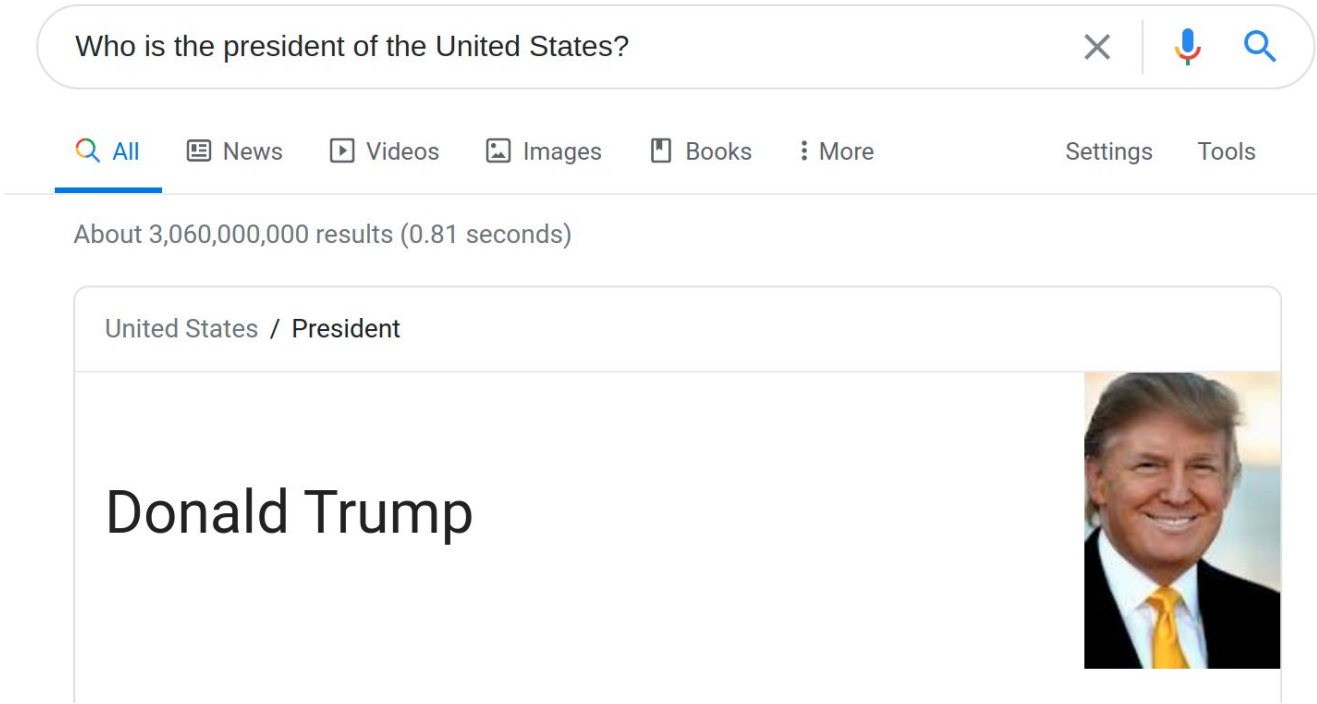
Machine Learning for Text using Latent Information

by Kechen Qin

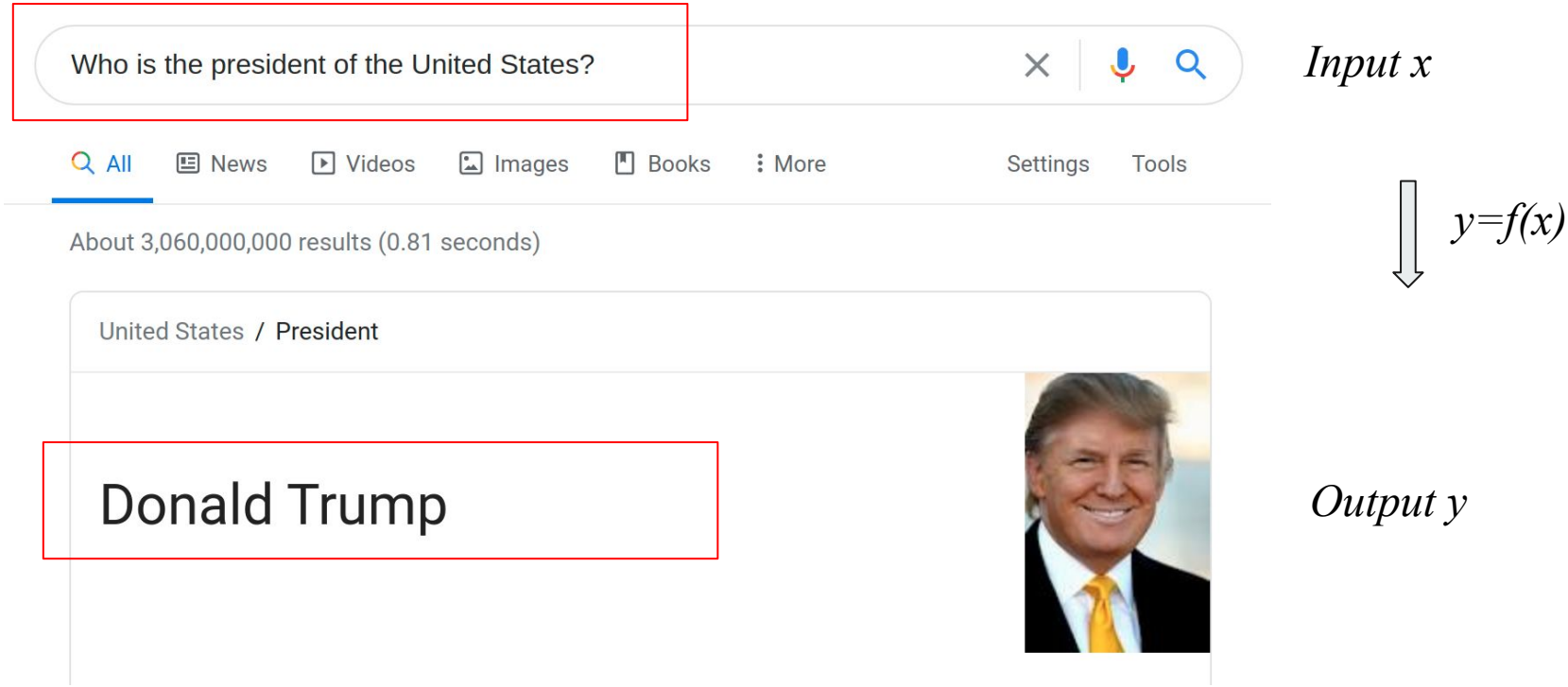
Khoury College of Computer Sciences
Northeastern University

Advisor: Javed A. Aslam
Committee: Virgil Pavlu, Christopher Amato, Xiaohui Cui

Supervised Learning



Supervised Learning



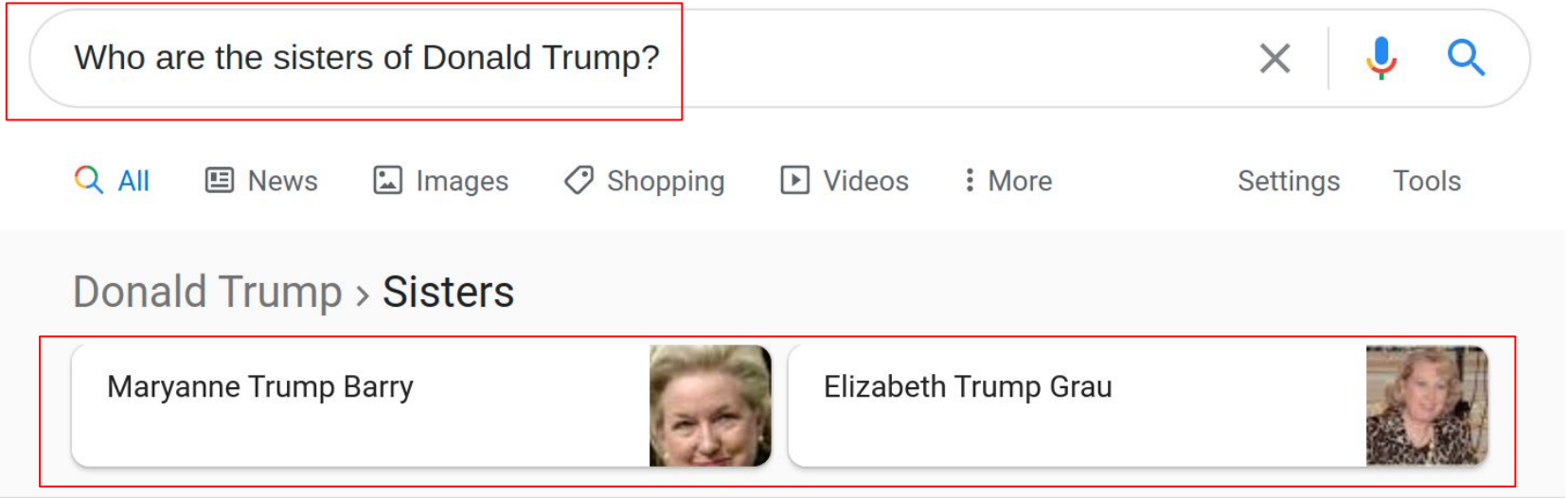
Input x

$y=f(x)$

Output y

Supervised Learning

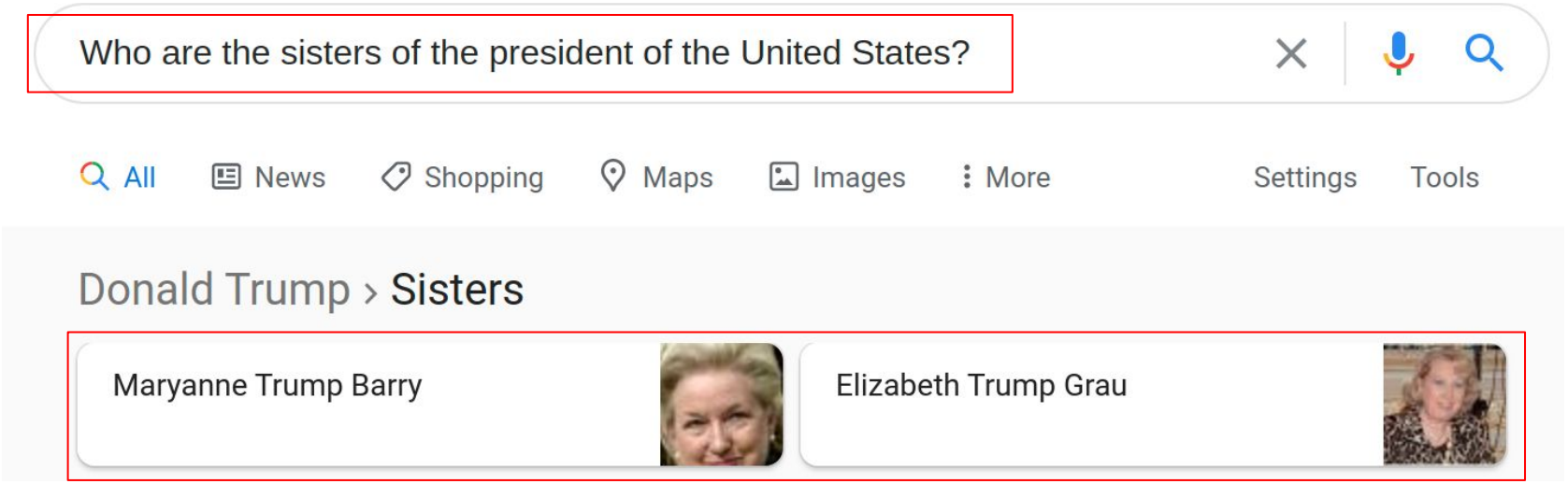
Input x



Output y

Supervised Learning

Input x



Output y

Supervised Learning

Who is the sister of the president of the United States? x

Who is the president of the United States? → Donald Trump

Who is the sister of Donald Trump?

$y=f(x)$

Maryanne Trump Barry and Elizabeth Trump Grau y

Supervised Learning with Latent Information

Who is the sister of the president of the United States? x

$z=f(x)$

Who is the president of the United States? → Donald Trump

z

Who is the sister of Donald Trump?

$y=f(z)$

Maryanne Trump Barry and Elizabeth Trump Grau y

Latent information z = *latent questions answer pairs*

Supervised Learning with Latent Information

Who is the sister of the president of the United States? x

$z=f(x)$

Who is the president of the United States?→ Donald Trump

Who is the mother of Donald Trump?→ Mary Anne Trump

Who are the daughters of Mary Anne Trump?

z

$y=f(z)$

Maryanne Trump Barry and Elizabeth Trump Grau y

Latent information z = *latent questions answer pairs*

Supervised Learning

A: I was just wondering if we want to have a rubber cover instead of a plastic one.

B: Yeah.

D: Alright. That could be a good idea.

C: So instead of the fascia that comes off being plastic, the fascia that comes off would be the ruber.

E: Alright. That could be a good idea. It would be comfortable to hold on also.

B: Well that's been really popular with mobile phones so I don't see why not.

Supervised Learning

A: I was just wondering if we want to have a rubber cover instead of a plastic one.

B: Yeah.

D: Alright. That could be a good idea.

