Machine Reading & Question Answering

Jacob Andreas / MIT 6.804-6.864 / Spring 2020

Admin

Homework 3 & midterm released.
Recommended deadline 2 weeks from posting date.

Little fixes:
(1b) minimized -> maximized
(eq 4) b -> exp{ b...

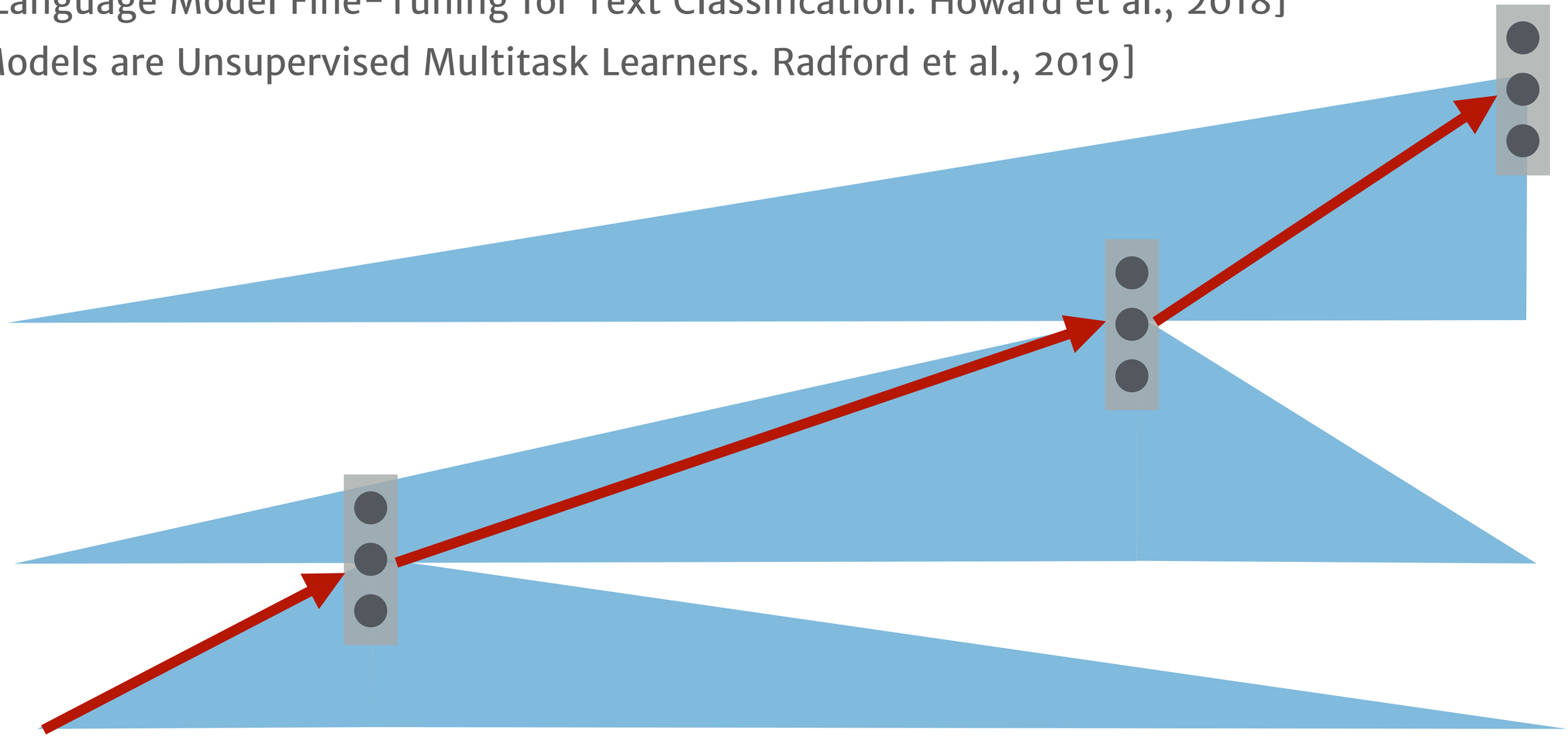
6.864: Start thinking about project topics!
Use Piazza to coordinate groups/project ideas.

Recap: language models & QA

GPT/ULMFit: Language modeling with neural sequence models

[Universal Language Model Fine-Tuning for Text Classification. Howard et al., 2018]

[Laguage Models are Unsupervised Multitask Learners. Radford et al., 2019]



John has a book. Mary has an apple. He gave her his

Fine-tuning LMs: text output

1. Pretrain on a language modeling task on billions to 10billions of words
2. Make a new “language modeling” dataset with your input-output pairs
3. Fine-tune everyhing together:

Pretrain:

The following year she published a paper called Idealtheorie in Ringbereichen, analyzing ascending chain conditions with regard to (mathematical) ideals. Noted algebraist Irving Kaplansky called this work "revolutionary"; the publication gave rise to the term "Noetherian ring" and the naming of several other mathematical objects as Noetherian.

Fine-tune:

for Fitting's theorem and the Fitting lemma; and Zeng Jiongzhi (also rendered "Chiungtze C. Tsen" in English), who proved Tsen's theorem. ***Who was Zeng Jiongzhi’s doctoral advisor? Emmy Noether.***

Challenges in QA with sequence models

In 1917, he passed the entrance examination and was admitted to Jiangxi Provincial First Normal College in Nanchang. He was subsidised by Lei Heng's son Tsebu S. Lee (雷子布), who was studying in Japan on government scholarship. After graduation in 1920, Tsen taught in primary school for two years. In 1922, Tsen entered National Wuchang Senior Normal College[n 3] to study undergraduate mathematics, and he graduated in 1926. After graduation, he passed a scholarship examination for studying in the West, and he worked as teacher in high schools for two years to perform the mandatory teaching service of his degree.[3]

In 1928, he went to Berlin University for language training for a year, and then he started studying mathematics at University of Göttingen in the summer semester of 1929. He studied algebra under Emmy Noether. Tsen received his doctoral degree in February 1934 under the supervision of Emmy Noether and Friedrich Karl Schmidt, and he dedicated his dissertation to his elder cousin Tsebu S. Lee. Having fled to the US, Noether evaluated the dissertation in a letter as "sehr gut" (very good). As a research fellow sponsored by the China Foundation for the Promotion of Education and Culture, Tsen did a postdoctoral research with Emil Artin at Hamburg University for a year. There he became friend with Shiing-Shen Chern, who was a graduate student back then. Chern remembered him as a cordial and open-minded person well-liked by everyone.[2]

Who was Zeng Jiongzhi's doctoral adviser? Emmy Noether.

Challenges in QA with sequence models

In 1917, he passed the entrance examination and was admitted to Jiangxi Provincial First Normal College in Nanchang. He was subsidised by Lei Heng's son Tsebu S. Lee (雷子布), who was studying in Japan on government scholarship. After graduation in 1920, Tsen taught in primary school for two years. In 1922, Tsen entered National Wuchang Senior Normal College[n 3] to study undergraduate mathematics, and he graduated in 1926. After graduation, he passed a scholarship examination for studying in the West, and he worked as teacher in high schools for two years to perform the mandatory teaching service of his degree.[3]

In 1928, he went to Berlin University for language training for a year, and then he started studying mathematics at University of Göttingen in the summer semester of 1929. He studied algebra under Emmy Noether. Tsen received his doctoral degree in February 1934 under the supervision of Emmy Noether and Friedrich Karl Schmidt, and he dedicated his dissertation to his elder cousin Tsebu S. Lee. Having fled to the US, Noether evaluated the dissertation in a letter as "sehr gut" (very good). As a research fellow sponsored by the China Foundation for the Promotion of Education and Culture, Tsen did a postdoctoral research with Emil Artin at Hamburg University for a year. There he became friend with Shiing-Shen Chern, who was a graduate student back then. Chern remembered him as a cordial and open-minded person well-liked by everyone.[2]

Who was Zeng Jiongzhi's doctoral adviser? Emmy Noether.

