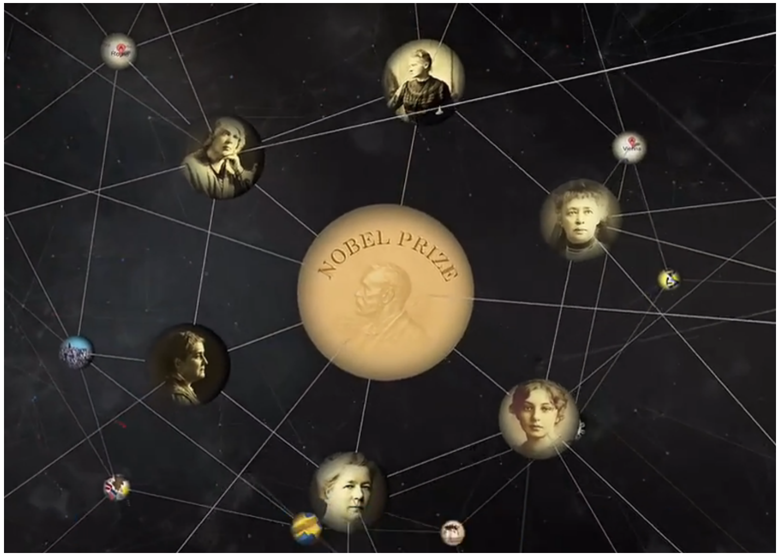
Knowledge Representation and Acquisition

Yankai Lin
Natural Language Processing Lab
linyk14@mails.tsinghua.edu.cn

Knowledge Graph

- Organize knowledge as a graph
  - node: entity
  - edge: relation
- Relation Facts
  - represent as triples (head, relation, tail)



Typical Knowledge Graph

Freebase™

DBpedia

yago
select knowledge

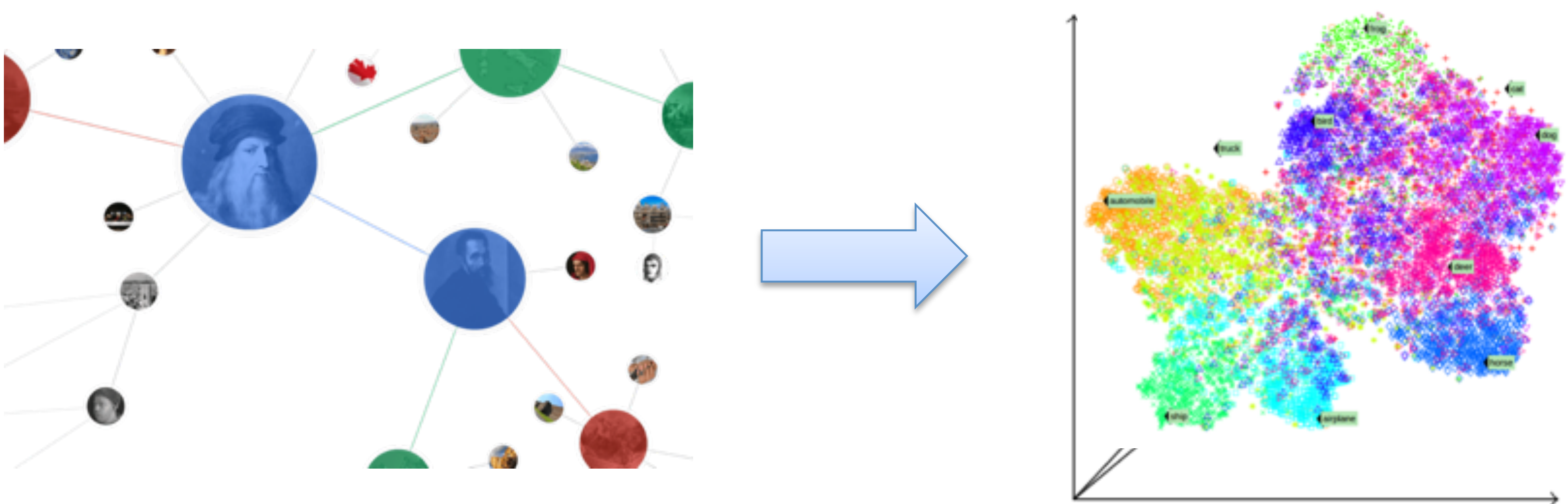
WordNet
A lexical database for English

Outline

- Knowledge Representation
- Knowledge Acquisition

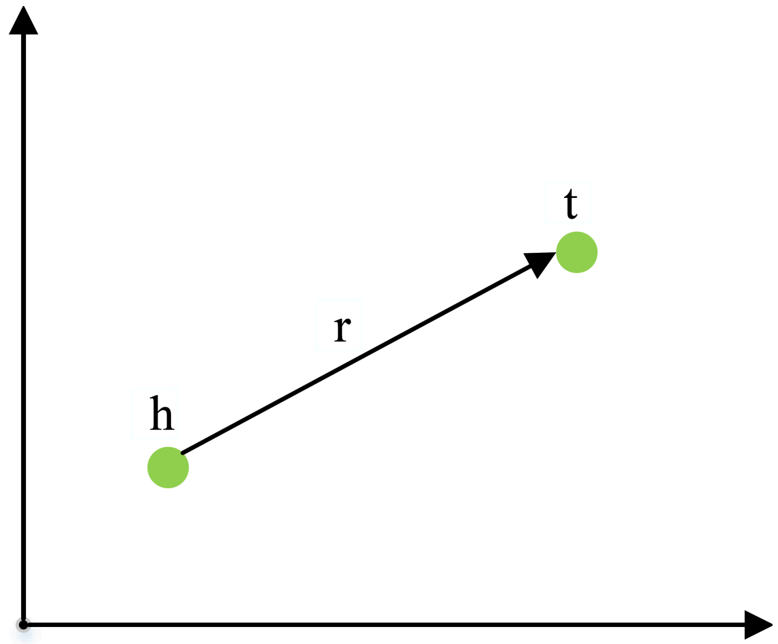
Knowledge Representation

- Traditional Knowledge Representation
  - Symbol-based Triples (such as RDF format)
  - Cannot capture the semantic relatedness between entities
- Solution: distributed knowledge representation



TransE

- Regard Relations as Translations between Entities



- Objective: **h + r = t**

Entity Prediction

WALL-E _has_genre ?

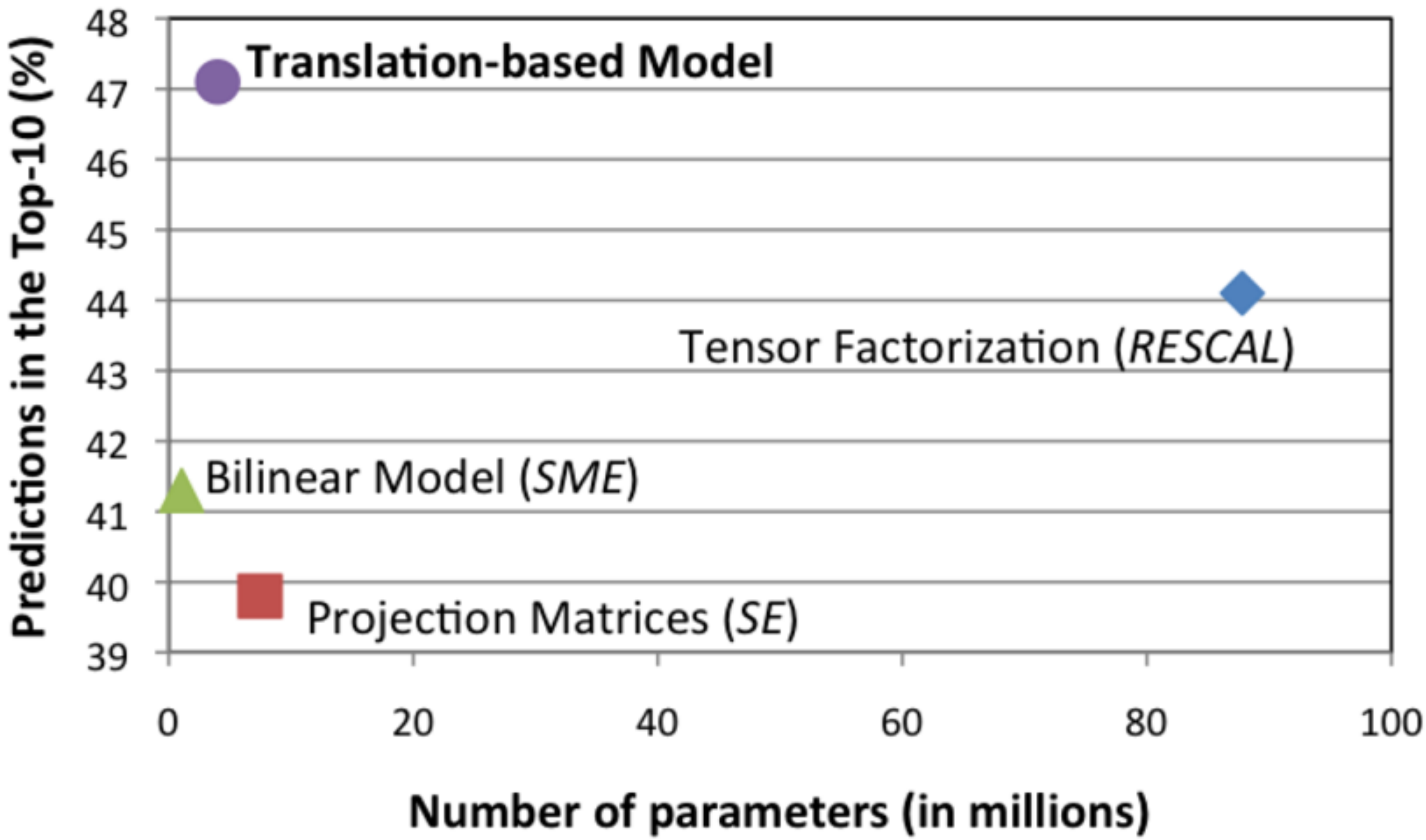
Entity Prediction

WALL-E _has_genre

Animation
Computer animation
Comedy film
Adventure film
Science Fiction
Fantasy
Stop motion
Satire
Drama
Connecting

Performance of Different Models

Freebase15K



Example of TransE

Entity	Tsinghua_University	A.C._Milan
1	University_of_Victoria	Inter_Milan
2	St._Stephen's_College,_Delhi	Celtic_F.C.
3	University_of_Ottawa	FC_Barcelona
4	University_of_British_Columbia	Genoa_C.F.C.
5	Peking_University	Udinese_Calcio
6	Utrecht_University	Real_Madrid_C.F.
7	Dalhousie_University	FC_Bayern_Munich
8	Brasenose_College,_Oxford	Bolton_Wanderers_F.C.
9	Cardiff_University	Borussia_Dortmund
10	Memorial_University_of_Newfoundland	Hertha_BSC_Berlin