C: So instead of the fascia that comes off being plastic, the fascia that comes off would be the ruber.

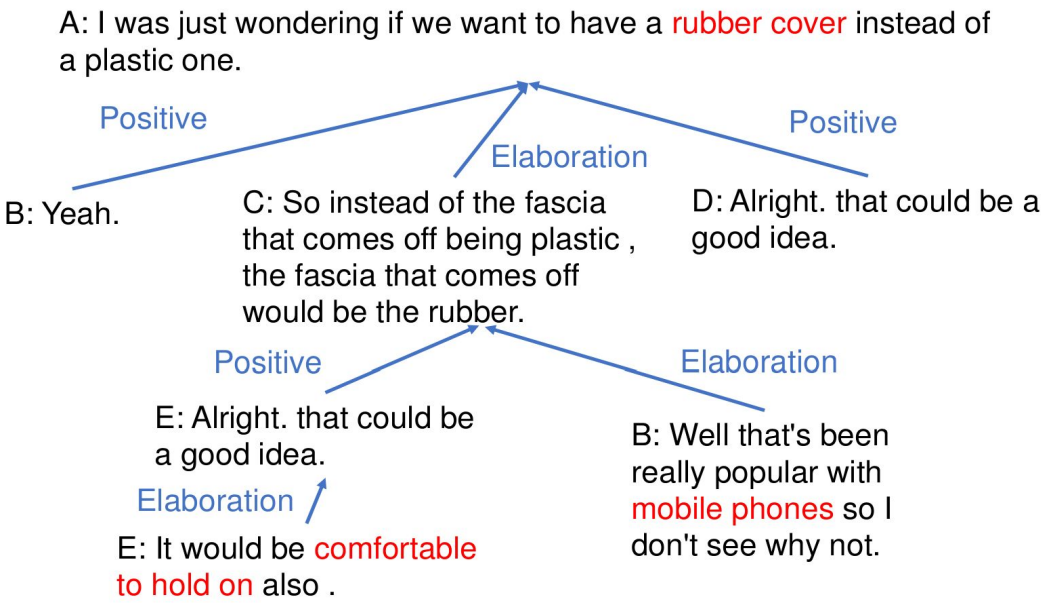
E: Alright. That could be a good idea. It would be comfortable to hold on also.

B: Well that's been really popular with mobile phones so I don't see why not.

x: meeting transcript

y: summary-worthy text spans

Supervised Learning with Latent Information



x: meeting transcript

z: latent conversation structure

y: summary-worthy text spans

More examples of Latent Variable Applications

- **Face recognition:** the gender of the person, the orientation of the face.
- **Object recognition:** the pose parameters of the object (location, orientation, scale), the lighting conditions.
- **Machine translation:** the word-to-word correspondences (word alignment).
- **Parts of Speech Tagging:** the segmentation of the sentence into syntactic units, the parse tree.
- **Speech Recognition:** the segmentation of the sentence into phonemes or phones.
- **Handwriting Recognition:** the segmentation of the line into characters.
- ...

Latent Variable Model

Instead of modeling a distribution $p(y|x)$ directly, we can introduce an unobserved latent variable z to represent the intermediate state. The joint distribution over the target and latent variable conditioned on the observed input can be written down as:

$$p(y, z|x) = p(y|z)p(z|x)$$

Usually, we can assume that z is a discrete variable. By marginalizing out all possible state of z, we obtain the desired data distribution $p(y|x)$:

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

Latent Variable Model

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

x=input, y=output, z=latent variable

Latent Variable Model

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

x=input, y=output, z=latent variable

- What latent information can bridge x and y → What is z ?

Latent Variable Model

$$p(y|x) = \sum_z p(y|z)p(z|x)$$

x=input, y=output, z=latent variable

- What latent information can bridge x and y → What is z ?
- How to model the relationship between latent information and observed variables → How to model $p(y|z)$ and $p(z|x)$?

Latent Variable Model

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

x=input, y=output, z=latent variable

- What latent information can bridge x and y → What is z ?
- How to model the relationship between latent information and observed variables → How to model $p(y|z)$ and $p(z|x)$?
- How to train the model if z is not observed → How to estimate the values of z during training ?

Machine Learning for Text using Latent Information

- Introduction to latent information
- Application 1: Latent label order in **multi-label classification**
- Application 2: Latent reasoning path in **knowledge based question answering**
- Timeline

What is Multi-Label Classification

***x*:**

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$y=f(x)$

***y*:**

{Champions league,
Sportblog,
José mourinho,
Internazionale,
Real madrid,
Bayern munich,
Champions league 2009-10}

Binary Relevance (BR)

x:

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$$b_1(y_1|x)$$
$$b_2(y_2|x)$$
$$b_3(y_3|x)$$
$$\dots$$
$$b_n(y_n|x)$$

y:

Champions league
Sportblog
José mourinho
Internazionale
Real madrid
Bayern munich
Champions league 2009-10

Binary Relevance (BR)

***x*:**

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$b_1(y_1|x)$
$b_2(y_2|x)$
$b_3(y_3|x)$
...
$b_n(y_n|x)$

***y*:**

Champions league
Sportblog
José mourinho
Internazionale
Real madrid
Bayern munich
Champions league 2009-10

n equals to the number of candidate tags, which could be more than a million!

Binary Relevance (BR)

x:

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$$b_1(y_1|x)$$
$$b_2(y_2|x)$$
$$b_3(y_3|x)$$
$$\dots$$
$$b_n(y_n|x)$$

y:

Champions league
Sportblog
José mourinho
Internazionale
Real madrid
Bayern munich
Champions league 2009-10

v.s.

final

Binary Relevance (BR)

x:

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$$b_1(y_1|x)$$
$$b_2(y_2|x)$$
$$b_3(y_3|x)$$
$$\dots$$
$$b_n(y_n|x)$$

y:

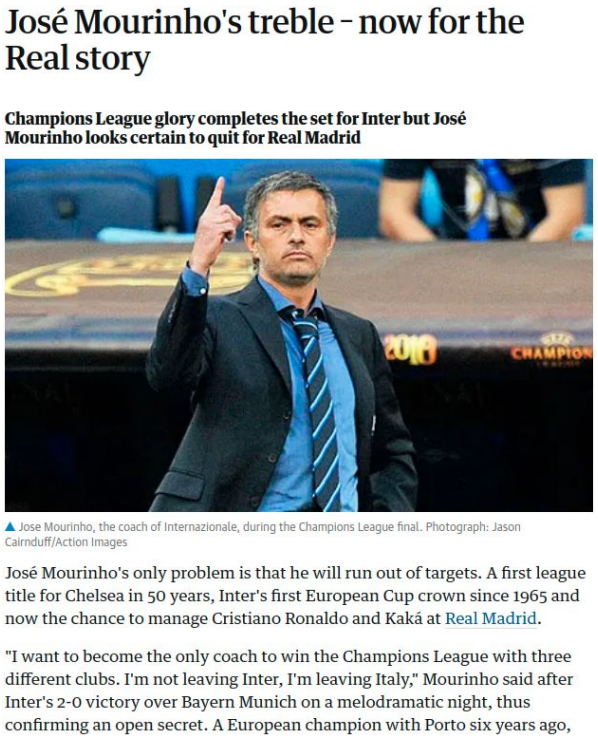
Champions league
Sportblog
José mourinho
Internazionale
Real madrid
Bayern munich
Champions league 2009-10
~~Champions league 2015-16~~

v.s.

final

Recurrent Neural Network (RNN)

x:



José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

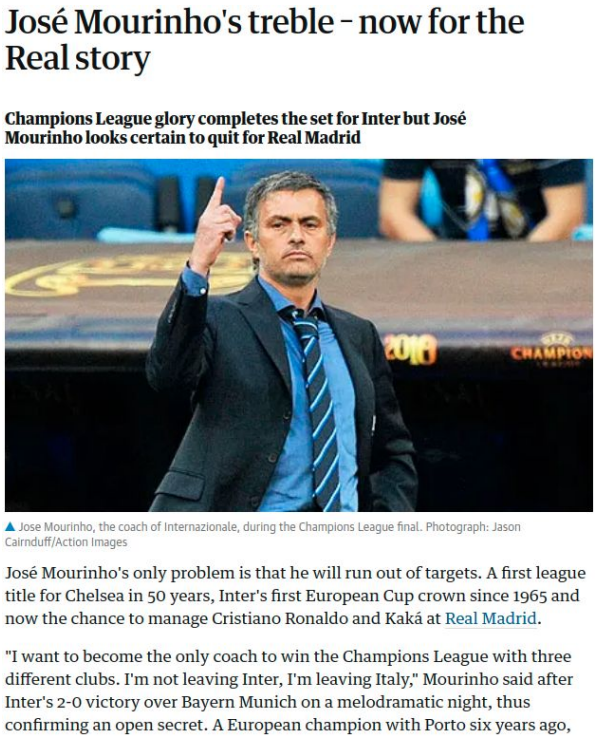
$$f(y_1|x)$$
$$f(y_2|x,y_1)$$
$$f(y_3|x,y_1,y_2)$$
$$\ldots$$
$$f(y_t|x,y_1,\ldots,y_{t-1})$$

y:

Sportblog→Champions league→
Champions league 2009-10→
Bayern munich→Internazionale→
José mourinho→Real madrid

Recurrent Neural Network (RNN)

x:



José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$$f(y_1|x)$$
$$f(y_2|x,y_1)$$
$$f(y_3|x,y_1,y_2)$$
$$\ldots$$
$$f(y_t|x,y_1,\ldots,y_{t-1})$$

y:

Sportblog→Champions league→
Champions league 2009-10→
Bayern munich→Internazionale→
José mourinho→Real madrid

Problem of Using a Predefined Label Order

x:

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter. I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$y_1,\cdots,y_t$:

Frequency:

Sportblog→Champions league→Real madrid→ José mourinho→y_t=?

Hierarchy:

Sportblog→Champions league→ Champions league 2009-10→Bayern munich→ y_t=?

Problem of Using a Predefined Label Order

x:

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter. I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$y_1,\cdots,y_t$:

Frequency:

Sportblog→Champions league→Real madrid→ José mourinho→y_t=Cristiano Ronaldo

Hierarchy:

Sportblog→Champions league→ Champions league 2009-10→Bayern munich→ y_t=Internazionale

Problem of Using a Predefined Label Order

x:

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter. I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$y_1,\dots,y_t$:

Frequency:

Sportblog→Champions league→Real madrid→ José mourinho→y_t=Cristiano Ronaldo

Hierarchy:

Sportblog→Champions league→ Champions league 2009-10→Bayern munich→ y_t=Internazionale

ORDER MATTERS!

That is our latent information!

Label Order as Latent Variable

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

x: raw text in document (no image feature)

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid….

y: a set of labels

{Sportblog, Champions league, Champions league 2009-10, Bayern munich, Internazionale, José mourinho, Real_madrid}

z: each label set permutation represent a way to sort labels

$[z_1, z_2, z_3, \ldots, z_n]$ = [Sportblog, Champions league, Champions league 2009-10, Bayern munich, Internazionale, José mourinho, Real_madrid]

Label Order as Latent Variable

$$p(y|x) = \sum_z p(y|z)p(z|x) \xrightarrow{p(y|z)=1} p(y|x) = \sum_z p(z|x)$$

x: raw text in document

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid….

y: a set of labels

{Sportblog, Champions league, Champions league 2009-10, Bayern munich, Internazionale, José mourinho, Real_madrid}

z: each label set permutation represent a way to sort labels

$[z_1, z_2, z_3, \ldots, z_n]$ = [Sportblog, Champions league, Champions league 2009-10, Bayern munich, Internazionale, José mourinho, Real_madrid]

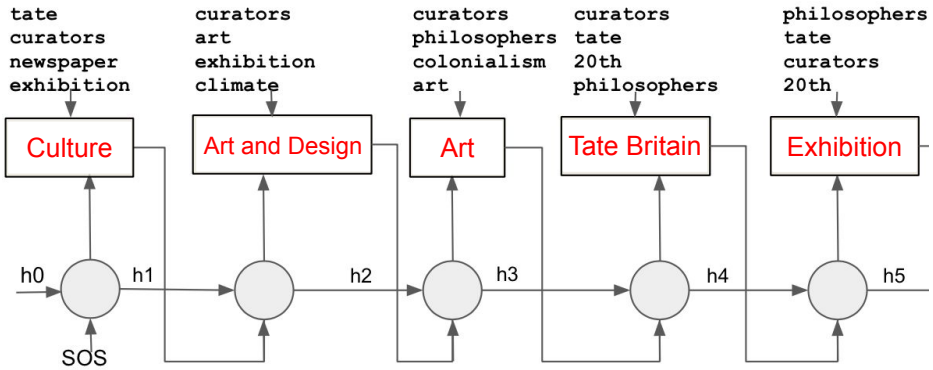
Set Permutation Probability $p(z|x)$

$$p(z|x)$$
$$=p(z_1,z_2,\ldots,z_n|x)$$
$$=p(z_1|x)p(z_2|x,z_1)\ldots p(z_n|x,z_1,z_2,\ldots,z_{n-1})$$

At each timestep t, we estimate $p(z_t|\ldots)$ using a multi-class classification model:

$$p(z_t|x,z_1,\ldots,z_{t-1})=softmax(\ [f(z_1,\ldots,z_{t-1});\ f(x)]\ )$$

Where $f(*)$ is a mapping function from random variable to its vector representation.