Challenges in QA with sequence models

In 1917, he passed the entrance examination and was admitted to Jiangxi Provincial First Normal College in Nanchang. He was subsidised by Lei Heng's son Tsebu S. Lee (雷子布), who was studying in Japan on government scholarship. After graduation in 1920, Tsen taught in primary school for two years. In 1922, Tsen entered National Wuchang Senior Normal College[n 3] to study undergraduate mathematics, and he graduated in 1926. After graduation, he passed a scholarship examination for studying in the West, and he worked as teacher in high schools for two years to perform the mandatory teaching service of his degree.[3]

In 1928, he went to Berlin University for language training for a year, and then he started studying mathematics at University of Göttingen in the summer semester of 1929. He studied algebra under Emmy Noether. Tsen received his doctoral degree in February 1934 under the supervision of Emmy Noether and Friedrich Karl Schmidt, and he dedicated his dissertation to his elder cousin Tsebu S. Lee. Having fled to the US, Noether evaluated the dissertation in a letter as "sehr gut" (very good). As a research fellow sponsored by the China Foundation for the Promotion of Education and Culture, Tsen did a postdoctoral research with Emil Artin at Hamburg University for a year. There he became friend with Shiing-Shen Chern, who was a graduate student back then. Chern remembered him as a cordial and open-minded person well-liked by everyone.[2]

Why are Zeng Jiongzhi's mathematical contributions significant?

Structured outputs

Person	Graduation Year	Doctoral Adviser	Dissertation Title
Emmy Noether			
Friedrich Schmidt			
Robert Berger			

Multi-hop QA

In 1917, he passed the entrance examination and was admitted to Jiangxi Provincial First Normal College in Nanchang. He was subsidised by Lei Heng's son Tsebu S. Lee (雷子布), who was studying in Japan on government scholarship. After graduation in 1920, Tsen taught in primary school for two years. In 1922, Tsen entered National Wuchang Senior Normal College[n 3] to study undergraduate mathematics, and he graduated in 1926. After graduation, he passed a scholarship examination for studying in the West, and he worked as teacher in high schools for two years to perform the mandatory teaching service of his degree.[3]

How many years did he spend teaching below the university level? Four years.

Information retrieval

Long-form question answering

Q: *What caused the U.S. civil war?*

Long-form question answering

Q: *What caused the U.S. civil war?*

Do we even need machine learning for this problem?

Long-form question answering

Q: *What caused the U.S. civil war?*

Do we even need machine learning for this problem?

While many
James McPh
between the
in the territo
as the first P

Origins of the American Civil War

From Wikipedia, the free encyclopedia

For events following the birth of the Confederacy (South Carolina's 1860 declaration of secession from the Union), see Battle of Fort Sumter and American Civil War.
See also: Historiographic issues about the American Civil War

Historians debating the **origins of the American Civil War** focus on the reasons why seven Southern states (followed by four more after the

s
ery
60

Long-form question answering

Let's just *find* existing documents with answers!

While many
James McPh
between the
in the territo
as the first P

Origins of the American Civil War

From Wikipedia, the free encyclopedia

For events following the birth of the Confederacy (South Carolina's 1860 declaration of secession from the Union), see Battle of Fort Sumter and American Civil War.
See also: Historiographic issues about the American Civil War

Historians debating the **origins of the American Civil War** focus on the reasons why seven Southern states (followed by four more after the

s
ery
60

Information retrieval

General form: define some

f(question, document)

that is large for related (question, doc) pairs and low for unrelated ones

Word co-occurrence

$$f(\text{question, document}) = \sum_{w} 1[w \in \text{question}] \cdot 1[w \in \text{document}]$$

does w appear in document?

Q: *What caused the U.S. civil war?*

While many still debate the ultimate causes of the Civil War, Pulitzer Prize-winning author James McPherson writes that "The Civil War started because of uncompromising differences between the free and slave states over the power of the national government to prohibit slavery in the territories that had not yet become states. When Abraham Lincoln won election in 1860 as the first Republican president on a platform pledging to keep slavery out of the territories, seven slave states in the deep South seceded and formed a new nation, the Confederate States of America. The incoming Lincoln administration and most of the Northern people refused to recognize the legitimacy of secession. They feared that it would discredit democracy and create

Problem: sparsity

$$f(\text{question, document}) = \sum_{w} 1[w \in \text{question}] \cdot 1[w \in \text{document}]$$

Q: *What caused the U.S. civil war?*

same content,
different token

While many still debate the ultimate causes of the Civil War, Pulitzer Prize-winning author James McPherson writes that "The Civil War started because of uncompromising differences between the free and slave states over the power of the national government to prohibit slavery in the territories that had not yet become states. When Abraham Lincoln won election in 1860 as the first Republican president on a platform pledging to keep slavery out of the territories, seven slave states in the deep South seceded and formed a new nation, the Confederate States of America. The incoming Lincoln administration and most of the Northern people refused to recognize the legitimacy of secession. They feared that it would discredit democracy and create

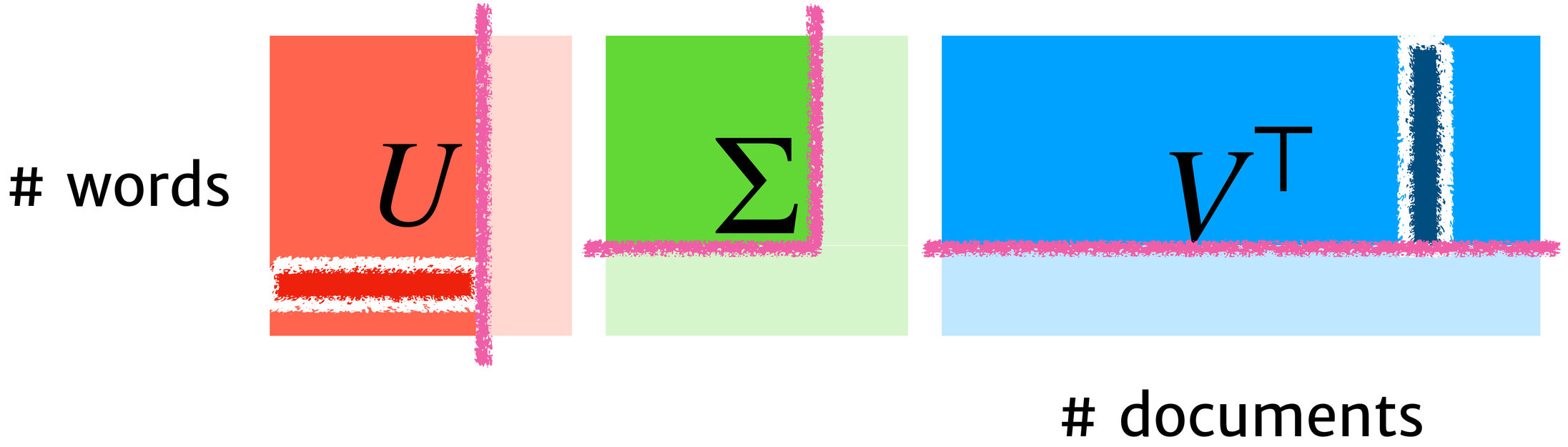
LSA to the rescue!

The term-document matrix:

rows are words
columns are contexts
entries indicate how many times word *i* appears in context *j*

$$W_{td} = \begin{array}{c} \\ cat \\ dog \\ the \end{array} \begin{array}{c} \begin{array}{ccccccc} d_1 & d_2 & d_3 & d_4 & d_5 & d_6 & d_7 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 & 1 & 1 & 0 \\ 20 & 13 & 18 & 22 & 15 & 4 & 20 \end{bmatrix} \end{array}$$

Latent Semantic Analysis: Intuition



Most words don’t appear in most documents, so dimensionality reduction techniques cluster words with similar contexts even when they don't co-occur.