Example of TransE

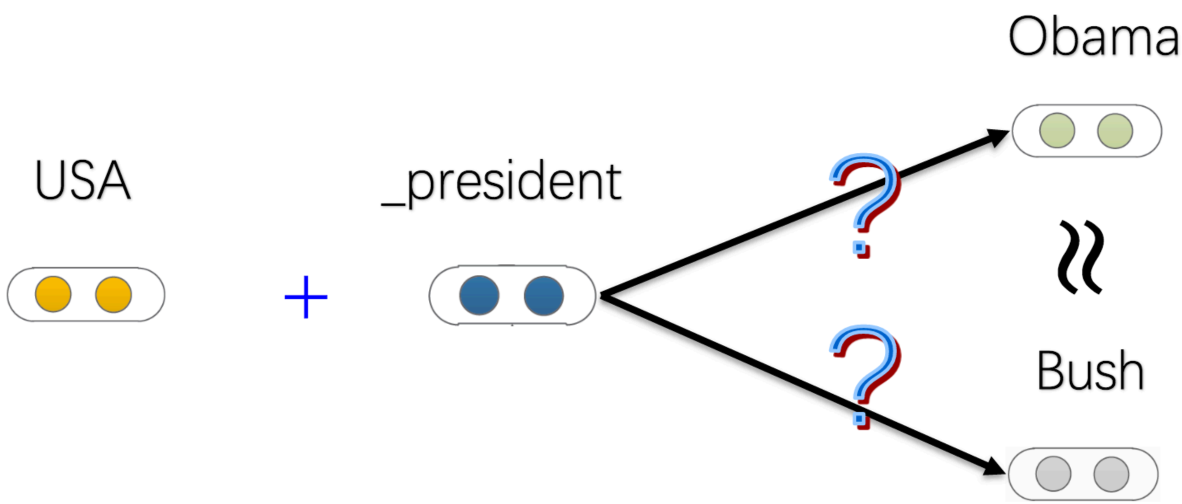
Head	China	Barack_Obama
Relation	**/location/location/adjoin**	**/education/education/institution**
1	Japan	Harvard_College
2	Taiwan	Massachusetts_Institute_of_Technology
3	Israel	American_University
4	South_Korea	University_of_Michigan
5	Argentina	Columbia_University
6	France	Princeton_University
7	Philippines	Emory_University
8	Hungary	Vanderbilt_University
9	North_Korea	University_of_Notre_Dame
10	Hong_Kong	Texas_A&M_University

Remaining Challenges

- Complex Relation Learning
- Relational Path Modeling

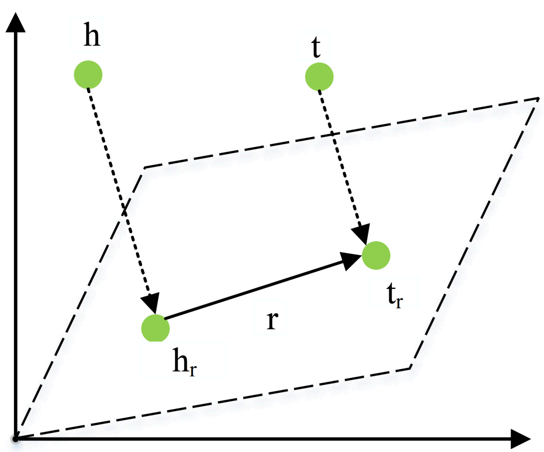
Complex Relation Learning

- 1-to-n, n-to-1 and n-to-n relations
  - (USA, _president, Obama)
  - (USA, _president, Bush)

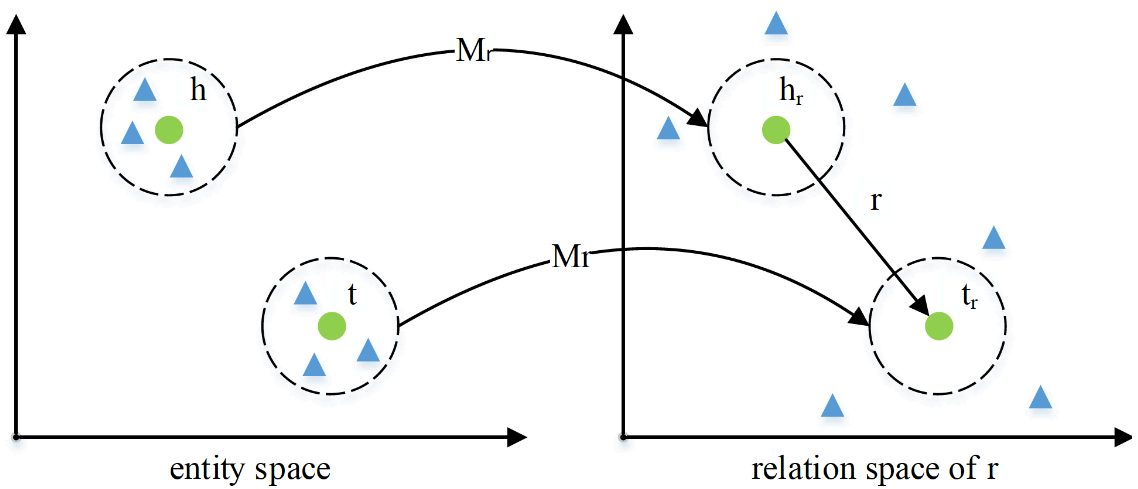


Complex Relation Learning

- Build relation-specific entity embeddings



TransH



TransR

Wang, et. al. (2014). Knowledge graph embedding by translating on hyperplanes. AAAI.
Lin, et. al.(2015). Learning entity and relation embeddings for knowledge graph completion. AAAI.

Entity Prediction

- Result of TransR

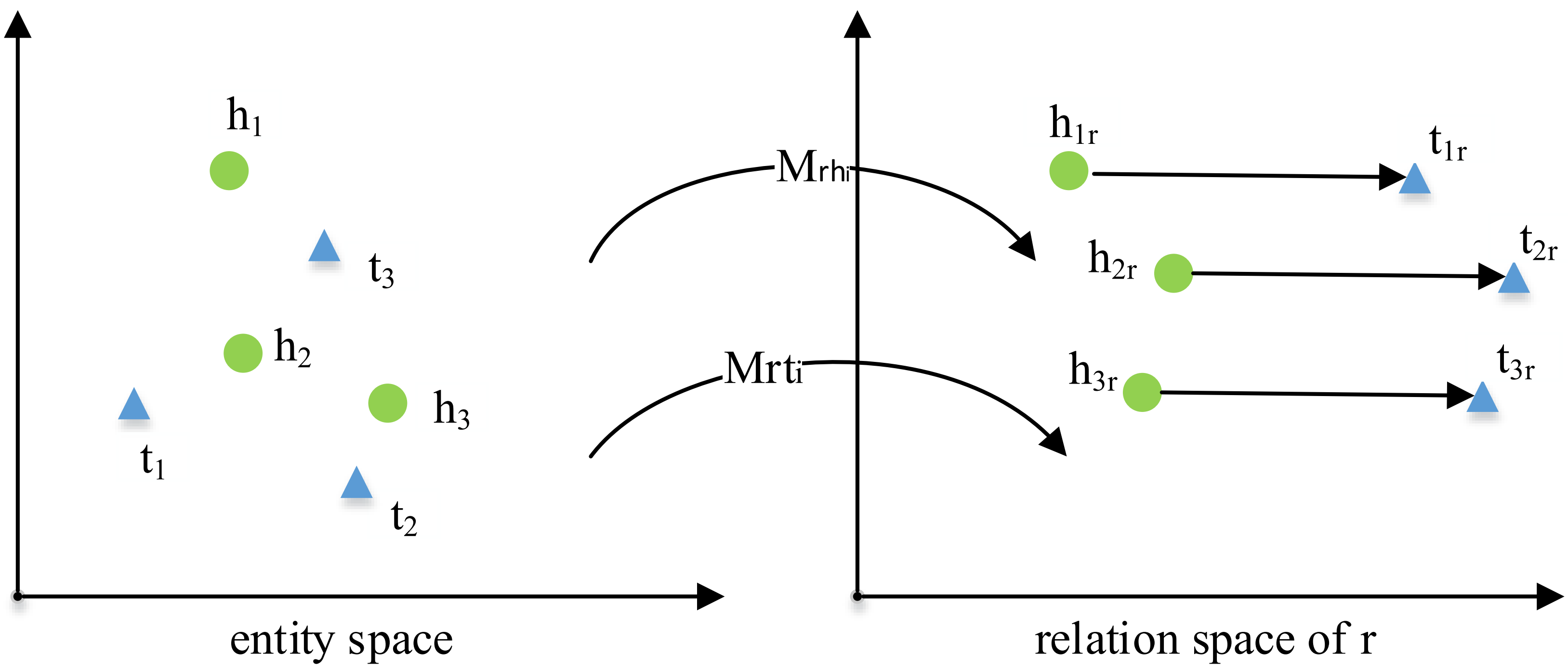
Data Sets	WN18				FB15K			
Metric	Mean Rank		Hits@10 (%)		Mean Rank		Hits@10 (%)	
	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter
Unstructured (Bordes et al. 2012)	315	304	35.3	38.2	1,074	979	4.5	6.3
RESCAL (Nickel, Tresp, and Kriegel 2011)	1,180	1,163	37.2	52.8	828	683	28.4	44.1
SE (Bordes et al. 2011)	1,011	985	68.5	80.5	273	162	28.8	39.8
SME (linear) (Bordes et al. 2012)	545	533	65.1	74.1	274	154	30.7	40.8
SME (bilinear) (Bordes et al. 2012)	526	509	54.7	61.3	284	158	31.3	41.3
LFM (Jenatton et al. 2012)	469	456	71.4	81.6	283	164	26.0	33.1
TransE (Bordes et al. 2013)	263	251	75.4	89.2	243	125	34.9	47.1
TransH (unif) (Wang et al. 2014)	318	303	75.4	86.7	211	84	42.5	58.5
TransH (bern) (Wang et al. 2014)	401	388	73.0	82.3	212	87	45.7	64.4
TransR (unif)	232	219	78.3	91.7	226	78	43.8	65.5
TransR (bern)	238	225	**79.8**	92.0	**198**	77	48.2	68.7
CTransR (unif)	243	230	78.9	**92.3**	233	82	44	66.3
CTransR (bern)	**231**	**218**	79.4	**92.3**	199	**75**	**48.4**	**70.2**

Example of TransR

Head Entity	Titanic		
Relation	/film/film/genre		
Model	TransE	TransH	TransR
1	War_film	Drama	Costume_drama
2	Period_piece	Romance_Film	Drama
3	Drama	Costume_drama	Romance_Film
4	History	Film_adaptation	Period_piece
5	Biography	Period_piece	Epic_film
6	Film_adaptation	Adventure_Film	Adventure_Film
7	Adventure_Film	LGBT	LGBT
8	Action_Film	Existentialism	Film_adaptation
9	Political_drama	Epic_film	Existentialism
10	Costume_drama	War_film	War_film