Train the model without supervision on z

$$p(y|x) = \sum_{z} p(z|x)$$

The number of different set permutations = the factorial of number of labels in the set

$$|z| = |y|!$$

7! = 5040 = run and update model 5040 times for 1 sample!

Train the model without supervision on z

$$p(y|x) = \sum_{z} p(z|x)$$

We use easily learnable label orders to train the model.

Select the Most Probable Label Orders

José Mourinho's treble – now for the Real story

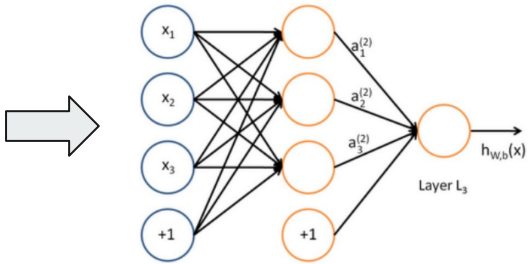
Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

{Champions league, Sportblog, José mourinho, Internazionale, Real madrid, Bayern munich, Champions league 2009-10}



p(z=[Champions league, ..., Real_madrid]|x) = 0.221
p(z=[Champions league, ..., Bayern munich]|x) = 0.143
p(z=[Champions league, ..., Sportblog]|x) = 0.082
...
p(z=[Sportblog, ..., Real madrid]|x) = 0.001

Select the Most Probable Label Orders

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

{Champions league, Sportblog, José mourinho, Internazionale, Real madrid, Bayern munich, Champions league 2009-10}

Most recent trained sequence prediction model (i.e. the model used to calculate p(z|x))

$p(z=[Champions\ league, ..., Real\_madrid]|x) = 0.221$
$p(z=[Champions\ league, ..., Bayern\ munich]|x) = 0.143$
$p(z=[Champions\ league, ..., Sportblog]|x) = 0.082$
...
$p(z=[Sportblog, ..., Real\ madrid]|x) = 0.001$

Select the Most Probable Label Orders

Step 1: Initialize model parameters.

Step 2: Get top permutations sorted by $p(z|x)$.

Step 3: Update model parameters by maximizing $p(y|x)$.

Repeat step 2 and step 3 until model converges.

Experimental Results

Properties:

- **Capture label dependencies** - predict label in a sequential manner.
- **Good scalability** - solve multi-classification task at each step.
- **Label order as latent variable** - No need to predefine label order.
- **Easy to implement** - fit with any base models (we use RNN structure).

Experimental Results

Properties:

- **Capture label dependencies** - predict label in a sequential manner.
- **Good scalability** - solve multi-classification task at each step.
- **Label order as latent variable** - No need to predefine label order.
- **Easy to implement** - fit with any base models (we use RNN structure).

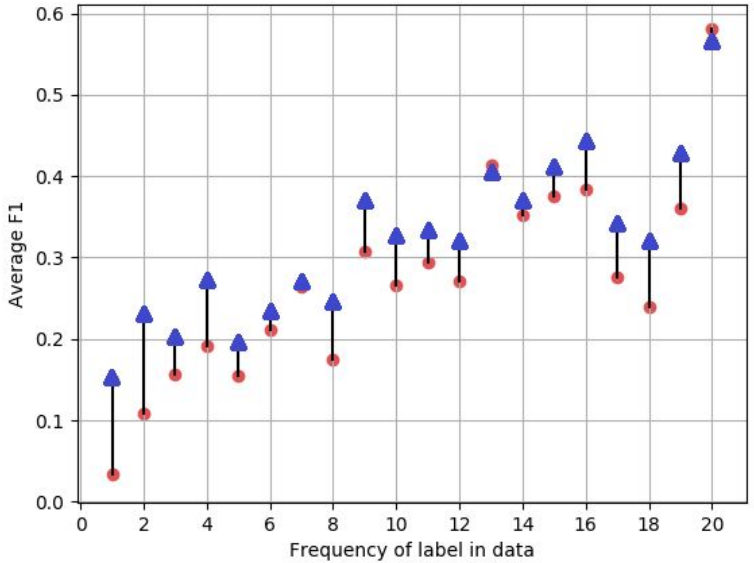
Methods	Slashdot		RCV1-v2		TheGuardian	
	label-F1	instance-F1	label-F1	instance-F1	label-F1	instance-F1
BR	.271	.484	.486	.802	.292	.572
PCC	.279	.480	.595	.818	-	-
standard RNN	.270	.528	.561	.824	.331	.603
Vinyals	.293	.530	.588	.829	.343	.599
Our Method	**.310**	**.538**	**.607**	**.838**	**.361**	**.607**

Better Performance on Rare Labels

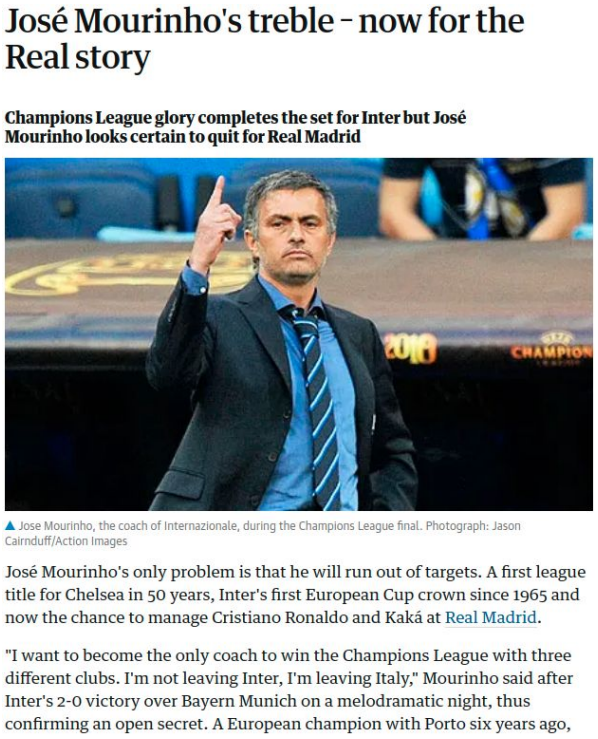
Datasets	label-F1		instance-F1	
	RNN	Our Method	RNN	Our Method
Slashdot	.270	.310	.528	.538
RCV1-V2	.561	.607	.824	.838
TheGuardian	.331	.361	.603	.607

$$\text{label-F1} = \frac{1}{L}\sum_{\ell=1}^{L}\frac{2\sum_{n=1}^{N} y_\ell^{(n)}\hat{y}_\ell^{(n)}}{\sum_{n=1}^{N} y_\ell^{(n)} + \sum_{n=1}^{N}\hat{y}_\ell^{(n)}}$$

$$\text{instance-F1} = \frac{1}{N}\sum_{n=1}^{N}\frac{2\sum_{\ell=1}^{L} y_\ell^{(n)}\hat{y}_\ell^{(n)}}{\sum_{\ell=1}^{L} y_\ell^{(n)} + \sum_{\ell=1}^{L}\hat{y}_\ell^{(n)}}$$



Proposed Work: Model Noisy Bookmarks



José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

User bookmarks:

guardian, article, real madrid, mourinho, championsleague, inter milan

Labeled tags:

Champions league, Sportblog, José mourinho, Internazionale, Real madrid, Bayern munich, Champions league 2009-10

Proposed Work: Model User Bookmarks

- Learn to clean noises from user bookmarks.

User bookmarks:

guardian, article, real madrid, mourinho, championsleague, inter milan

Labeled tags:

Champions league, Sportblog, José mourinho, Internazionale, Real madrid, Bayern munich, Champions league 2009-10

Proposed Work: Model User Bookmarks

- Learn to clean noises from user bookmarks.

User bookmarks:

~~guardian~~, ~~article~~, real madrid, mourinho, championsleague, inter milan

Labeled tags:

Champions league, Sportblog, José mourinho, Internazionale, Real madrid, Bayern munich, Champions league 2009-10

Proposed Work: Model User Bookmarks

- Learn to clean noises from user bookmarks.

User bookmarks:

~~guardian~~, ~~article~~, real madrid, mourinho, championsleague, inter milan

Labeled tags:

Champions league, Sportblog, José mourinho, Internazionale, Real madrid, Bayern munich, Champions league 2009-10

Proposed Work: Model User Bookmarks

- Model clean tags as latent variable.

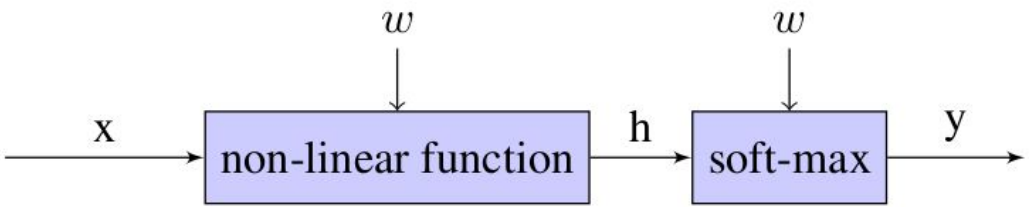
User bookmarks:

guardian, article, real madrid, mourinho, championsleague, inter milan

Labeled tags:

Champions league, Sportblog, José mourinho, Internazionale, Real madrid, Bayern munich, Champions league 2009-10

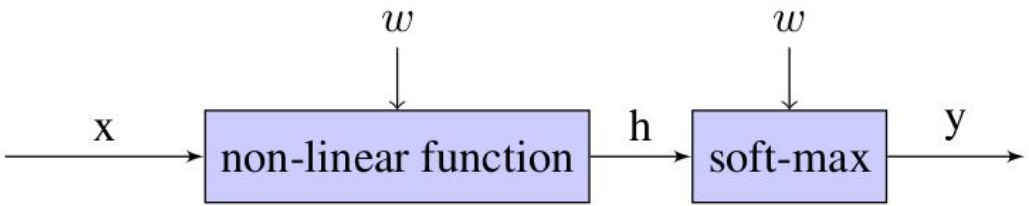
Proposed Work: Model User Bookmarks

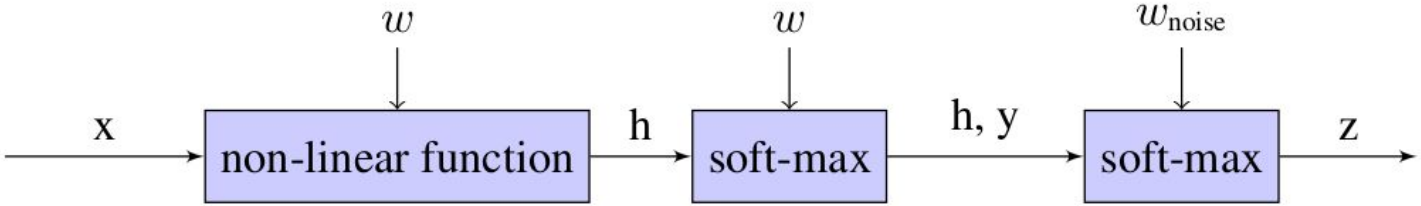


x: document
y: a set of clean tag
w: coefficients
h: hidden representation

Goldberger, Jacob, and Ehud Ben-Reuven. "Training deep neural-networks using a noise adaptation layer." (2016).

Proposed Work: Model User Bookmarks





x: document
z: a set of bookmarks
y: a set of clean tags
w: coefficients
h: hidden representation

Goldberger, Jacob, and Ehud Ben-Reuven. "Training deep neural-networks using a noise adaptation layer." (2016).

Machine Learning for Text using Latent Information

- Introduction to latent information
- Application 1: Latent label order in **multi-label classification**
- Application 2: Latent reasoning path in **knowledge based question answering**
- Timeline

What is Knowledge based Question Answering

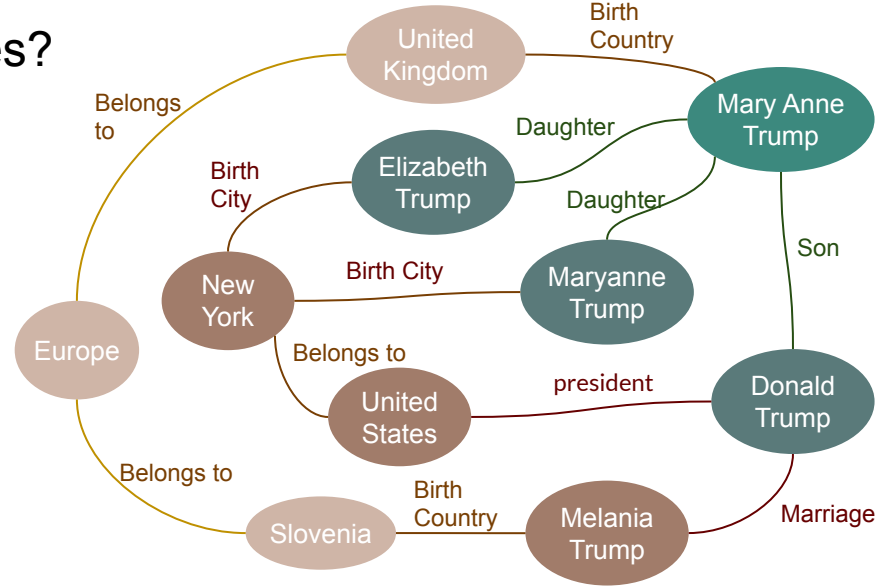
Question:
Who is the president of the United States?

