Context Dependent Fine Grained Entity Typing

Rajarshi Das

Reading Group Presentation

June 29, 2016

Entity Type Classification

Entity Type Classification

- can be defined as the task of assigning semantic labels to mentions of entities in documents (such as '*person*', '*location*', '*organization*', '*misc.*')

Entity Type Classification

- can be defined as the task of assigning semantic labels to mentions of entities in documents (such as '*person*', '*location*', '*organization*', '*misc.*')

- Entity types are useful for a variety of related natural language tasks such as coreference resolution (Recasens et al., 2013), relation extraction (Yao et al. 2010; Ling and Weld 2012)

Entity Type Classification

- can be defined as the task of assigning semantic labels to mentions of entities in documents (such as ‘*person*’, ‘*location*’, ‘*organization*’, ‘*misc.*’)

- Entity types are useful for a variety of related natural language tasks such as coreference resolution (Recasens et al., 2013), relation extraction (Yao et al. 2010; Ling and Weld 2012)

- They have also increased the performance of several downstream applications such as Question Answering (Lin et al., 2012) and Knowledge Base Completion (Carlson et al., 2010; Das et al., 2016)

Fine-grained Entity Types

Fine-grained Entity Types

- Standard type classification task just use few coarse labels (such as person, location, organization, etc)

Fine-grained Entity Types

- Standard type classification task just use few coarse labels (such as person, location, organization, etc)
- Recent work has focused on much larger set of fine grained labels (Ling and Weld, 2012; Yosef et al., 2012; Del Corro et al, 2015; *inter-alia*)

Fine-grained Entity Types

- Standard type classification task just use few coarse labels (such as person, location, organization, etc)
- Recent work has focused on much larger set of fine grained labels (Ling and Weld, 2012; Yosef et al., 2012; Del Corro et al, 2015; *inter-alia*)
- Today I will discuss few recent paper about context dependent fine grained entity types…

Context-Dependent Fine-Grained Entity Type Tagging

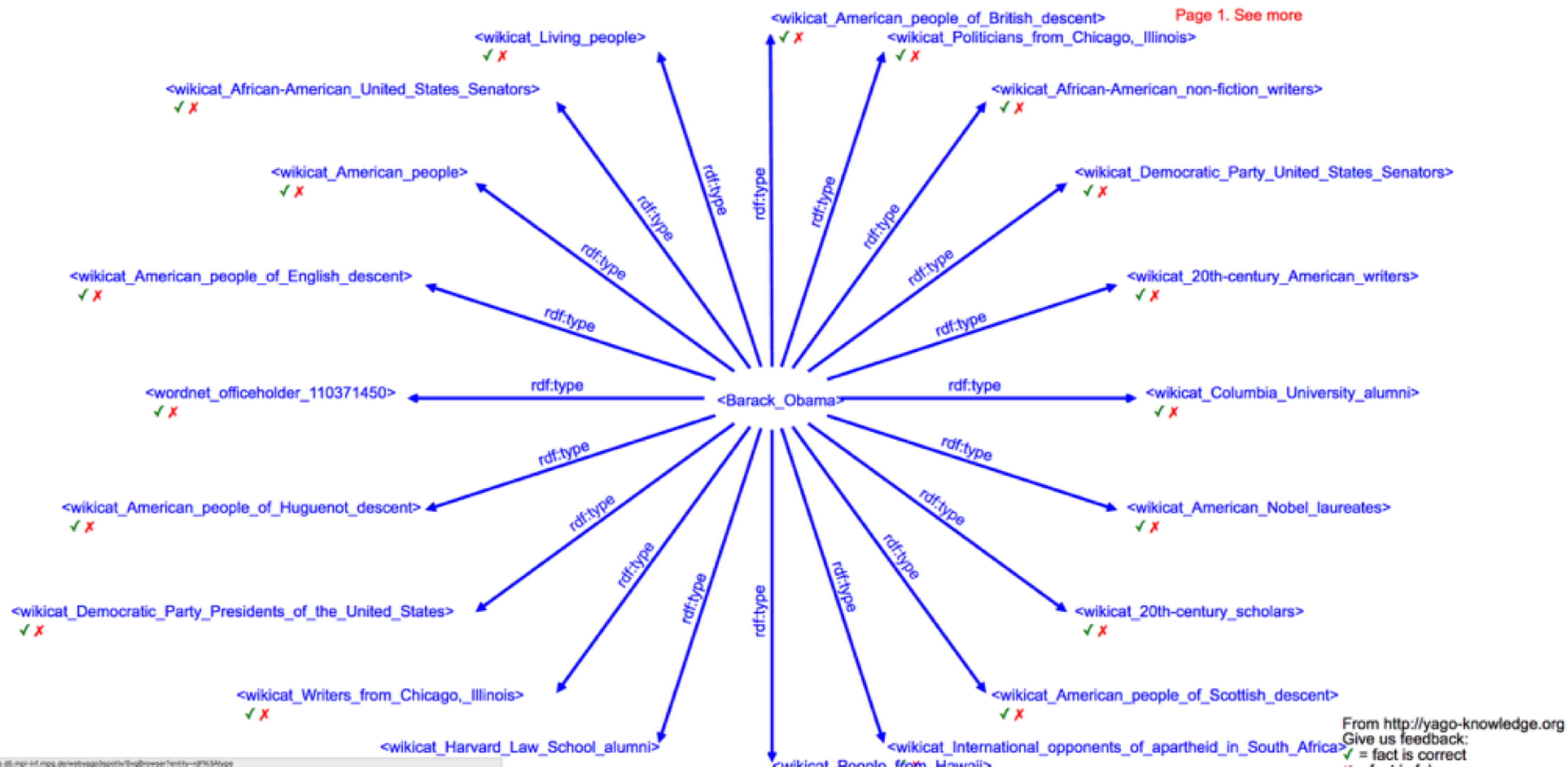
Dan Gillick, Nevena Lazic, Kuzman Ganchev, Jesse Kirchner, David Huynh
Google

Why context dependent types?

Why context dependent types?

Why context dependent types?

Types in YAGO



Why context dependent types?

Why context dependent types?

Why context dependent types?

Types in YAGO

Why context dependent types?

Why context dependent types?

Are all of these types always relevant?

Why context dependent types?

Are all of these types always relevant?

In April 1986, he won election as mayor (a nonpartisan position) of his adopted hometown, Carmel-by-the-Sea, California

Why context dependent types?

Are all of these types always relevant?

In April 1986, he won election as mayor (a nonpartisan position) of his adopted hometown, Carmel-by-the-Sea, California

Relevant Types:
mayor, politician, legislator

Why context dependent types?

Are all of these types always relevant?

In April 1986, he won election as mayor (a nonpartisan position) of his adopted hometown, Carmel-by-the-Sea, California

Relevant Types:
mayor, politician, legislator

Eastwood has been recognized with multiple awards and nominations for his work in film, television, and music.

Why context dependent types?

Are all of these types always relevant?

In April 1986, he won election as mayor (a nonpartisan position) of his adopted hometown, Carmel-by-the-Sea, California

Relevant Types:
mayor, politician, legislator

Eastwood has been recognized with multiple awards and nominations for his work in film, television, and music.

Relevant Types:
film director, producer, entertainer

Context Dependent Types

Context Dependent Types

- Context Dependent Fine Type tagging: set of acceptable types for an entity mention are only those that are relevant to the local context

Context Dependent Types

- Context Dependent Fine Type tagging: set of acceptable types for an entity mention are only those that are relevant to the local context

> If a hostile predator emerges for Saatchi & Saatchi Co., cofounders
>
> Charles and Maurice Saatchi will lead…..

Context Dependent Types

- Context Dependent Fine Type tagging: set of acceptable types for an entity mention are only those that are relevant to the local context

If a hostile predator emerges for Saatchi & Saatchi Co., cofounders

Charles and Maurice Saatchi will lead…..

Entity Mentions: predators, Saatchi & Saatchi, Charles & Mau…

Context Dependent Types

- Context Dependent Fine Type tagging: set of acceptable types for an entity mention are only those that are relevant to the local context

If a hostile predator emerges for Saatchi & Saatchi Co., cofounders Charles and Maurice Saatchi will lead…..

Entity Mentions: predators, Saatchi & Saatchi, Charles & Mau…

Fine Types:
predator, Saatchi & Saatchi Co. : ***organization/company***
Charles and Maurice Saatchi : ***person/business***

Context Dependent Types

- Context Dependent Fine Type tagging: set of acceptable types for an entity mention are only those that are relevant to the local context

If a hostile predator emerges for Saatchi & Saatchi Co., cofounders Charles and Maurice Saatchi will lead…..

Entity Ment tchi, Charles & Mau…

Fine Types

predator, Saatchi & Saatchi Co. : *organization/company*

Charles and Maurice Saatchi : ***person/business***

Note: Without the context "predator" would not be identified as an "organization"

Type Labels and Manual Annotation

Type Labels and Manual Annotation

- The type labels are derived from Freebase
- They are also organized into a hierarchy. For examples an "*athlete*" type is "*person/athlete*"

Type Labels and Manual Annotation

- The type labels are derived from Freebase

- They are also organized into a hierarchy. For examples an "*athlete*" type is "*person/athlete*"

PERSON

- **artist**
  - actor
  - author
  - director
  - music
- **education**
  - student
  - teacher
- **athlete**
- **business**
- **coach**
- **doctor**
- **legal**
- **military**
- **political figure**
- **religious leader**
- **title**

LOCATION

- **structure**
  - airport
  - government
  - hospital
  - hotel
  - restaurant
  - sports facility
  - theatre
- **geography**
  - body of water
  - island
  - mountain
- **transit**
  - bridge
  - railway
  - road
- **celestial**
- **city**
- **country**
- **park**

ORGANIZATION

- **company**
  - broadcast
  - news
- **education**
- **government**
- **military**
- **music**
- **political party**
- **sports league**
- **sports team**
- **stock exchange**
- **transit**

