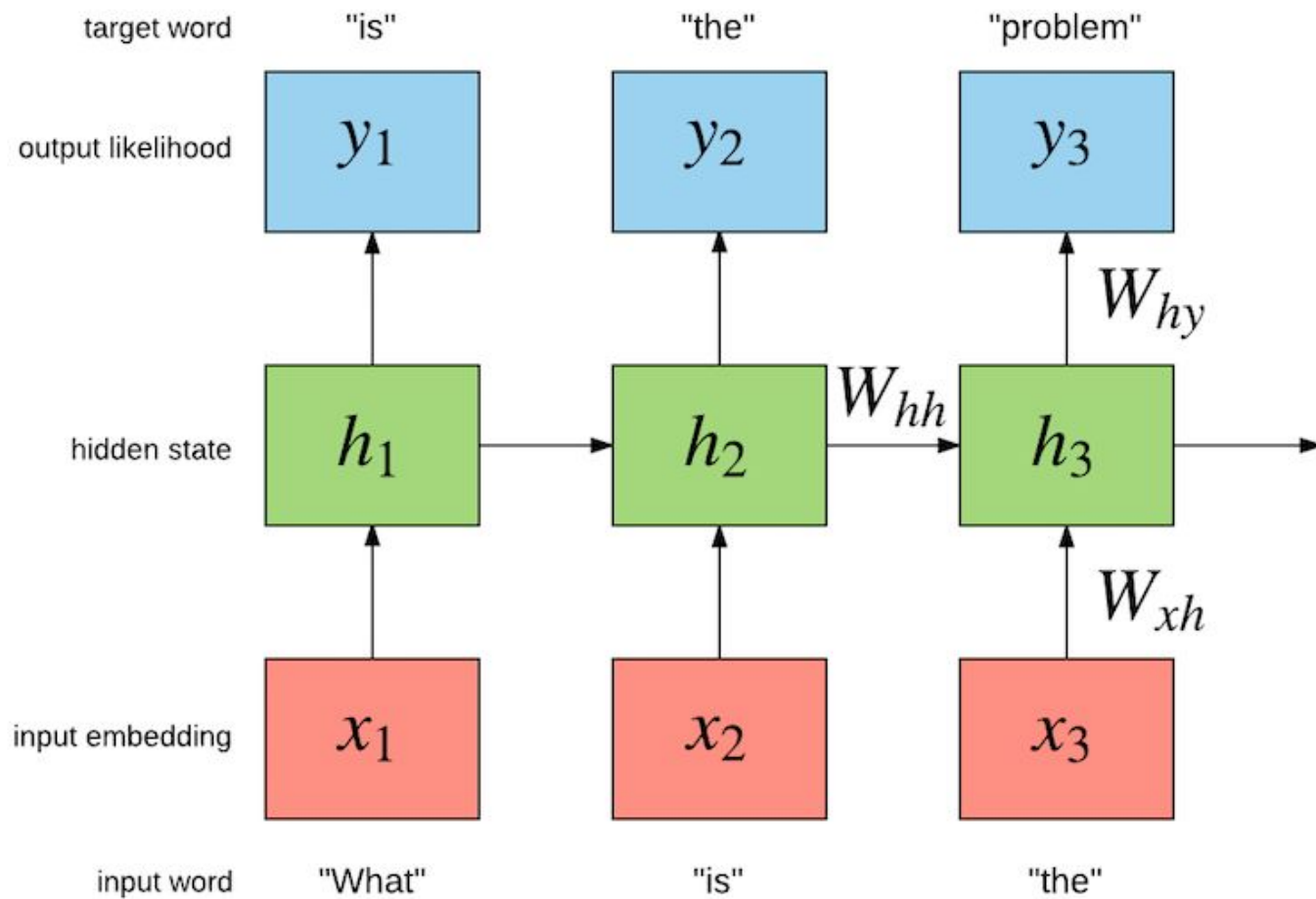




Download from Github
(thanks friends!)

RNN Language Models

From words to sentences



Once upon a _____

Open AI GPT-2: Too good to release to the public

As early as the 1880s, Some Union Club members had considered putting in an offer for the University Club's 54th Street corner lot, but the more conservative faction of the club had nixed the location as too far uptown.