Answer:
Donald Trump

What is Knowledge based Question Answering

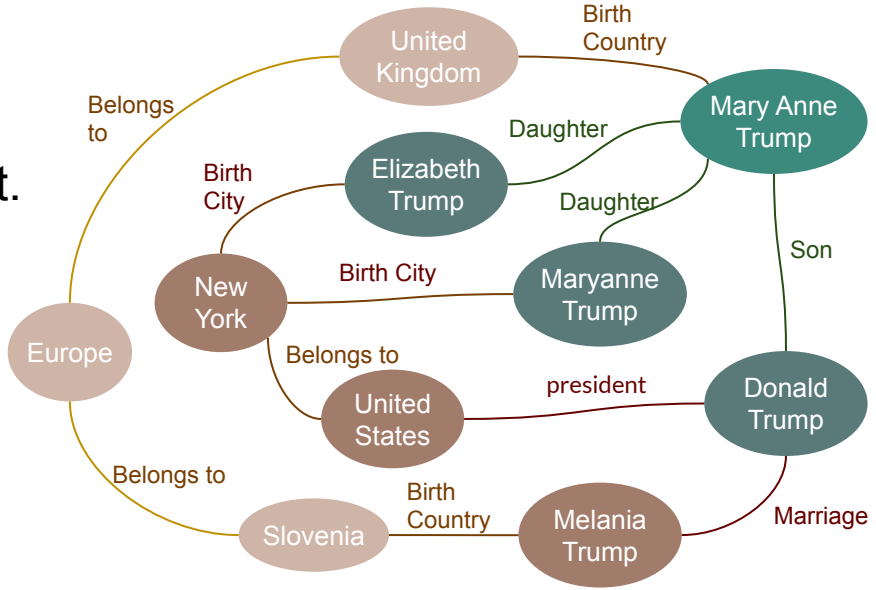
Question:
Who is the president of the United States?

Answer:
Donald Trump



Knowledge Graph

- Each node e is an entity.
- Each edge r represents a relation between two connected entities.
- A triplet $(e_{head}; r; e_{tail})$ is called a fact.

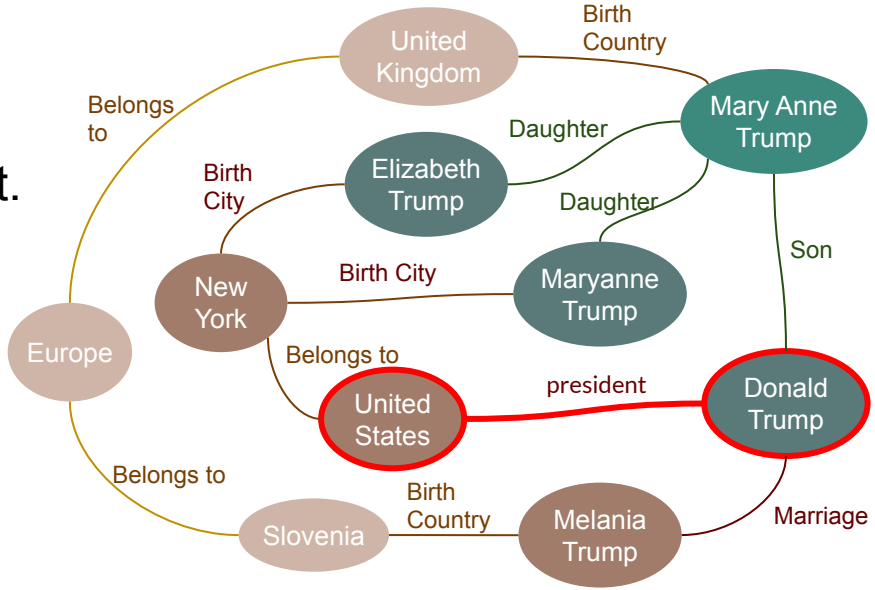


Knowledge Graph

- Each node e is an entity.
- Each edge r represents a relation between two connected entities.
- A triplet $(e_{head}; r; e_{tail})$ is called a fact.

Fact:

(United States, President, Donald Trump)



What is Knowledge based Question Answering

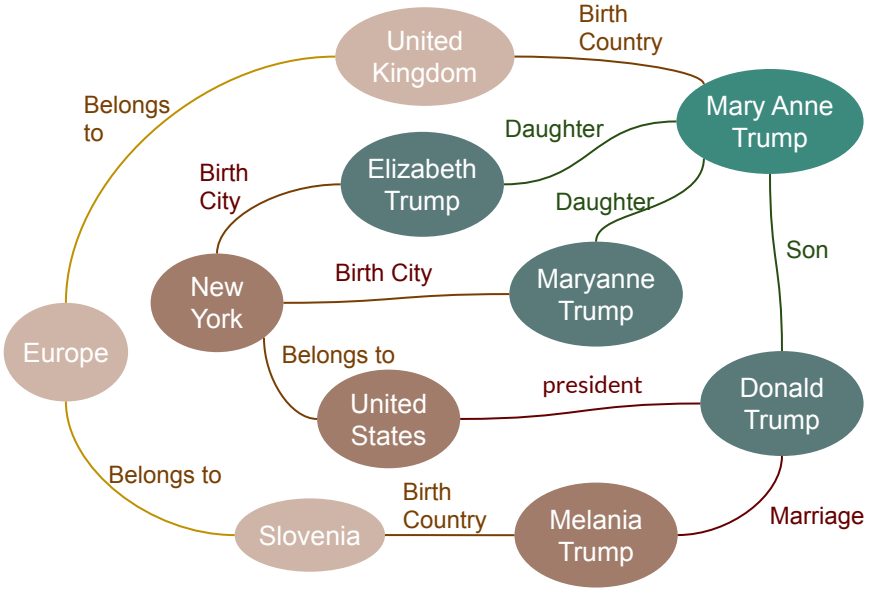
Question:
Who is the sister of the president of the United States?

Who is the president of the United States?

Who is the mother of Donald Trump?

Who are the daughters of Mary MacLeod?

Answer:
Maryanne Trump / Elizabeth Trump



What is Knowledge based Question Answering

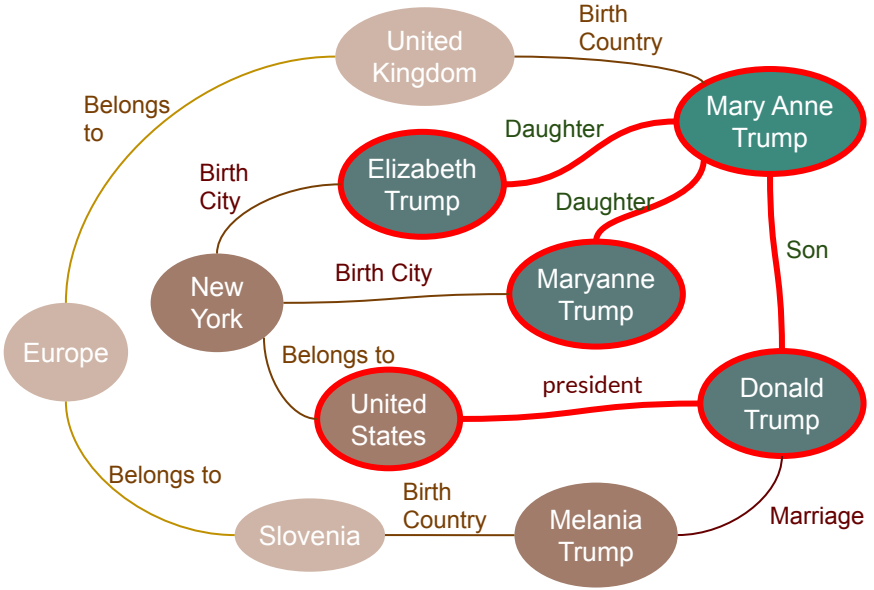
Question:
Who is the sister of the president of the United States?

(United States, President, Donald Trump)

(Donald Trump, Mother, Mary Anne Trump)

(Mary Anne Trump, Daughter, Maryanne/Elizabeth)

Answer:
Maryanne Trump / Elizabeth Trump



What is Knowledge based Question Answering

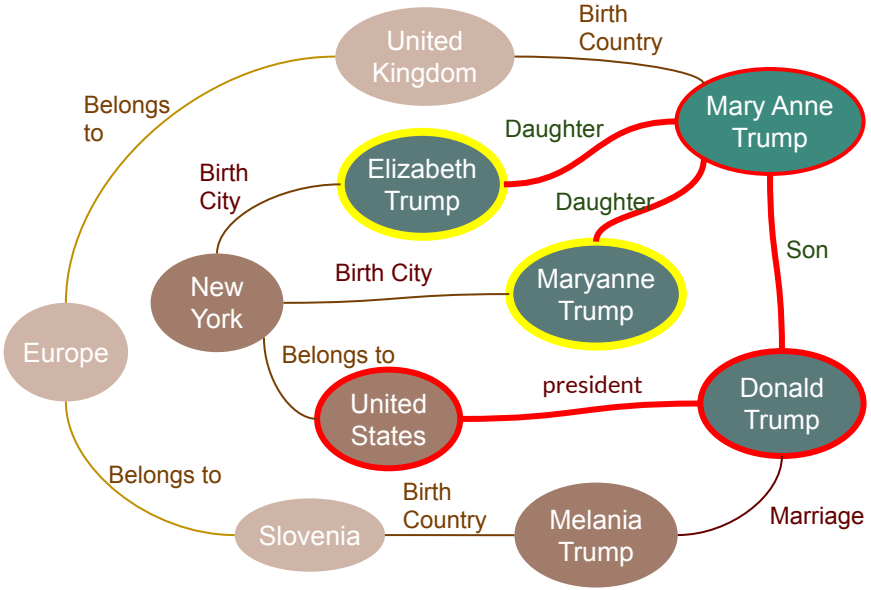
Question:
Who is the sister of the president of the United States?

(United States, President, Donald Trump)

(Donald Trump, Mother, Mary Anne Trump)

(Mary Anne Trump, Daughter, Maryanne/Elizabeth)

Answer:
Maryanne Trump / Elizabeth Trump



What is Knowledge based Question Answering

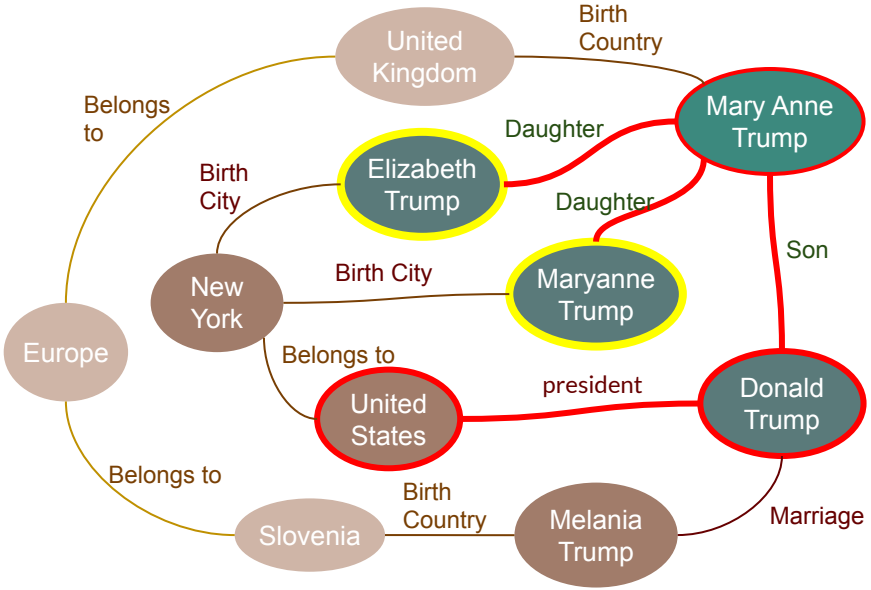
Question:
Who is the sister of the president of the United States?

(United States, President, Donald Trump)

(Donald Trump, Mother, Mary Anne Trump)

(Mary Anne Trump, Daughter, Maryanne/Elizabeth)

Answer:
Maryanne Trump / Elizabeth Trump



Latent info = reasoning path (Highlight in red)

Reasoning Path as Latent Variable

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

x: question

Who is the sister of the president of the United States?

z: reasoning path

United States→President→Donald Trump→Mother→Mary Anne Trump→Daughter→

y: answer

Maryanne Trump / Elizabeth Trump

Notations

For a given question x, a reasoning path z is a sequence in the form:

$$z = e_0 \rightarrow r_1 \rightarrow e_1 \rightarrow ,\ldots, \rightarrow e_{T-1} \rightarrow r_T$$

that points to the answer:

$$z \rightarrow (e_T = y)$$

Notations

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

$p(y|z)=p(e_T|e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T)$

$p(z|x)=p(e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T|x)=p(e_0|x)p(r_1|x,e_0)p(e_1|x,e_0,r_1)\ldots p(r_T|x,e_0,r_1,\ldots,e_{T-1})$

Notations

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

$p(y|z)=p(e_T|e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T)$

$p(z|x)=p(e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T|x)=p(e_0|x)p(r_1|x,e_0)p(e_1|x,e_0,r_1)\ldots p(r_T|x,e_0,r_1,\ldots,e_{T-1})$

Notations

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

$p(y|z)=p(e_T|e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T)=p(e_T|e_{T-1},r_T)$

$p(z|x)=p(e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T|x)=p(e_0|x)p(r_1|x,e_0)p(e_1|x,e_0,r_1)\ldots p(r_T|x,e_0,r_1,\ldots,e_{T-1})$

We just need to model two terms $p(e|*)$ and $p(r|*)$.