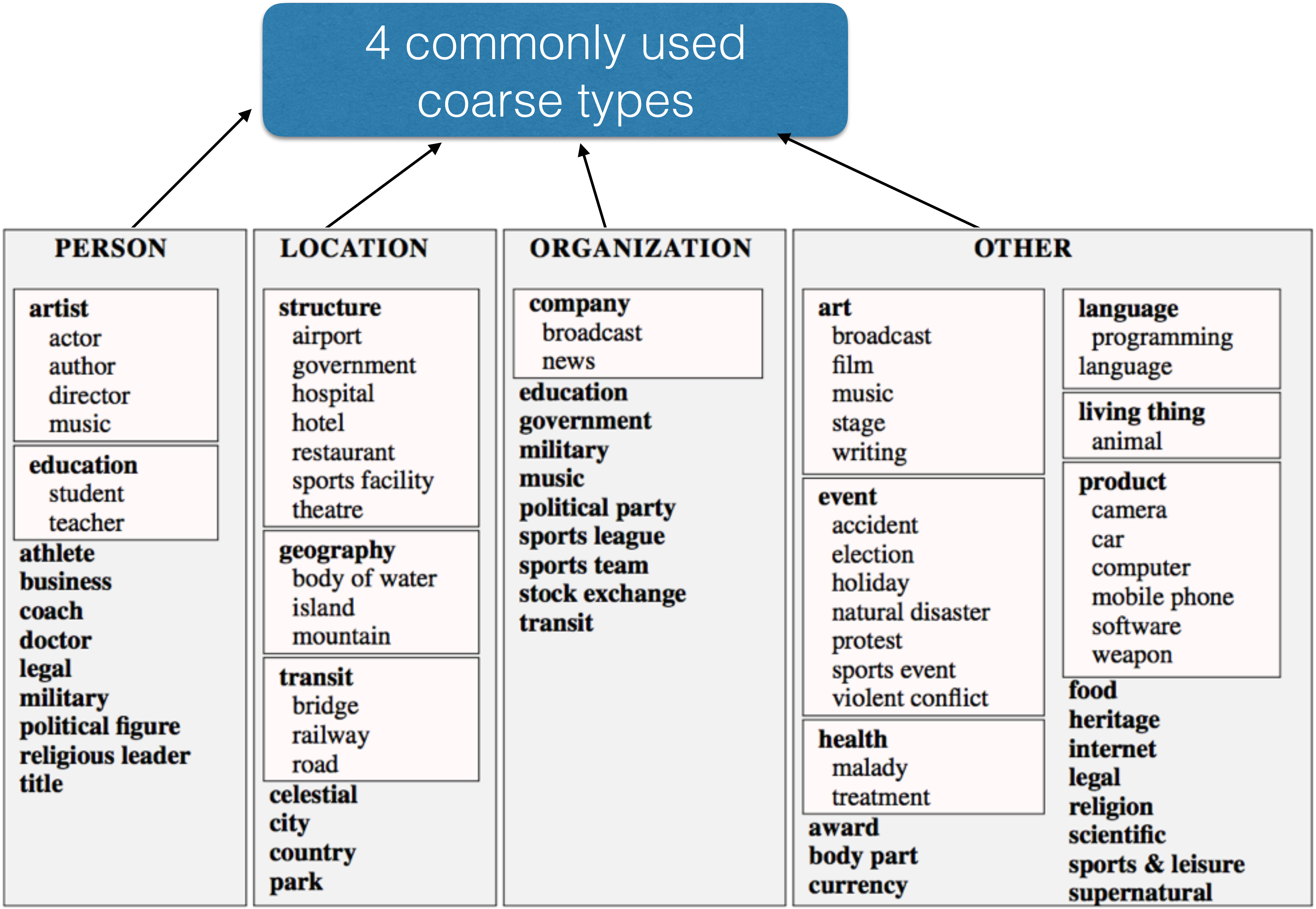
OTHER

- **art**
  - broadcast
  - film
  - music
  - stage
  - writing
- **event**
  - accident
  - election
  - holiday
  - natural disaster
  - protest
  - sports event
  - violent conflict
- **health**
  - malady
  - treatment
- **award**
- **body part**
- **currency**
- **language**
  - programming language
- **living thing**
  - animal
- **product**
  - camera
  - car
  - computer
  - mobile phone
  - software
  - weapon
- **food**
- **heritage**
- **internet**
- **legal**
- **religion**
- **scientific**
- **sports & leisure**
- **supernatural**

Type Labels and Manual Annotation

PERSON	LOCATION	ORGANIZATION	OTHER	
artist actor author director music	**structure** airport government hospital hotel restaurant sports facility theatre	**company** broadcast news	**art** broadcast film music stage writing	**language** programming language
education student teacher	**geography** body of water island mountain	**education**	**event** accident election holiday natural disaster protest sports event violent conflict	**living thing** animal
athlete	**transit** bridge railway road	**government**	**health** malady treatment	**product** camera car computer mobile phone software weapon
business	**celestial**	**military**	**award**	**food**
coach	**city**	**music**	**body part**	**heritage**
doctor	**country**	**political party**	**currency**	**internet**
legal	**park**	**sports league**		**legal**
military		**sports team**		**religion**
political figure		**stock exchange**		**scientific**
religious leader		**transit**		**sports & leisure**
title				**supernatural**

Type Labels and Manual Annotation



Type Labels and Manual Annotation

PERSON

- **artist**
  - actor
  - author
  - director
  - music
- **education**
  - student
  - teacher
- **athlete**
- **business**
- **coach**
- **doctor**
- **legal**
- **military**
- **political figure**
- **religious leader**
- **title**

LOCATION

- **structure**
  - airport
  - government
  - hospital
  - hotel
  - restaurant
  - sports facility
  - theatre
- **geography**
  - body of water
  - island
  - mountain
- **transit**
  - bridge
  - railway
  - road
- **celestial**
- **city**
- **country**
- **park**

ORGANIZATION

- **company**
  - broadcast
  - news
- **education**
- **government**
- **military**
- **music**
- **political party**
- **sports league**
- **sports team**
- **stock exchange**
- **transit**

OTHER

- **art**
  - broadcast
  - film
  - music
  - stage
  - writing
- **event**
  - accident
  - election
  - holiday
  - natural disaster
  - protest
  - sports event
  - violent conflict
- **health**
  - malady
  - treatment
- **award**
- **body part**
- **currency**
- **language**
  - programming language
- **living thing**
  - animal
- **product**
  - camera
  - car
  - computer
  - mobile phone
  - software
  - weapon
- **food**
- **heritage**
- **internet**
- **legal**
- **religion**
- **scientific**
- **sports & leisure**
- **supernatural**

Type Labels and Manual Annotation

Level II types -
person/athlete
organiztion/company

PERSON

- **artist**
  - actor
  - author
  - director
  - music
- **education**
  - student
  - teacher
- **athlete**
- **business**
- **coach**
- **doctor**
- **legal**
- **military**
- **political figure**
- **religious leader**
- **title**

LOCATION

- **structure**
  - airport
  - government
  - hospital
  - hotel
  - restaurant
  - sports facility
  - theatre
- **geography**
  - body of water
  - island
  - mountain
- **transit**
  - bridge
  - railway
  - road
- **celestial**
- **city**
- **country**
- **park**

ORGANIZATION

- **company**
  - broadcast
  - news
- **education**
- **government**
- **military**
- **music**
- **political party**
- **sports league**
- **sports team**
- **stock exchange**
- **transit**

OTHER

- **art**
  - broadcast
  - film
  - music
  - stage
  - writing
- **event**
  - accident
  - election
  - holiday
  - natural disaster
  - protest
  - sports event
  - violent conflict
- **health**
  - malady
  - treatment
- **award**
- **body part**
- **currency**
- **language**
  - programming language
- **living thing**
  - animal
- **product**
  - camera
  - car
  - computer
  - mobile phone
  - software
  - weapon
- **food**
- **heritage**
- **internet**
- **legal**
- **religion**
- **scientific**
- **sports & leisure**
- **supernatural**

Type Labels and Manual Annotation

PERSON

- **artist**
  - actor
  - author
  - director
  - music
- **education**
  - student
  - teacher
- **athlete**
- **business**
- **coach**
- **doctor**
- **legal**
- **military**
- **political figure**
- **religious leader**
- **title**

LOCATION

- **structure**
  - airport
  - government
  - hospital
  - hotel
  - restaurant
  - sports facility
  - theatre
- **geography**
  - body of water
  - island
  - mountain
- **transit**
  - bridge
  - railway
  - road
- **celestial**
- **city**
- **country**
- **park**

ORGANIZATION

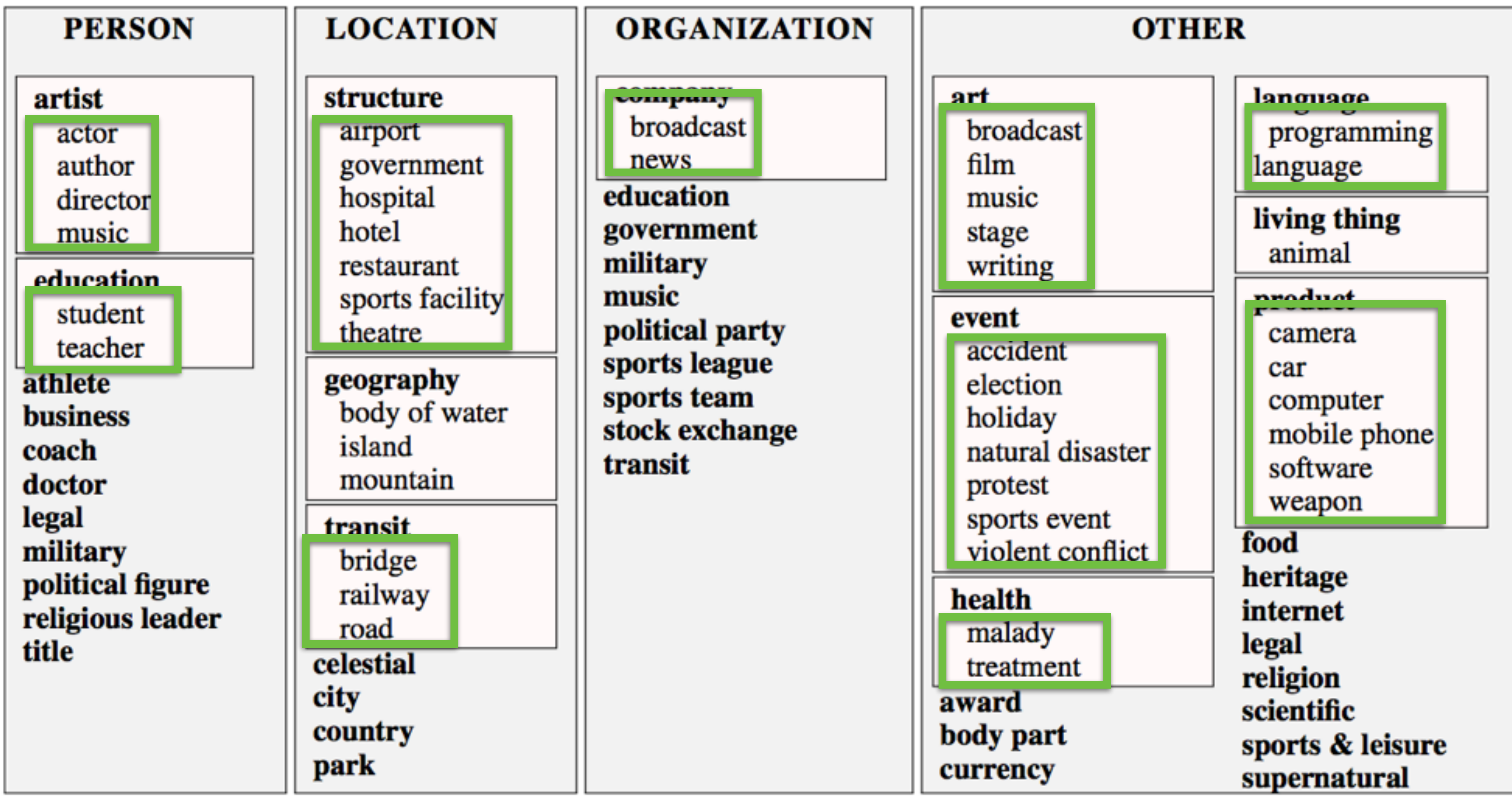
- **company**
  - broadcast
  - news
- **education**
- **government**
- **military**
- **music**
- **political party**
- **sports league**
- **sports team**
- **stock exchange**
- **transit**

OTHER

- **art**
  - broadcast
  - film
  - music
  - stage
  - writing
- **event**
  - accident
  - election
  - holiday
  - natural disaster
  - protest
  - sports event
  - violent conflict
- **health**
  - malady
  - treatment
- **award**
- **body part**
- **currency**
- **language**
  - programming language
- **living thing**
  - animal
- **product**
  - camera
  - car
  - computer
  - mobile phone
  - software
  - weapon
- **food**
- **heritage**
- **internet**
- **legal**
- **religion**
- **scientific**
- **sports & leisure**
- **supernatural**

Type Labels and Manual Annotation

Level III types -
person/artist/actor
other/event/election



Manual Annotation

Manual Annotation

- All news documents from the OntoNotes test set were manually annotated for context dependent types.