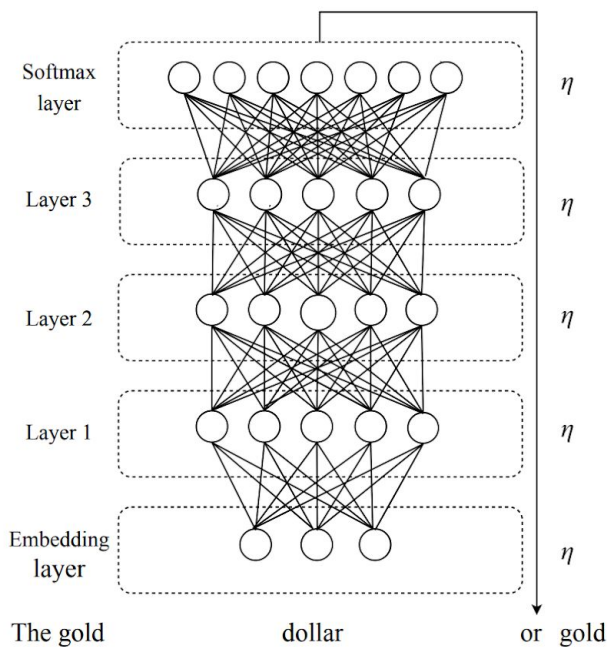
A decade later, with commercial enterprises surrounding their 21st Street home, the club finally decided to move, purchasing a large lot between 51st and 52nd streets on Fifth Avenue. Deciding to move was hard enough—then deciding what to build at the new location was even harder. A significant portion of the membership wanted to make a carbon copy of the club's 21st Street home, enlarging it to fit the club's new location. More progressive members commissioned a design from Cass Gilbert, whose splendid Custom House was then rising on Bowling Green. Unable to make a decision, the building committee called on Charles McKim to break the tie. McKim gave the nod to the rebuilt version of the old clubhouse—so the building committee, oddly, chose to move forward with

the new one.

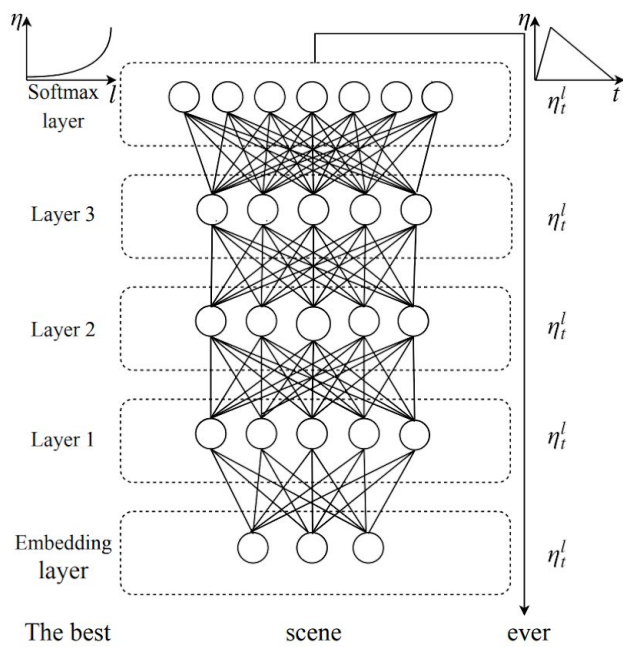
The new club house was a massive, open footprint that extended to the roof. The original Union Club had been built on a scale that was more grandiose than anything Tavernier had imagined, and McKim's scale was no different. The new building, by contrast, was a post-war marvel that was not only elegant, but also extremely efficient. When Richard et. al. arrived in New York, the club had just moved into its new home, and they were still figuring out how to run it. As one member recalls, "There was a considerable amount of confusion in the beginning, and we were all in a state of bewilderment."

The new club house was a massive, open footprint that extended to the roof. The original Union Club had been built on a scale that was more grandiose than anything Tavernier had imagined, and McKim's scale was no different. The new building, by contrast, was a post-war marvel that was not only elegant, but also extremely efficient. As one member recalls, "There was a considerable amount of confusion in the beginning, and we were all in a state of bewilderment."