Retrieval with vector similarity

$$f(\text{question, document}) = u_q^\top v_d$$

sum of word embeddings
OR document embedding

Q: *What caused the U.S. civil war?*

While many still debate the ultimate causes of the Civil War, Pulitzer Prize-winning author James McPherson writes that "The Civil War started because of uncompromising differences between the free and slave states over the power of the national government to prohibit slavery in the territories that had not yet become states. When Abraham Lincoln won election in 1860 as the first Republican president on a platform pledging to keep slavery out of the territories, seven slave states in the deep South seceded and formed a new nation, the Confederate States of America. The incoming Lincoln administration and most of the Northern people refused to recognize the legitimacy of secession. They feared that it would discredit democracy and create

More tricks

tf-idf weighting:

Q: *What caused the U.S. civil war?*

Modern solutions: pretrained representations

$f(\textbf{question, document}) = \texttt{BERT}(\textbf{question})^\top \texttt{BERT}(\textbf{document})$

fine-tune with supervision:

(what caused the civil war?, d_1, `true`)

(why is the sky blue?, d_1, `false`)

(why is the sky blue?, d_2, `true`)

does doc. d_1 answer the question?

Retrieving answers

$$\text{argmax}_{\text{doc}} \quad f(\text{question, doc}) = u_{\text{q}}^{\top} v_{\text{d}}$$

might be millions of these!

Lots of specialized algorithms / data structures for solving this problem:

[Gionis, Indyk & Motwani, 1999. Similarity Search in High Dimensions via Hashing]

[Charikar, 2002. Similarity Estimation Techniques from Rounding Algorithms]

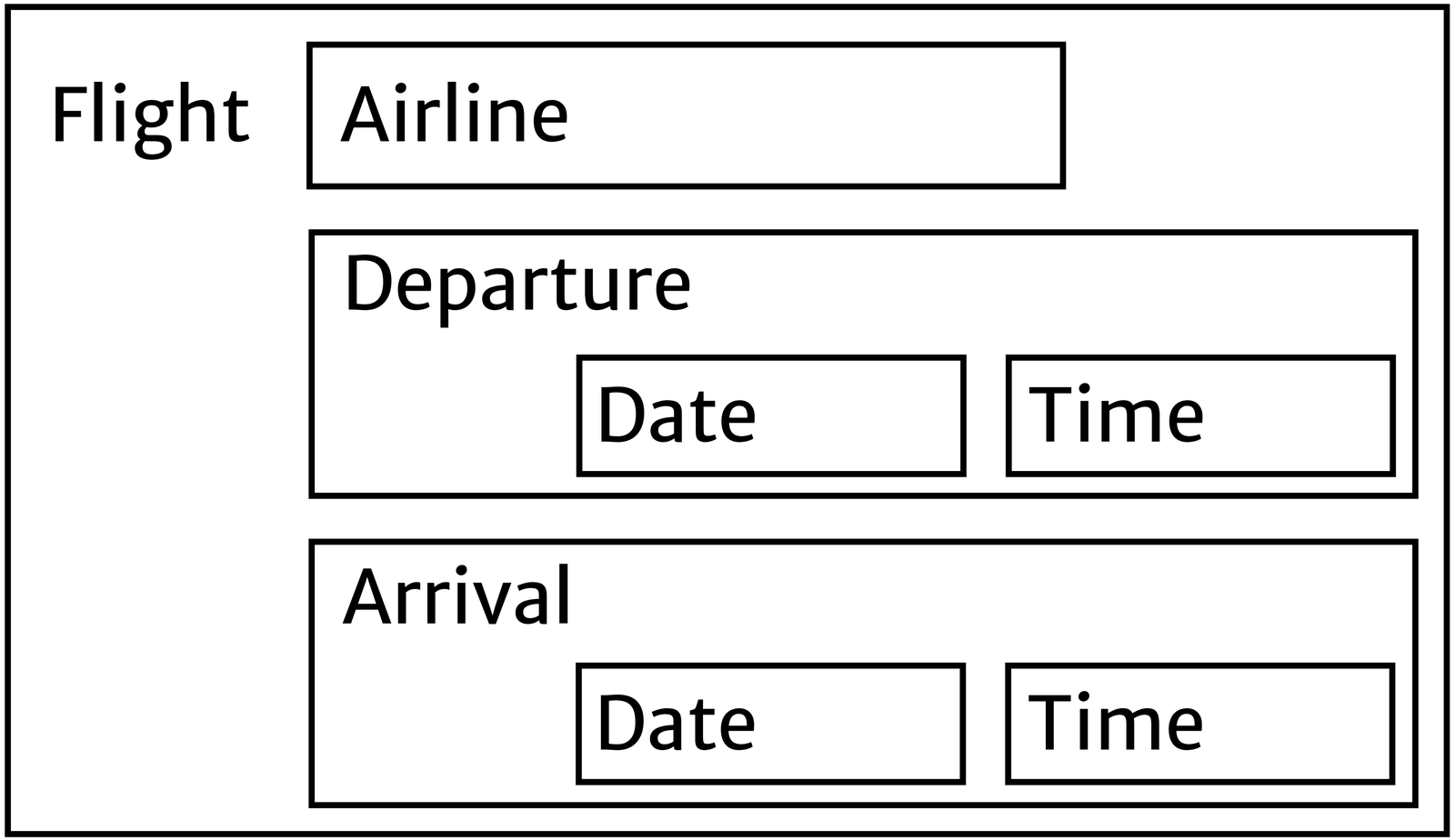
[Shrivastava & Li, 14. Asymmetric LSH for Sublinear Time Maximum Inner Product Search]