Manual Annotation

- All news documents from the OntoNotes test set were manually annotated for context dependent types.

Statistic	Value
Documents	77
Entity mentions	11304
Labels	17704
Labels at Level 1	11909
Labels at Level 2	5209
Labels at Level 3	586

Manual Annotation

- All news documents from the OntoNotes test set were manually annotated for context dependent types.

Statistic	**Value**
Documents	77
Entity mentions	11304
Labels	17704
Labels at Level 1	11909
Labels at Level 2	5209
Labels at Level 3	586

- Each document annotated by 6 annotators
- More annotations at the top level
- More disagreement at the bottom level (Specificity)
- Some disagreements at type level too (Type)

Distant Supervision for Training

Distant Supervision for Training

- The data described before was the *test* data

Distant Supervision for Training

- The data described before was the *test* data
- For training:

Distant Supervision for Training

- The data described before was the *test* data
- For training:
    - 133K news docs as training corpus.

Distant Supervision for Training

- The data described before was the *test* data
- For training:
  - 133K news docs as training corpus.
  - Run standard NLP pipeline to detect entity mentions.

Distant Supervision for Training

- The data described before was the <u>*test*</u> data
- For training:
    - 133K news docs as training corpus.
    - Run standard NLP pipeline to detect entity mentions.
        - POS tagger -> Dependency Parser -> NP extractor

Distant Supervision for Training

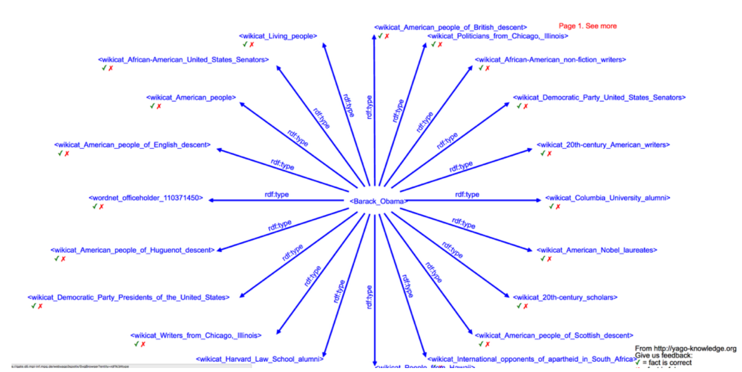
- The data described before was the *test* data
- For training:
  - 133K news docs as training corpus.
  - Run standard NLP pipeline to detect entity mentions.
    - POS tagger -> Dependency Parser -> NP extractor
  - Assign Freebase types to the resolved entities.

Distant Supervision for Training

- The data described before was the *test* data
- For training:
  - 133K news docs as training corpus.
  - Run standard NLP pipeline to detect entity mentions.
    - POS tagger -> Dependency Parser -> NP extractor
  - Assign Freebase types to the resolved entities.
- Problem with this approach?

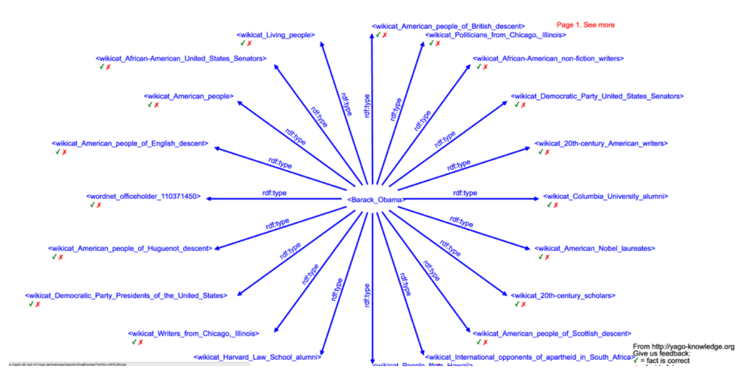
Distant Supervision for Training

- The data described before was the <u>*test*</u> data
- For training:
  - 133K news docs as training corpus.
  - Run standard NLP pipeline to detect entity mentions.
    - POS tagger -> Dependency Parser -> NP extractor
  - Assign Freebase types to the resolved entities.
- Problem with this approach?



Distant Supervision for Training

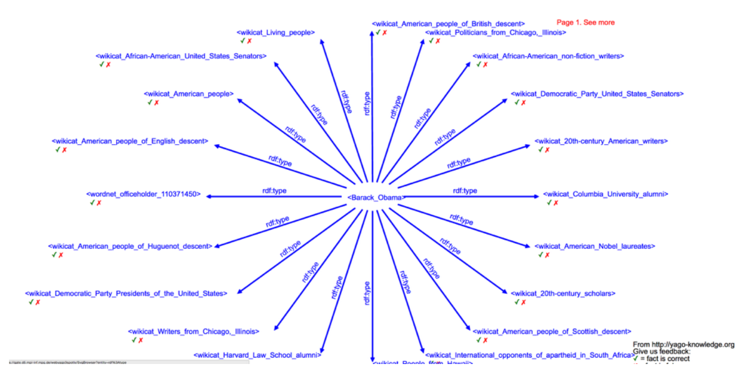
- The data described before was the *test* data
- For training:
  - 133K news docs as training corpus.
  - Run standard NLP pipeline to detect entity mentions.
    - POS tagger -> Dependency Parser -> NP extractor
  - Assign Freebase types to the resolved entities.
- Problem with this approach?



- Barack Obama will have all the types associated with him i.e. mentions will *not* have context dependent types

Distant Supervision for Training

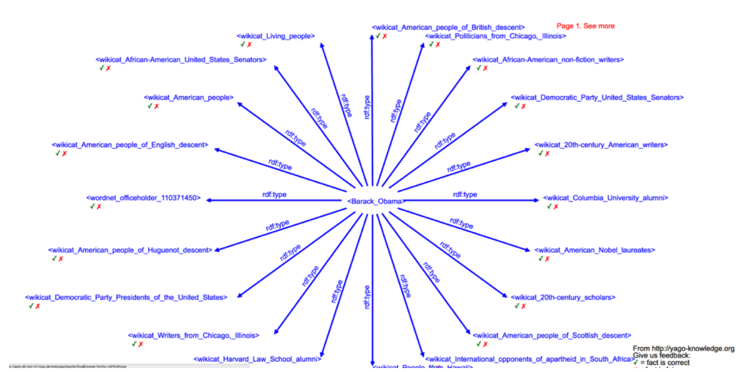
- The data described before was the ***test*** data
- For training:
  - 133K news docs as training corpus.
  - Run standard NLP pipeline to detect entity mentions.
    - POS tagger -> Dependency Parser -> NP extractor
  - Assign Freebase types to the resolved entities.
- Problem with this approach?



- Barack Obama will have all the types associated with him i.e. mentions will ***not*** have context dependent types
- Reduce the mismatch between training and test set by applying some heuristics.

Distant Supervision for Training

- The data described before was the *test* data
- For training:
    - 133K news docs as training corpus.
    - Run standard NLP pipeline to detect entity mentions.
        - POS tagger -> Dependency Parser -> NP extractor
    - Assign Freebase types to the resolved entities.
- Problem with this approach?



- Barack Obama will have all the types associated with him i.e. mentions will *not* have context dependent types
- Reduce the mismatch between training and test set by applying some heuristics.

Heuristics

Heuristics

- Sibling Pruning: remove sibling types associated with a single entity, leaving only the parent type i.e. a mention with types '*person/athlete*' and '*person/political-figure'* will be pruned to just '*person'.*

Heuristics

- <u>Sibling Pruning</u>: remove sibling types associated with a single entity, leaving only the parent type i.e. a mention with types '*person/athlete*' and '*person/political-figure'* will be pruned to just '*person'.*
    - Motivation: Uncommon for several sibling types to be relevant in the same context.

Heuristics

- <u>Sibling Pruning</u>: remove sibling types associated with a single entity, leaving only the parent type i.e. a mention with types '*person/athlete*' and '*person/political-figure'* will be pruned to just '*person'.*
    - Motivation: Uncommon for several sibling types to be relevant in the same context.
    - Less common entities associated with few freebase types are better as training data as they are annotated with types relevant from context.

Heuristics

- Sibling Pruning: remove sibling types associated with a single entity, leaving only the parent type i.e. a mention with types ‘*person/athlete*’ and ‘*person/political-figure’* will be pruned to just ‘*person’.*
  - Motivation: Uncommon for several sibling types to be relevant in the same context.
  - Less common entities associated with few freebase types are better as training data as they are annotated with types relevant from context.
- Coarse Type Pruning: Remove types that do not agree with the output of a standard coarse grained classifier trained on {‘person’, ‘location’, ‘organization’, ‘other’}

Heuristics

- <u>Sibling Pruning</u>: remove sibling types associated with a single entity, leaving only the parent type i.e. a mention with types '*person/athlete*' and '*person/political-figure'* will be pruned to just '*person'.*
    - Motivation: Uncommon for several sibling types to be relevant in the same context.
    - Less common entities associated with few freebase types are better as training data as they are annotated with types relevant from context.
- <u>Coarse Type Pruning</u>: Remove types that do not agree with the output of a standard coarse grained classifier trained on {'person', 'location', 'organization', 'other'}
    - Motivation: Removes conflicting types of mentions such as location & organization

Heuristics

- <u>Sibling Pruning</u>: remove sibling types associated with a single entity, leaving only the parent type i.e. a mention with types '*person/athlete*' and '*person/political-figure'* will be pruned to just '*person'.*
  - Motivation: Uncommon for several sibling types to be relevant in the same context.
  - Less common entities associated with few freebase types are better as training data as they are annotated with types relevant from context.
- <u>Coarse Type Pruning</u>: Remove types that do not agree with the output of a standard coarse grained classifier trained on {'person', 'location', 'organization', 'other'}
  - Motivation: Removes conflicting types of mentions such as location & organization
- <u>Minimum Count Pruning</u>: Remove types which appear less then 'k' times.

Features, Model & Inference

Features, Model & Inference

- Sparse feature based model

Features, Model & Inference

- Sparse feature based model

She is married to the 44th and current President of the United States, Barack Obama, and is the first African-American First Lady…..

Features, Model & Inference

- Sparse feature based model

She is married to the 44th and current President of the United States, Barack Obama, and is the first African-American First Lady…..

Features, Model & Inference

- Sparse feature based model

She is married to the 44th and current President of the United States, Barack Obama, and is the first African-American First Lady…..

Features, Model & Inference

- Sparse feature based model
- For each mention, we have a sparse feature vector

She is married to the 44th and current President of the United States, Barack Obama, and is the first African-American First Lady…..

x

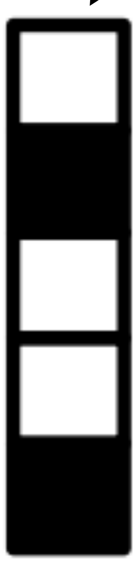
Feature	Description	Example
Head	The syntactic head of the mention phrase	"Obama"
Non-head	Each non-head word in the mention phrase	"Barack", "H."
Cluster	Word cluster id for the head word	"59"
Characters	Each character trigram in the mention head	":ob", "oba", "bam", "ama", "ma:"
Shape	The word shape of the words in the mention phrase	"Aa A. Aa"
Role	Dependency label on the mention head	"nsubj"
Context	Words before and after the mention phrase	"B:who", "A:first"
Parent	The head's lexical parent in the dependency tree	"picked"
Topic	The most likely topic label for the document	"politics"

Table 4: List of features used in type tagging. Features are extracted from each mention. Context used for example features: "... who [Barack H. Obama] first picked ..."