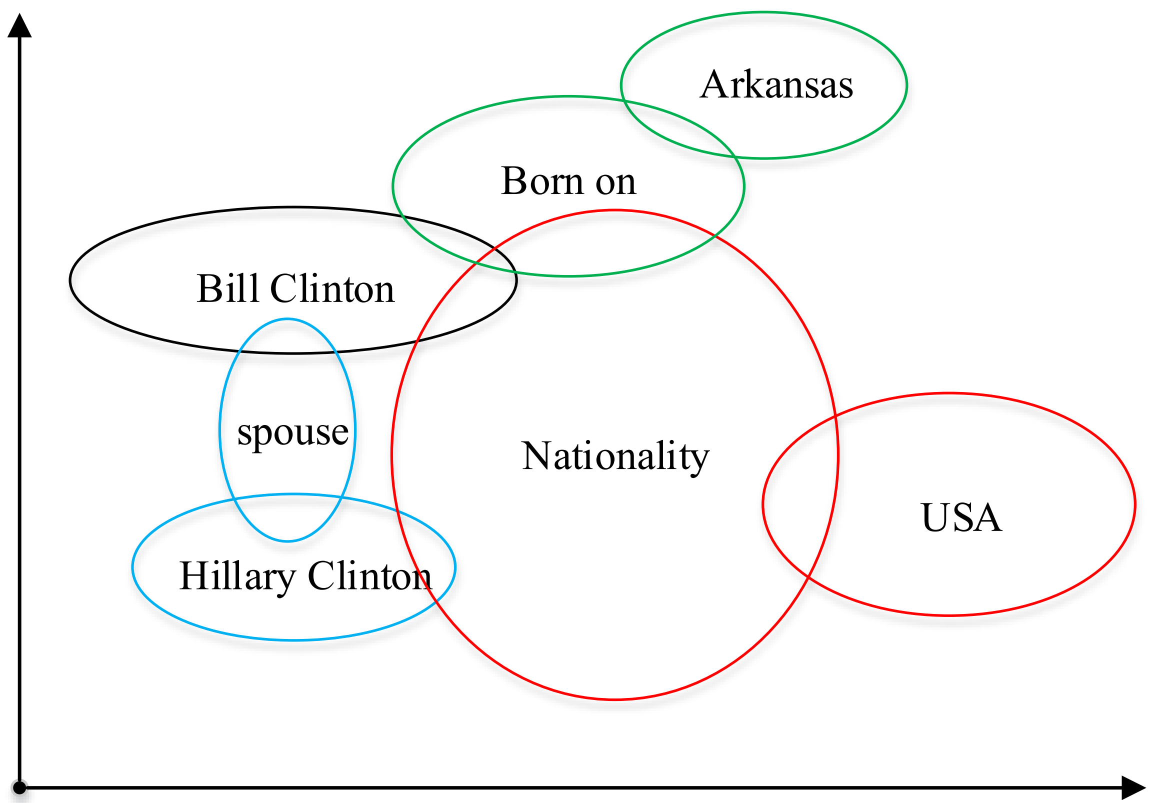
TransD

- Projection matrices related not only to relation but also head/tail entities



KG2E

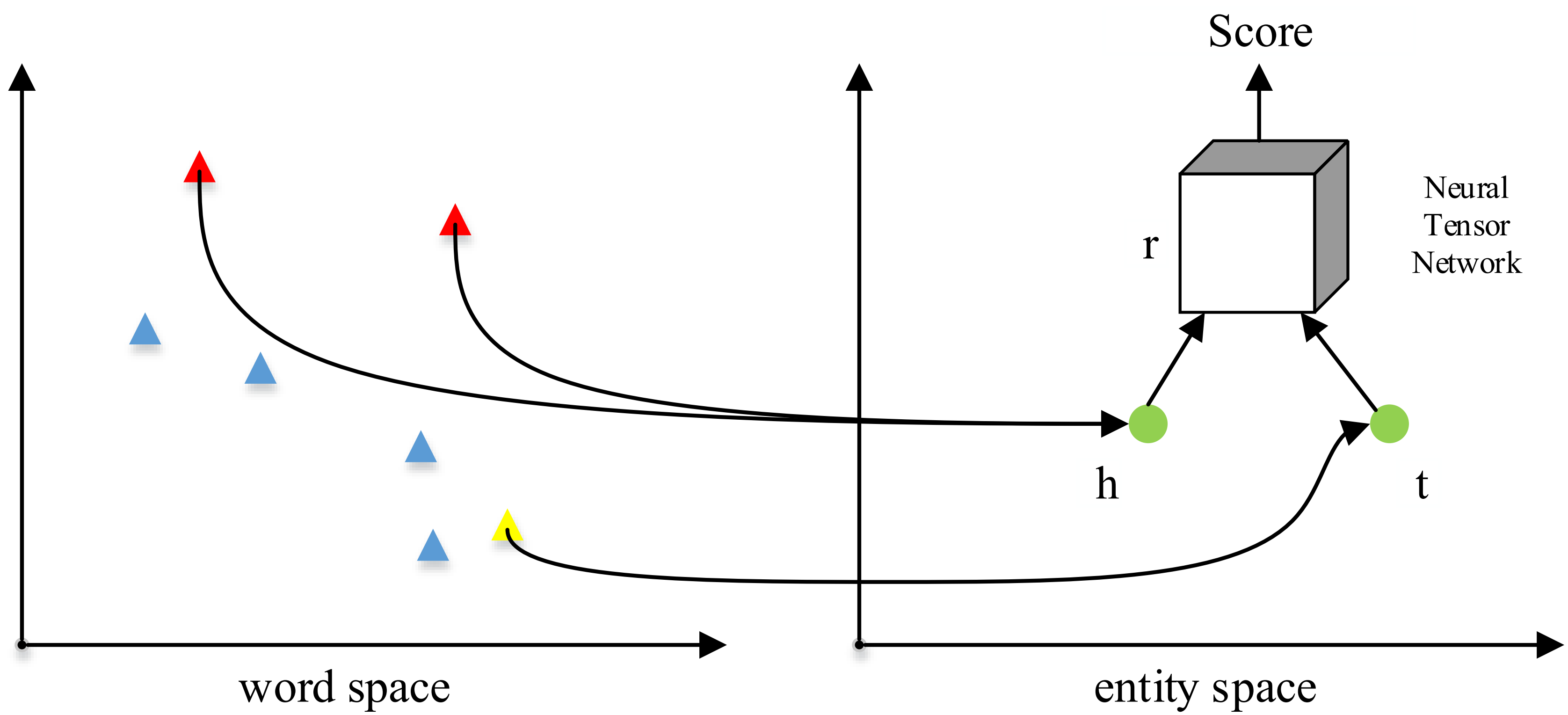
- Consider the (un)certainties of entities and relations
- Models relations/entities with Gaussian Distribution.



He,et al. (2015) Learning to Represent Knowledge Graphs with Gaussian Embedding. CIKM

NTN

- NTN models KG with a Neural Tensor Network and represents entities via word vectors.



Socher, et al. (2013) Reasoning with neural tensor networks for knowledge base completion. NIPS.

Other Models

- **TranSparse** uses sparse projection matrices to deal with the issue of entities and relations are heterogeneous and unbanlanced

- **Holographic Embeddings** (Hole) uses the circular correlation to combine the expressive power of the tensor product with the efficiency and simplicity of TransE.

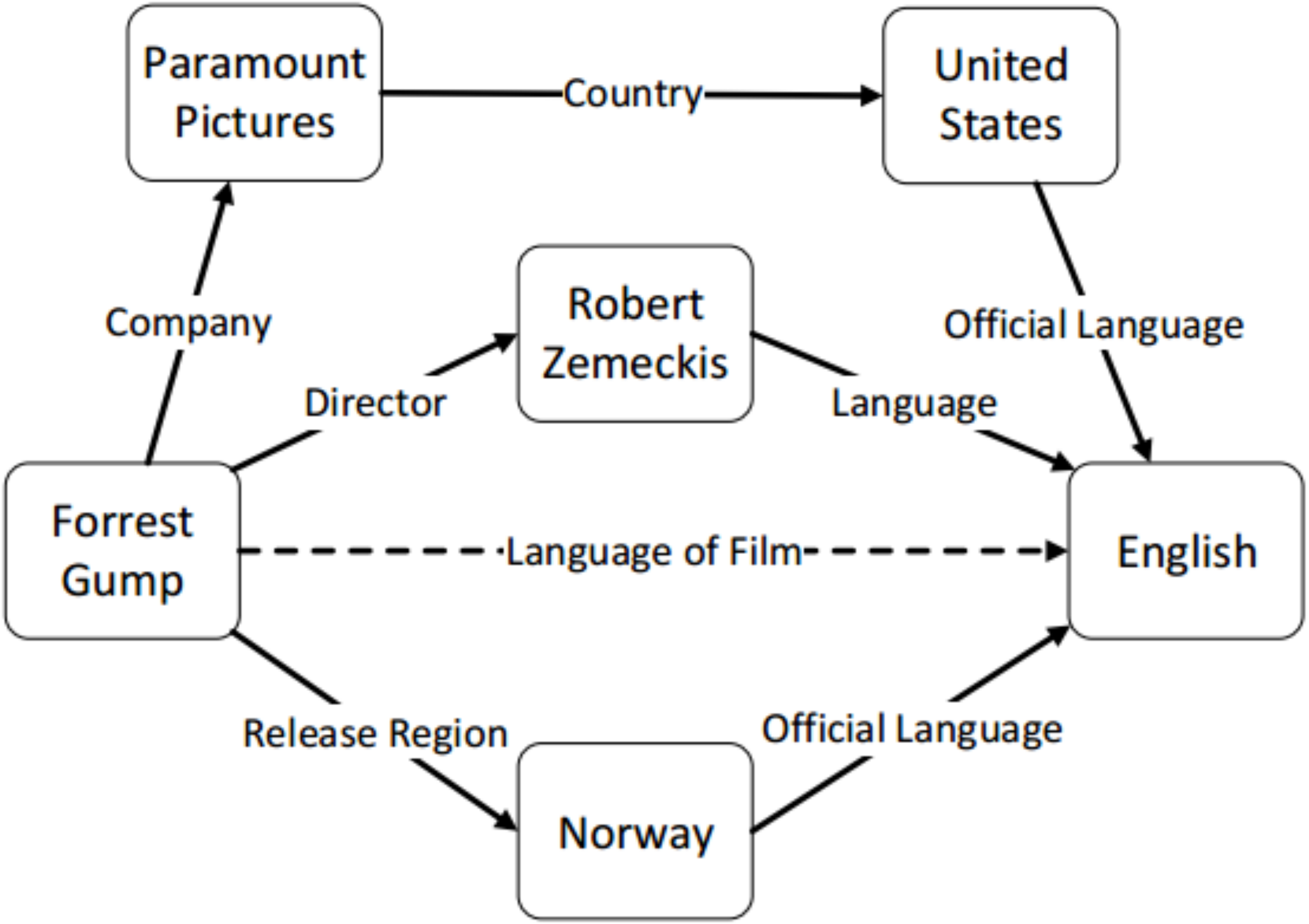
- **Complex Embeddings** employs eigenvalue decomposition model which makes use of complex valued embeddings.

