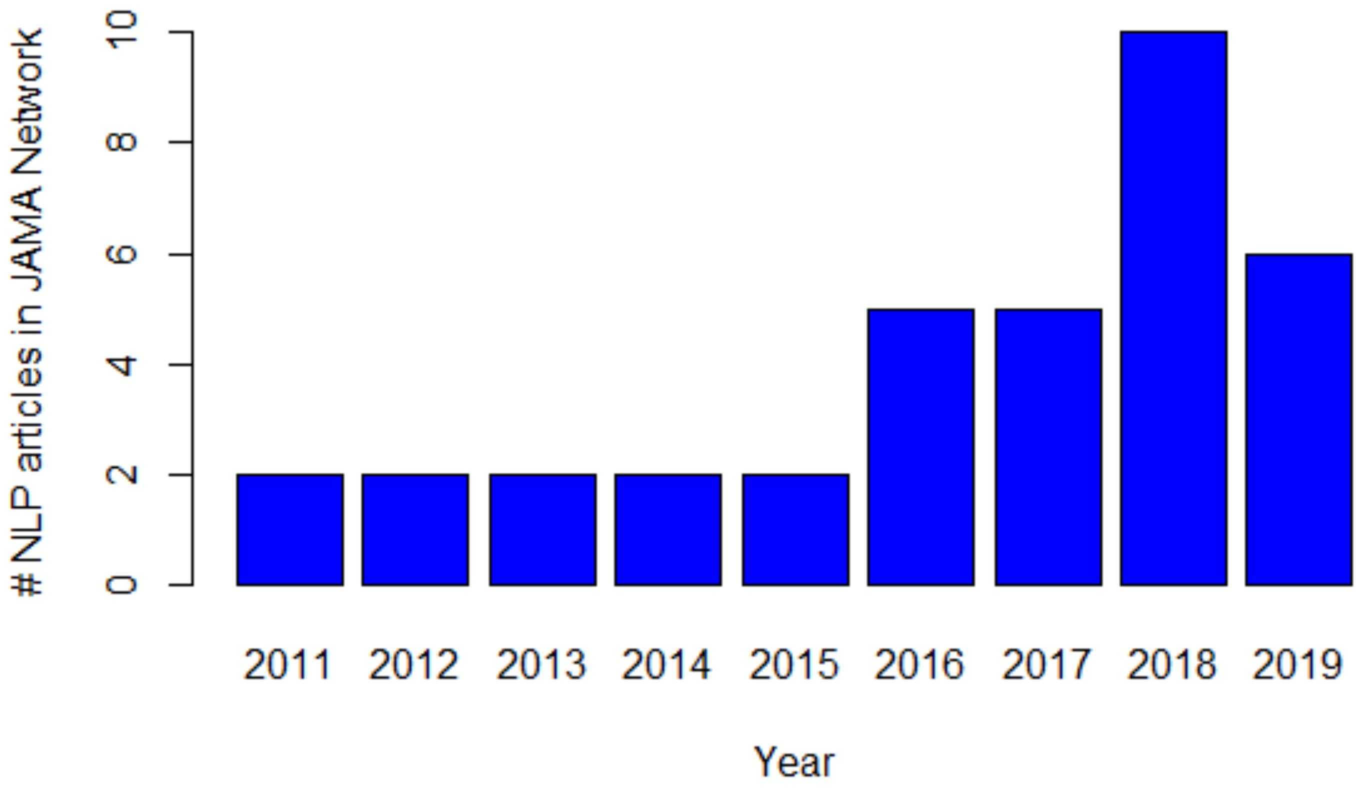
Biomedical NLP in Practice

Matthew Engelhard

In 2019, I did a brief survey of NLP in JAMA…

Of 28 articles:

- 27 were based entirely on word counts
- Most focused on identifying specific diagnoses or events within notes
- Even today, these simple approaches often work best in practice

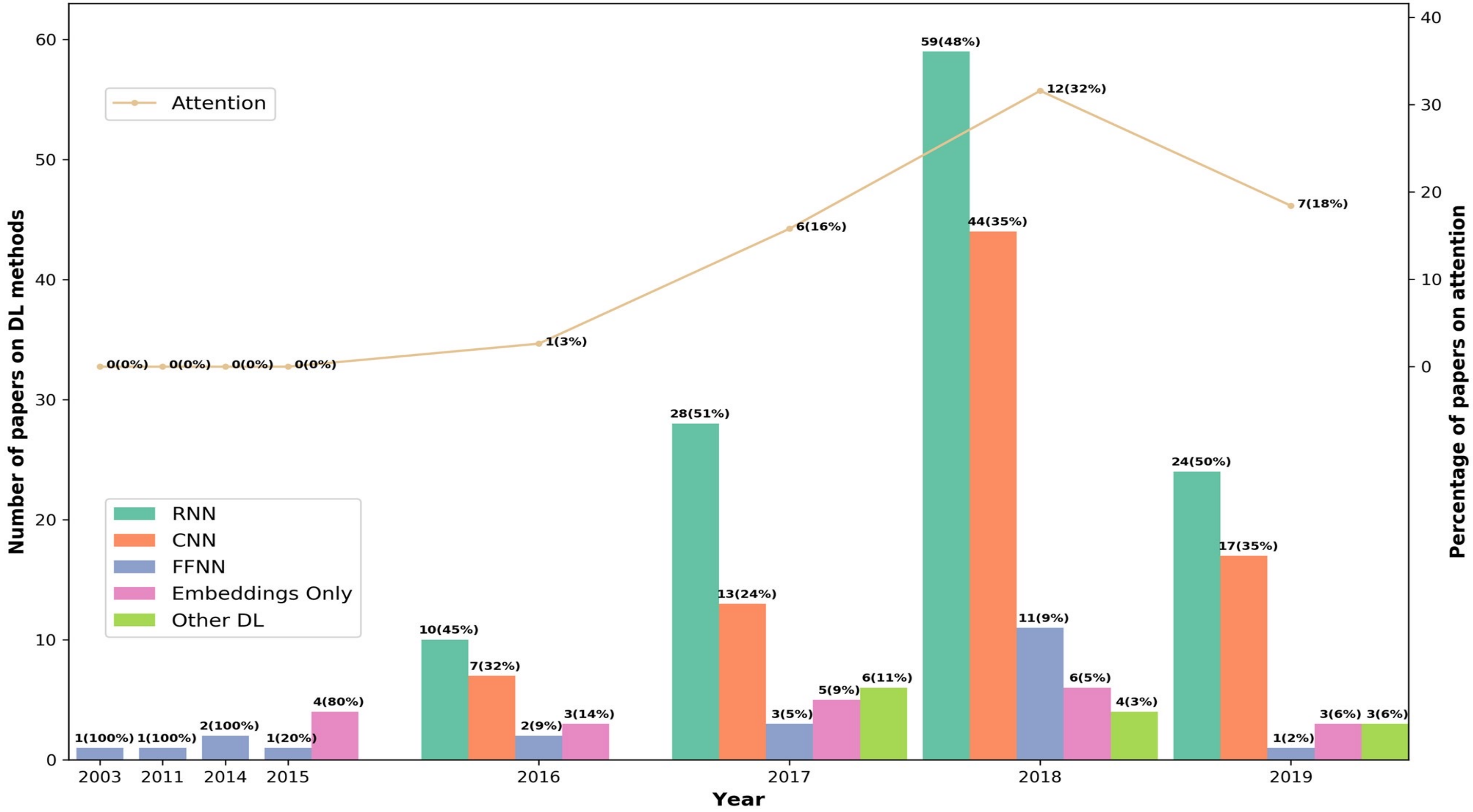


In 2022, deep learning is becoming dominant

Why?