Features, Model & Inference

- Sparse feature based model
- For each mention, we have a sparse feature vector
- and a sparse label vector

She is married to the 44th and current President of the United States, $\boxed{\text{Barack Obama,}}$ and is the first African-American First Lady…..



Feature	Description	Example
Head	The syntactic head of the mention phrase	"Obama"
Non-head	Each non-head word in the mention phrase	"Barack", "H."
Cluster	Word cluster id for the head word	"59"
Characters	Each character trigram in the mention head	":ob", "oba", "bam", "ama", "ma:"
Shape	The word shape of the words in the mention phrase	"Aa A. Aa"
Role	Dependency label on the mention head	"nsubj"
Context	Words before and after the mention phrase	"B:who", "A:first"
Parent	The head's lexical parent in the dependency tree	"picked"
Topic	The most likely topic label for the document	"politics"

Table 4: List of features used in type tagging. Features are extracted from each mention. Context used for example features: "... who [Barack H. Obama] first picked ..."

$\mathbf{x}$

$\mathbf{y}_t \in \mathbb{R}^{\mathsf{T}}$

T equals the total number of types.

Models & Inference

Models & Inference

- The problem can be viewed as structured multi-label classification problem.

Models & Inference

- The problem can be viewed as structured multi-label classification problem.

$$\left[\begin{array}{cl} 0 & \text{/location/city} \\ 1 & \text{/person} \\ 1 & \text{/person/political\_figure} \\ 0 & \text{/other/food} \\ 0 & \text{/organization/business} \\ & \\ & \end{array}\right]^{\top}$$

Models & Inference

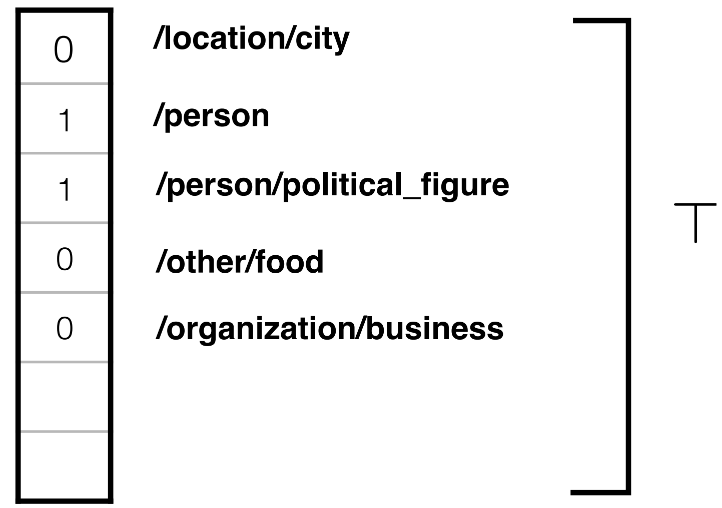
- The problem can be viewed as structured multi-label classification problem.

$$\begin{bmatrix} 0 & \text{/location/city} \\ 1 & \text{/person} \\ 1 & \text{/person/political\_figure} \\ 0 & \text{/other/food} \\ 0 & \text{/organization/business} \\ & \\ & \end{bmatrix}^{T}$$

Training Data

Models & Inference

- The problem can be viewed as structured multi-label classification problem.



Training Data

- +ve examples: For a type, +ve examples are all mentions marked as itself and also its descendants i.e. a mention labeled '*person/artist*' is a +ve example for '*person*'

Models & Inference

- The problem can be viewed as structured multi-label classification problem.

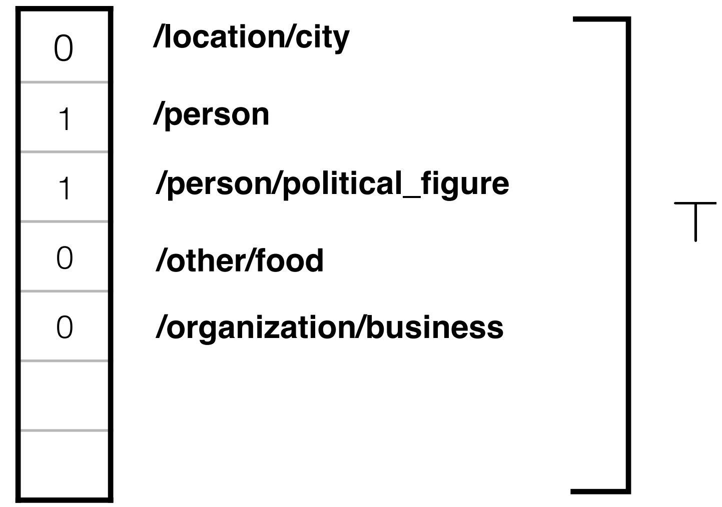
$$\left[\begin{array}{cl} 0 & \textbf{/location/city} \\ 1 & \textbf{/person} \\ 1 & \textbf{/person/political\_figure} \\ 0 & \textbf{/other/food} \\ 0 & \textbf{/organization/business} \\ & \\ & \end{array}\right]^{\top}$$

Training Data

- +ve examples: For a type, +ve examples are all mentions marked as itself and also its descendants i.e. a mention labeled ‘*person/artist’* is a +ve example for ‘*person*’
- -ve examples:

Models & Inference

- The problem can be viewed as structured multi-label classification problem.



Training Data

- +ve examples: For a type, +ve examples are all mentions marked as itself and also its descendants i.e. a mention labeled '*person/artist*' is a +ve example for '*person*'
- -ve examples:
    1. all other types with the same parent ('*person/artist' will be -ve for* '*person/author*')

Models & Inference

- The problem can be viewed as structured multi-label classification problem.

$$\left[\begin{array}{cl} 0 & \textbf{/location/city} \\ 1 & \textbf{/person} \\ 1 & \textbf{/person/political\_figure} \\ 0 & \textbf{/other/food} \\ 0 & \textbf{/organization/business} \\ & \\ & \end{array}\right]^{\top}$$

Training Data

- +ve examples: For a type, +ve examples are all mentions marked as itself and also its descendants i.e. a mention labeled '*person/artist*' is a +ve example for '*person*'
- -ve examples:
  1. all other types with the same parent ('*person/artist' will be -ve for* '*person/author*')
  2. all other types at the same depth ('*organization/business*' *will be -ve for* '*person/author*')

Models & Inference

- The problem can be viewed as structured multi-label classification problem.

$$\left[\begin{array}{cl} 0 & \text{/location/city} \\ 1 & \text{/person} \\ 1 & \text{/person/political\_figure} \\ 0 & \text{/other/food} \\ 0 & \text{/organization/business} \\ & \\ & \end{array}\right]^{\top}$$

Training Data

- +ve examples: For a type, +ve examples are all mentions marked as itself and also its descendants i.e. a mention labeled ‘*person/artist*’ is a +ve example for ‘*person*’
- -ve examples:
  1. all other types with the same parent (‘*person/artist’ will be -ve for* ‘*person/author*’)
  2. all other types at the same depth (‘*organization/business*’ *will be -ve for* ‘*person/author*’)
  3. all other types

Models & Inference
Local Model

Models & Inference
Local Model

- Local classifiers: A binary logistic regression classifier is trained for each label (i.e. there are T classifiers) and label consistency is enforced at inference time

Models & Inference
Local Model

- Local classifiers: A binary logistic regression classifier is trained for each label (i.e. there are T classifiers) and label consistency is enforced at inference time
- During inference:
    - Independent: Assign all types that exceeds some decision threshold
    - Conditional: Multiply each label with the probability of its parent. This strategy ensures that if a label is selected then its parent would also be selected
    - Marginalization: (if I understand correctly) Probability of a label is the sum of configurations in which it appears.

Models & Inference
Global/Flat Model

Models & Inference
Global/Flat Model

- Flat classifier: A flat softmax classifier to discriminate between all types

Models & Inference
Global/Flat Model

- Flat classifier: A flat softmax classifier to discriminate between all types
  - Classifier expects single type label to each instance.

Models & Inference
Global/Flat Model

- Flat classifier: A flat softmax classifier to discriminate between all types
  - Classifier expects single type label to each instance.
  - To account for that each multi-label instance, will be treated as multiple single label instance

Models & Inference
Global/Flat Model

- Flat classifier: A flat softmax classifier to discriminate between all types
  - Classifier expects single type label to each instance.
  - To account for that each multi-label instance, will be treated as multiple single label instance

Models & Inference
Global/Flat Model

- Flat classifier: A flat softmax classifier to discriminate between all types
  - Classifier expects single type label to each instance.
  - To account for that each multi-label instance, will be treated as multiple single label instance

- Inference time: assign all labels whose probability exceeds a threshold; not just the max.

Results

Results

Different -ve example strategies:

Results

Different -ve example strategies:

Negatives	Prec	Rec	F1	AUC
All	77.98	59.55	67.53	66.56
Sibling	79.93	58.94	67.85	66.50
Depth	**80.05**	**62.20**	**70.01**	**69.29**

Results

Different -ve example strategies:

Negatives	**Prec**	**Rec**	**F1**	**AUC**
All	77.98	59.55	67.53	66.56
Sibling	79.93	58.94	67.85	66.50
Depth	**80.05**	**62.20**	**70.01**	**69.29**

Different inference schemes:

Results

Different -ve example strategies:

Negatives	**Prec**	**Rec**	**F1**	**AUC**
All	77.98	59.55	67.53	66.56
Sibling	79.93	58.94	67.85	66.50
Depth	**80.05**	**62.20**	**70.01**	**69.29**

Different inference schemes:

Inference	**Prec**	**Rec**	**F1**	**AUC**
Independent	77.06	61.54	68.43	67.74
Conditional	77.89	**63.30**	69.84	**70.04**
Marginals	**80.05**	62.20	**70.01**	69.29

Results

Results

Comparison between local and flat classifiers among different levels

Results

Comparison between local and flat classifiers among different levels

Classifier	Precision	Recall	F1
Level 1 Flat	84.39	79.01	81.61
Level 1 Local	87.12	78.84	**82.80**
Level 2 Flat	46.61	25.99	33.37
Level 2 Local	56.76	30.88	**40.00**
Level 3 Flat	75.00	1.78	3.47
Level 3 Local	24.00	8.28	**12.32**

Table 6: Precision, recall, and F-Score given by the flat and local classifiers at each level of the type taxonomy. We use all heuristics and Depth negative examples for the local classifiers. Level 1 are the labels immediately below the root of our tree: *person*, *location*, *organization*, and *other*. Level 2 are the labels below them such as *person/artist* while Level 3 are one level lower such as *person/artist/actor*.

Conclusions

- Strives to make fine grained typing meaningful by requiring context dependence
- Introduce several distant supervision heuristics aimed at pruning irrelevant labels from the training data and match the gold data.
- Introduce new dataset 11,304 manually annotated mentions in 77 OntoNotes news documents.

Embedding Methods for Fine Grained Entity Type Classification

-Dani Yogatama, Dan Gillick, Nevena Lazic
CMU, Google

Key Contribution

Key Contribution

- Learn low dimensional embeddings of entity types and mentions instead of sparse binary embeddings.

Key Contribution

- Learn low dimensional embeddings of entity types and mentions instead of sparse binary embeddings.
- Previously, each mention was one big sparse feature vector

Key Contribution

- Learn low dimensional embeddings of entity types and mentions instead of sparse binary embeddings.

- Previously, each mention was one big sparse feature vector

If a hostile predator emerges for Saatchi & Saatchi Co., cofounders

Charles and Maurice Saatchi will lead…..