Information extraction

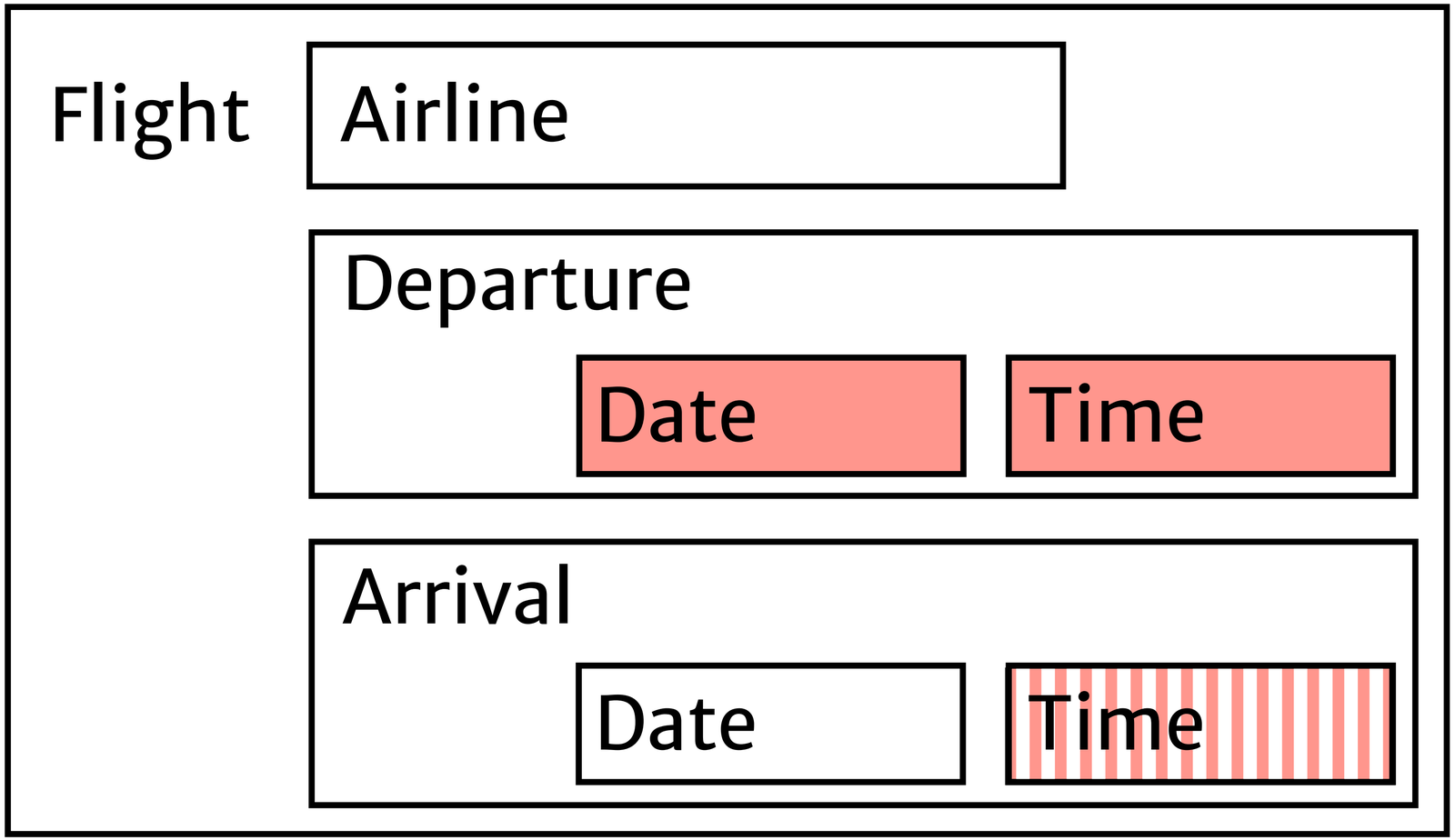
Structured semantic queries

Person	Graduation Year	Doctoral Adviser	Dissertation Title
Emmy Noether			
Friedrich Schmidt			

Structured semantic queries



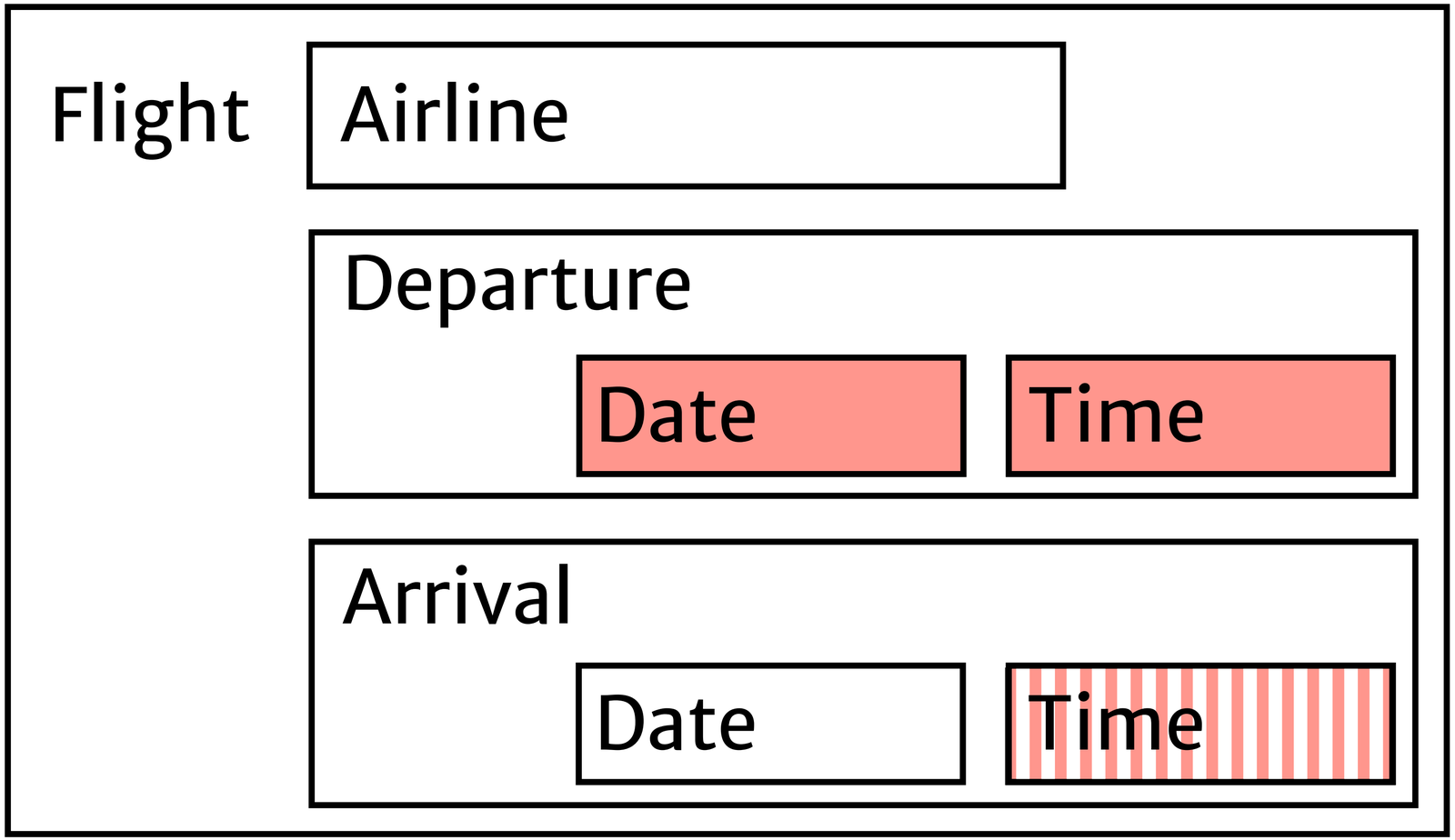
Structured semantic queries



Hi Mom,
Just booked my flight home
for the holiday! I'll be leavi
Boston at noon on Sunday t
18th and getting in at 3:30.
Looking forward to seeing e
soon!

Structured semantic queries

Key advantage: can perform **automated reasoning** on structured meaning reps.!



Hi Mom,
Just booked my flight home
for the holiday! I'll be leavi
Boston at noon on Sunday t
18th and getting in at 3:30.
Looking forward to seeing e
soon!

Entity recognition

Hi Mom,
Just booked my flight home for the holiday! I'll be leaving Boston at noon on Sunday the 18th and getting in at 3:30. Looking forward to seeing everyone soon!

Entity recognition

Hi Mom,
Just booked my flight home for the holiday! I'll be leaving Boston at noon on Sunday the 18th and getting in at 3:30. Looking forward to seeing everyone soon!

Entity recognition

$p(\texttt{TIME} \mid noon\ , \ldots)$

Classify tokens
(e.g. w/ contextual representations)

∅ PLACE ∅ TIME ∅ DATE DATE DATE ∅

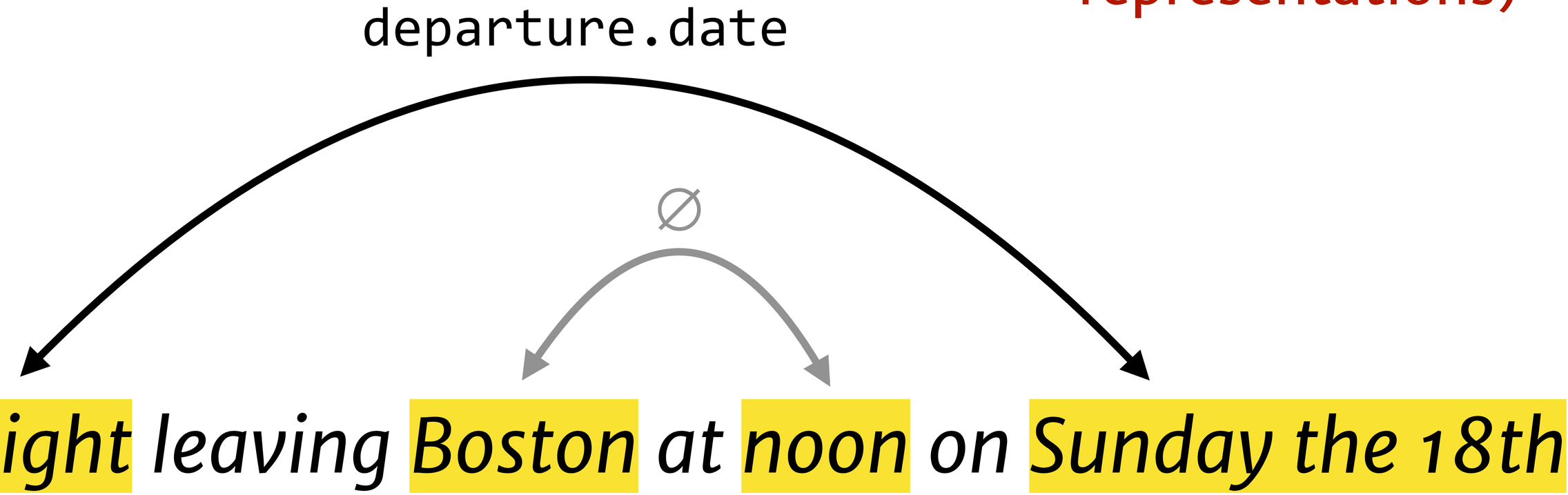
CRF / transformer

leaving Boston at noon on Sunday the 18th and

Relation classification

$p(\texttt{departure.date} \mid \textit{flight}, \textit{Sunday}, \ldots)$

Classify entity pairs (e.g. with ctx. representations)