Ji, et al. (2016) Knowledge Graph Completion with Adaptive Sparse Transfer Matrix. AAAI.
Nichkel, et al. (2015) Holographic Embeddings of Knowledge Graphs. Arxiv.
Trouillon, et al. (2016) Complex embeddings for simple link prediction. Arxir.

Remaining Challenges

- Complex Relation Learning
- Relational Path Modeling

Utilize Relational Path



Utilize Relational Path

- Path Ranking Algorithm

ID	PRA Path (Comment)
	athletePlaysForTeam
1	$c \xrightarrow{\text{athletePlaysInLeague}} c \xrightarrow{\text{leaguePlayers}} c \xrightarrow{\text{athletePlaysForTeam}} c$ (teams with many players in the athlete's league)
2	$c \xrightarrow{\text{athletePlaysInLeague}} c \xrightarrow{\text{leagueTeams}} c \xrightarrow{\text{teamAgainstTeam}} c$ (teams that play against many teams in the athlete's league)
	athletePlaysInLeague
3	$c \xrightarrow{\text{athletePlaysSport}} c \xrightarrow{\text{players}} c \xrightarrow{\text{athletePlaysInLeague}} c$ (the league that players of a certain sport belong to)
4	$c \xrightarrow{\text{isa}} c \xrightarrow{\text{isa}^{-1}} c \xrightarrow{\text{athletePlaysInLeague}} c$ (popular leagues with many players)
	athletePlaysSport
5	$c \xrightarrow{\text{isa}} c \xrightarrow{\text{isa}^{-1}} c \xrightarrow{\text{athletePlaysSport}} c$ (popular sports of all the athletes)
6	$c \xrightarrow{\text{athletePlaysInLeague}} c \xrightarrow{\text{superpartOfOrganization}} c \xrightarrow{\text{teamPlaysSport}} c$ (popular sports of a certain league)
	stadiumLocatedInCity
7	$c \xrightarrow{\text{stadiumHomeTeam}} c \xrightarrow{\text{teamHomeStadium}} c \xrightarrow{\text{stadiumLocatedInCity}} c$ (city of the stadium with the same team)
8	$c \xrightarrow{\text{latitudeLongitude}} c \xrightarrow{\text{latitudeLongitudeOf}} c \xrightarrow{\text{stadiumLocatedInCity}} c$ (city of the stadium with the same location)
	teamHomeStadium
9	$c \xrightarrow{\text{teamPlaysInCity}} c \xrightarrow{\text{cityStadiums}} c$ (stadiums located in the same city with the query team)
10	$c \xrightarrow{\text{teamMember}} c \xrightarrow{\text{athletePlaysForTeam}} c \xrightarrow{\text{teamHomeStadium}} c$ (home stadium of teams which share players with the query)
	teamPlaysInCity
11	$c \xrightarrow{\text{teamHomeStadium}} c \xrightarrow{\text{stadiumLocatedInCity}} c$ (city of the team's home stadium)
12	$c \xrightarrow{\text{teamHomeStadium}} c \xrightarrow{\text{stadiumHomeTeam}} c \xrightarrow{\text{teamPlaysInCity}} c$ (city of teams with the same home stadium as the query)
	teamPlaysInLeague
13	$c \xrightarrow{\text{teamPlaysSport}} c \xrightarrow{\text{players}} c \xrightarrow{\text{athletePlaysInLeague}} c$ (the league that the query team's members belong to)
14	$c \xrightarrow{\text{teamPlaysAgainstTeam}} c \xrightarrow{\text{teamPlaysInLeague}} c$ (the league that the query team's competing team belongs to)
	teamPlaysSport
15	$c \xrightarrow{\text{isa}} c \xrightarrow{\text{isa}^{-1}} c \xrightarrow{\text{teamPlaysSport}} c$ (sports played by many teams)
16	$c \xrightarrow{\text{teamPlaysInLeague}} c \xrightarrow{\text{leagueTeams}} c \xrightarrow{\text{teamPlaysSport}} c$ (the sport played by other teams in the league)

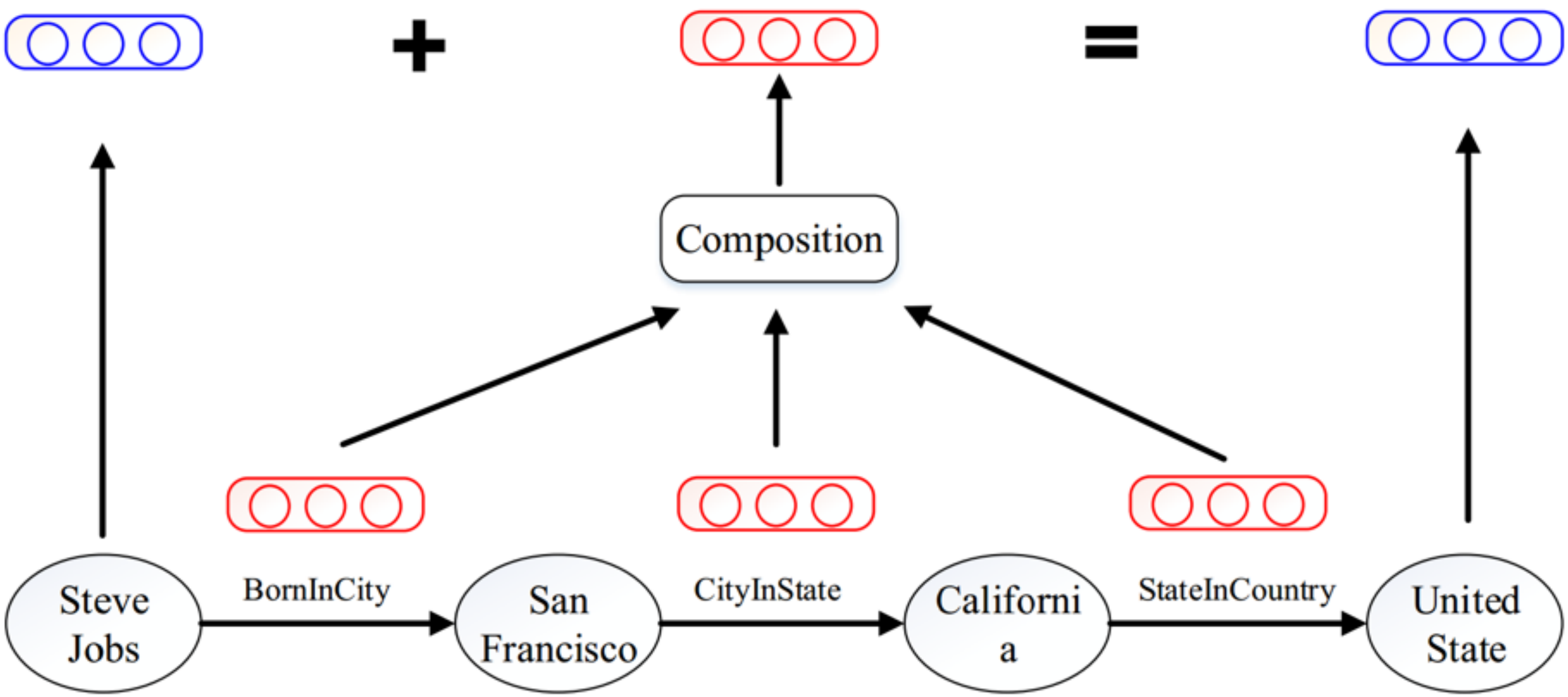
Lao, et al. (2011). Random walk inference and learning in a large scale knowledge base. EMNLP.

PTransE：Path-based TransE

	TransE	PTransE
KB	$h \xrightarrow{r} t$	$h \xrightarrow{r_1} e_1 \xrightarrow{r_2} t$
Triples	(h, r, t)	(h, r_1, e_1) (e_1, r_2, t) $(h, r_1 \circ r_2, t)$
Objectives	$\mathbf{h} + \mathbf{r} = \mathbf{t}$	$\mathbf{h} + \mathbf{r}_1 = \mathbf{e}_1$ $\mathbf{e}_1 + \mathbf{r}_2 = \mathbf{t}$ $\mathbf{h} + (\mathbf{r}_1 \circ \mathbf{r}_2) = \mathbf{t}$

Lin, et al. (2015). Modeling Relation Paths for Representation Learning of Knowledge Bases. EMNLP.

PTransE: Path-based TransE



Entity Prediction

Metric	Mean Rank		Hits@10 (%)	
	Raw	Filter	Raw	Filter
RESCAL	828	683	28.4	44.1
SE	273	162	28.8	39.8
SME (linear)	274	154	30.7	40.8
SME (bilinear)	284	158	31.3	41.3
LFM	283	164	26.0	33.1
TransE	243	125	34.9	47.1
TransH	212	87	45.7	64.4
TransR	**198**	77	48.2	68.7
TransE (Our)	205	63	47.9	70.2
PTransE (ADD, 2-step)	200	**54**	**51.8**	83.4
PTransE (MUL, 2-step)	216	67	47.4	77.7
PTransE (RNN, 2-step)	242	92	50.6	82.2
PTransE (ADD, 3-step)	207	58	51.4	**84.6**

+35%

Relation Prediction

Metric	Mean Rank		Hits@1 (%)	
	Raw	Filter	Raw	Filter
TransE	2.8	2.5	65.1	$\boxed{84.3}$
+Rev	2.6	2.3	67.1	86.7
+Rev+Path	2.4	1.9	65.2	89.0
PTransE (ADD, 2-step)	**1.7**	**1.2**	69.5	93.6
-TransE	135.8	135.3	51.4	78.0
-Path	2.0	1.6	**69.7**	89.0
PTransE (MUL, 2-step)	2.5	2.0	66.3	89.0
PTransE (RNN, 2-step)	1.9	1.4	68.3	93.2
PTransE (ADD, 3-step)	1.8	1.4	68.5	$\boxed{\mathbf{94.0}}$

+10%

Example of PTransE

Head Entity	Barack_Obama	
Relation	/education/education/institution	
Model	TransE	PTransE
1	Harvard_College	Columbia_University
2	Massachusetts_Institute_of_Technology	Occidental_College
3	American_University	Punahou_School
4	University_of_Michigan	University_of_Chicago
5	Columbia_University	Stanford_University
6	Princeton_University	Princeton_University
7	Emory_University	University_of_Pennsylvania
8	Vanderbilt_University	University_of_Virginia
9	University_of_Notre_Dame	University_of_Michigan
10	Texas_A&M_University	Yale_University