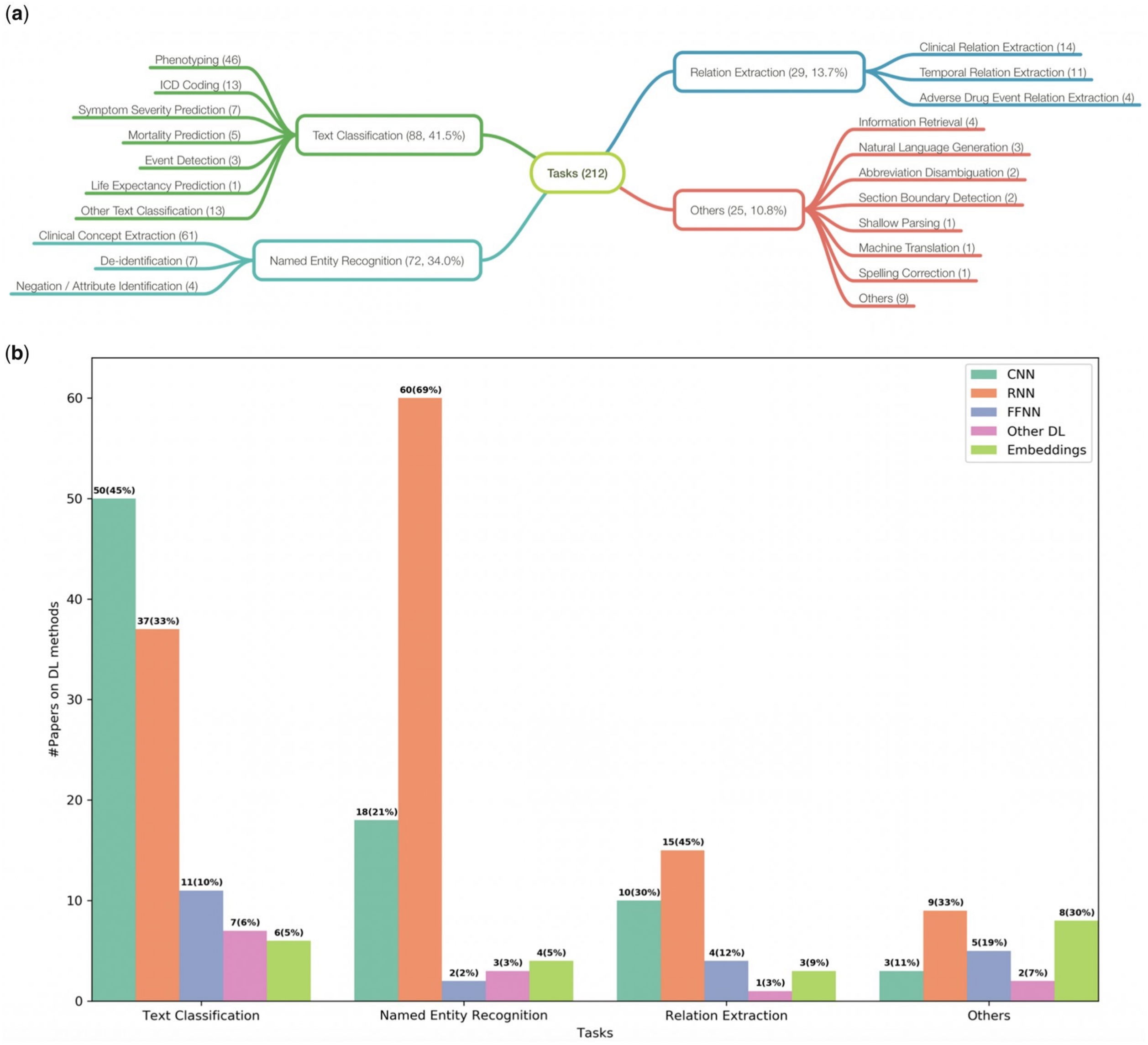
- Rarely worse than BoW and often better

- Can be very easy to use pre-trained models



J Am Med Inform Assoc, Volume 27, Issue 3, March 2020, Pages 457–470, https://doi.org/10.1093/jamia/ocz200

In 2022, deep learning is becoming dominant



J Am Med Inform Assoc, Volume 27, Issue 3, March 2020, Pages 457–470, https://doi.org/10.1093/jamia/ocz200

Common tasks:

- Text classification
  (classify a note)
- Named entity recognition
  (identify clinical concepts within a note)
- Relation Extraction
  (identify relationships between pairs of concepts within notes)
- Others…

How does “deep” NLP work?