Feature	Description	Example
Head	The syntactic head of the mention phrase	"Obama"
Non-head	Each non-head word in the mention phrase	"Barack", "H."
Cluster	Word cluster id for the head word	"59"
Characters	Each character trigram in the mention head	":ob", "oba", "bam", "ama", "ma:"
Shape	The word shape of the words in the mention phrase	"Aa A. Aa"
Role	Dependency label on the mention head	"nsubj"
Context	Words before and after the mention phrase	"B:who", "A:first"
Parent	The head's lexical parent in the dependency tree	"picked"
Topic	The most likely topic label for the document	"politics"

Table 4: List of features used in type tagging. Features are extracted from each mention. Context used for example features: "... who [Barack H. Obama] first picked ..."

Key Contribution

- Learn low dimensional embeddings of entity types and mentions instead of sparse binary embeddings.

- Previously, each mention was one big sparse feature vector

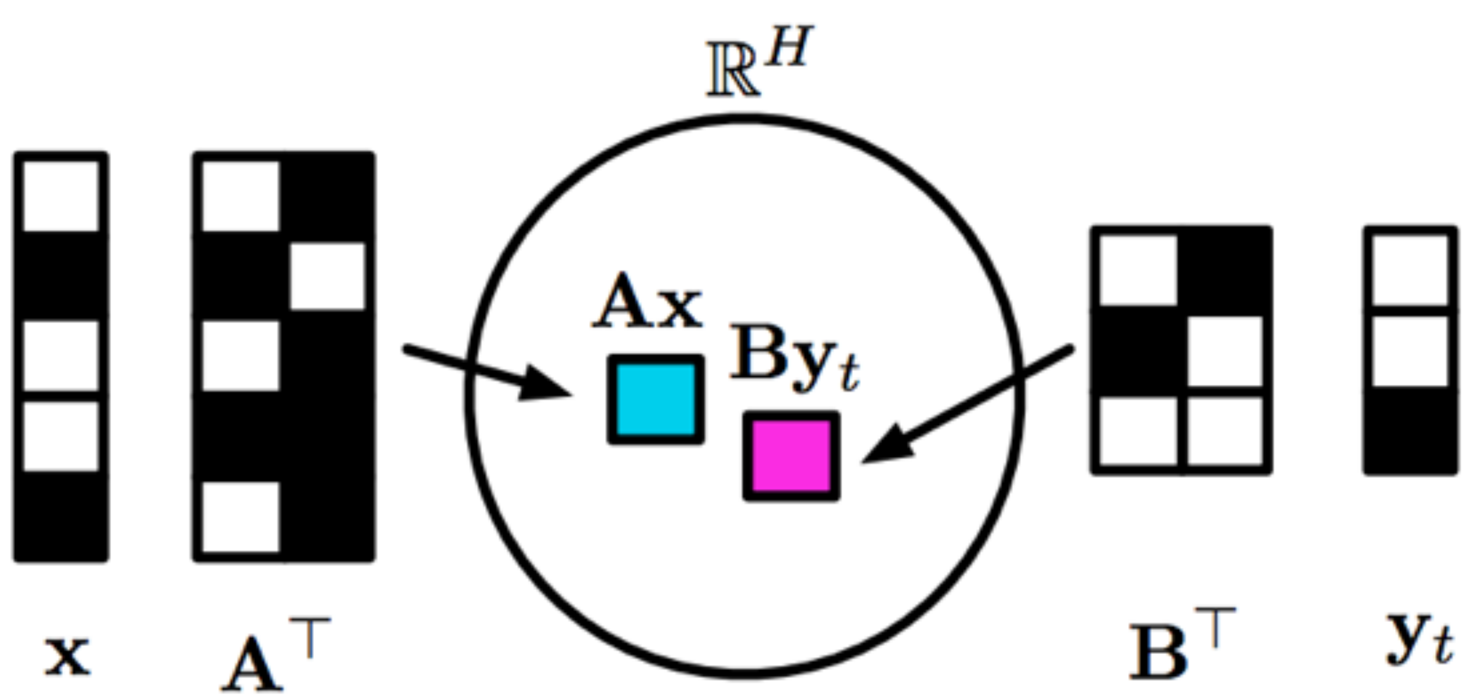
Key Contribution

- Learn low dimensional embeddings of entity types and mentions instead of sparse binary embeddings.

- Previously, each mention was one big sparse feature vector
- Now instead, learn a low dimensional representation for each mention and types.

Key Contribution

- Learn low dimensional embeddings of entity types and mentions instead of sparse binary embeddings.
- Previously, each mention was one big sparse feature vector
- Now instead, learn a low dimensional representation for each mention and types.



Key Contribution

- Learn low dimensional embeddings of entity types and mentions instead of sparse binary embeddings.

- Previously, each mention was one big sparse feature vector
- Now instead, learn a low dimensional representation for each mention and types.

Key Contribution

- Learn low dimensional embeddings of entity types and mentions instead of sparse binary embeddings.

- Previously, each mention was one big sparse feature vector
- Now instead, learn a low dimensional representation for each mention and types.
- Motivation: Learning low dimensional embeddings allows information sharing among related labels. For example: person/author would be more closer to person/artist than location/city.

Model

Model

Interested in learning mapping function f and g

$$f(x) : \mathbb{R}^D \rightarrow \mathbb{R}^H$$

$$\forall t \in \{1, 2, \ldots, \top\}, g(y_t) : \{0, 1\}^\top \rightarrow \mathbb{R}^H$$

Model

Interested in learning r

Low dim. embedding for mentions

f and g

$$f(x) : \mathbb{R}^D \to \mathbb{R}^H$$

$$\forall t \in \{1, 2, \ldots, \top\}, g(y_t) : \{0, 1\}^\top \to \mathbb{R}^H$$

Model

Interested in learning r Low dim. embedding for mentions f and g

$$f(x) : \mathbb{R}^D \rightarrow \mathbb{R}^H$$

$$\forall t \in \{1, 2, \ldots, \top\}, g(y_t) : \{0, 1\}^\top \rightarrow \mathbb{R}^H$$

Low dim. embedding for each entity type (label)

Model

Interested in learning r … f and g

Low dim. embedding for mentions

$$f(x) : \mathbb{R}^D \rightarrow \mathbb{R}^H$$

$$\forall t \in \{1, 2, \ldots, \top\}, g(y_t) : \{0, 1\}^\top \rightarrow \mathbb{R}^H$$

Low dim. embedding for each entity type (label)

Learning:

Score of a label t (represented as one hot vector y_t) and a feature vector x

Model

Interested in learning r Low dim. embedding for mentions f and g

$$f(x) : \mathbb{R}^D \rightarrow \mathbb{R}^H$$

$$\forall t \in \{1, 2, \ldots, \top\}, g(y_t) : \{0, 1\}^\top \rightarrow \mathbb{R}^H$$

Low dim. embedding for each entity type (label)

Learning:
Score of a label t (represented as one hot vector y_t) and a feature vector x

$$s(\mathbf{x}, \mathbf{y_t},; \mathbf{A}, \mathbf{B}) = \mathbf{f}(\mathbf{x}, \mathbf{A}) \cdot \mathbf{g}(\mathbf{y_t}, \mathbf{B}) = \mathbf{A}\mathbf{x} \cdot \mathbf{B}\mathbf{y_t}$$

$$\mathbf{A} \in \mathbb{R}^{\mathbf{H} \times \mathbf{D}}, \mathbf{B} \in \mathbb{R}^{\mathbf{H} \times \top}$$

Model

Interested in learning r f and g

Low dim. embedding for mentions

$$f(x) : \mathbb{R}^D \to \mathbb{R}^H$$

$$\forall t \in \{1, 2, \ldots, \top\}, g(y_t) : \{0, 1\}^\top \to \mathbb{R}^H$$

Low dim. embedding for each entity type (label)

Learning:
Score of a label t (represented as one hot vector y_t) and a feature vector x

$$s(\mathbf{x}, \mathbf{y_t},; \mathbf{A}, \mathbf{B}) = \mathbf{f}(\mathbf{x}, \mathbf{A}) \cdot \mathbf{g}(\mathbf{y_t}, \mathbf{B}) = \mathbf{Ax} \cdot \mathbf{By_t}$$

$$\mathbf{A} \in \mathbb{R}^{\mathbf{H} \times \mathbf{D}}, \mathbf{B} \in \mathbb{R}^{\mathbf{H} \times \top}$$

f and g are linear mappings and score is calculated as dot product

Kernel Extension

Kernel Extension

$$\begin{aligned} s(x, y_t) &= Ax \cdot By_t \\ &= \sum_d (A_d x_d)^\top B_t \end{aligned}$$

Kernel Extension

$$
\begin{aligned}
s(x, y_t) &= Ax \cdot By_t \\
&= \sum_d (A_d x_d)^\top B_t
\end{aligned}
$$

Kernel Extension:

$$
\sum_d K_{d,t} (A_d x_d)^\top B_t
$$

Kernel Extension

$$s(x, y_t) = Ax \cdot By_t$$
$$= \sum_d (A_d x_d)^\top B_t$$

Kernel Extension:

$$\sum_d K_{d,t}(A_d x_d)^\top B_t$$

K is the kernel matrix

$K_{d,t} = 1$, if A_d is one of the nearest neighbor of B_t

Kernel Extension

$$s(x, y_t) = Ax \cdot By_t$$
$$= \sum_d (A_d x_d)^\top B_t$$

Kernel Extension:

$$\sum_d K_{d,t} (A_d x_d)^\top B_t$$

K is the kernel matrix

$K_{d,t} = 1$, if A_d is one of the nearest neighbor of B_t

- It implicitly plays the role of label-dependent feature selector, learning which features can interact with which labels and turning the appropriate noisy ones off

Kernel Extension

$$\begin{aligned} s(x, y_t) &= Ax \cdot By_t \\ &= \sum_d (A_d x_d)^\top B_t \end{aligned}$$

Kernel Extension:

$$\sum_d K_{d,t} (A_d x_d)^\top B_t$$

K is the kernel matrix

$K_{d,t} = 1$, if A_d is one of the nearest neighbor of B_t

- It implicitly plays the role of label-dependent feature selector, learning which features can interact with which labels and turning the appropriate noisy ones off
- It is also a way of introducing non-linearity!

Results

Results

Method	P	R	F1
FLAT	79.22	60.18	68.40
BINARY	80.05	62.20	70.01
WSABIE	**80.58**	66.20	72.68
K-WSABIE	80.11	**67.01**	**72.98**

Table 3: Precision (P), Recall (R), and F1-score on the GFT test dataset for four competing models. The improvements for WSABIE and K-WSABIE over both baselines are statistically significant ($p < 0.01$).

Results

Method	P	R	F1
FLAT	79.22	60.18	68.40
BINARY	80.05	62.20	70.01
WSABIE	**80.58**	66.20	72.68
K-WSABIE	80.11	**67.01**	**72.98**

Table 3: Precision (P), Recall (R), and F1-score on the GFT test dataset for four competing models. The improvements for WSABIE and K-WSABIE over both baselines are statistically significant ($p < 0.01$).