Entity Probability *p(e|*)*

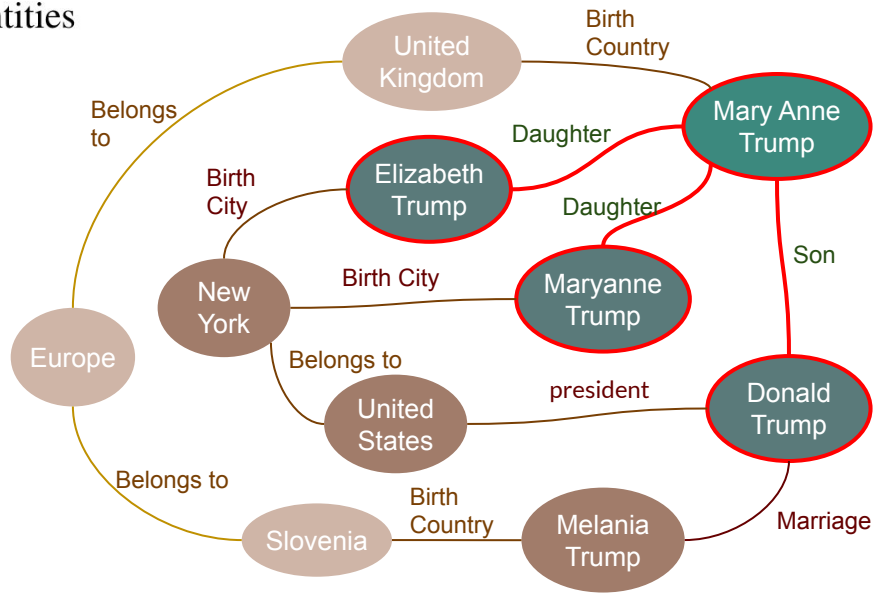
$$p(e_t|e_{t-1}, r_t) = \begin{cases} 1/M & \text{if } e_t \text{ is one of the } M \text{ matched entities} \\ 0 & \text{if } e_t \text{ is not a matched entity} \end{cases}$$

p(Elizabeth_Trump|...,Daughter,Mary Anne)=½

p(Maryanne_Trump|...,Daughter,Mary Anne)=½

p(Donald_Trump|...,Daughter,Mary Anne)=0

p(Donald_Trump|...,Son,Mary Anne)=1



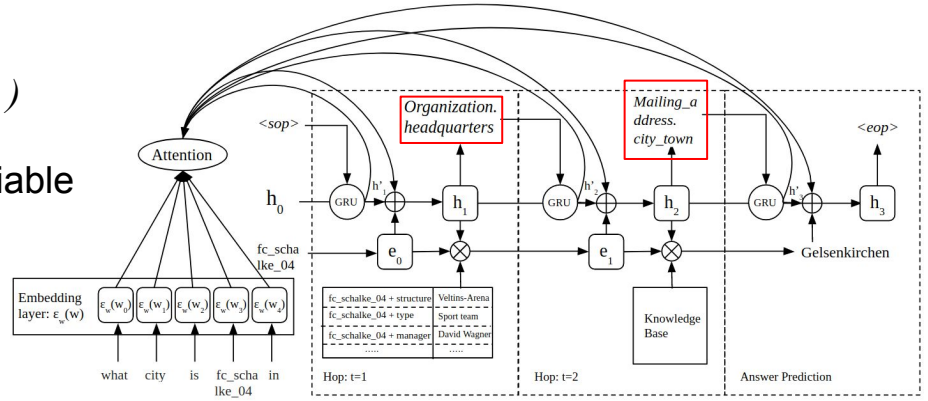
Relation Probability $p(r|*)$

At each timestep t, given r_{t-1} and e_{t-1}, we estimate $p(r_t|...)$ using a recurrent structure:

$$p(r_t|e_0,r_1,...,e_{t-1})=softmax([f(e_0,...,e_{t-1}); f(r_1,...,r_{t-1}); f(x)])$$

Where $f(*)$ is a mapping function from random variable to its vector representation.

Therefore $f(e_{t-1})$, $f(r_{t-1})$, and $f(x)$ are vector representations of the previous entity, previous relation, and the input query.

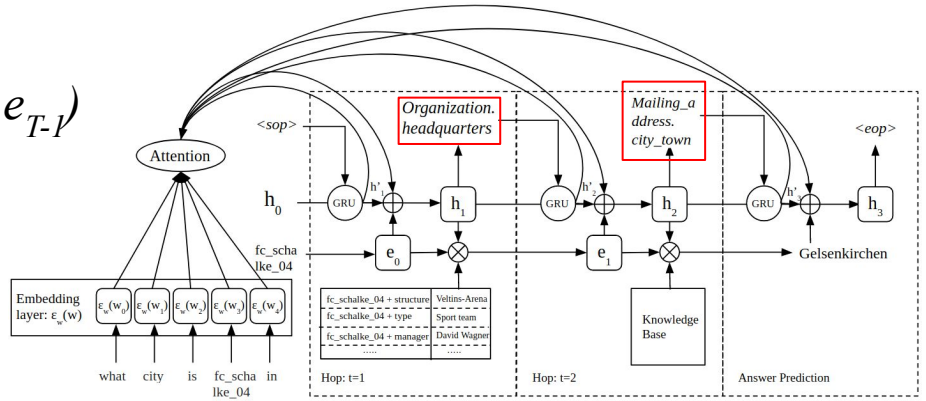


Latent Reasoning Path Prediction $p(z|x)$

$$p(z|x)$$
$$=p(e_0,r_1,e_1,r_2,\ldots,e_{T-1},r_T|x)$$
$$=p(e_0)p(r_1|e_0)p(e_1|e_0,r_1)\ldots p(r_T|e_0,r_1,e_1,r_2,\ldots,e_{T-1})$$

1. e_0 is identified by entity linking tool.

2. At each timestep t, we estimate $p(r_t|*)$ and $p(e_t|*)$ as discussed.



Estimate Values of z in Preprocessing

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

To train the model without using labeled z, we use graph algorithm to select reasoning paths from the graph.

Preliminary Experimental Results

Properties:

- **Model multiple reasoning paths:** consider multiple reasoning paths for each question answer pair make the model more stable than using a single path in most existing work.
- **Reasoning path as latent variable:** our model can be trained without using labeled reasoning paths.
- **Easy to implement:** fit with any base models (we use RNN structure).

	Extra Supervision	Model $p(e)$	Different Setup	WQSP	CWQ
STAGG_SP	Y		Semantic Parsing	**71.7**	-
HR-BiLSTM	Y			62.3	31.2
KBQA-GST	Y	Y		67.9	36.5
NSM	Y		Neural Program Generation	**69.0**	-
KV-MemNN				38.6	-
STAGG_Answer			Semantic Parsing	66.8	-
GRAFT-Net		Y		62.8	26.0
Our Method		Y		67.9	**41.9**

Proposed Work: Advanced Path Selection

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

The summation makes training process intractable.

We need to consider all valid paths between e_0 and e_{answer}.

A real-world knowledge graph contains **billions** of entity-relation facts. Between two nodes, there are a very large number of valid paths!

More importantly, not all the valid paths are good enough to serve as a reasoning path.

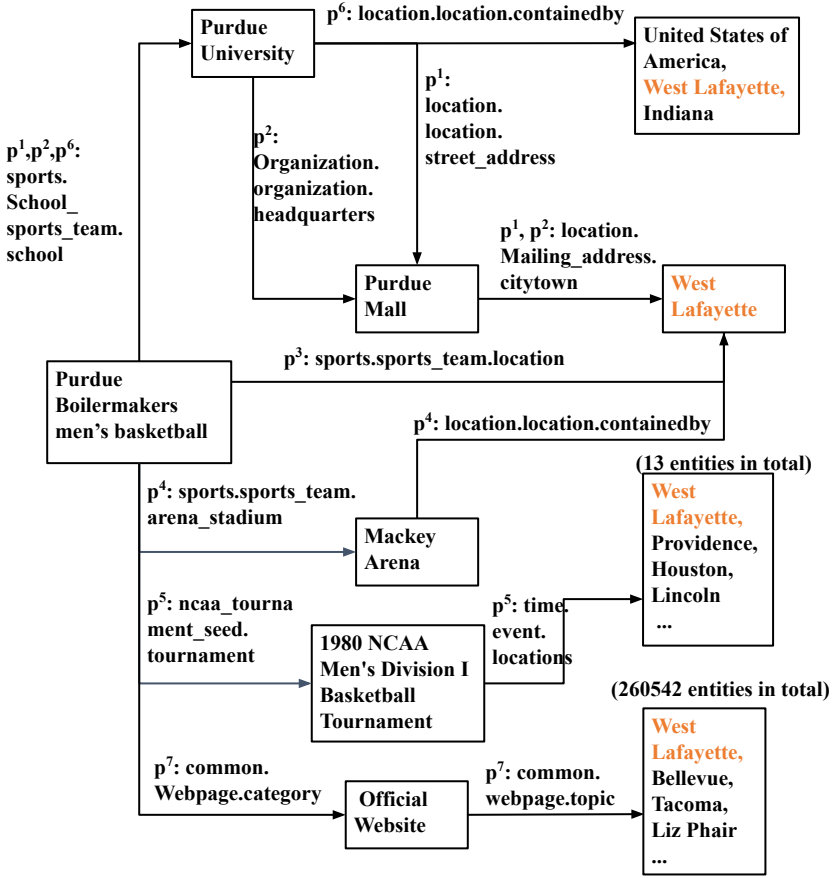
Path Selection - Rule #1

Question:

What city is home to the University that is known for Purdue Boilermakers men's basketball?

Answer:

West Lafayette



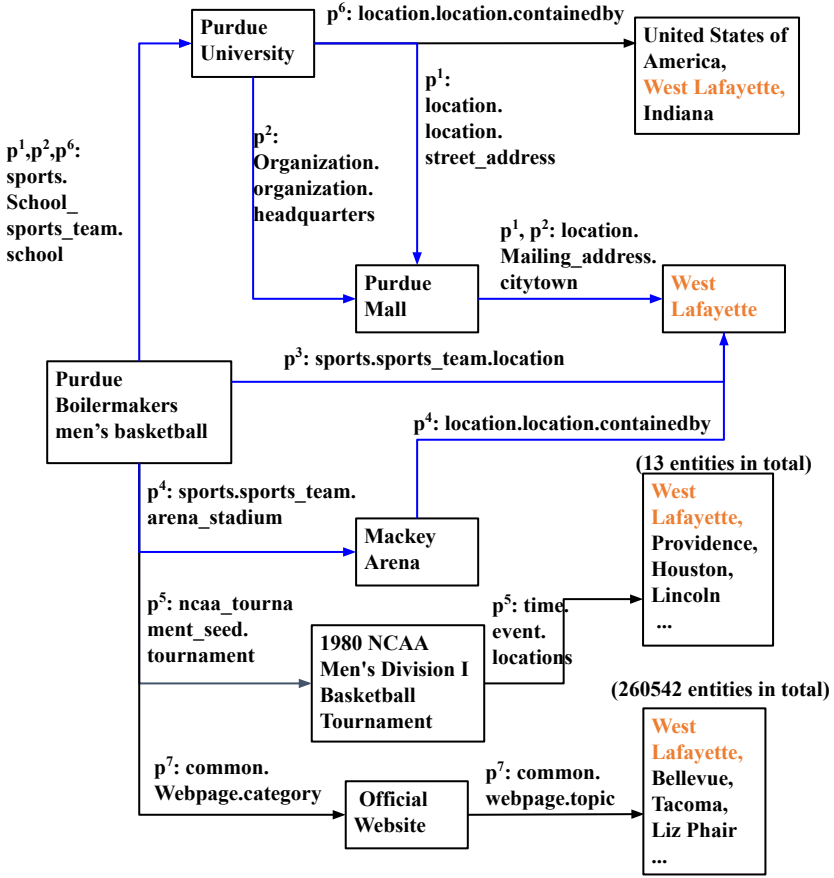
Path Selection - Rule #1

Question:

What city is home to the University that is known for Purdue Boilermakers men's basketball?