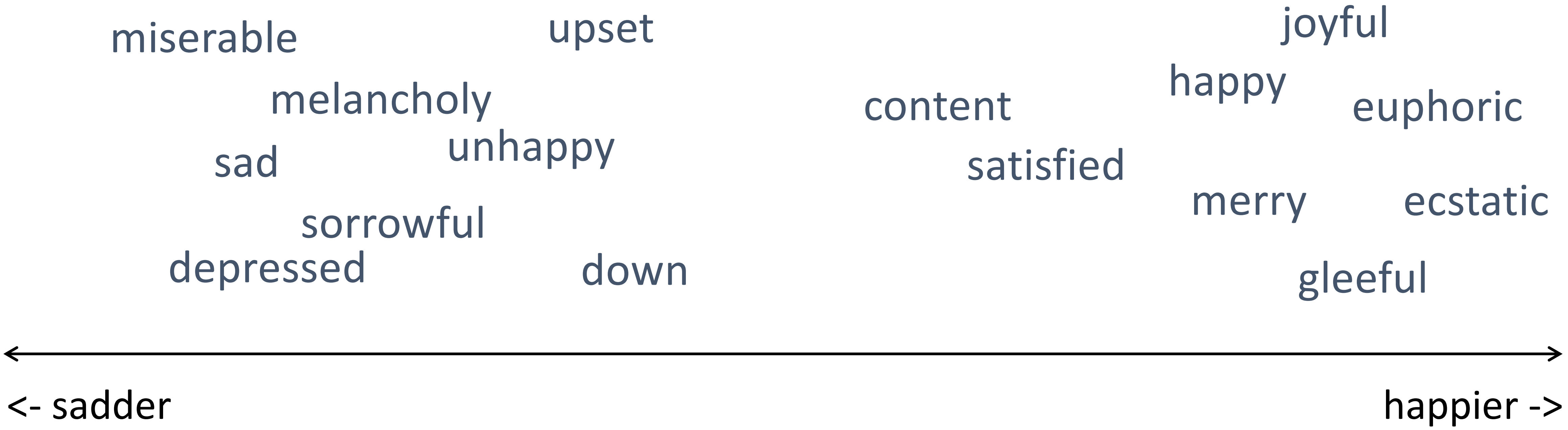
Answer, part 1: word vectors

~~Old~~ Tried and true approach: word counts

I passed out and Mom said I was shaking

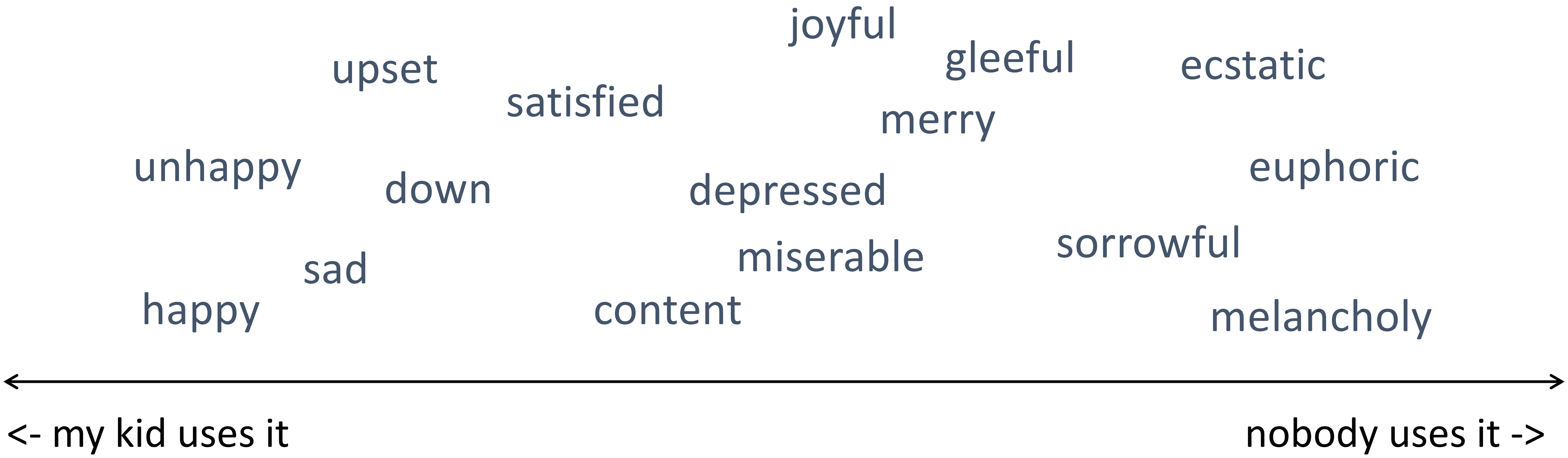
	shaking	what	clinic	how	helps	was	nearest	many	with	said	months	the	morning	mom	should	sickness	and	I	is	how	out	breastfeed	passed	where
x_3	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	2	0	0	1	0	1	0

New approach: we'd like to *encode the meaning* of each word by assigning numeric attributes



Attribute 1: how happy or sad is the word?

New approach: we'd like to *encode the meaning* of each word by assigning numeric attributes



Attribute 2: how highfalutin is the word?

Training a robot to buy groceries

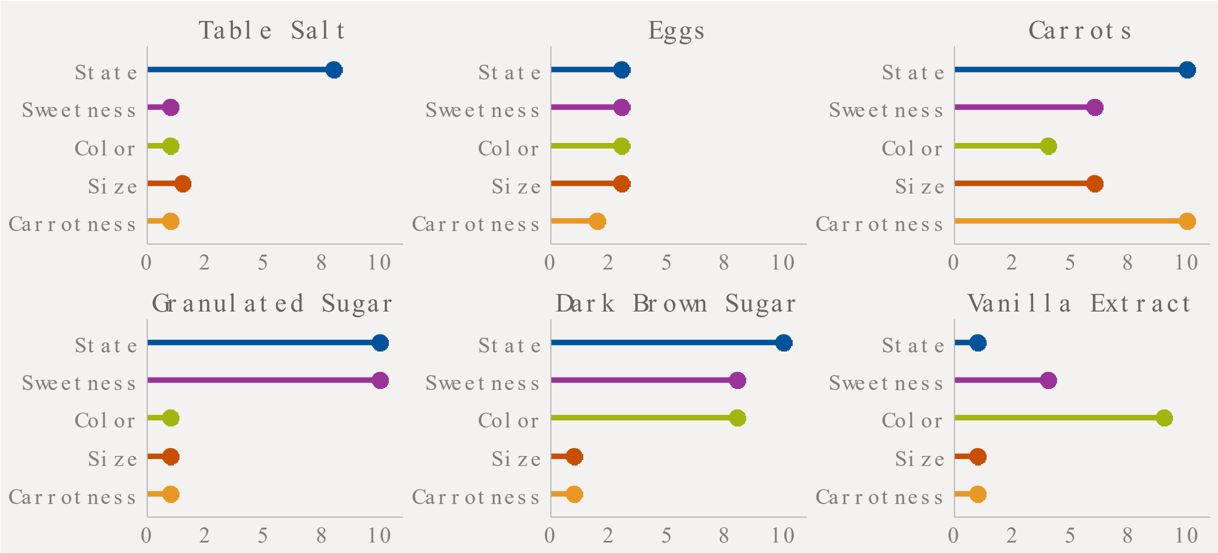
Example from Anand Chowdhury, MMCi 2019

Grocery List

- ❑ granulated sugar
- ❑ vanilla extract
- ❑ dark brown sugar
- ❑ carrots
- ❑ table salt
- ❑ eggs

Identify items by their attributes
(including previously unseen items)

Dimension	1	10
State	Liquid	Solid
Sweetness	Bland	Sweet
Color	Light	Dark
Size	Small	Large
Carrotness	Not really	Platonic essence of carrot



We’re putting all our words on a map…

- We’ll use a few hundred attributes, not just two.
- The closer together two words are on this map, the more similar their meaning.

Why does this help us?