Example of PTransE

Head Entity	Stanford_University	
Relation	/education/educational_institution/students_graduates	
Model	TransE	PTransE
1	Steven_Spielberg	Raymond_Burr
2	Ron_Howard	Ted_Danson
3	Stan_Lee	Delmer_Daves
4	Barack_Obama	D.W._Moffett
5	Milton_Friedman	Gale_Anne_Hurd
6	Walter_F._Parkes	Jack_Palance
7	Michael_Cimino	Kal_Penn
8	Gale_Anne_Hurd	Kurtwood_Smith
9	Bryan_Singer	Alexander_Payne
10	Aaron_Sorkin	Richard_D._Zanuck

Other Challenges

- Utilize Multi-source Information
  - Textual Information
  - Visual Information
  - Type Information

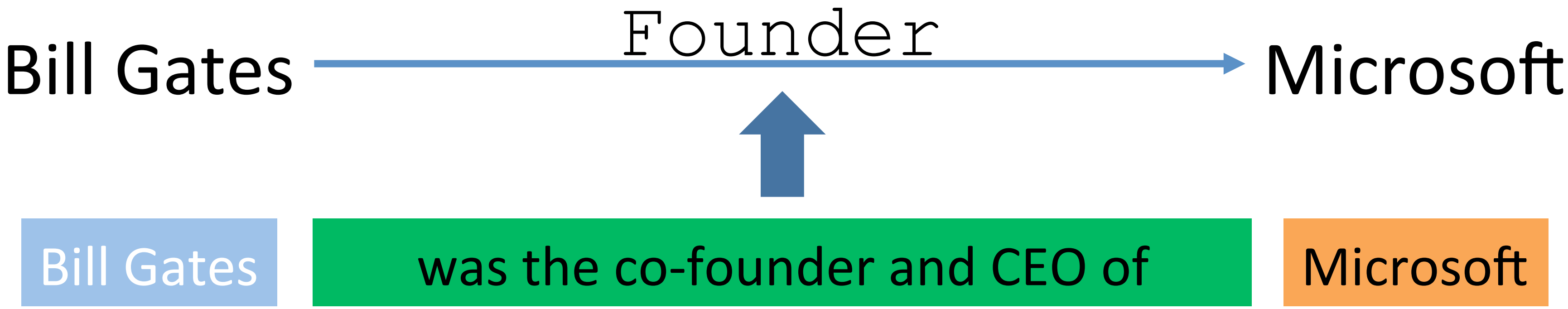
- Consider Logic Rules
  - Implication
  - Inference

Outline

- Knowledge Representation
- Knowledge Acquisition

Relation Extraction

- Extract Relational Facts from plain texts



Remaining Challenge

- Lack of Labeling Data

- Utilize Multi-lingual Data

Distant Supervised Relation Extraction

- Wrong Label Issue

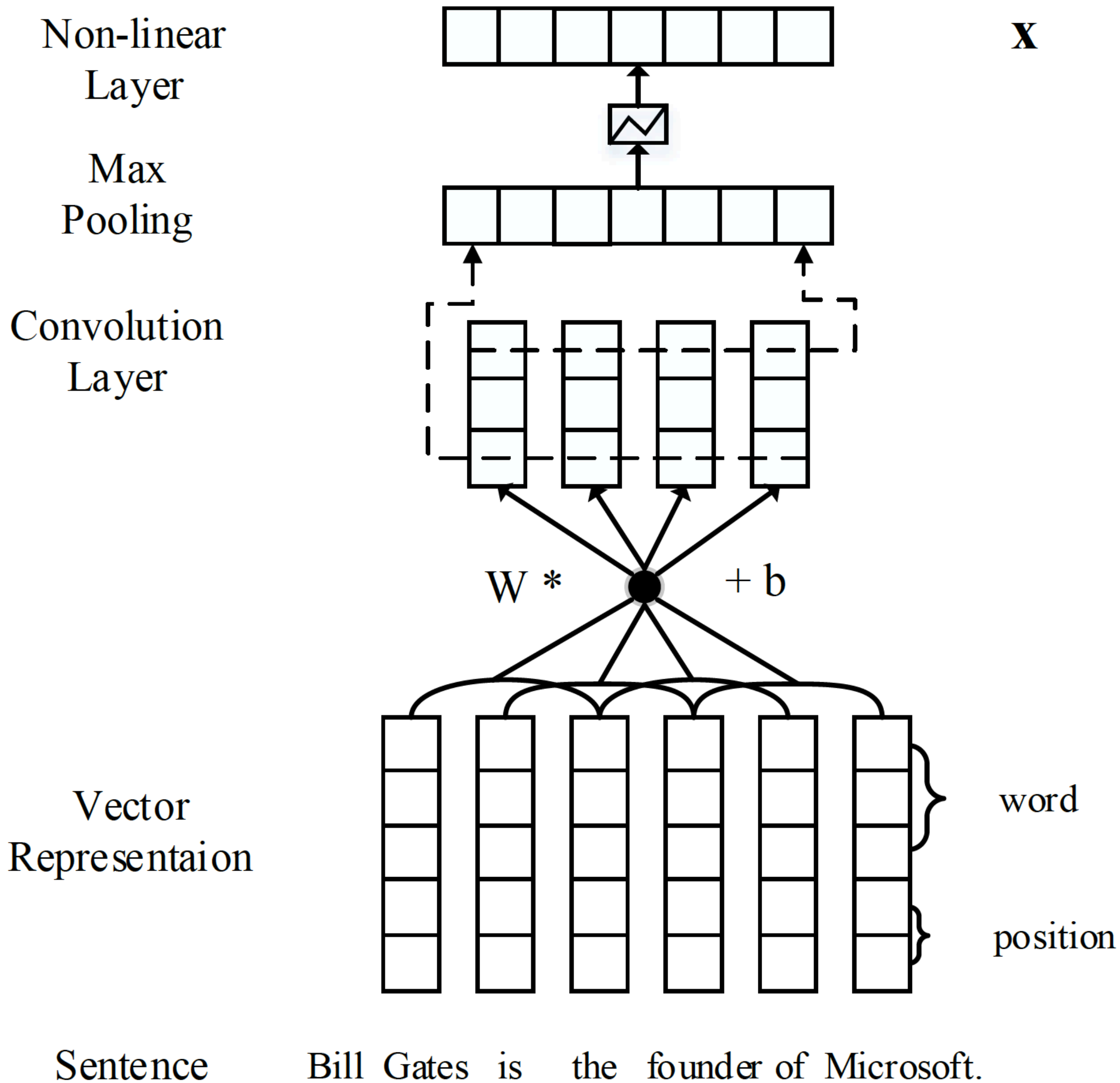
Bill Gates — Founder — Microsoft

Bill Gates was the co-founder and CEO of Microsoft

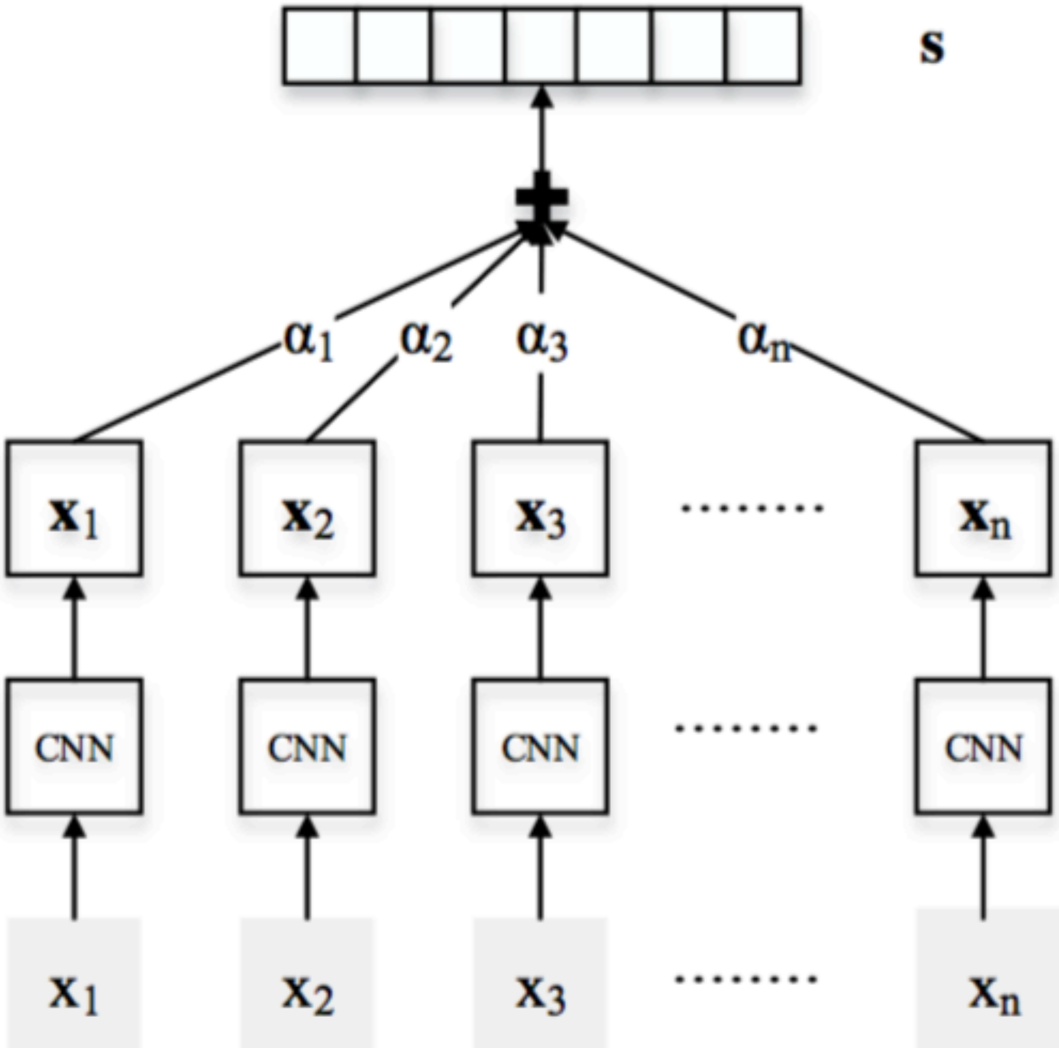
Bill Gates announced to retire from Microsoft