Answer:

West Lafayette



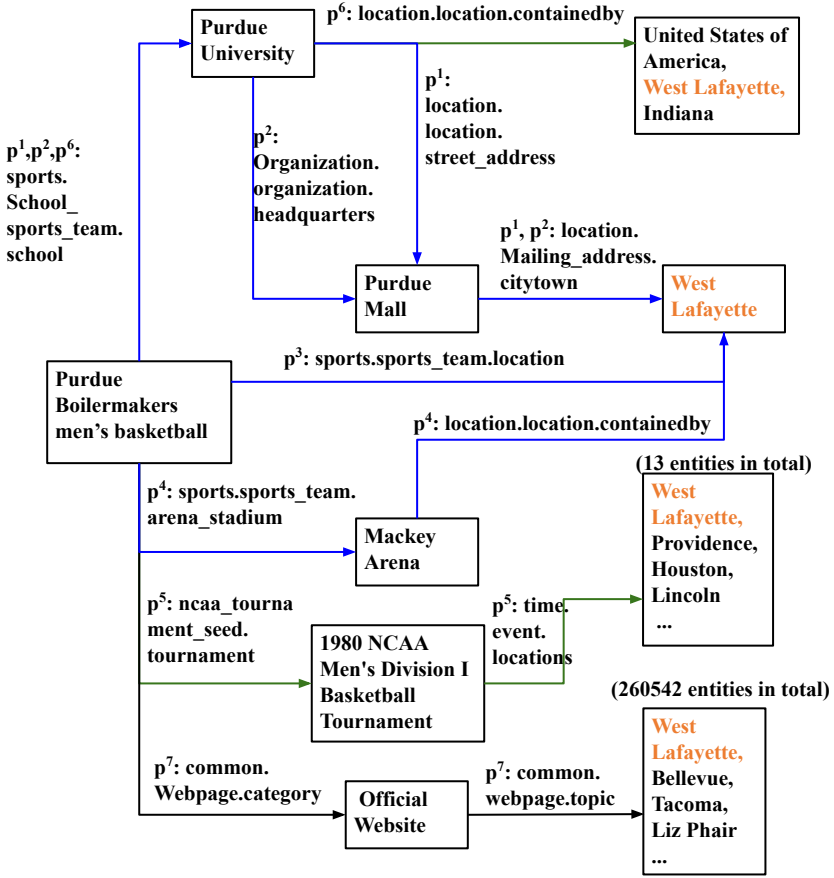
Path Selection - Rule #1

Question:

What city is home to the University that is known for Purdue Boilermakers men's basketball?

Answer:

West Lafayette



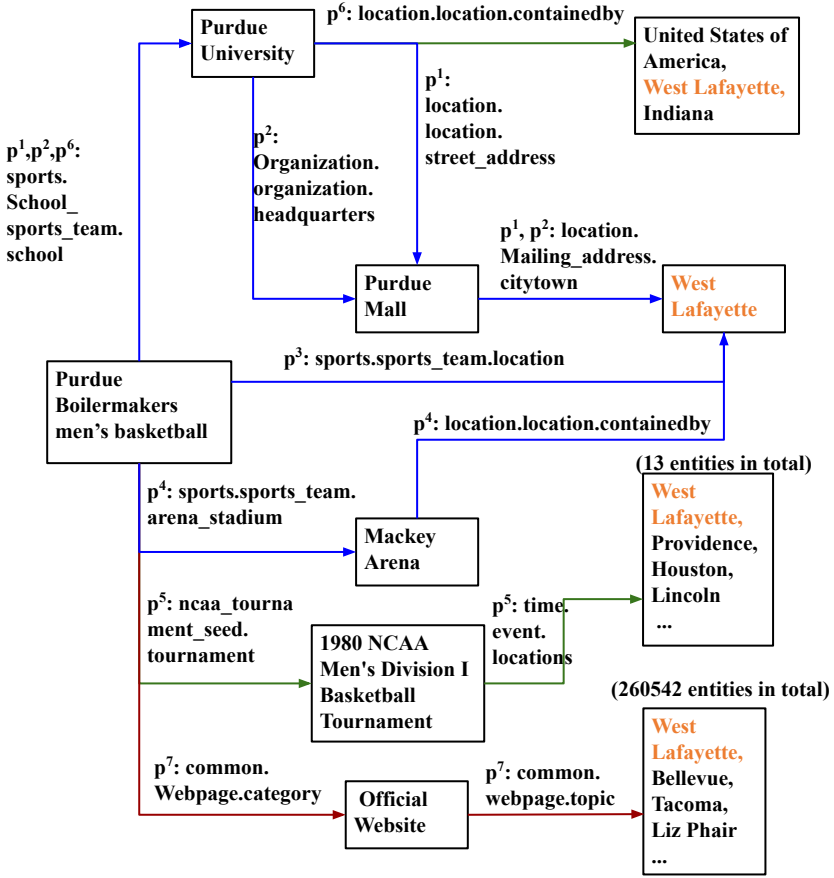
Path Selection - Rule #1

Question:

What city is home to the University that is known for Purdue Boilermakers men's basketball?

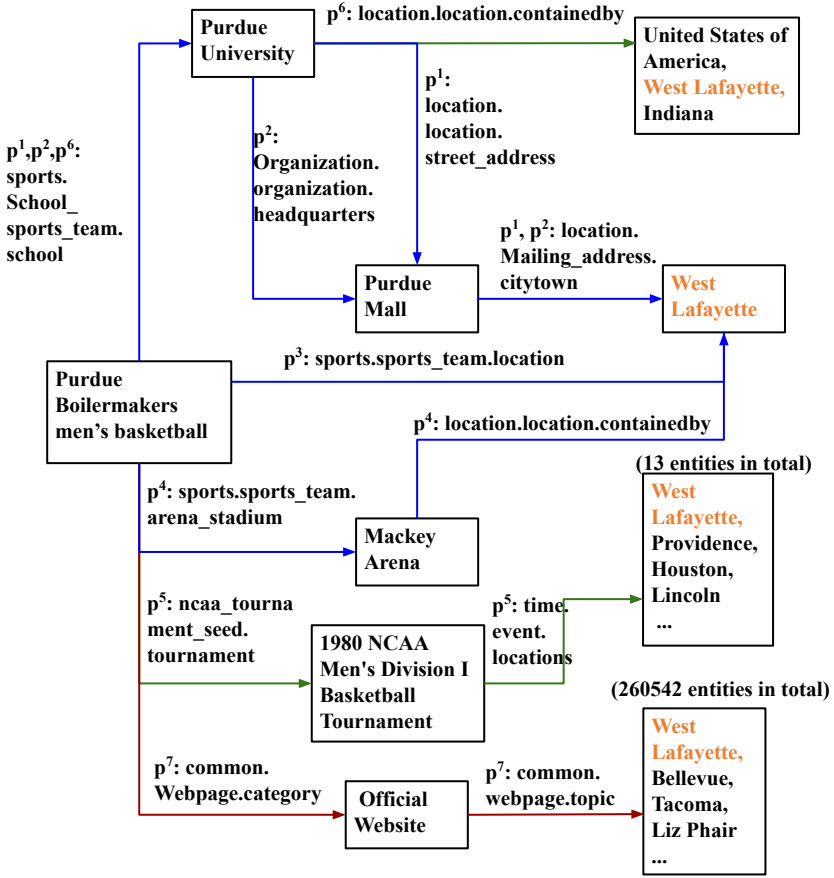
Answer:

West Lafayette



Path Selection - Rule #1

Rule 1: We want to filter out paths pointing to too many entities.



Path Selection - Rule #2

Question: Who was the owner of kfc?

Answer: Colonel Sanders

Path 1: kfc→organization.organization.founders→Colonel Sanders

Path 2: kfc→advertising_characters.product.advertising_characters→Colonel Sanders

Path Selection - Rule #2

Question: Who was the owner of kfc?

Answer: Colonel Sanders

Path 1: kfc→organization.organization.founders→Colonel Sanders

~~Path 2: kfc →advertising_characters.product.advertising_characters →Colonel Sanders~~

Rule 2: We want to filter out paths that are not relevant to the question.

Reasoning Path as Latent Variable

Step 1: Use graph algorithm to collect all valid paths between topic entity e_0 and answer e_{answer}.

Step 2: Select paths based on rule #1 and rule #2.

Step 3: Update model parameters by maximizing likelihood $p(y|x)$ based on selected paths.

Repeat step 2 and step 3 until the model converges.

Timeline

Timeline	Task
by June 2020	Designing evaluation experiments for QA task - Human identification - Major claim extraction - Discourse relation classification
by Winter 2020	Improving path selection - Use current trained model to select good paths - Use advanced bootstrapping methods to select good paths - Explore other directions to solve the problem - Evaluate performance of the proposed method
by Winter 2020	Refining model architecture - Neural Transformer - Memory Network - Propose novel model structures - Evaluate performance of the proposed model
by Summer 2021	Handling noisy tags in multi-label classification - Propose novel ideas to handle noisy tags - Propose novel model structures - Evaluate performance of the proposed model
by Fall 2021	Thesis writing and defense.

Thank you!

Questions?

Other Work

Use latent topic to predict a winner in a debate:

Winning on the Merits: The Joint Effects of Content and Style on Debate Outcomes (TACL), 2017.

Use latent conversation structure information to generate meeting minutes:

Joint Modeling of Content and Discourse Relations in Dialogues (ACL), 2017.

Capture label dependencies in multi-label prediction task:

Learning to Calibrate and Rerank Multi-label Predictions (ECML PKDD), 2019.

Ranking-Based AutoEncoder for Extreme Multi-label Classification (NAACL-HLT), 2019.