One-shot relation extraction

$p(\texttt{departure.date} \mid \textit{flight}, \textit{Sunday}, \ldots)$

need supervised data for every relation!

One-shot relation extraction

$p(\texttt{departure.date} \mid \textit{flight}, \textit{Sunday}, \ldots)$

need supervised data for every relation!

instead, model:

$p(\texttt{true} \mid \texttt{departure.date}, \textit{flight}, \textit{Sunday}, \ldots)$

[e.g. Levy et al. 2017]

One-shot relation extraction

$p(\texttt{departure.date} \mid \textit{flight}, \textit{Sunday}, \ldots)$

need supervised data for every relation!

instead, model:

$p(\texttt{true} \mid \texttt{departure.date}, \textit{flight}, \textit{Sunday}, \ldots)$

$p(\texttt{true} \mid \textit{departure date}, \textit{flight}, \textit{Sunday}, \ldots)$

$p(\texttt{true} \mid \textit{my \_ leaves on \_}, \textit{flight}, \textit{Sunday}, \ldots)$

Open challenges

General-purpose knowledge base construction *without* a predefined schema?

Robust automated reasoning about noisily extracted facts?

Extractive question answering

Answering questions about documents

Key idea: for lots of question answering problems, the answer occurs as a span in the document.

Just treat this as a span classification problem!

In 1928, he went to Berlin University for language training for a year, and then he started studying mathematics at University of Göttingen in the summer semester of 1929. He studied algebra under Emmy Noether. Tsen received his doctoral degree in February 1934 under the supervision of Emmy Noether and Friedrich Karl Schmidt, and he dedicated his dissertation to his elder cousin Tsebu S. Lee. Having fled to the US, Noether evaluated the dissertation in a letter as "sehr gut" (very good). As a research fellow sponsored by the China Foundation for the Promotion of Education and Culture, Tsen did a postdoctoral research with Emil Artin at Hamburg University for a year. There he became friends with Shiing-Shen Chern, who was a graduate student back then. Chern remembered him as a cordial and open-minded person well-liked by everyone.[2]

Who was Zeng Jiongzhi's doctoral adviser? Emmy Noether.

Training data

Jacksonville is the largest city by population in the U.S. state of Florida, and the largest city by area in the contiguous United States. It is the county seat of Duval County, with which the city government consolidated in 1968. Consolidation gave Jacksonville its great size and placed most of its metropolitan population within the city limits; with an estimated population of 853,382 in 2014, it is the most populous city proper in Florida and the Southeast, and the 12th most populous in the United States. Jacksonville is the principal city in the Jacksonville metropolitan area, with a population of 1,345,596 in 2010.

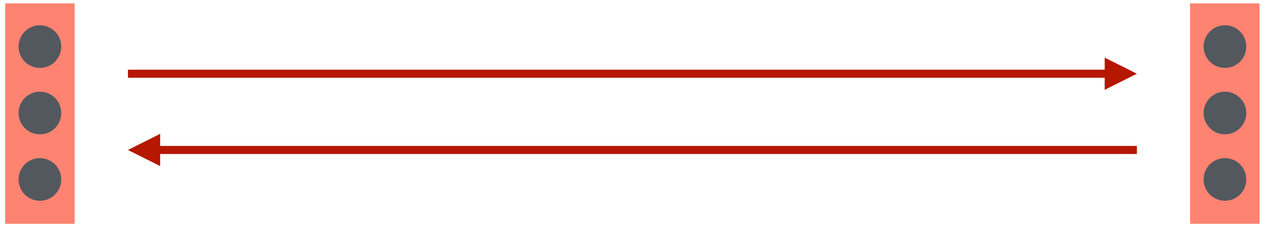
Which Florida city has the biggest population? Jacksonville

What was the population of Jacksonville in 2010? 1,345,596

In which county does Jacksonville reside? Duval County

[Rajpurkar et al. 2016]

Simple attentional models

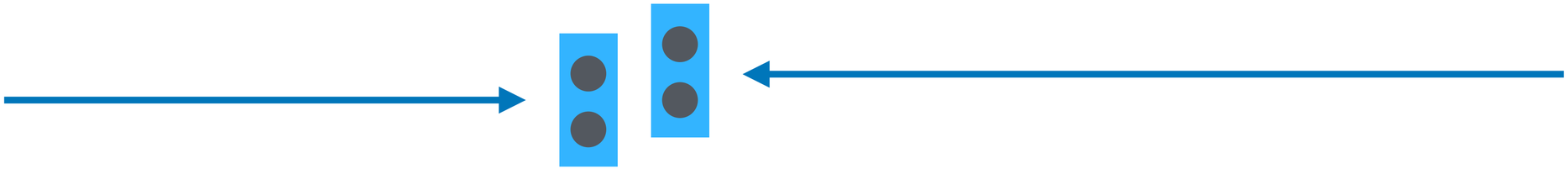


v_q: question rep.

In which county does Jacksonville reside?

v_d: document rep.

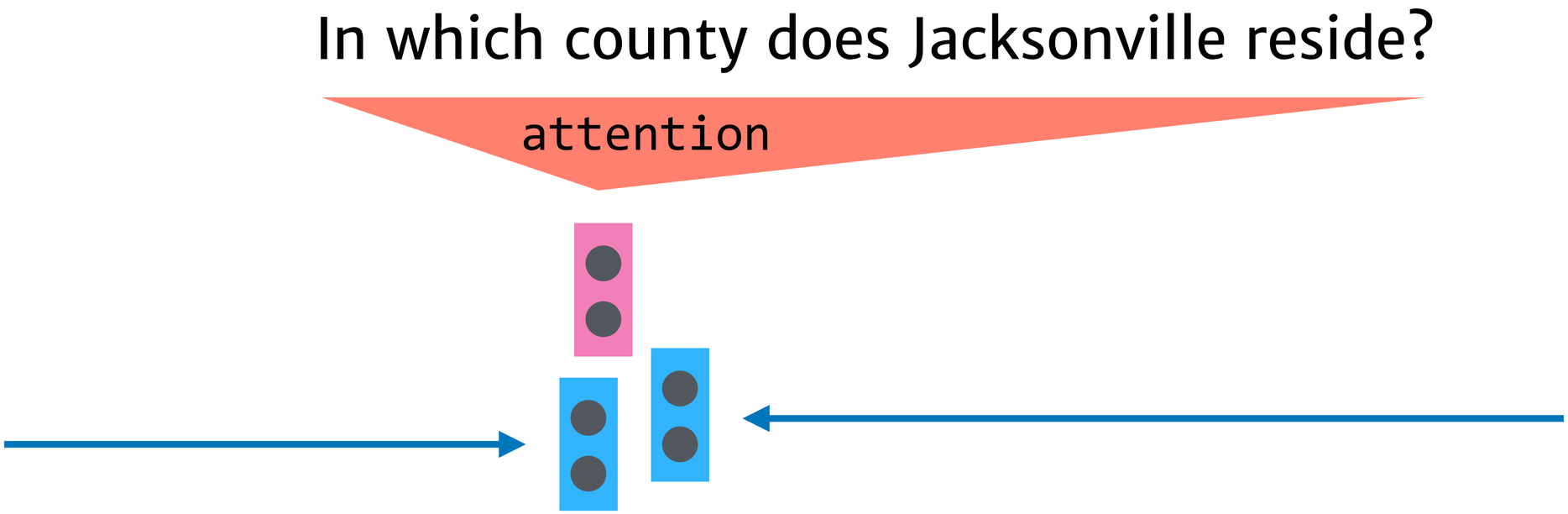
$$p(\text{start} = i) \propto v_q^\top v_d$$