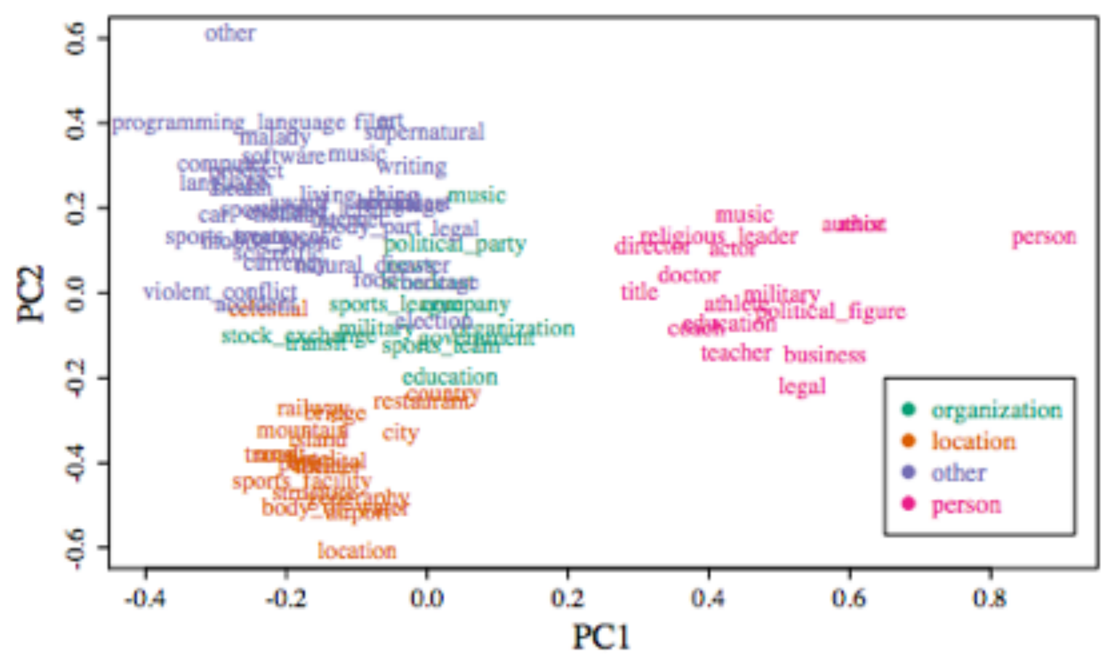


Figure 3: Two-dimensional projections of label embeddings for GFT dataset. See text for details.

An Attentive Neural Architecture for Fine-grained Entity Type Classification

Sonse Shimaoka, Pontus Stenetorp, Kentaro Inui, Sebastian Riedel
Tohoku University, University College London

Key Contribution

Key Contribution

- Previous work modeled context as sparse features

Key Contribution

- Previous work modeled context as sparse features

Feature	Description	Example
Head	The syntactic head of the mention phrase	“Obama”
Non-head	Each non-head word in the mention phrase	“Barack”, “H.”
Cluster	Word cluster id for the head word	“59”
Characters	Each character trigram in the mention head	“:ob”, “oba”, “bam”, “ama”, “ma:”
Shape	The word shape of the words in the mention phrase	“Aa A. Aa”
Role	Dependency label on the mention head	“nsubj”
Context	Words before and after the mention phrase	“B:who”, “A:first”
Parent	The head’s lexical parent in the dependency tree	“picked”
Topic	The most likely topic label for the document	“politics”

Table 4: List of features used in type tagging. Features are extracted from each mention. Context used for example features: “... who [Barack H. Obama] first picked ...”

Key Contribution

- Previous work modeled context as sparse features

Feature	Description	Example
Head	The syntactic head of the mention phrase	"Obama"
Non-head	Each non-head word in the mention phrase	"Barack", "H."
Cluster	Word cluster id for the head word	"59"
Characters	Each character trigram in the mention head	":ob", "oba", "bam", "ama", "ma:"
Shape	The word shape of the words in the mention phrase	"Aa A. Aa"
Role	Dependency label on the mention head	"nsubj"
Context	Words before and after the mention phrase	"B:who", "A:first"
Parent	The head's lexical parent in the dependency tree	"picked"
Topic	The most likely topic label for the document	"politics"

Table 4: List of features used in type tagging. Features are extracted from each mention. Context used for example features: "... who [Barack H. Obama] first picked ..."

- This has severe limitations:
  - Cannot model larger context; will lead to feature explosion.

Key Contribution

- Previous work modeled context as sparse features

Feature	Description	Example
Head	The syntactic head of the mention phrase	"Obama"
Non-head	Each non-head word in the mention phrase	"Barack", "H."
Cluster	Word cluster id for the head word	"59"
Characters	Each character trigram in the mention head	":ob", "oba", "bam", "ama", "ma:"
Shape	The word shape of the words in the mention phrase	"Aa A. Aa"
Role	Dependency label on the mention head	"nsubj"
Context	Words before and after the mention phrase	"B:who", "A:first"
Parent	The head's lexical parent in the dependency tree	"picked"
Topic	The most likely topic label for the document	"politics"

Table 4: List of features used in type tagging. Features are extracted from each mention. Context used for example features: "... who [Barack H. Obama] first picked ..."

- This has severe limitations:
  - Cannot model larger context; will lead to feature explosion.
  - Similar context do not share statistical strengths

Key Contribution

- Previous work modeled context as sparse features

Feature	Description	Example
Head	The syntactic head of the mention phrase	"Obama"
Non-head	Each non-head word in the mention phrase	"Barack", "H."
Cluster	Word cluster id for the head word	"59"
Characters	Each character trigram in the mention head	":ob", "oba", "bam", "ama", "ma:"
Shape	The word shape of the words in the mention phrase	"Aa A. Aa"
Role	Dependency label on the mention head	"nsubj"
Context	Words before and after the mention phrase	"B:who", "A:first"
Parent	The head's lexical parent in the dependency tree	"picked"
Topic	The most likely topic label for the document	"politics"

Table 4: List of features used in type tagging. Features are extracted from each mention. Context used for example features: "... who [Barack H. Obama] first picked ..."

- This has severe limitations:
  - Cannot model larger context; will lead to feature explosion.
  - Similar context do not share statistical strengths

River Monongahela <u>flows through</u> Pittsburgh

Pittsburgh has 3 rivers Allegheny, Monongahela, Ohio <u>running through</u> it

Key Contribution

Key Contribution

- Recurrent Neural Networks to the rescue!

Key Contribution

- Recurrent Neural Networks to the rescue!
- Given an entity mention and its left and right context words….

$$l_1, \ldots, l_C, m_1, \ldots, m_M, r_1, \ldots, r_C$$

Key Contribution

- Recurrent Neural Networks to the rescue!
- Given an entity mention and its left and right context words….

$$l_1, \ldots, l_C, m_1, \ldots, m_M, r_1, \ldots, r_C$$

- Encode the left and right context using a LSTM.

$$\overrightarrow{l_1, \ldots, l_C}^{\text{LSTM}}, m_1, \ldots, m_M, \overleftarrow{r_1, \ldots, r_C}^{\text{LSTM}}$$

Key Contribution

- Recurrent Neural Networks to the rescue!
- Given an entity mention and its left and right context words….

$$l_1, \ldots, l_C, m_1, \ldots, m_M, r_1, \ldots, r_C$$

- Encode the left and right context using a LSTM.

$$\overset{\text{LSTM}}{\overrightarrow{l_1, \ldots, l_C}}, m_1, \ldots, m_M, \overset{\text{LSTM}}{\overleftarrow{r_1, \ldots, r_C}}$$

- Side Note: They averaged the mention embeddings to get one vector

$$v_m = \frac{\sum_{i=1}^{M} u(m_i)}{M}$$

Attentive Encoder

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

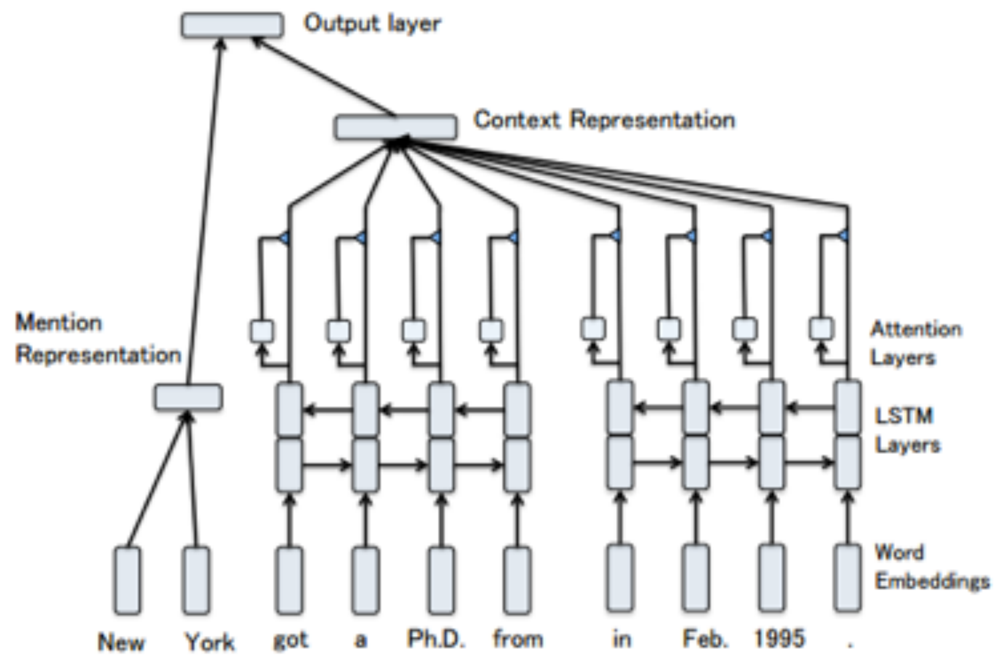


Figure 1: An illustration of our proposed model predicting fine-grained semantic types for the mention "New York" in the sentence "She got a Ph.D from New York in Feb. 1995.".

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

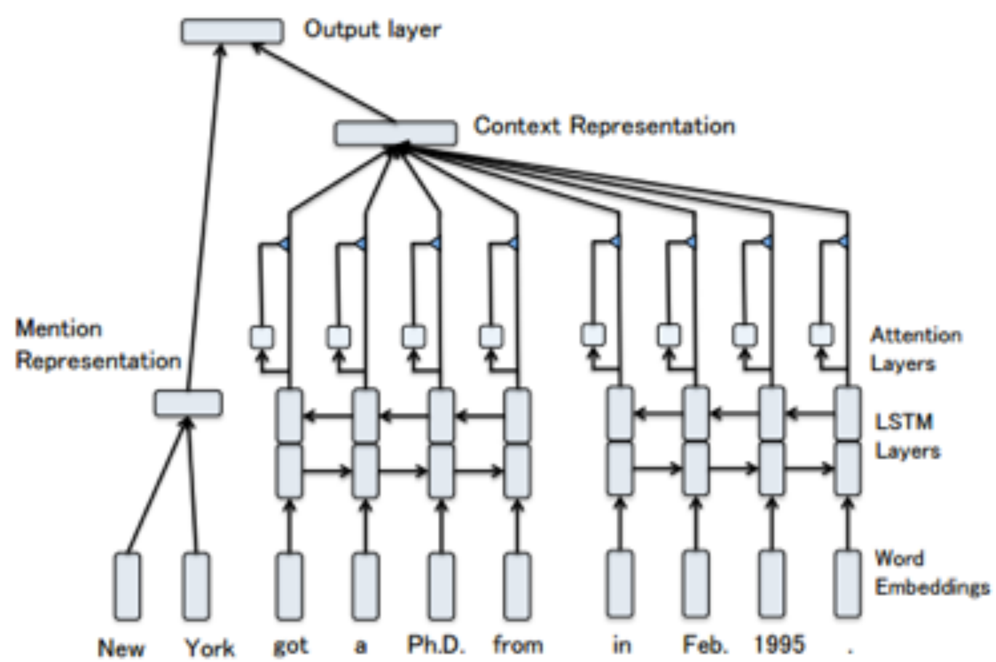


Figure 1: An illustration of our proposed model predicting fine-grained semantic types for the mention "New York" in the sentence "She got a Ph.D from New York in Feb. 1995.".