Bill Gates and Paul Allen co-founded the IT giant Microsoft

Sentence Encoder



Sentence-Level Selective Attention

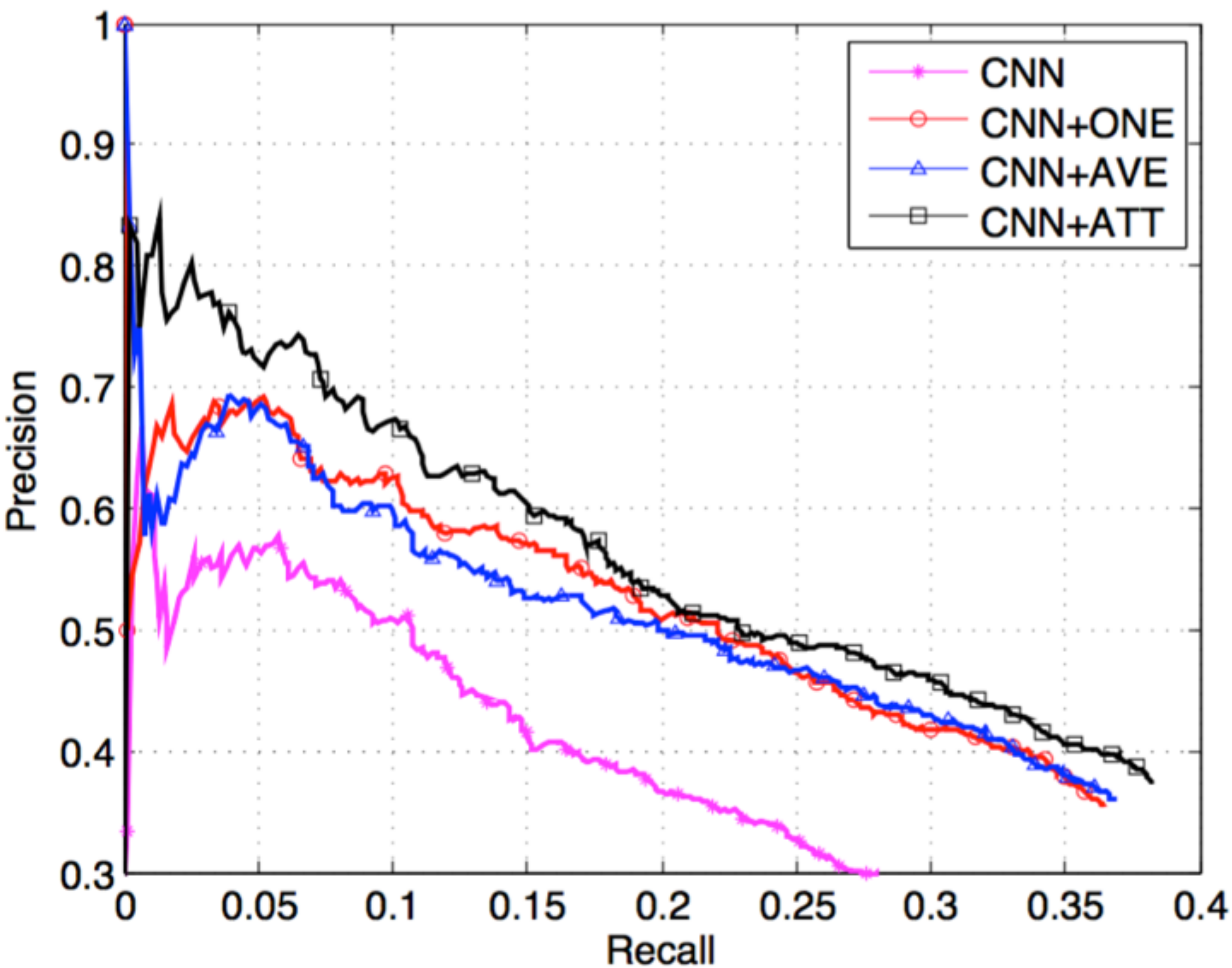


$$\alpha_i = \frac{\exp(e_i)}{\sum_k \exp(e_k)}$$

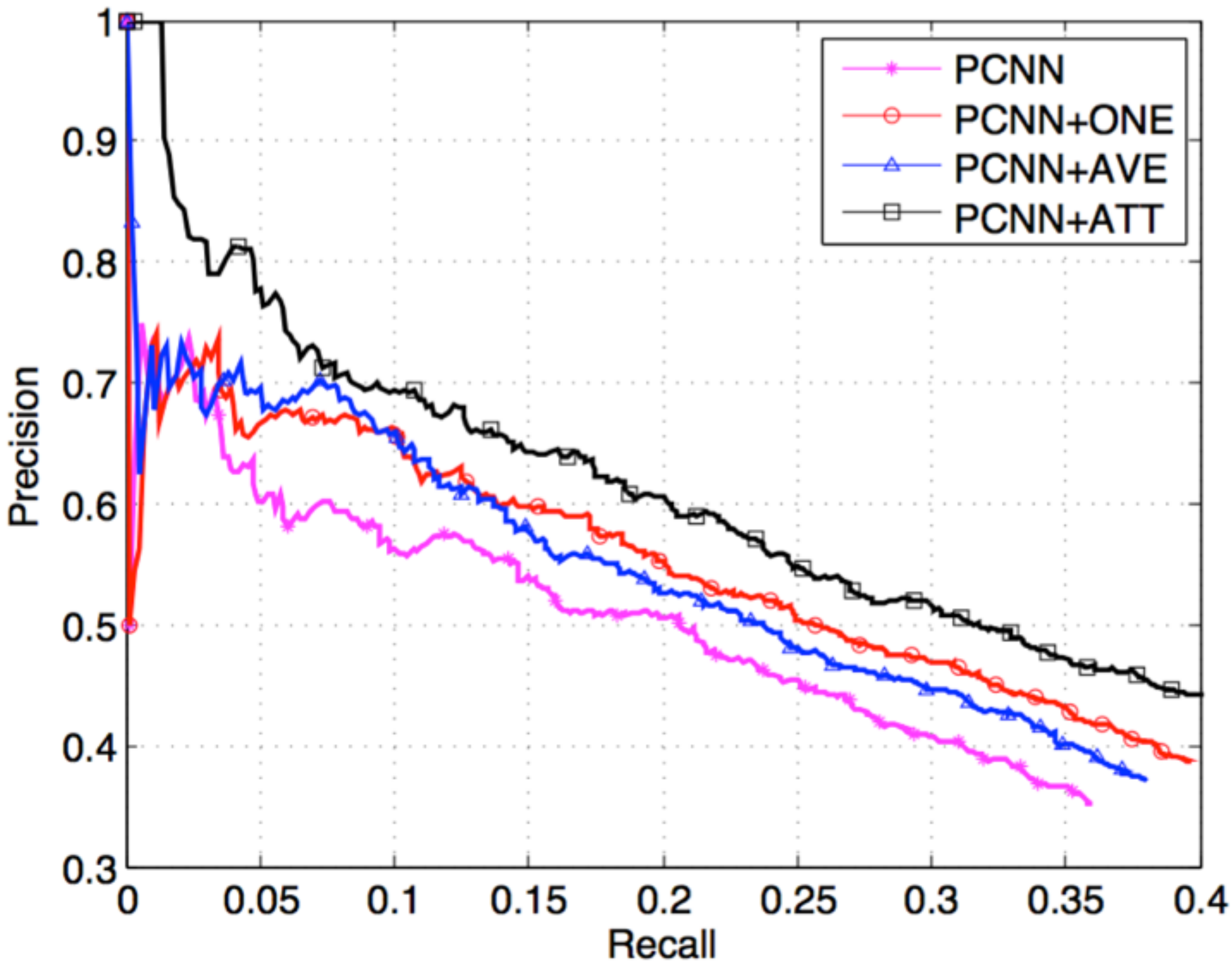
$$e_i = \mathbf{x}_i \mathbf{A} \mathbf{r}$$

Lin, et al. (2016). Neural Relation Extraction with Selective Attention over Instances. ACL.

Effect of Selective Attention



Effect of Selective Attention



Zeng, et al. (2015). Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks. EMNLP.

Effect of Sentence Number

- Setting
  - One
  - Two
  - All

Setting	One				Two				All			
P@N(%)	100	200	300	Ave	100	200	300	Ave	100	200	300	Ave
CNN+One	68.0	60.7	53.8	60.9	70.0	62.7	55.8	62.9	67.0	64.7	58.1	63.4
+Two	75.0	67.2	58.8	67.1	69.0	63.2	60.5	64.0	64.0	60.2	60.1	60.4
+All	**76.0**	65.2	**60.8**	67.4	76.0	65.7	62.1	68.0	76.0	68.6	59.8	68.2
PCNN+One	73.0	64.8	56.8	65.0	70.0	67.2	63.1	66.9	72.0	69.7	64.1	68.7
+Two	71.0	63.7	57.8	64.3	73.0	65.2	62.1	66.9	73.0	66.7	62.8	67.6
+All	73.0	**69.2**	**60.8**	**67.8**	**77.0**	**71.6**	**66.1**	**71.6**	76.0	73.1	67.4	72.2

Case Study

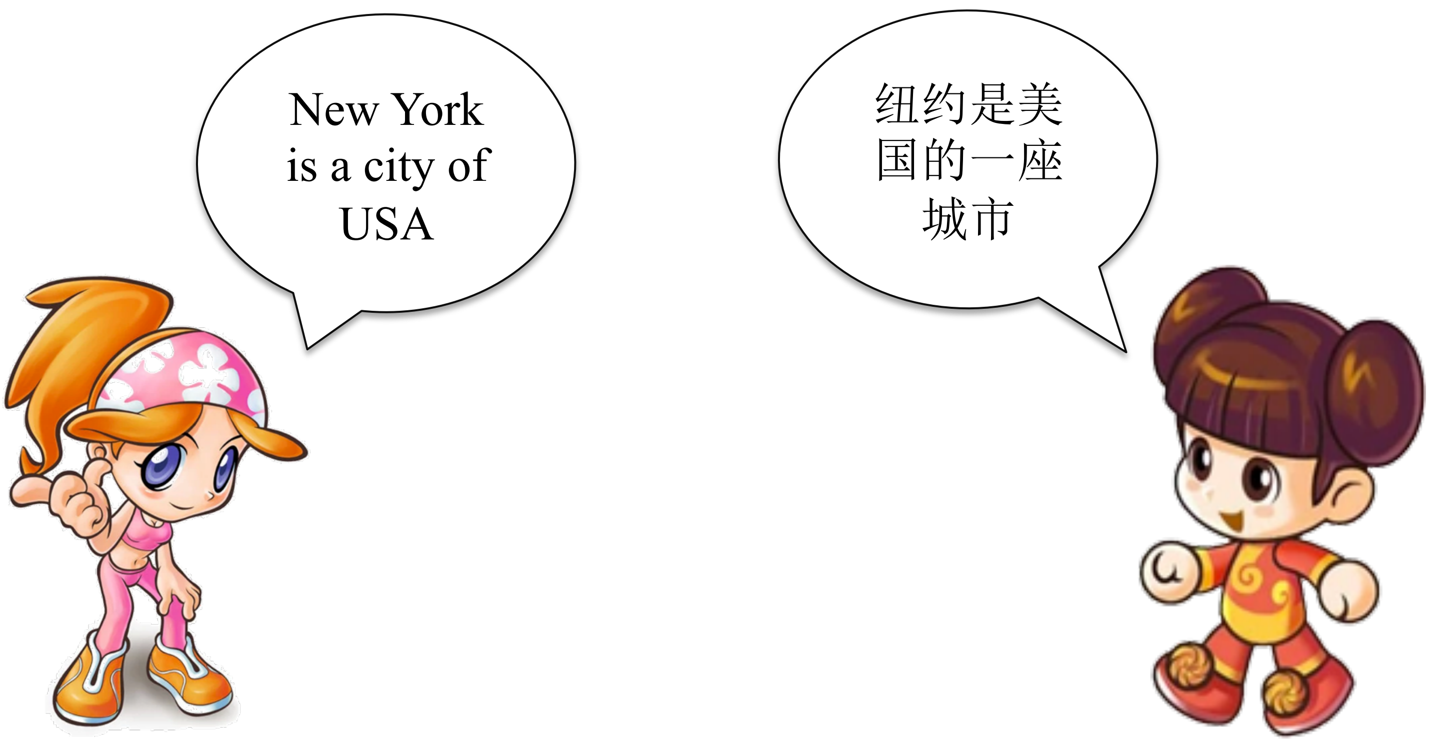
Relation	employer_of
Bad	When Howard Stern was preparing to take his talk show to **Sirius Satellite Radio**, following his former boss, **Mel Karmazin**, Mr. Hollander argued that ...
Good	**Mel Karmazin**, the chief executive of **Sirius Satellite Radio**, made a lot of phone calls ...
Relation	**place_of_birth**
Bad	**Ernst Haefliger**, a Swiss tenor who ... roles , died on Saturday in **Davos**, Switzerland, where he maintained a second home
Good	**Ernst Haefliger** was born in **Davos** on July 6, 1919, and studied at the Wettinger Seminary ...