It is the county seat of Duval County, with which the city government

[e.g. Chen & Manning 2017]

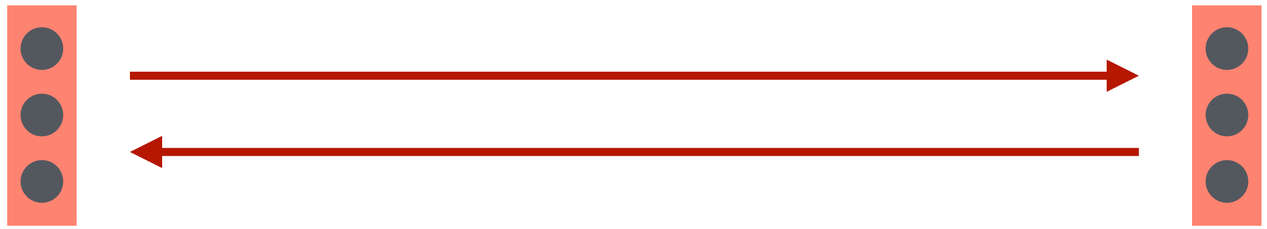
“Attention flow”-based models



It is the county seat of Duval County, with which the city government

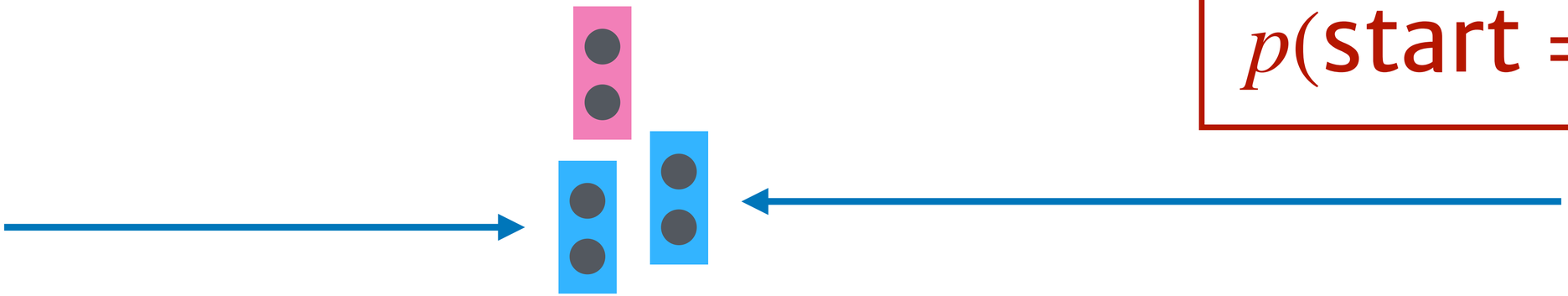
[e.g. Lee et al. 2016, Seo et al. 2016]

“Attention flow”-based models



In which county does Jacksonville reside?

$$p(\text{start} = i) \propto v_q^\top v_d$$



It is the county seat of Duval County, with which the city government

[e.g. Lee et al. 2016, Seo et al. 2016]

What is this good for?

Successes for passage-based QA:

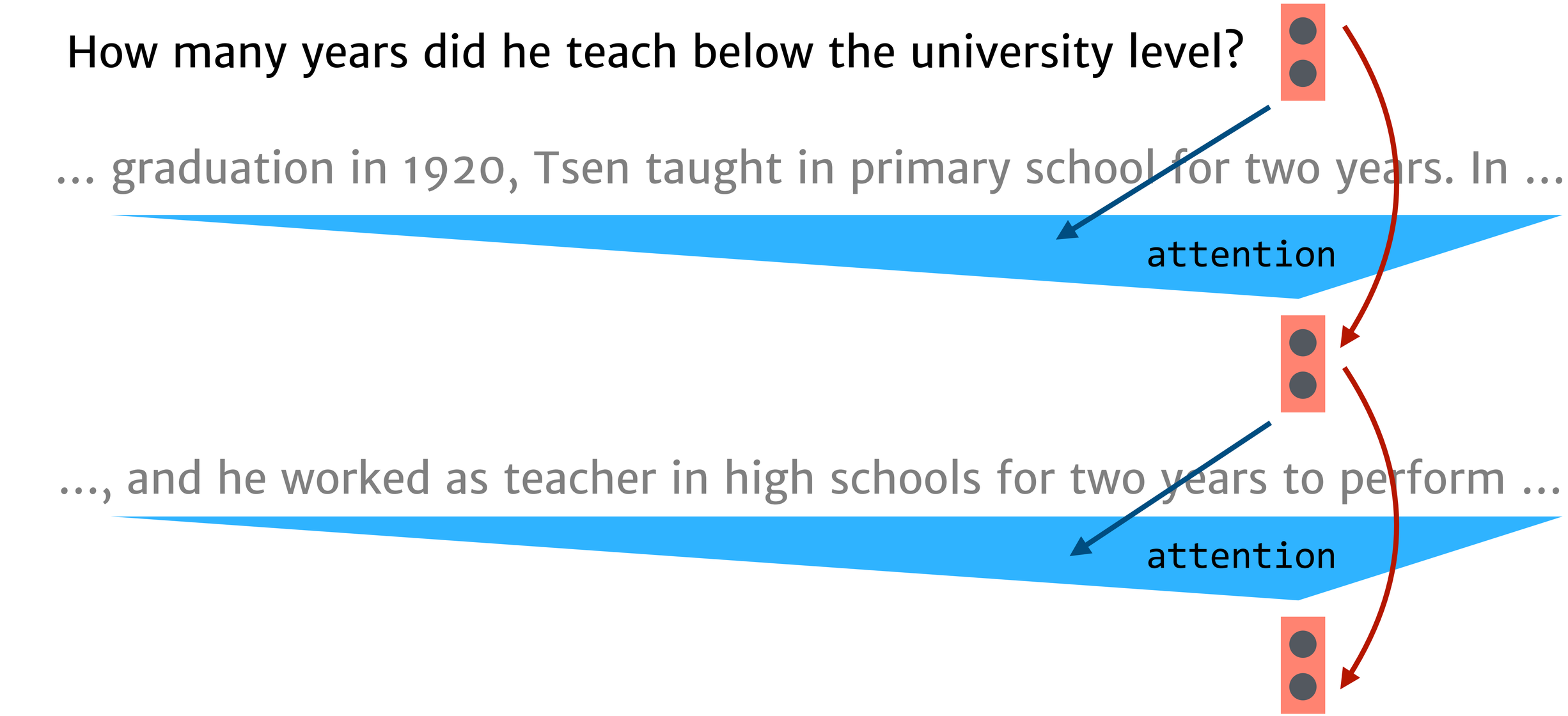
- short source documents (1 or 2 paragraphs)
- simple factoid questions
- accuracy on standard English benchmarks:
  ~65% for simple neural models
  ~86% human accuracy
  ~90% fanciest neural models (pretraining etc.)