- For each of the intermediate context vector,

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

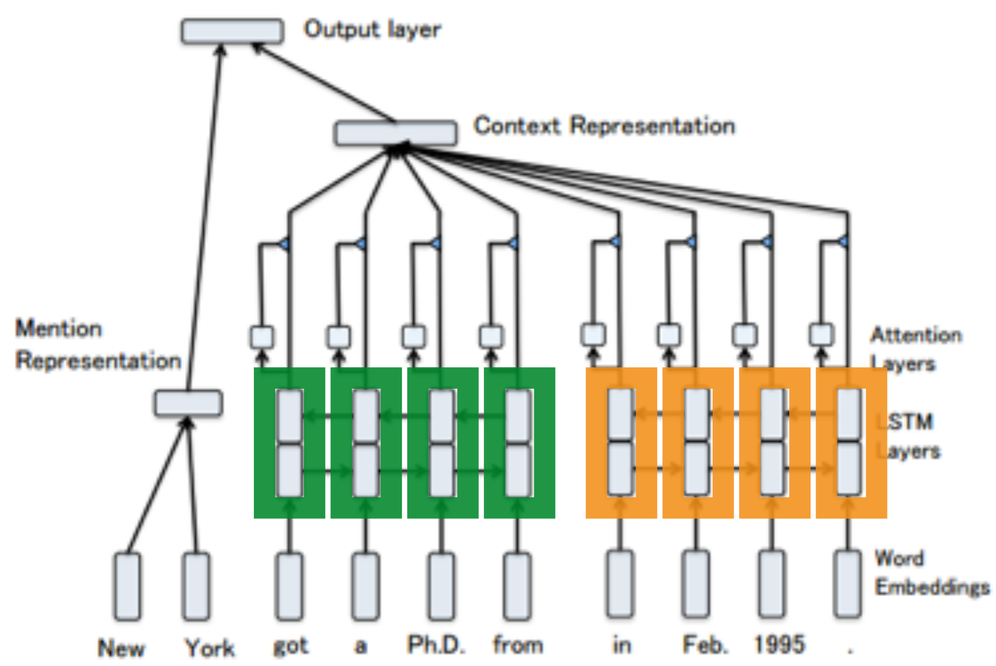


Figure 1: An illustration of our proposed model predicting fine-grained semantic types for the mention "New York" in the sentence "She got a Ph.D from New York in Feb. 1995.".

- For each of the intermediate context vector,

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

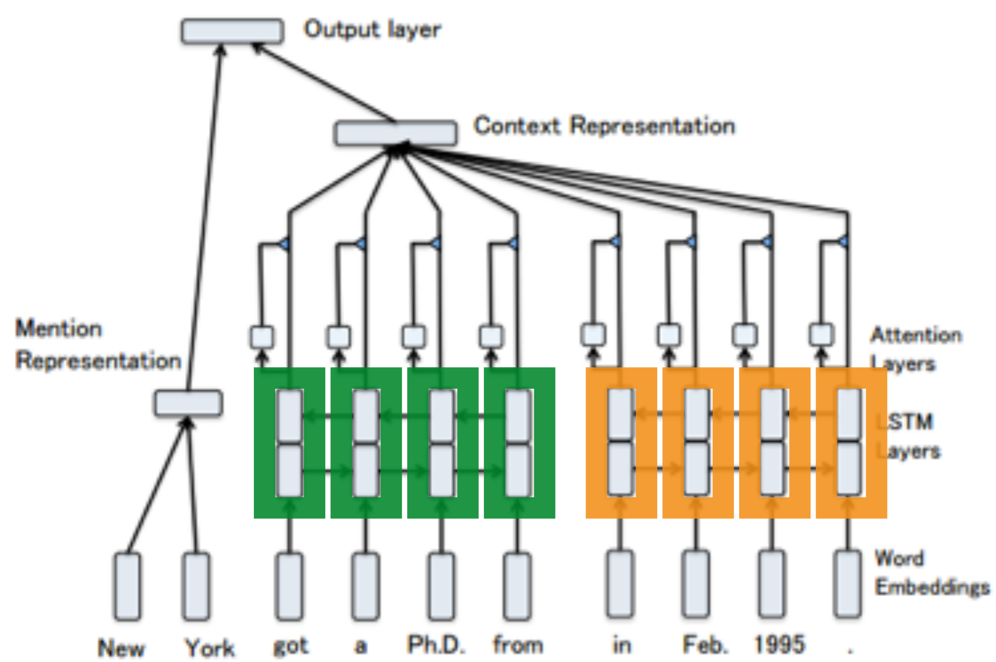


Figure 1: An illustration of our proposed model predicting fine-grained semantic types for the mention "New York" in the sentence "She got a Ph.D from New York in Feb. 1995.".

- For each of the intermediate context vector, compute an attention weight.

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

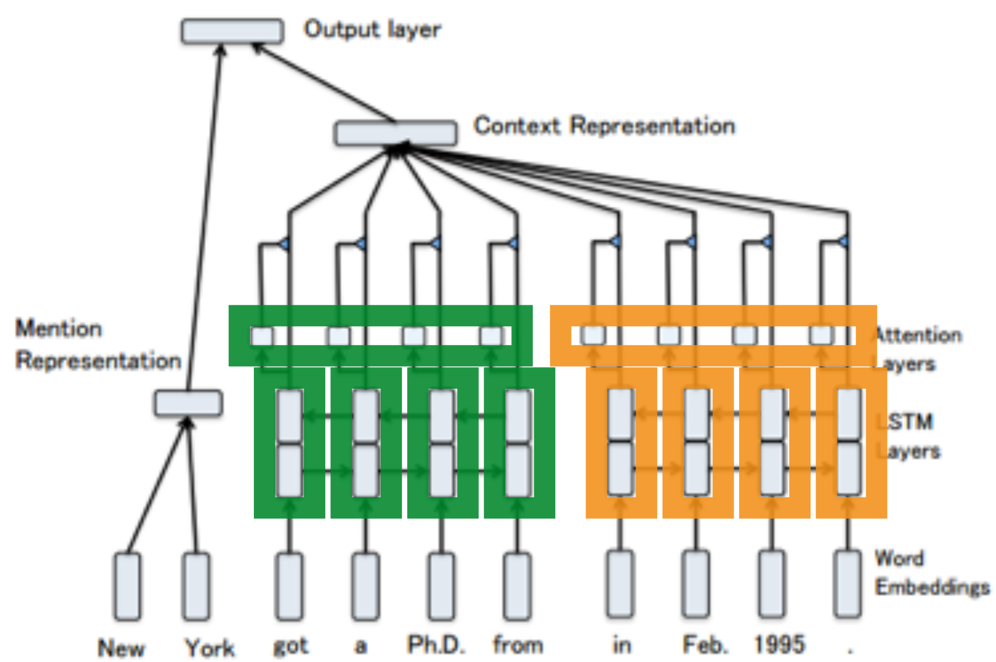


Figure 1: An illustration of our proposed model predicting fine-grained semantic types for the mention "New York" in the sentence "She got a Ph.D from New York in Feb. 1995.".

- For each of the intermediate context vector, compute an attention weight.

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

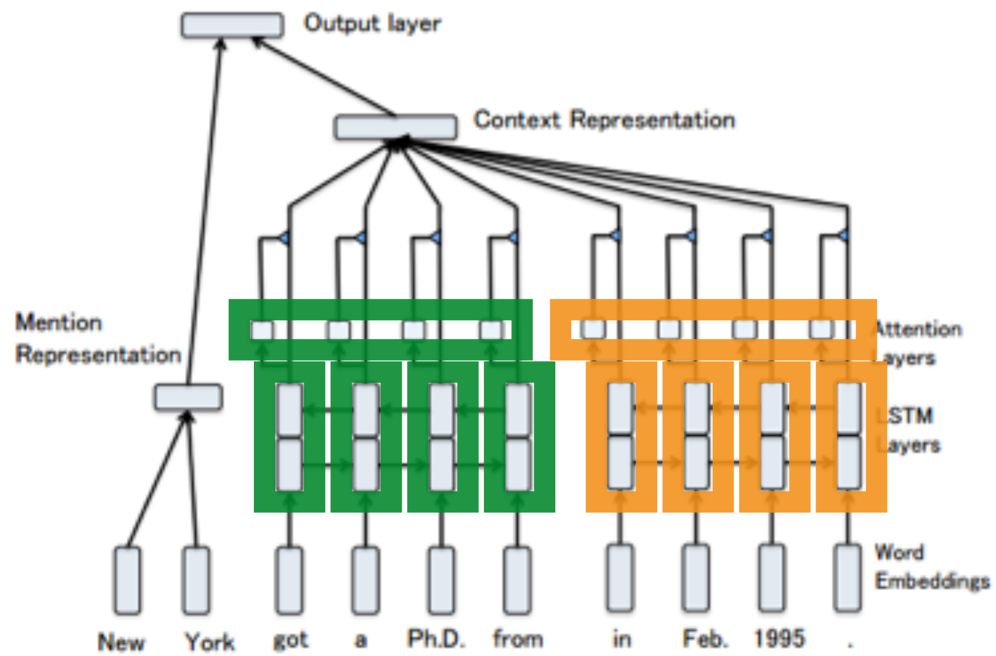


Figure 1: An illustration of our proposed model predicting fine-grained semantic types for the mention "New York" in the sentence "She got a Ph.D from New York in Feb. 1995.".

- For each of the intermediate context vector, compute an attention weight.

- The context representation is the weighted sum of the intermediate representations.

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

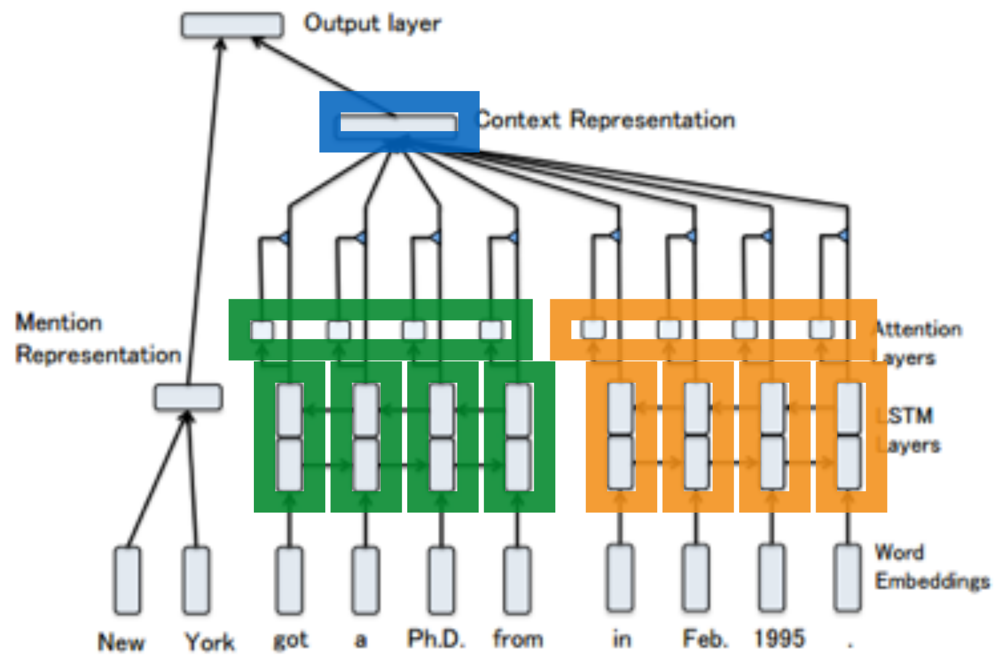


Figure 1: An illustration of our proposed model predicting fine-grained semantic types for the mention "New York" in the sentence "She got a Ph.D from New York in Feb. 1995.".