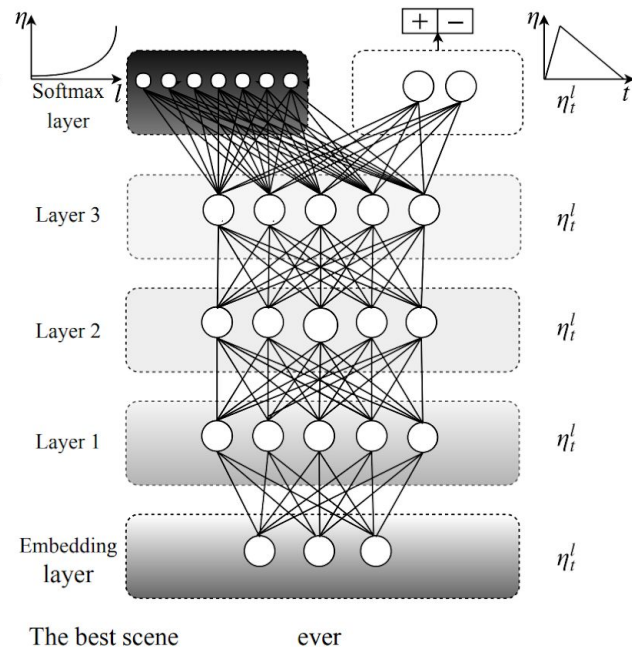
General-purpose (pre-trained LM) vs.
task-specific (fine-tuned LM)



(a) LM pre-training

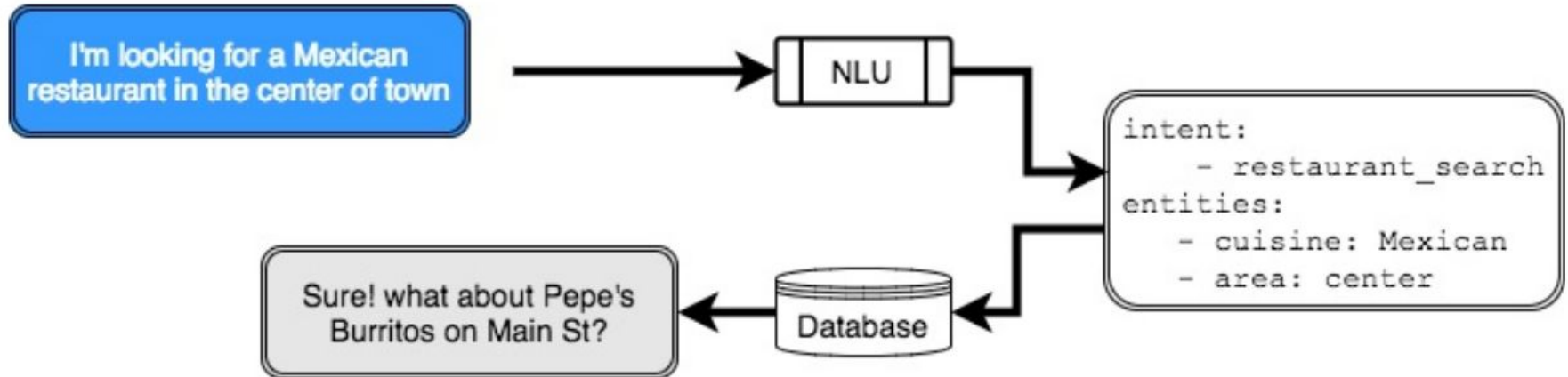


(b) LM fine-tuning



(c) Classifier fine-tuning

Intent Classification



Intent example

In summary

Bag of words (ignores structure) to embeddings(co-occurrences) to RNN LMs
(entire sentence/paragraph)

More complex models require more data and/or more time to train

Transfer learning can be applied for NLP tasks.

Choose the model that's most appropriate for your question

About me

PhD in Neurobiology

Left academia for data science, worked in startups for a few years

- Computer vision projects with BMW and other automotive OEMs
- NLP projects for hiring at Safeway, UberEATS and other large retailers
- Technical product management, integrating ML engineering with business

Currently running my own consulting shop

- Would love to discuss how machine learning can be applied to your business or research

Questions?

twitter: @nikhildes

email: nikhil.deshmukh@pm.me