- The model can make sense of words it hasn’t seen before (weren’t used in training)
- Similar words (e.g. synonyms) will have similar attributes, and therefore will have similar effect on model predictions
- (more complicated) Now we can convert text to a sequence of vectors; and we were already very good at making predictions from sequences of vectors

How do we learn these attributes?

-> there's an additional, optional lecture on this

KEY IDEA: words are *defined* by the <u>context</u> in which they appear

A man strolls down the street

A woman strolls down the street

A child strolls down the street

A crocodile strolls down the street

A banana strolls down the street

A concept strolls down the street

How do we learn these attributes?

KEY IDEA: words are *defined* by the <u>context</u> in which they appear

-> if words are always exchangeable, they must have very similar meaning

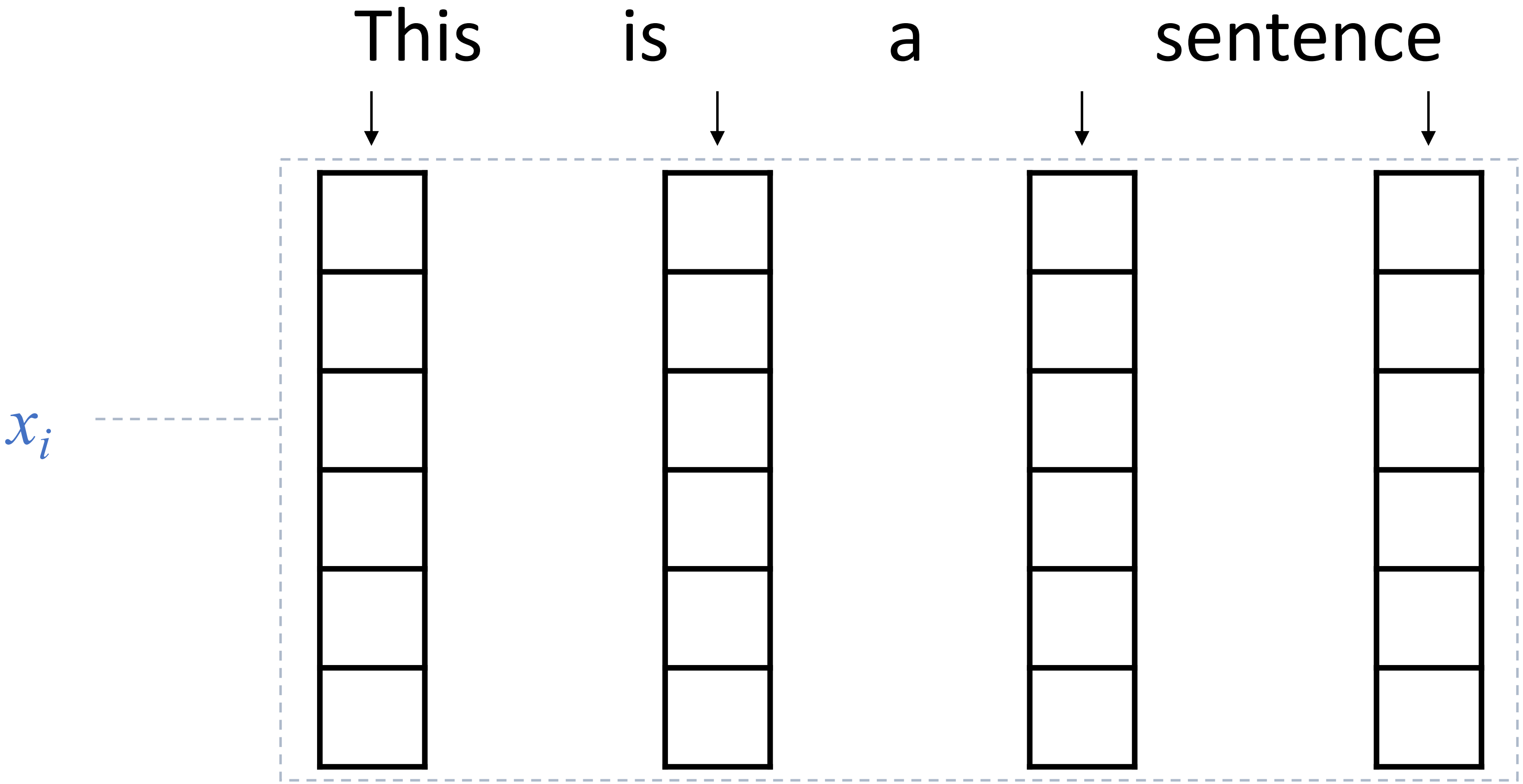
https://www.parenting.com/activities/baby/teach-baby-to-talk/

learn word meaning like an adult:
explicit definitions

learn word meaning like an child:
<u>implicit definitions from context</u>

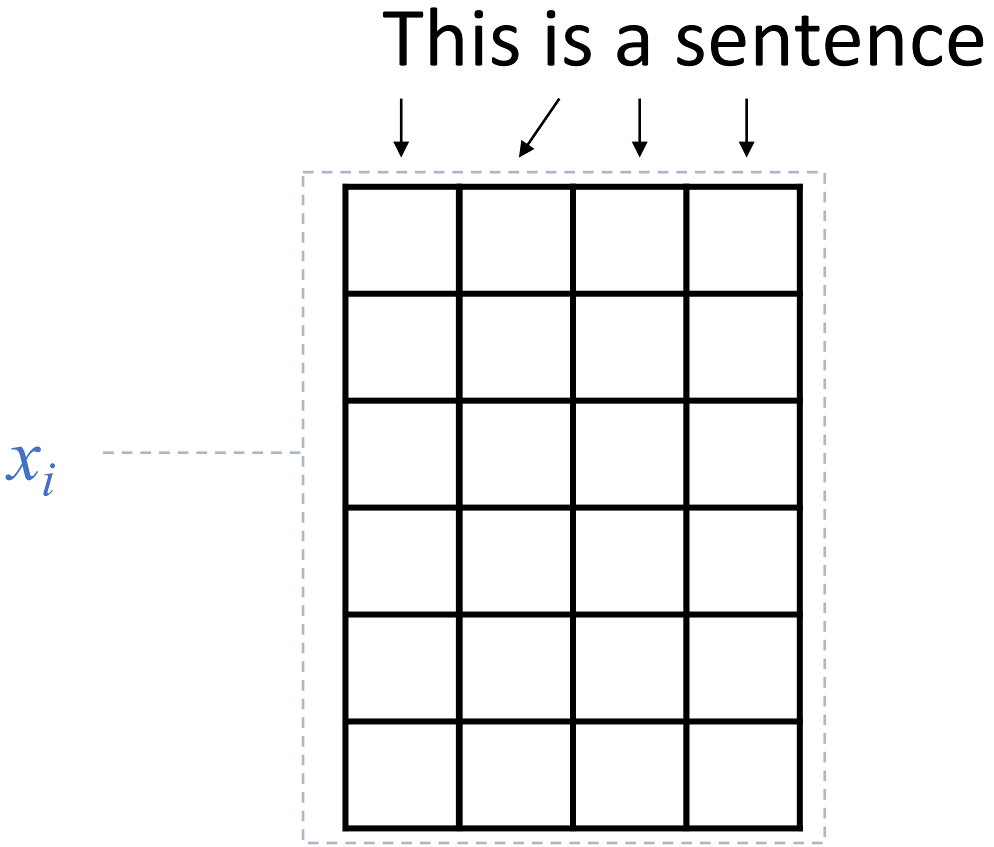
Now that we have word vectors, how do we use them?