Supervised Learning

#Car=1

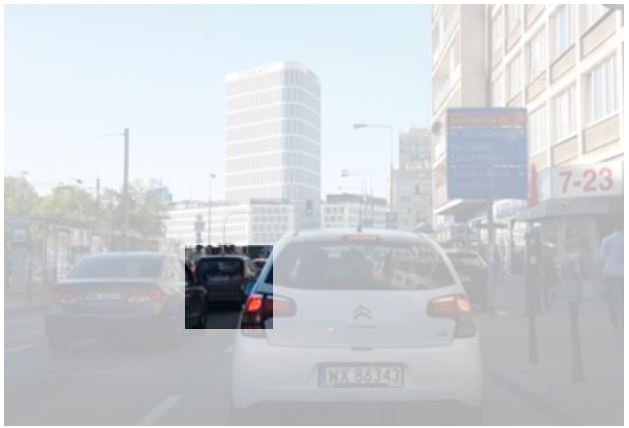
Input x *y=f(x)* *Output y*

Supervised Learning

#Car=3

Input x *y=f(x)* *Output y*

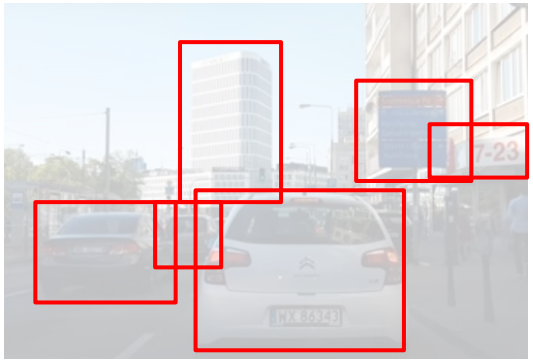
Supervised Learning



#Car=3

Input x *y=f(x)* *Output y*

Supervised Learning with Latent Information



#Car=3

Input x

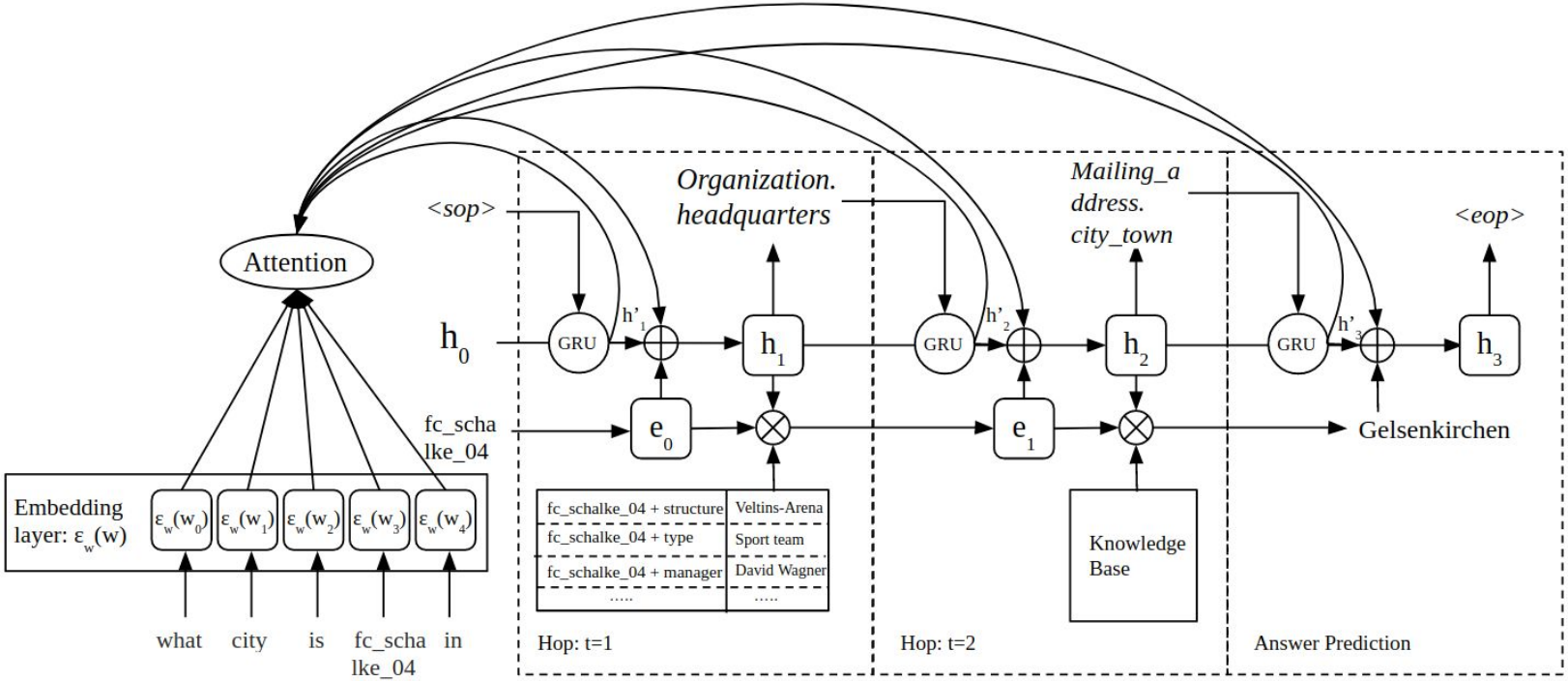
z=possible locations

Output y

$z=f(x)$

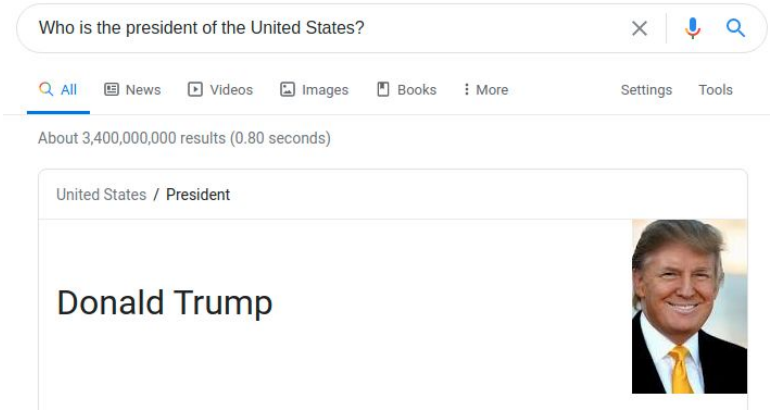
$y=f(z)$

Model Structure



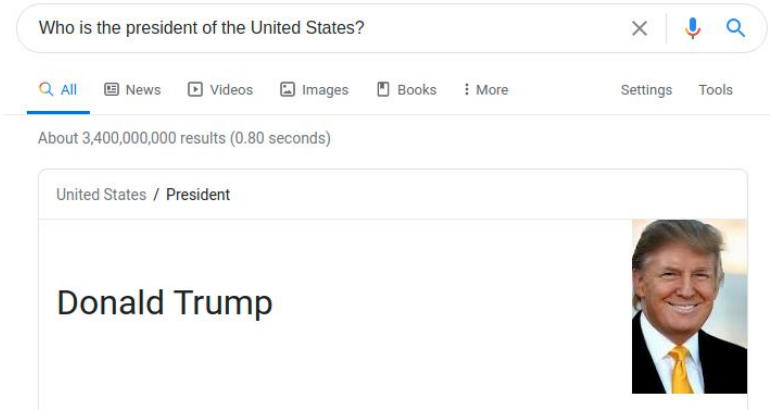
Supervised Learning

Who is the president of the United States?



Supervised Learning

Who is the president of the United States? ⇨ Feature x

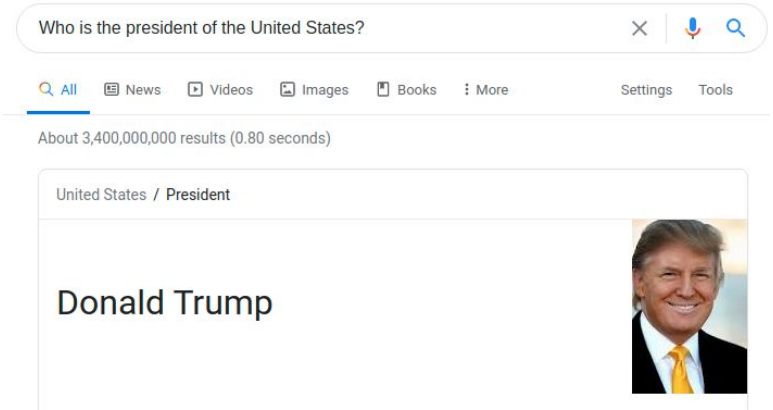


⇨ Target y

Supervised Learning

Who is the president of the United States? ⇨ Feature x

$y=f(x)$

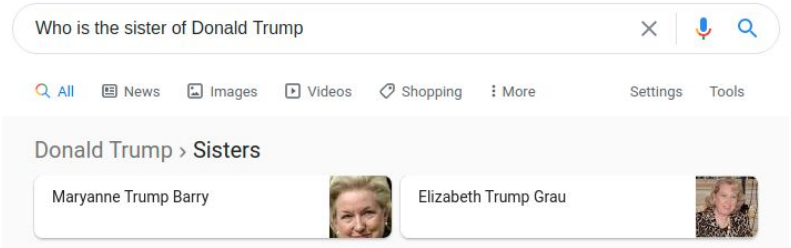


⇨ Target y

Supervised Learning

Who is the sister of Donald Trump?

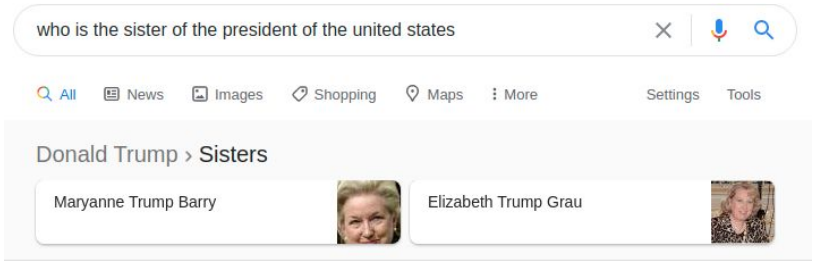
$y=f(x)$



Supervised Learning

Who is the sister of the president of the United States?

$y=f(x)$



Latent Information

Who is the sister of the president of the United States? x

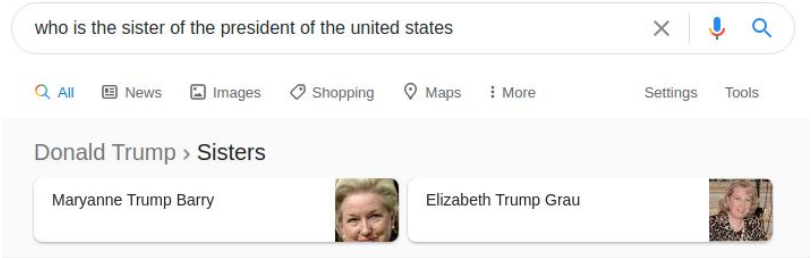
$z=f(x)$

Who is the president of the United States? → Donald Trump

Who is the sister of Donald Trump? → Latent information z

$y=f(z)$

y



Latent Information

Who is the sister of the president of the United States? x

Who is the president of the United States?→ Donald Trump

Who are the parents of Donald Trump?→ XXX

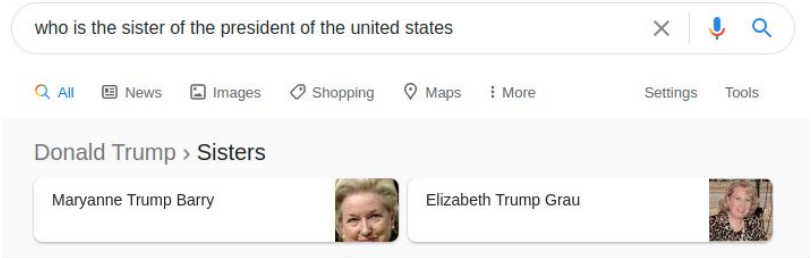
Who are the daughters of XXX?

$z=f(x)$

Latent information z

$y=f(z)$

y

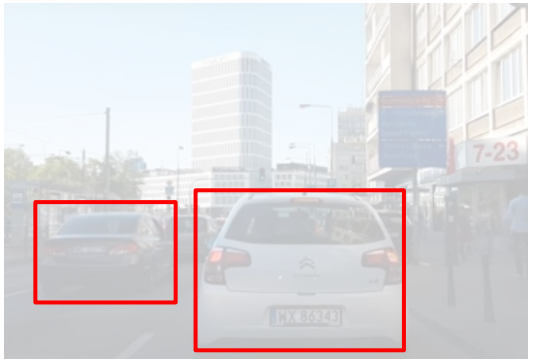


Supervised Learning

Car

Feature x $y=f(x)$ Target y

Supervised Learning with Latent Information



Car

Feature x

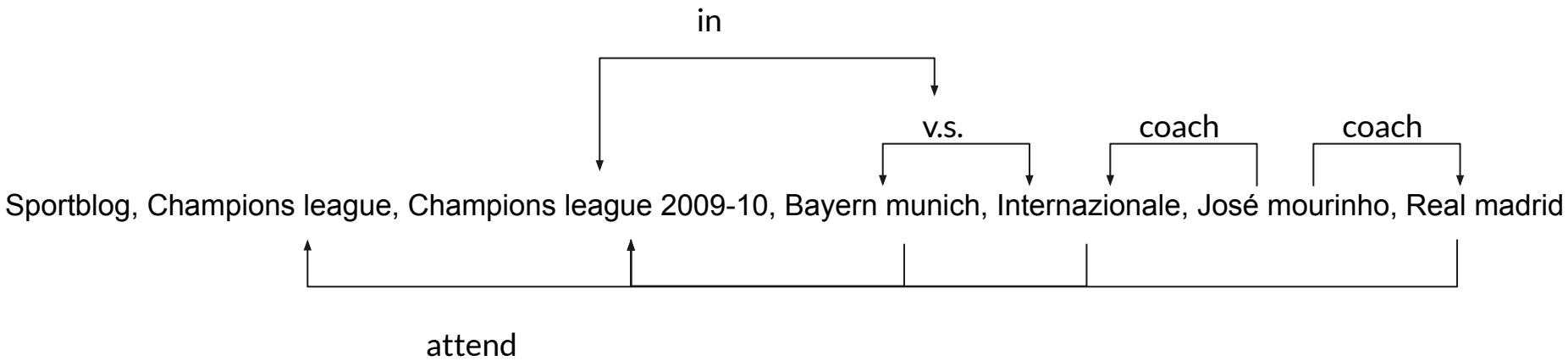
Latent information z

Target y

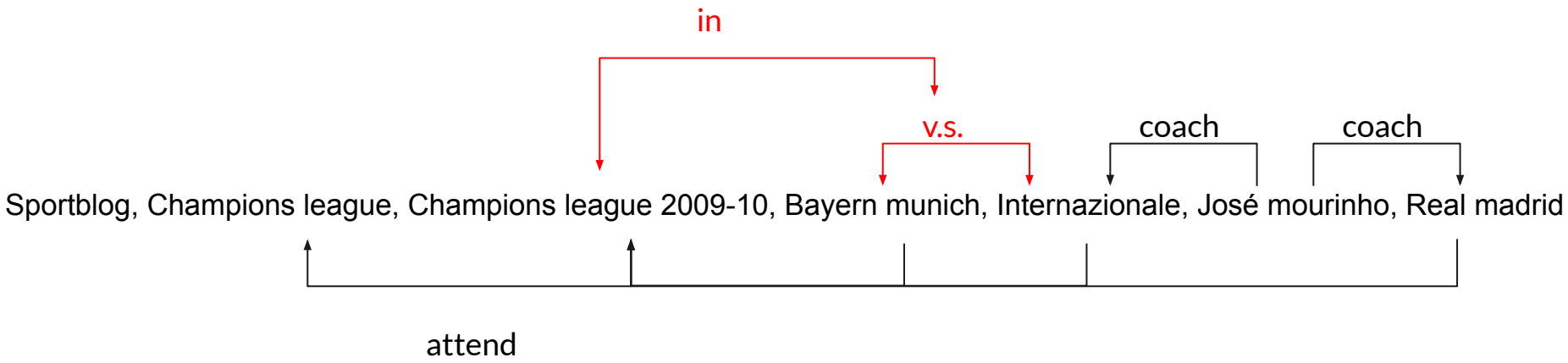
$z=f(x)$

$y=f(z)$

Label Dependencies

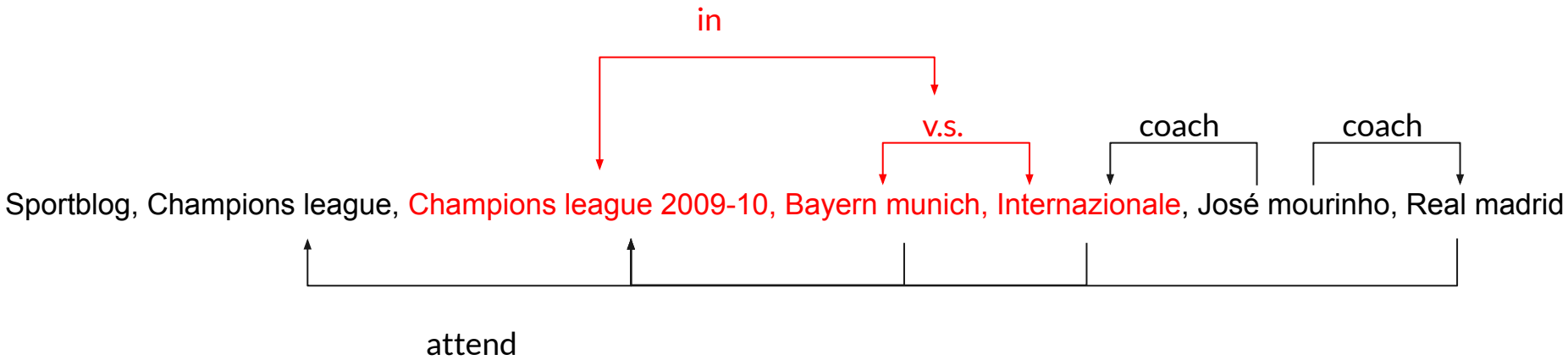


Label Dependencies



Label Dependencies

~~Champions league 2010-11~~
~~World Cup 2010~~



Different Ways to Sort Labels (classifiers)

Frequency:
Sportblog→Champions league→Real_madrid→José mourinho→Internazionale→Champions league 2009-10→Bayern munich

Hierarchy:
Sportblog→Champions league→Champions league 2009-10→Bayern munich→Internazionale→Real_madrid→José mourinho

Alphabeta:
Bayern munich→Champions league→Champions league 2009-10→Internazionale→José mourinho→Real_madrid→Sportblog

What is latent information?