Multi-hop questions

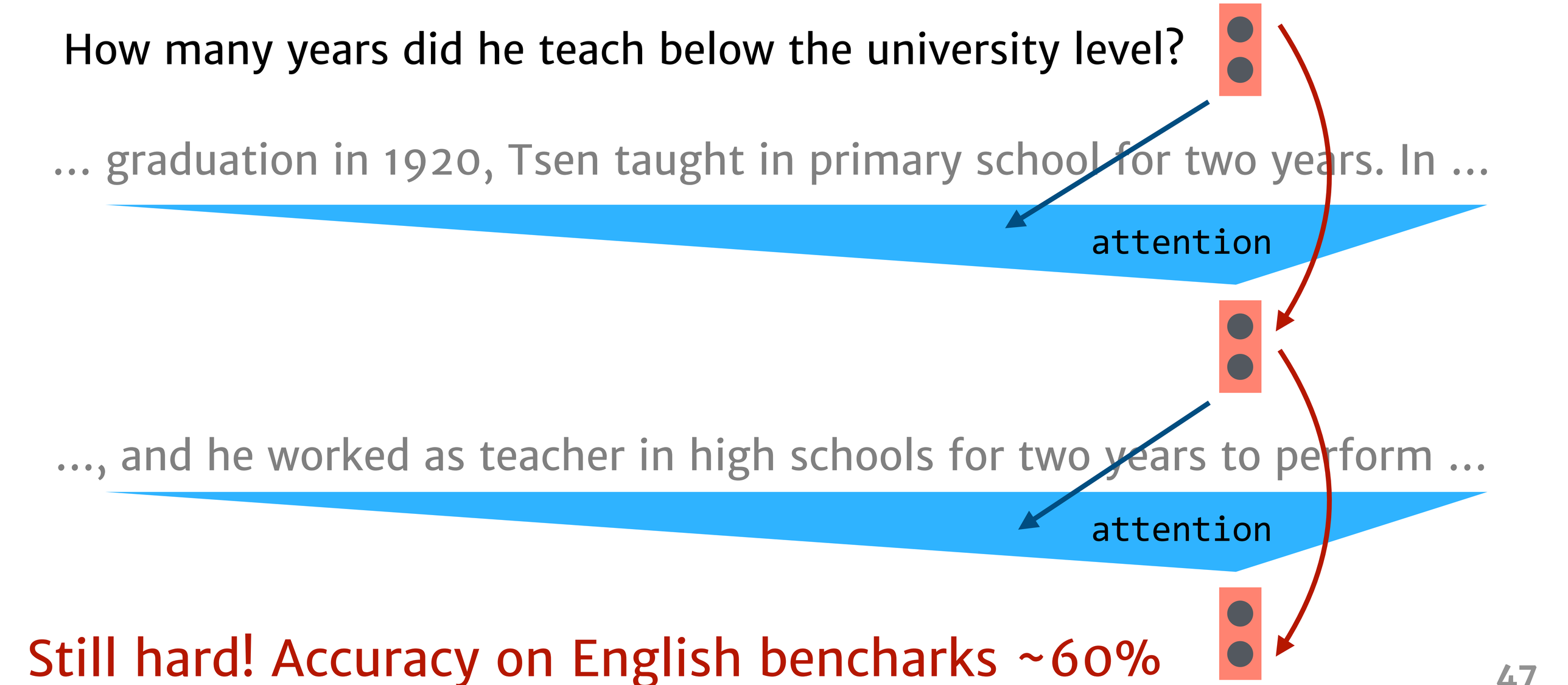
In 1917, he passed the entrance examination and was admitted to Jiangxi Provincial First Normal College in Nanchang. He was subsidised by Lei Heng's son Tsebu S. Lee (雷子布), who was studying in Japan on government scholarship. After graduation in 1920, Tsen taught in primary school for two years. In 1922, Tsen entered National Wuchang Senior Normal College[n 3] to study undergraduate mathematics, and he graduated in 1926. After graduation, he passed a scholarship examination for studying in the West, and he worked as teacher in high schools for two years to perform the mandatory teaching service of his degree.[3]

How many years did he spend teaching below the university level? Four years.

Models for multi-hop QA



Models for multi-hop QA



Still hard! Accuracy on English bencharks ~60%

Retrieving documents

In 1917, he passed the entrance examination and was admitted to Jiangxi Provincial First Normal College in Nanchang. He was subsidised by Lei Heng's son Tsebu S. Lee (雷子布), who was studying in Japan on government scholarship. After graduation in 1920, Tsen taught in primary school for two years. In 1922, Tsen entered National Wuchang Senior Normal College[n 3] to study undergraduate mathematics, and he graduated in 1926. After graduation, he passed a scholarship examination for studying in the West, and he worked as teacher in high schools for two years to perform the mandatory teaching service of his degree.[3]

How many years did he spend teaching below the university level? Four years.

All models so far: assume passage is given!

What if we could compute $\sum_{\text{doc,answer}} p(\text{doc})\ p(\text{answer} \mid \text{doc})$?

Pipelined approaches

Easy: just retrieve passages with an information retrieval system like from the beginning of this lecture.

Then run models trained with passage supervision to answer open-ended questions!

[e.g. Chen & Manning 2017]

End-to-end approaches

Learn the information retrieval step
to optimize answer accuracy!

either:

(1) fix a retrieval engine and use reinforcement learning to pick the right key [Buck et al. 2018]

(2) build an ENORMOUS sparse transformer and train end-to-end [Guu et al. 2020]

Harder problems

Abstractive question answereing

Can we train models to generate novel answer text?

Q: What exactly are vitamins?

Human Answer

Vitamins are a kind of nutrition that your body needs to work properly . Just like you need to eat a variety of different foods , you need a bunch of different vitamins .

[Fan et al. 2019]

Abstractive question answering

Q: What exactly are vitamins?

Extractive Short

Vitamins Facts Exactly what Nutritional supplements ? What Do Minerals and vitamins Complete ? All vitamins are essential . P > The fat soluble vitamins include Vitamin A , Vitamin D , Vitamin E , and Vitamin K. These all dissolve in fat . This being said , any diet that excludes a certain food group is excluding the vitamins supplied by this food group , and can lead to vitamin deficiency .

(model: sentences with lots of words that overlap w question)

[Fan et al. 2019]

Abstractive question answering

Q: What exactly are vitamins?

Generative Short

There are a lot of different types of vitamins and minerals . For example , vitamin A , vitamin C , and vitamin C are all vitamins , but they are all different . They are all made up of a bunch of different things . The body needs many things like vitamins , minerals , and proteins . These can be broken down into amino acids . These vitamins are then broken down by the body to make proteins , which can then be used to build proteins . The body can then use these amino acids to build the proteins . This process is called ' building blocks ' , and it can be done in a variety of ways .

[Fan et al. 2019]

(model: big language model)

Abstractive question answering

Still a *very* hard problem!

From retrieval to reasoning

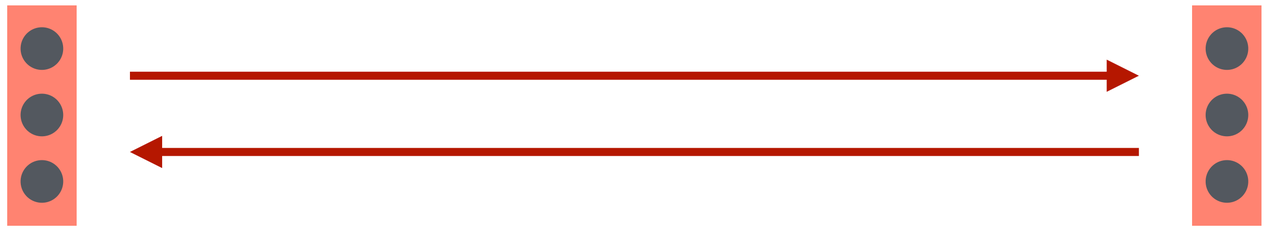
Can we integrate logical and numerical reasoning?

Q: how many yards was the second longest passing touchdown?