Remaining Challenge

- Lack of Labeling Data
- Utilize Multi-lingual Data

Multi-lingual Relation Extraction

- Only consider mono-lingual data → People speaking different languages also share similar knowledge



Lin, et al. (2017). Neural Relation Extraction with Multi-lingual Attention. ACL.

Utilize Multi-lingual Data

- Mono-lingual RE for each languages

- Multi-lingual RE

Which one
is better?

Consistency

- Half of Chinese and English sentences are longer than 20 words

- Relation: City of
  - New York is a city in the northeastern United States.
  - 纽约位于美国纽约州东南部大西洋沿岸，是美国第一大城市及第一大港.
  - 纽约是美国人口最多的城市.

- Advantage: patterns expressing relations consist among languages

Complementarity

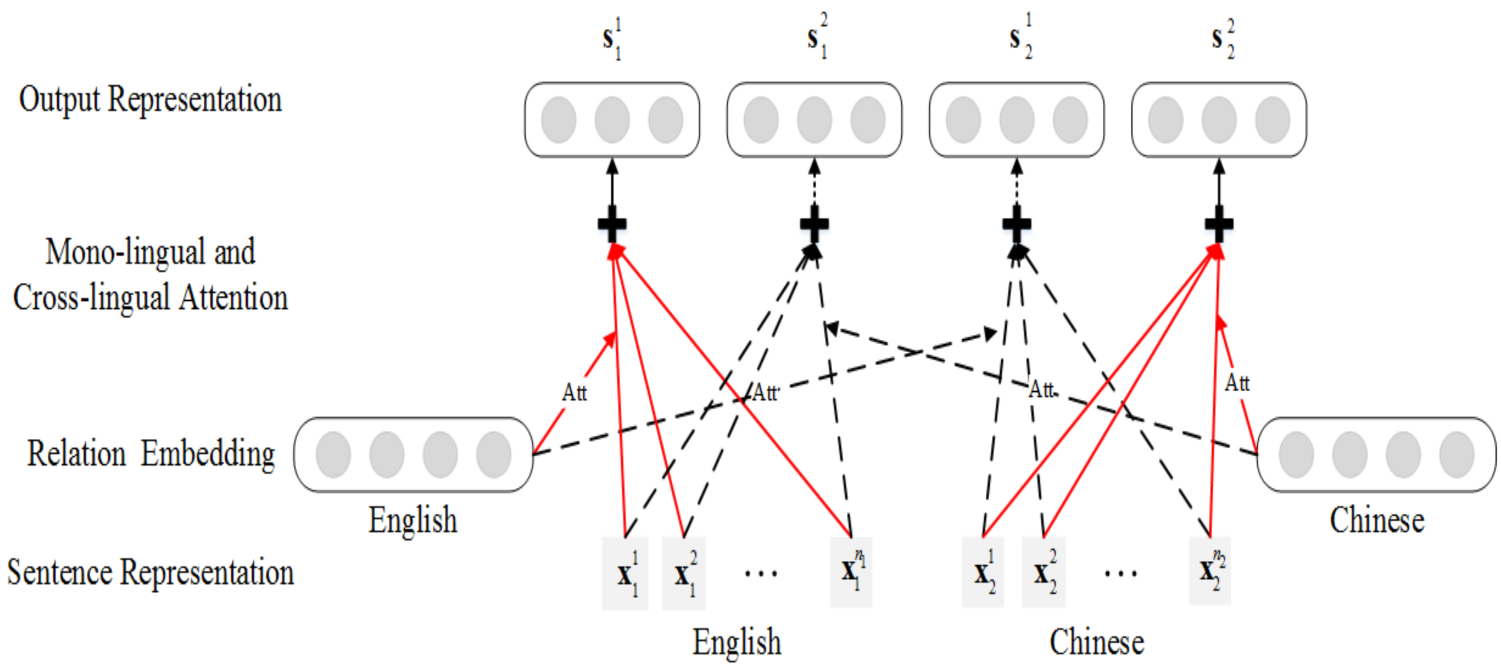
- Unique relational facts
  - 42.2% in English data
  - 41.6% in Chinese data
- The number of sentences expressing relational facts varies a lot in half of relations
- Advantage: texts in different languages can be complementary to each other

Methodology

- Sentence Encoder
- Multi-lingual Attention
- Relation extractor

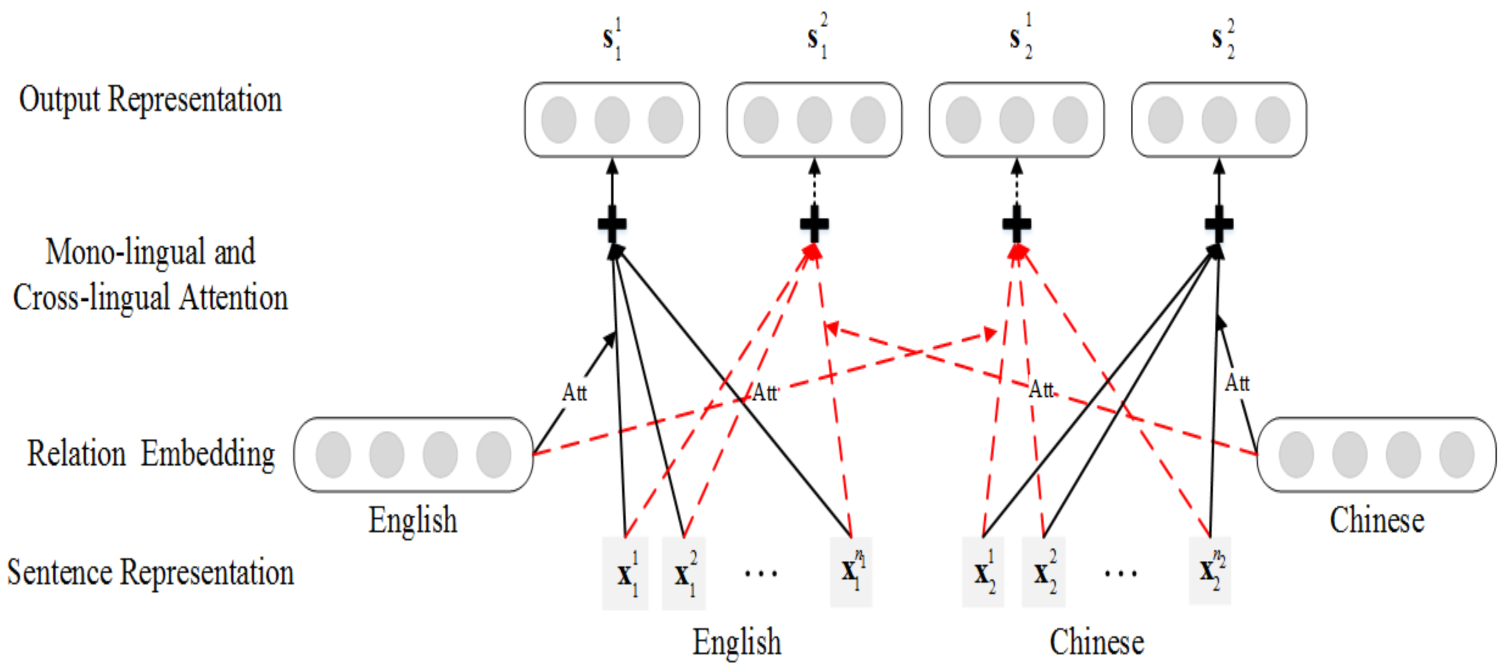
Multi-lingual Attention

- Mono-lingual Attention
- Cross-lingual Attention



Multi-lingual Attention

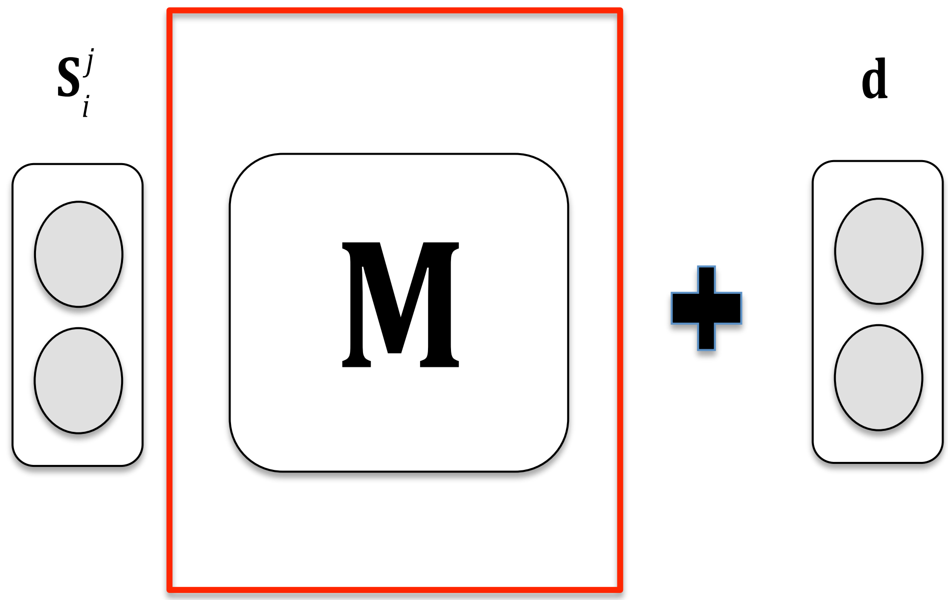
- Mono-lingual Attention
- Cross-lingual Attention