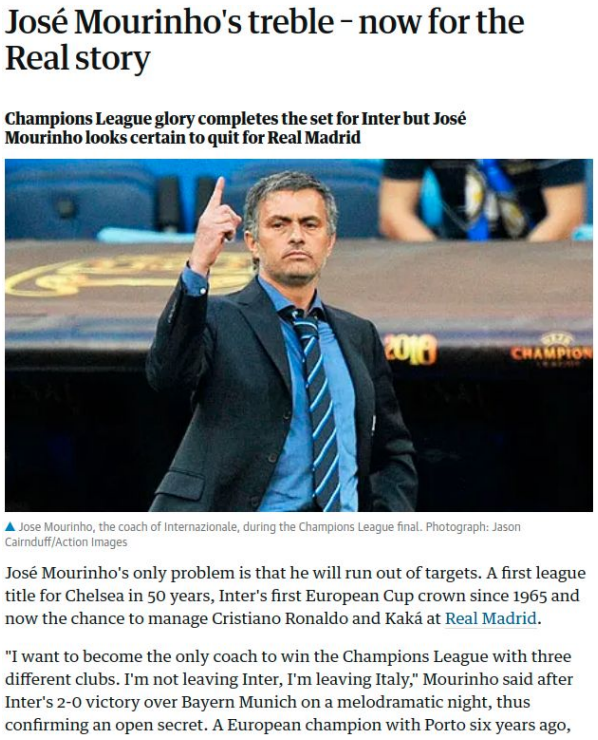
Answer Prediction $p(y|z)$

$(e_0, r_1, e_1, r_2, \ldots, e_{T-1}, r_T) \rightarrow e_{T-1} = y$, Our final goal is to estimate answer y.

$$p(e_t|e_{t-1}, r_t) = \begin{cases} 1/M & \text{if } e_t \text{ is one of the } M \text{ matched entities} \\ 0 & \text{if } e_t \text{ is not a matched entity} \end{cases}$$

Probabilistic Classifier Chain (PCC)

x:



José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

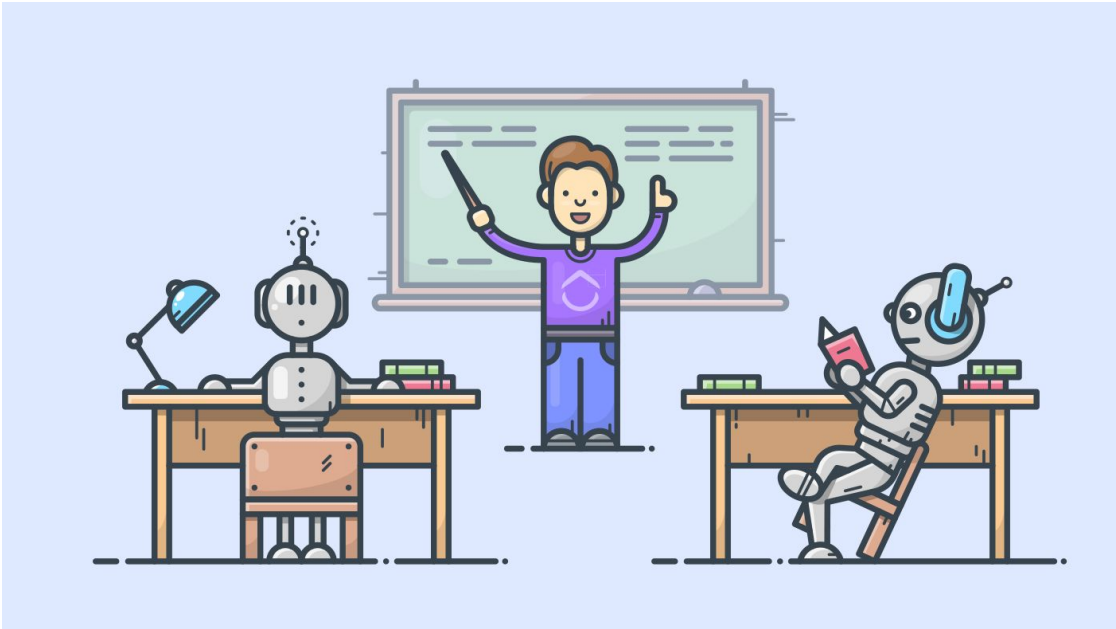
"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter, I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$$b_1(y_1|x)$$
$$b_2(y_2|x,y_1)$$
$$b_3(y_3|x,y_1,y_2)$$
$$\ldots$$
$$b_n(y_n|x,y_1,\ldots,y_{n-1})$$

y:

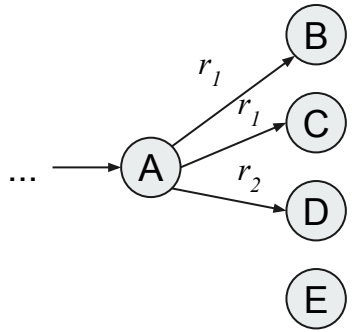
Champions league→Sportblog→
José mourinho→Internazionale→
Real_madrid→Bayern munich→
Champions league 2009-10

Teach Machines to Think like Humans



Entity Probability *p(e|*)*

$$p(e_t|e_{t-1}, r_t) = \begin{cases} 1/M & \text{if } e_t \text{ is one of the } M \text{ matched entities} \\ 0 & \text{if } e_t \text{ is not a matched entity} \end{cases}$$



$p(B|...,A,r_1)=1/2$
$p(C|...,A,r_1)=1/2$
$p(B|...,A,r_2)=0$
$p(D|...,A,r_2)=1$
$p(E|...,A,r_*)=0$

Different Ways to Sort Labels (classifiers)

Alphabeta:
Bayern munich→Champions league→Champions league 2009-10→Internazionale→José mourinho→Real_madrid→Sportblog

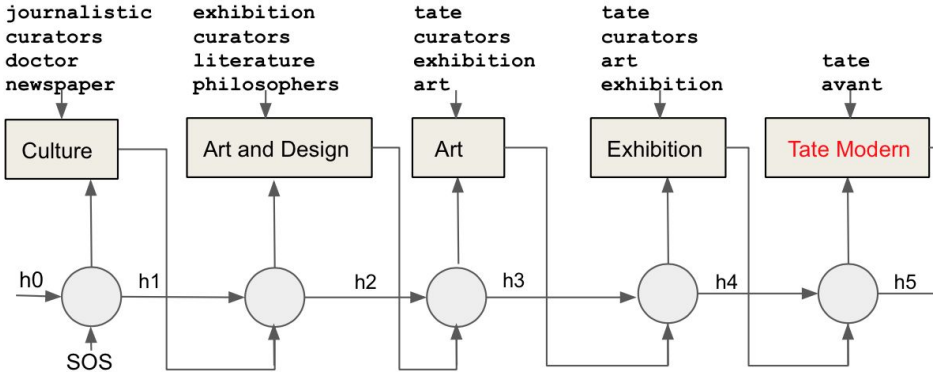
Frequency:
Sportblog→Champions league→Real_madrid→José mourinho→Internazionale→Champions league 2009-10→Bayern munich

Hierarchy:
Sportblog→Champions league→Champions league 2009-10→Bayern munich→Internazionale→Real_madrid→José mourinho

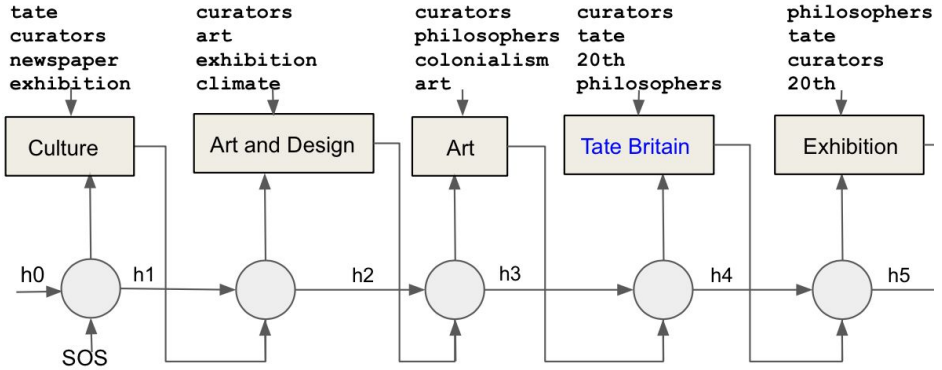
Manually:
Sportblog→Champions league→Champions league 2009-10→Bayern munich→Internazionale→José mourinho→Real_madrid

Case Study

RNN trained with fixed label order:



RNN trained with latent label order:



More examples of Latent Variable Models

- Gaussian Mixture Models (GMMs)
- Latent Dirichlet Allocation (LDA)
- Probabilistic Latent Semantic Analysis (pLSA)
- Hidden Markov Models (HMMs)
- Principal Component Analysis (PCA)
- ...

Problem of Using a Predefined Label Order

x:

José Mourinho's treble – now for the Real story

Champions League glory completes the set for Inter but José Mourinho looks certain to quit for Real Madrid

▲ Jose Mourinho, the coach of Internazionale, during the Champions League final. Photograph: Jason Cairnduff/Action Images

José Mourinho's only problem is that he will run out of targets. A first league title for Chelsea in 50 years, Inter's first European Cup crown since 1965 and now the chance to manage Cristiano Ronaldo and Kaká at Real Madrid.

"I want to become the only coach to win the Champions League with three different clubs. I'm not leaving Inter. I'm leaving Italy," Mourinho said after Inter's 2-0 victory over Bayern Munich on a melodramatic night, thus confirming an open secret. A European champion with Porto six years ago,

$y_1,\dots,y_t$:

Frequency:

Sportblog→Champions league→Real_madrid→ José mourinho→y_t=Cristiano Ronaldo

Hierarchy:

Sportblog→Champions league→ Champions league 2009-10→Bayern munich→ y_t=Internazionale (→Real_madrid→José mourinho)

ORDER MATTERS!

That is our latent information!

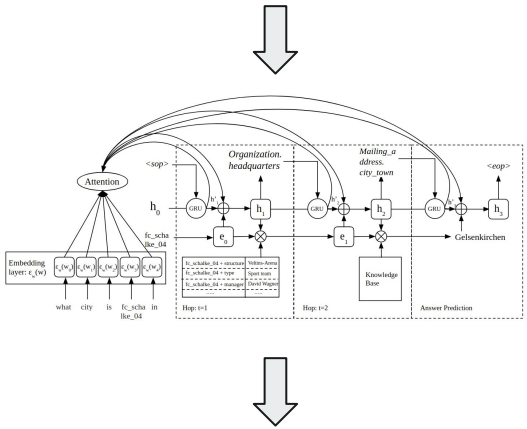
Rule 1: filter out paths leading to too many entities

$$p(y|x) = \sum_{z} p(y|z)p(z|x)$$

$$p(e_t|e_{t-1}, r_t) = \begin{cases} 1/M & \text{if } e_t \text{ is one of the } M \text{ matched entities} \\ 0 & \text{if } e_t \text{ is not a matched entity} \end{cases}$$

Rule 2: filter out irrelevant paths

Question: Who was the owner of kfc?

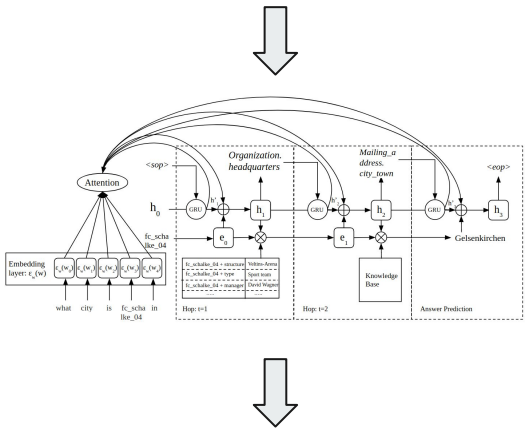


p(kfc→organization.organization.founders→Colonel Sanders|x) = 0.8

p(kfc→advertising_characters.product.advertising_characters→Colonel Sanders) = 0.2

Rule 2: filter out irrelevant paths

Question: Who was the owner of kfc?



p(kfc→organization.organization.founders→Colonel Sanders|x) = 0.8

~~*p(kfc→advertising_characters.product.advertising_characters→Colonel Sanders) = 0.2*~~