Hoping to rebound from their loss to the Patriots, the Raiders stayed at home for a Week 16 duel with the Houston Texans. Oakland would get the early lead in the first quarter as quarterback JaMarcus Russell completed a 20-yard touchdown pass to rookie wide receiver Chaz Schilens. The Texans would respond with fullback Vonta Leach getting a 1-yard touchdown run, yet the Raiders would answer with kicker Sebastian Janikowski getting a 33-yard and a 30-yard field goal. Houston would tie the game in the second quarter with kicker Kris Brown getting a 53-yard and a 24-yard field goal. Oakland would take the lead in the third quarter with wide receiver Johnnie Lee Higgins catching a 29-yard touchdown pass from Russell, followed up by an 80-yard punt return for a touchdown. The Texans tried to rally in the fourth quarter as Brown nailed a 40-yard field goal, yet the Raiders' defense would shut down any possible attempt.

[Dua et al. 2019]

Numerical reasoning



v_q: question rep.

What was the length difference between the two longest touchdowns?

label: –

label: +

a 29-yard touchdown pass from Russell, followed up by an 80-yard TD

"do arithmetic" by assigning signs to passage numbers

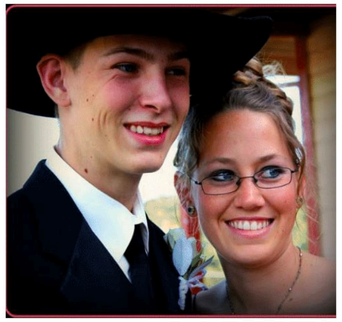
[Dua et al. 2019]

Visual question answering

Who is wearing glasses?

man

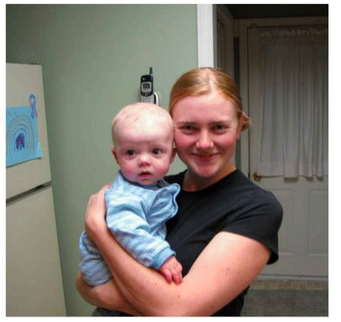
woman



Where is the child sitting?

fridge

arms



Is the umbrella upside down?

yes

no

How many children are in the bed?

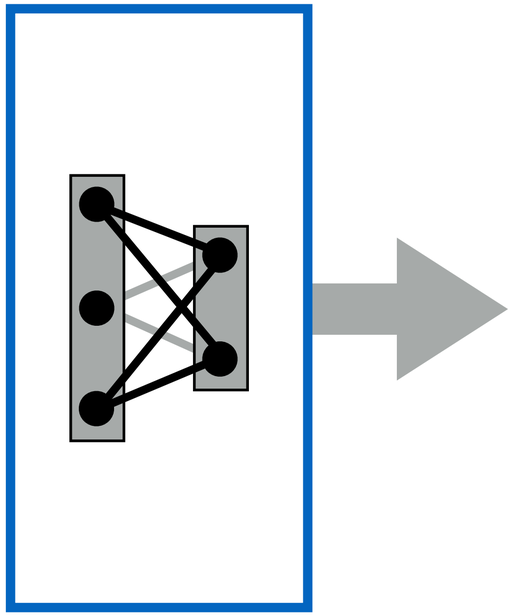
2

1

[Antol et al. 2016]

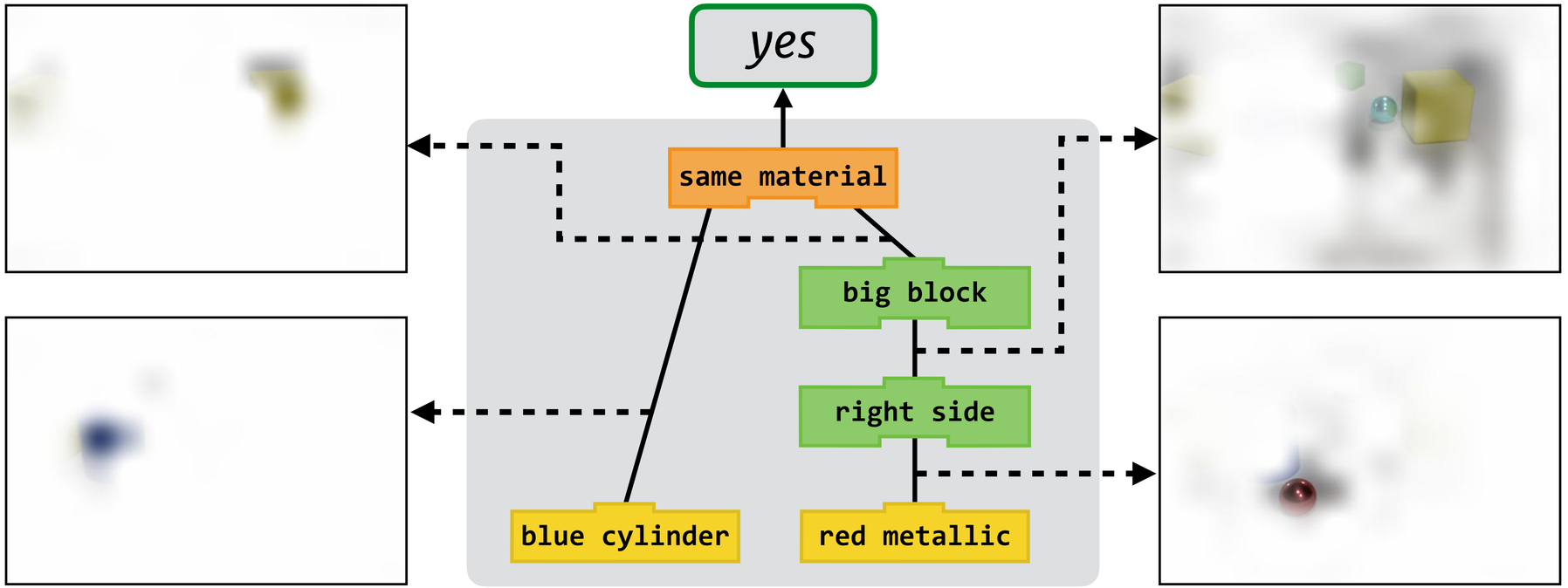
Models for visual QA

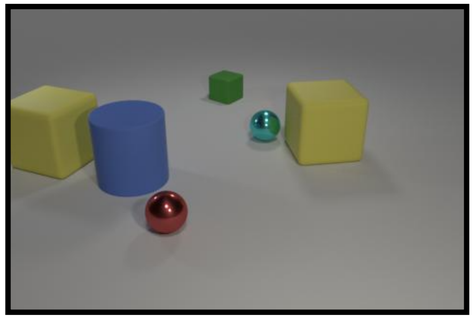
What color is the necktie?



yellow

Models for visual QA





Does the blue cylinder have the same material as the big block on the right side of the red metallic thing?

Question answering for the real world

Answer bias

Correct Response

Predicted A: 2

Incorrect Responses

Q: How many zebras

Predicted A: 2

Predicted A: 2

Predicted A: 2

All Correct Responses

Q: What covers the ground

Predicted A: snow

Predicted A: snow

Predicted A: snow

Predicted A: snow

Predicted A: snow

[Agrawal et al. 2016]

Structural bias

Article: Super Bowl 50
Paragraph: "*Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.*"
Question: "*What is the name of the quarterback who was 38 in Super Bowl XXXIII?*"
Original Prediction: John Elway
Prediction under adversary: Jeff Dean

[Jia & Liang 2017]

Challenges for question answering

Existing models achieve “superhuman” performance on standard benchmarks without much understanding.

(Learning very precise models of human annotators is a lot easier than learning to read!)

Challenges for question answering

Existing models achieve "superhuman" performance on standard benchmarks without much understanding.

(Learning very precise models of human annotators is a lot easier than learning to read!)

Most existing QA datasets are designed to be interesting to academic researchers, not useful to ordinary people.

Visual question answering revisited

VQA for blind users:

Q: Does this foundation have any sunscreen?

A: yes

Q: What is this?

A: 10 euros

Q: What color is this?

A: green

Q: Please can you tell me what this item is?

A: butternut squash red pepper soup

Totally different problem—mostly about OCR!

[Gurari et al. 2018]

Solutions

Technical solutions:

- auto-balance output labels
- train on adversarial examples

Human solutions:

- collect data from real people and organic processes!

Next week: Dialogue