- Look up words individually to obtain their vectors
- Construct a sequence of vectors



Now that we have word vectors, how do we use them?

- Look up words individually to obtain their vectors
- Construct a sequence of vectors



Now we have a grid of numbers

Similar in many ways to an image

How does “deep” NLP work?

Answer, part 2: hierarchical feature extraction (like in image processing)

Now we can use deep learning to build our hierarchy of features.

pixels

→ low-level motifs → high-level motifs → dog

label

low-level motifs:

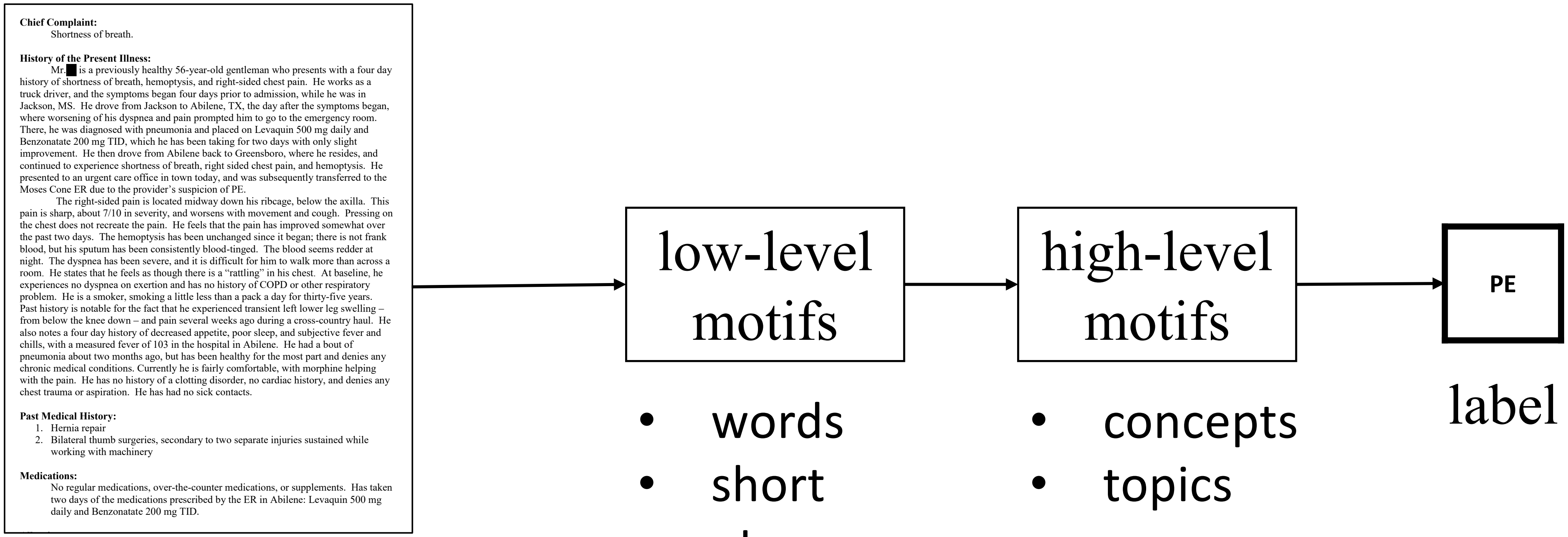
- edges
- shapes
- textures

high-level motifs:

- eyes
- ears
- paws

End goal: predict *dog* from *pixels*

Now we can use deep learning to build our hierarchy of *semantic* features.


Chief Complaint:
Shortness of breath.

History of the Present Illness:
Mr. [redacted] is a previously healthy 56-year-old gentleman who presents with a four day history of shortness of breath, hemoptysis, and right-sided chest pain. He works as a truck driver, and the symptoms began four days prior to admission, while he was in Jackson, MS. He drove from Jackson to Abilene, TX, the day after the symptoms began, where worsening of his dyspnea and pain prompted him to go to the emergency room. There, he was diagnosed with pneumonia and placed on Levaquin 500 mg daily and Benzonatate 200 mg TID, which he has been taking for two days with only slight improvement. He then drove from Abilene back to Greensboro, where he resides, and continued to experience shortness of breath, right sided chest pain, and hemoptysis. He presented to an urgent care office in town today, and was subsequently transferred to the Moses Cone ER due to the provider's suspicion of PE.

The right-sided pain is located midway down his ribcage, below the axilla. This pain is sharp, about 7/10 in severity, and worsens with movement and cough. Pressing on the chest does not recreate the pain. He feels that the pain has improved somewhat over the past two days. The hemoptysis has been unchanged since it began; there is not frank blood, but his sputum has been consistently blood-tinged. The blood seems redder at night. The dyspnea has been severe, and it is difficult for him to walk more than across a room. He states that he feels as though there is a "rattling" in his chest. At baseline, he experiences no dyspnea on exertion and has no history of COPD or other respiratory problem. He is a smoker, smoking a little less than a pack a day for thirty-five years. Past history is notable for the fact that he experienced transient left lower leg swelling – from below the knee down – and pain several weeks ago during a cross-country haul. He also notes a four day history of decreased appetite, poor sleep, and subjective fever and chills, with a measured fever of 103 in the hospital in Abilene. He had a bout of pneumonia about two months ago, but has been healthy for the most part and denies any chronic medical conditions. Currently he is fairly comfortable, with morphine helping with the pain. He has no history of a clotting disorder, no cardiac history, and denies any chest trauma or aspiration. He has had no sick contacts.