- For each of the intermediate context vector, compute an attention weight.
- The context representation is the weighted sum of the intermediate representations.

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

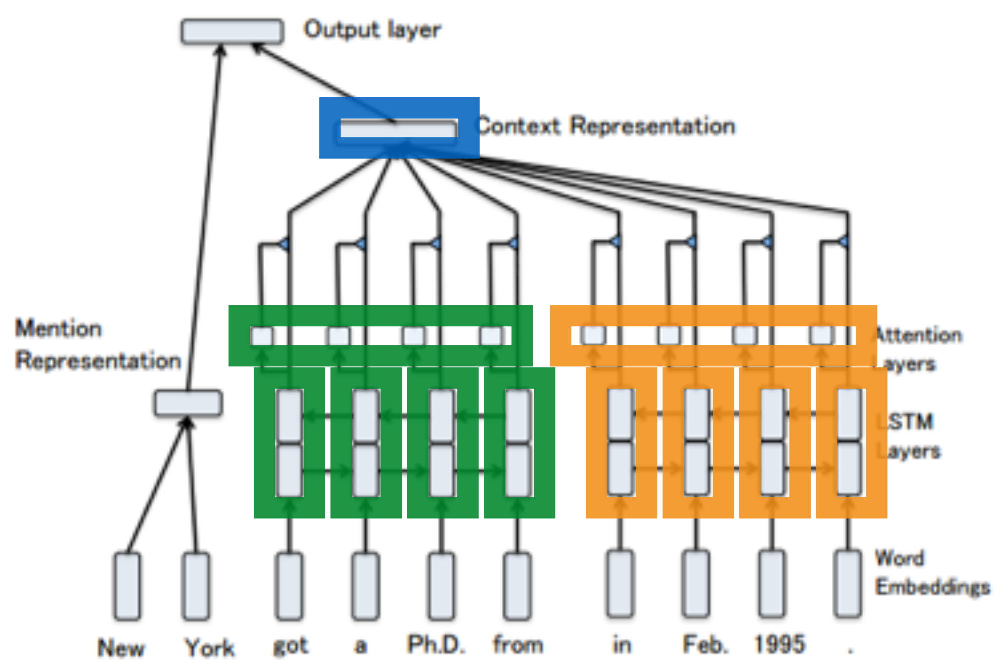
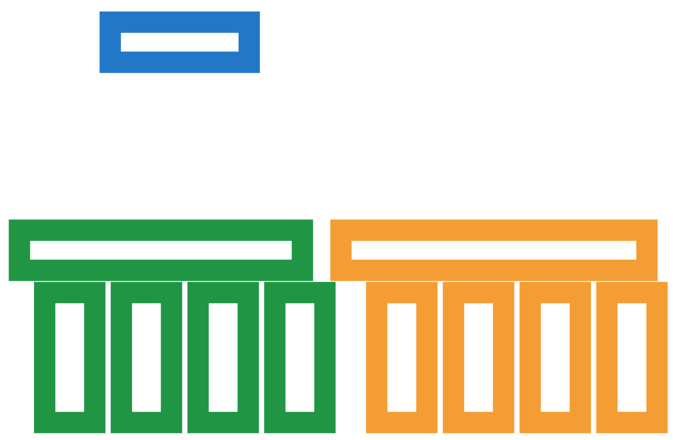


Figure 1: An illustration of our proposed model predicting fine-grained semantic types for the mention "New York" in the sentence "She got a Ph.D from New York in Feb. 1995.".

- For each of the intermediate context vector, compute an attention weight.
- The context representation is the weighted sum of the intermediate representations.
- The attention weights are computed using a 2 layer feed forward network.

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.



- For each of the intermediate context vector, compute an attention weight.
- The context representation is the weighted sum of the intermediate representations.
- The attention weights are computed using a 2 layer feed forward network.

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

$$
\begin{aligned}
e_i^l &= \tanh\left(W_e \begin{bmatrix} \overrightarrow{h_i^l} \\ \overleftarrow{h_i^l} \end{bmatrix}\right) && (7) \\
\tilde{a}_i^l &= \exp(W_a e_i^l) && (8) \\
a_i^l &= \frac{\tilde{a}_i^l}{\sum_{i=1}^{C} \tilde{a}_i^l + \tilde{a}_i^r} && (9) \\
v_c &= \sum_{i=1}^{C} a_i^l \begin{bmatrix} \overrightarrow{h_i^l} \\ \overleftarrow{h_i^l} \end{bmatrix} + a_i^r \begin{bmatrix} \overrightarrow{h_i^r} \\ \overleftarrow{h_i^r} \end{bmatrix} && (10)
\end{aligned}
$$

- For each of the intermediate context vector, compute an attention weight.
- The context representation is the weighted sum of the intermediate representations.
- The attention weights are computed using a 2 layer feed forward network.

Attentive Encoder

- The extend their LSTM model to incorporate attention over the context words.

$$e_i^l = \tanh\left(W_e \begin{bmatrix} \overrightarrow{h}_i^l \\ \overleftarrow{h}_i^l \end{bmatrix}\right) \quad (7)$$

$$\tilde{a}_i^l = \exp(W_a e_i^l) \quad (8)$$

$$a_i^l = \frac{\tilde{a}_i^l}{\sum_{i=1}^{C} \tilde{a}_i^l + \tilde{a}_i^r} \quad (9)$$

$$v_c = \sum_{i=1}^{C} a_i^l \begin{bmatrix} \overrightarrow{h}_i^l \\ \overleftarrow{h}_i^l \end{bmatrix} + a_i^r \begin{bmatrix} \overrightarrow{h}_i^r \\ \overleftarrow{h}_i^r \end{bmatrix} \quad (10)$$

Side Note: I am not
sure
why don't they
condition on the
mention embedding

- For each of the intermediate context vector, compute an attention weight.
- The context representation is the weighted sum of the intermediate representations.
- The attention weights are computed using a 2 layer feed forward network.

Training

Training

- Train in a multi-label setting. Each instance (mention, context) produces ‘K’ scores for each output types.

Training

- Train in a multi-label setting. Each instance (mention, context) produces ‘K’ scores for each output types.

$$\mathrm{y} = \sigma(\mathrm{W}[v_m;v_c]) \quad W \in \mathbb{R}^{K \times 2D} \quad v_m, v_c \in \mathbb{R}^D$$

Training

- Train in a multi-label setting. Each instance (mention, context) produces 'K' scores for each output types.

$$y = \sigma(\mathrm{W}[v_m;v_c]) \quad W \in \mathbb{R}^{K \times 2D} \quad v_m, v_c \in \mathbb{R}^D$$

- Loss is the usual cross entropy loss to maximize the likelihood of the training set.

$$L(y,t) = \sum_{k=1}^{K} -t_k \log(y_k) - (1-t_k)\log(1-y_k)$$

$$t \in \{0,1\}^K$$

Results

Results

Sentence	Prediction
... ... The film is a remake of [Secrets (1924)] , a silent film starring Norma Talmadge	/film 0.986 /art 0.982
The film is a remake of Secrets (1924) , a silent film starring [Norma Talmadge]	/person 0.999 /actor 0.987
... ... The festival brought together the foremost filmmakers , including Francois Truffaut , [Roman Polanski] , Robert Enrico , and others	/person 1.00 /director 0.963 /author 0.958 /artist 0.950 /actor 0.871
... ... Jim Hodges , the Democratic nominee , handily defeated Republican Governor [David Beasley] to become the 114th governor of South Carolina	/person 1.00 /politician 0.983
She is best known for roles in various TV Dramas and tokusatsu shows such as [Ultraseven X] and Kamen Rider Kiva	/broadcats_program 0.892

Figure 2: Examples of our model attending over contexts for a given mention.

Results

Sentence	Prediction
... ... The film is a remake of [Secrets (1924)] , a silent film starring Norma Talmadge	/film 0.986 /art 0.982
The film is a remake of Secrets (1924) , a silent film starring [Norma Talmadge]	/person 0.999 /actor 0.987
... ... The festival brought together the foremost filmmakers , including Francois Truffaut , [Roman Polanski] , Robert Enrico , and others	/person 1.00 /director 0.963 /author 0.958 /artist 0.950 /actor 0.871
... ... Jim Hodges , the Democratic nominee , handily defeated Republican Governor [David Beasley] to become the 114th governor of South Carolina	/person 1.00 /politician 0.983
She is best known for roles in various TV Dramas and tokusatsu shows such as [Ultraseven X] and Kamen Rider Kiva	/broadcats_program 0.892

Figure 2: Examples of our model attending over contexts for a given mention.

Models	Strict	Loose Macro	Loose Micro
Ling and Weld (2012)	52.30	69.90	69.30
Yogatama et al. (2015)	-	-	72.25
Averaging Encoder	51.89	72.24	68.65
LSTM Encoder	55.60	73.95	71.34
Attentive Encoder	**58.97**	**77.96**	**74.94**

Table 2: Strict, Loose Macro and Loose Micro F1-scores

Conclusion/Discussion

Conclusion/Discussion

- presented 3 recent work on ***context dependent*** fine grained entity typing.

Conclusion/Discussion

- presented 3 recent work on *context dependent* fine grained entity typing.
- The models range from sparse linear ones to dense low dimensional representation to the recent NN approaches.

Conclusion/Discussion

- presented 3 recent work on *context dependent* fine grained entity typing.
- The models range from sparse linear ones to dense low dimensional representation to the recent NN approaches.
- Open questions:

Conclusion/Discussion

- presented 3 recent work on *context dependent* fine grained entity typing.
- The models range from sparse linear ones to dense low dimensional representation to the recent NN approaches.
- Open questions:
    - Are the few hundred entity types enough?

Conclusion/Discussion

- presented 3 recent work on *context dependent* fine grained entity typing.
- The models range from sparse linear ones to dense low dimensional representation to the recent NN approaches.
- Open questions:
  - Are the few hundred entity types enough?
  - Do we need to follow any taxonomy at all? Why not learn a type representation for each entity.

Conclusion/Discussion

- presented 3 recent work on *context dependent* fine grained entity typing.
- The models range from sparse linear ones to dense low dimensional representation to the recent NN approaches.
- Open questions:
  - Are the few hundred entity types enough?
  - Do we need to follow any taxonomy at all? Why not learn a type representation for each entity.
  - Any other way of representing context?

References

Marta Recasens, Marie-Catherine de Marneffe, and Christopher Potts. 2013. The life and death of discourse entities: Identifying singleton mentions.

Limin Yao, Sebastian Riedel, and Andrew McCallum. 2010. Collective cross-document relation extraction without labelled data. In Proc. of EMNLP.

Thomas Lin, Mausam, and Oren Etzioni. 2012. No noun phrase left behind: Detecting and typing unlinkable entities.In Proc. of EMNLP.

Andrew Carlson, Justin Betteridge, Richard C. Wang, Estevam R. Hruschka Jr., and Tom M. Mitchell. 2010. Coupled semi-supervised learning for information extraction. In Proc. of WSDM.

Luciano Del Corro, Abdalghani Abujabal, Rainer Gemulla, and Gerhard Weikum. 2015. Finet: Context-aware fine-grained named entity typing. In EMNLP

Mohamed Amir Yosef, Sandro Bauer, Johannes Hoffart, Marc Spaniol, and Gerhard Weikum. 2012. Hyena: Hierarchical type classification for entity names In ACL

Xiao Ling and Daniel S Weld. 2012. Fine-grained entity recognition. In AAAI

Rajarshi Das, Arvind Neelakantan, David Belanger, Andrew McCallum, Incorporating Selectional Preferences in Multihop relation extraction. In NAACL, AKBC.