Past Medical History:
1. Hernia repair
2. Bilateral thumb surgeries, secondary to two separate injuries sustained while working with machinery

Medications:
No regular medications, over-the-counter medications, or supplements. Has taken two days of the medications prescribed by the ER in Abilene: Levaquin 500 mg daily and Benzonatate 200 mg TID.

grid of semantic attributes

- words
- short phrases

- concepts
- topics

label

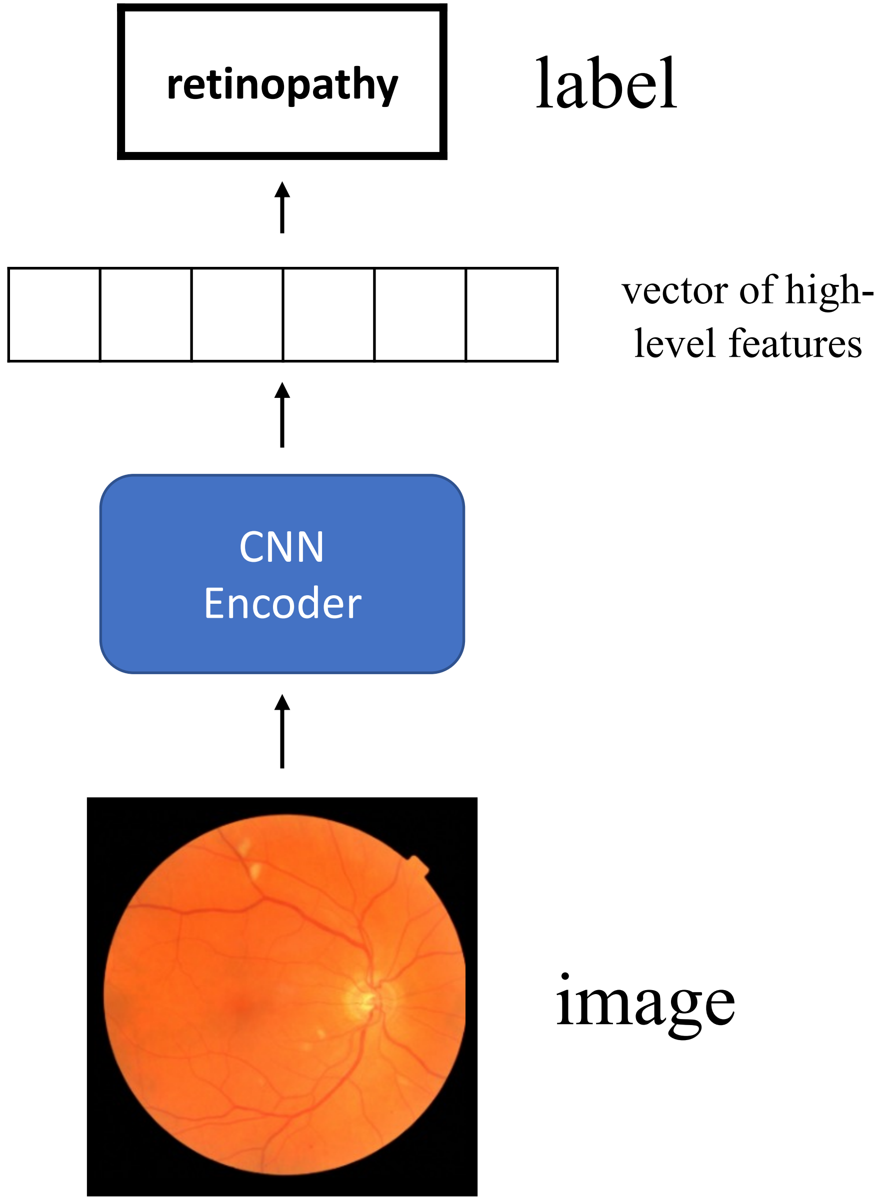
End goal: predict *pulmonary embolism* from *text*

Recall: in image processing, we start with a pre-trained *encoder*

1. A CNN *image encoder* that converts the raw image to a vector of high-level motifs / features.

2. A *final layer*, or *prediction head* – this is a <u>logistic regression</u> model – that makes predictions about the label from these high-level features.

- We will <u>reuse</u> the encoder but <u>replace</u> the prediction head, since it is specific to the previous (non-medical) task.

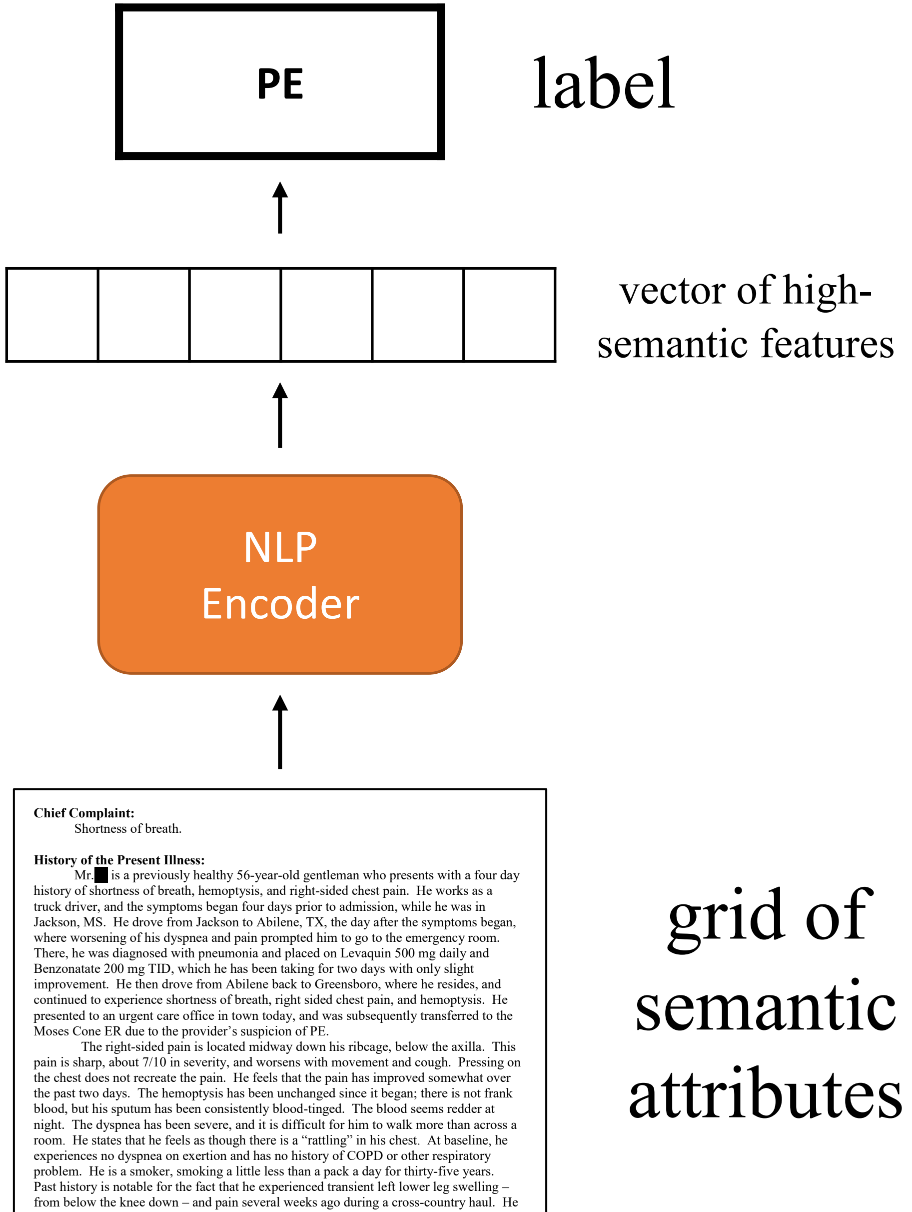


In modern (deep) NLP, we also start with a pre-trained *encoder*

1. A transformer network *image encoder* that converts the raw semantic attributes to a vector of high-level motifs / features.

2. A *final layer*, or *prediction head* – this is a logistic regression model – that makes predictions about the label from these high-level features.

- We will reuse the encoder but replace the prediction head, since it is specific to the previous task.



Our encoder (& word vectors) is pre-trained on biomedical corpora.

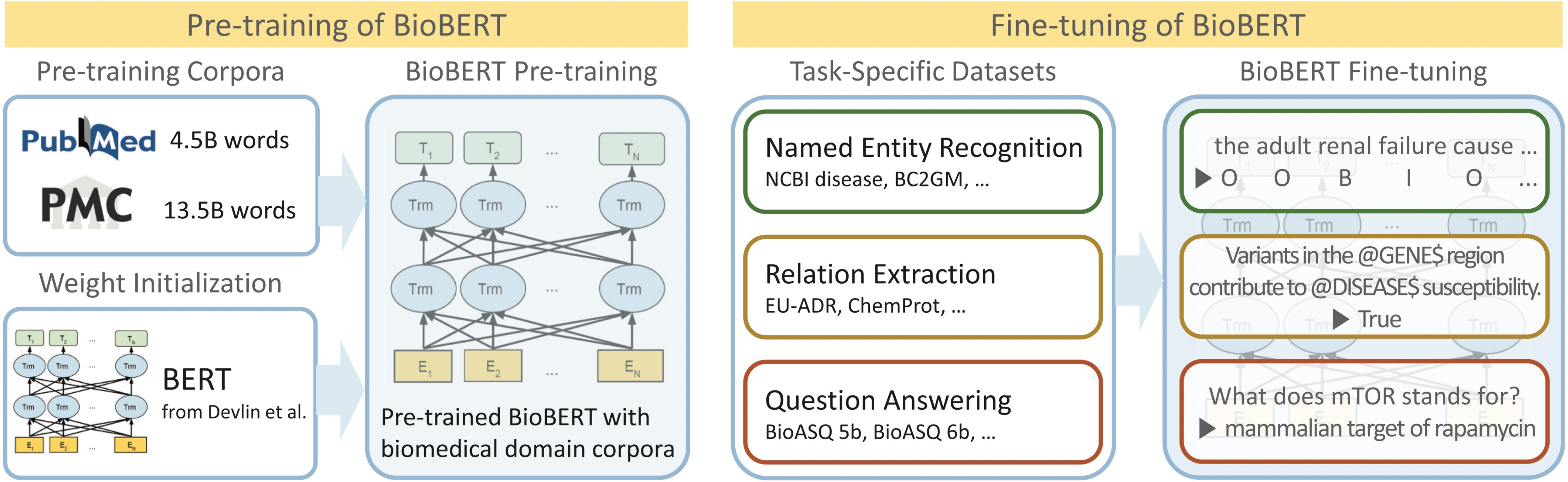


Fig. 1. Overview of the pre-training and fine-tuning of BioBERT

Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020 Feb 15;36(4):1234-40.

Our encoder (& word vectors) is pre-trained on biomedical corpora.

- Common encoders (e.g. BERT, GPT3) have millions or billions of parameters (up to 1T)
- However, the principles remain the same: neural networks performing hierarchical feature extraction
- Different tasks require slightly different final modifications to the encoder
- Deep NLP is becoming more accessible (and common in the clinical literature) as tools to acquire and use these encoders continue to improve

Named Entity Recognition

The ~~old~~ tried and true way:

- Unified Medical Language System (UMLS)
- Apache cTAKES
- Rules-based systems to extract medical concepts from free text
- Can then build predictive models based on presence or absence of specific medical concepts

Example: NILE

- Best approach in many cases
- Fast, easy to implement

Narrative Information Linear Extraction (NILE)

Introduction

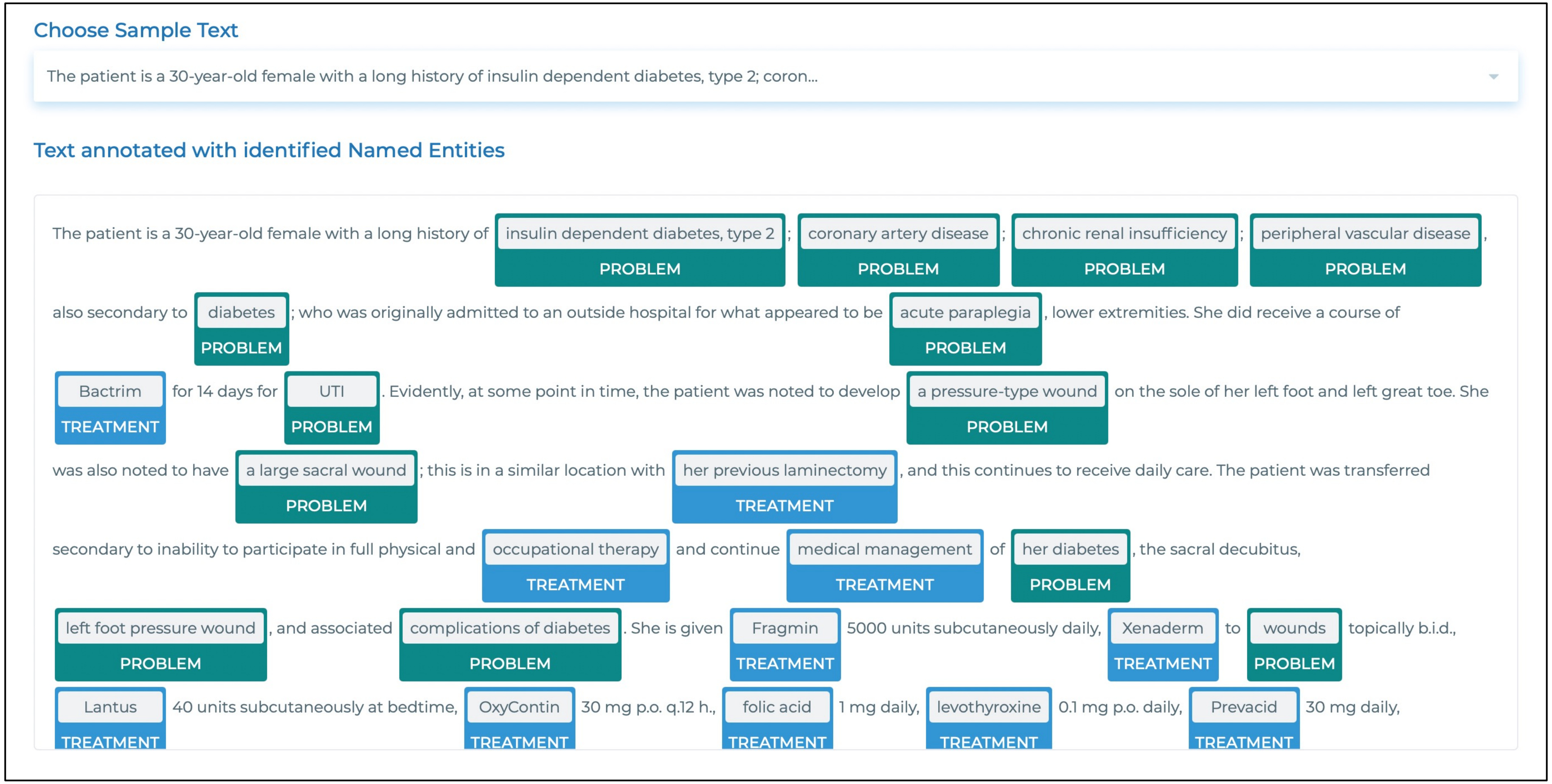
NILE is an efficient and effective software for natural language processing (NLP) of clinical narrative texts. It uses a prefix tree algorithm for named entity recognition, and finite-state machines for semantic analysis, both of which were inspired by the natural reading behavior of humans. The design aims to directly translate linguistic and clinical knowledge to code, allowing for the development of functions to parse complex language patterns.

The software was developed by Sheng Yu and Tianxi Cai at Harvard T.H. Chan School of Public Health and Tianrun Cai at The Brigham and Women's Hospital. It is distributed free of charge for academic and non-commercial research use by the President and Fellows of Harvard College.

NILE

Named Entity Recognition

The new way: deep NLP encoder



https://demo.johnsnowlabs.com/healthcare/NER_CLINICAL/

Conclusions

- Text data are central to clinical medicine, so the potential for NLP impact is high (but *not yet realized*)
- Simple, count-based NLP models are surprisingly effective in most clinical applications.
- Complex, deep learning NLP models have exceeded human performance. In these models, words are converted to vectors of semantic attributes, and increasingly complex, heirarchical semantic features are then extracted.
- Similar to image processing, we can take advantage of complex NLP models by repurposing them for a specific clinical task via fine-tuning of parameters.

A brief note on interpretability…

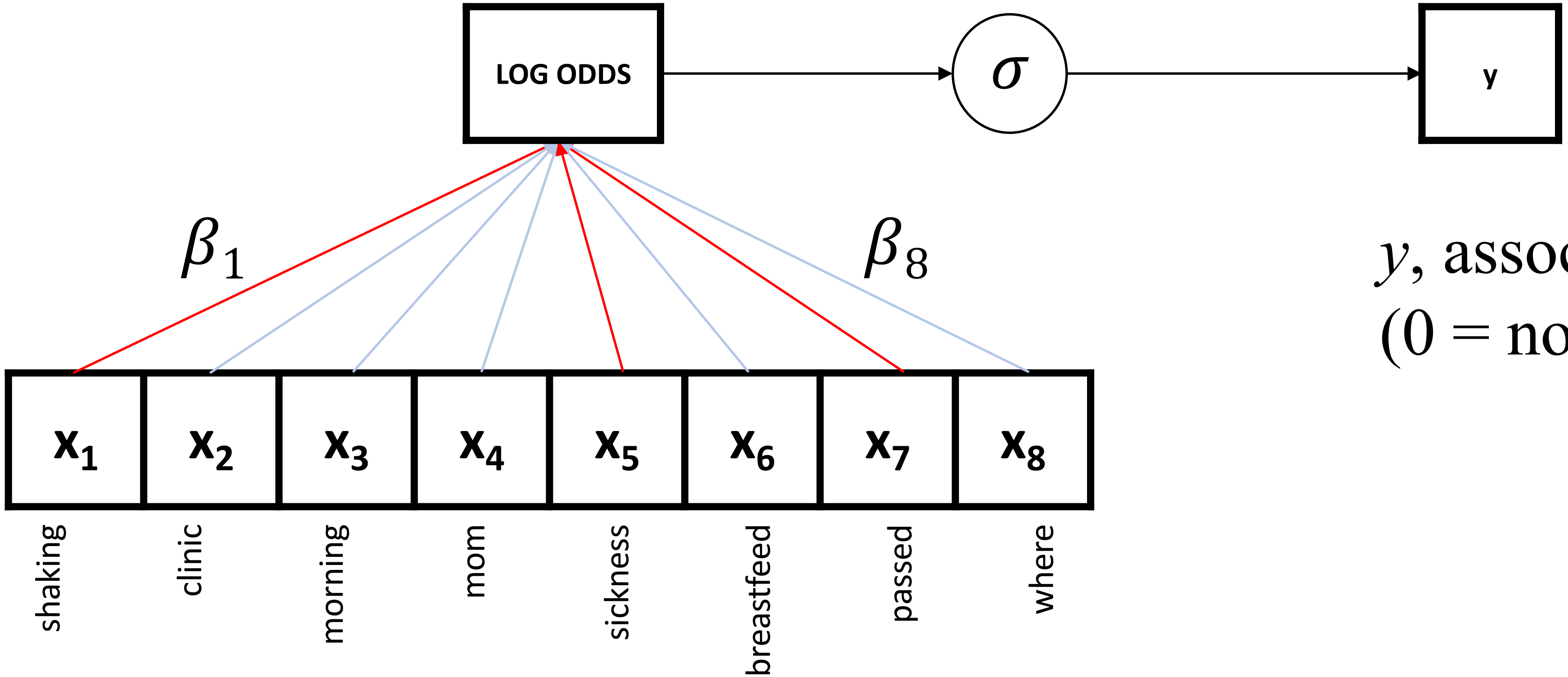
More on this next time.

We can *interpret* a count-based NLP model

- Suppose you use logistic regression with count-based features, and your model predicts that that an SMS you receive is urgent.
- **Q:** Is it hard to figure out why it made that prediction?

We can *interpret* a count-based NLP model

- Suppose you use logistic regression with count-based features, and your model predicts that that an SMS you receive is urgent.
- **Q:** Is it hard to figure out why it made that prediction?
- **A:** No. You can look at the coefficients to see which words increased and decreased the predicted probability.



y, associated label:
(0 = not urgent, 1 = urgent)

Can we interpret a deep learning NLP model?

- Suppose you apply a deep neural network to a sequence of word vectors, and your model predicts that that the SMS you receive is urgent.
- **Q:** Is it hard to figure out why it made that prediction?