Relation Extractor

- Mono-lingual

$$p(r \mid \mathbf{S}_i^j, \theta) =$$

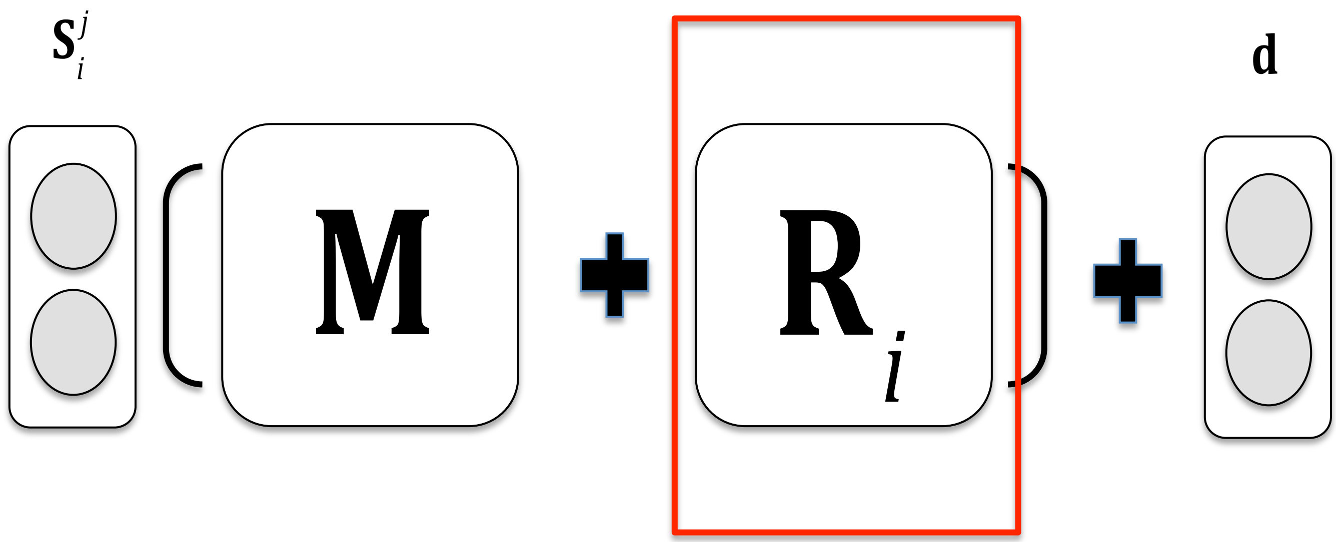


- Global relation matrix

Relation Extractor

- Multi-lingual

$$p(r \mid \mathbf{S}_i^j, \theta) =$$



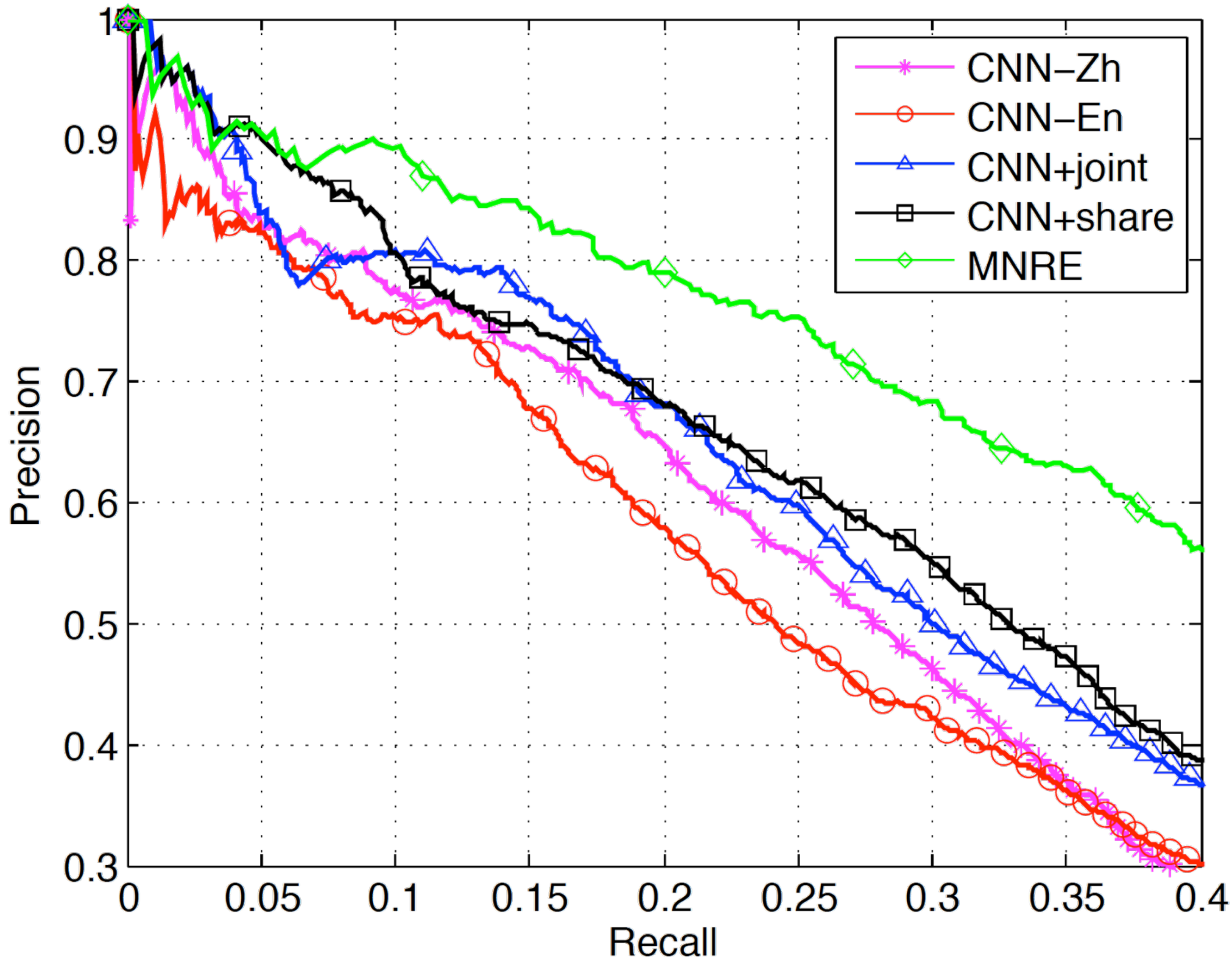
- Language specific relation matrix

Dataset

- Align between Wikidata and NYT

DataSet		#Rel	#Sent	#Fact
English	Train	176	1,022,239	47,638
	Valid		80,191	2,192
	Test		162,018	4,326
Chinese	Train	176	940,595	42,526
	Valid		82,699	2,192
	Test		167,224	4,326

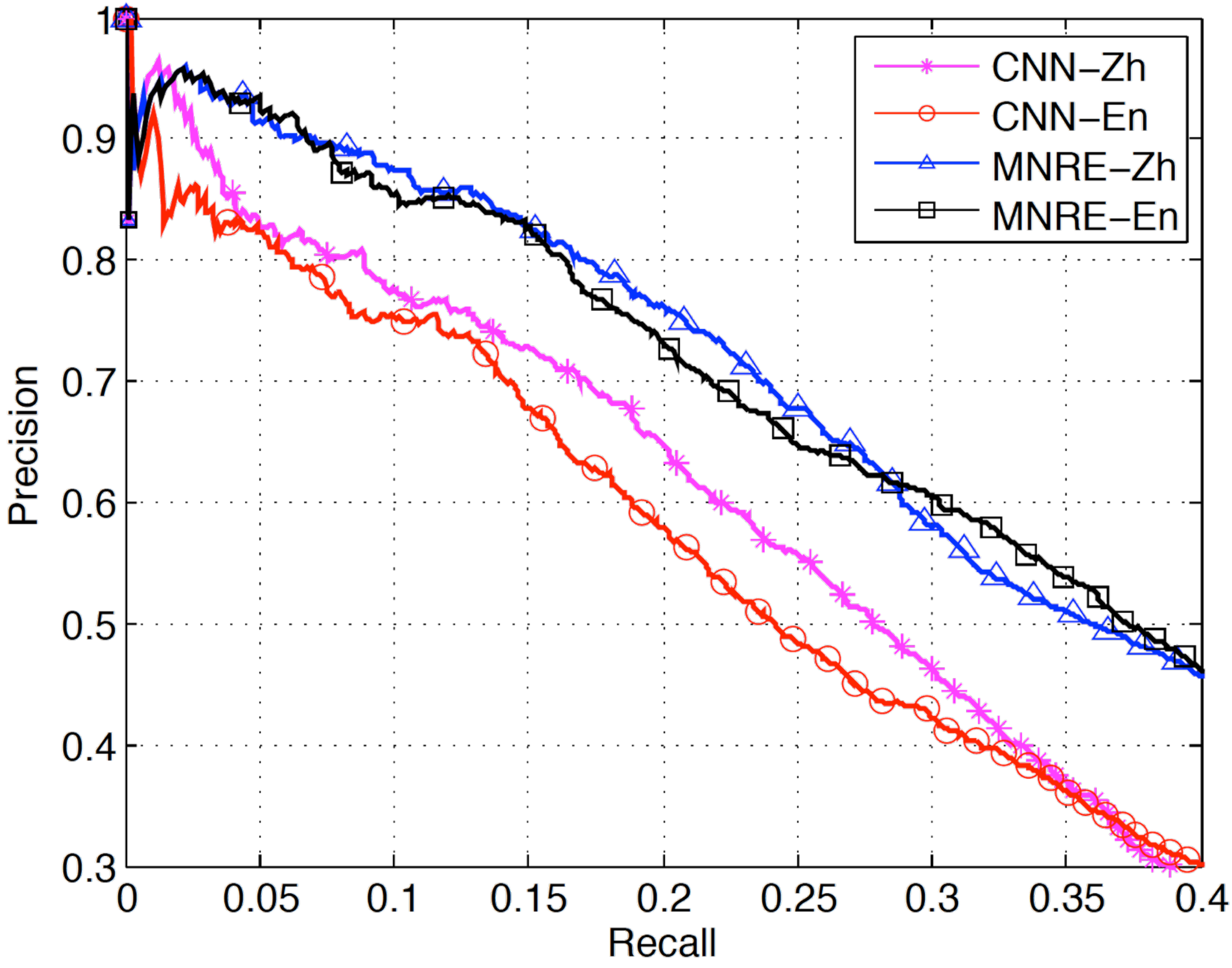
Effectiveness of Consistency



Effectiveness of Consistency

CNN +Zh	CNN +En	MNRE	Sentence
---	Medium	Low	**Barzun** is a commune in the Pyrénées-Atlantiques department in the Nouvelle-Aquitaine region of south-western **France**.
---	Medium	High	**Barzun** was born in Créteil , **France**
Medium	---	Low	作为从**法国**移民到美国来的顶尖知识分子，巴尔赞与莱昂内尔·特里林、德怀特·麦克唐纳等人一道，在冷战时期积极参与美国的公共知识生活…
Medium	---	High	巴尔赞于1907年出生于**法国**一个知识分子家庭，1920年赴美。

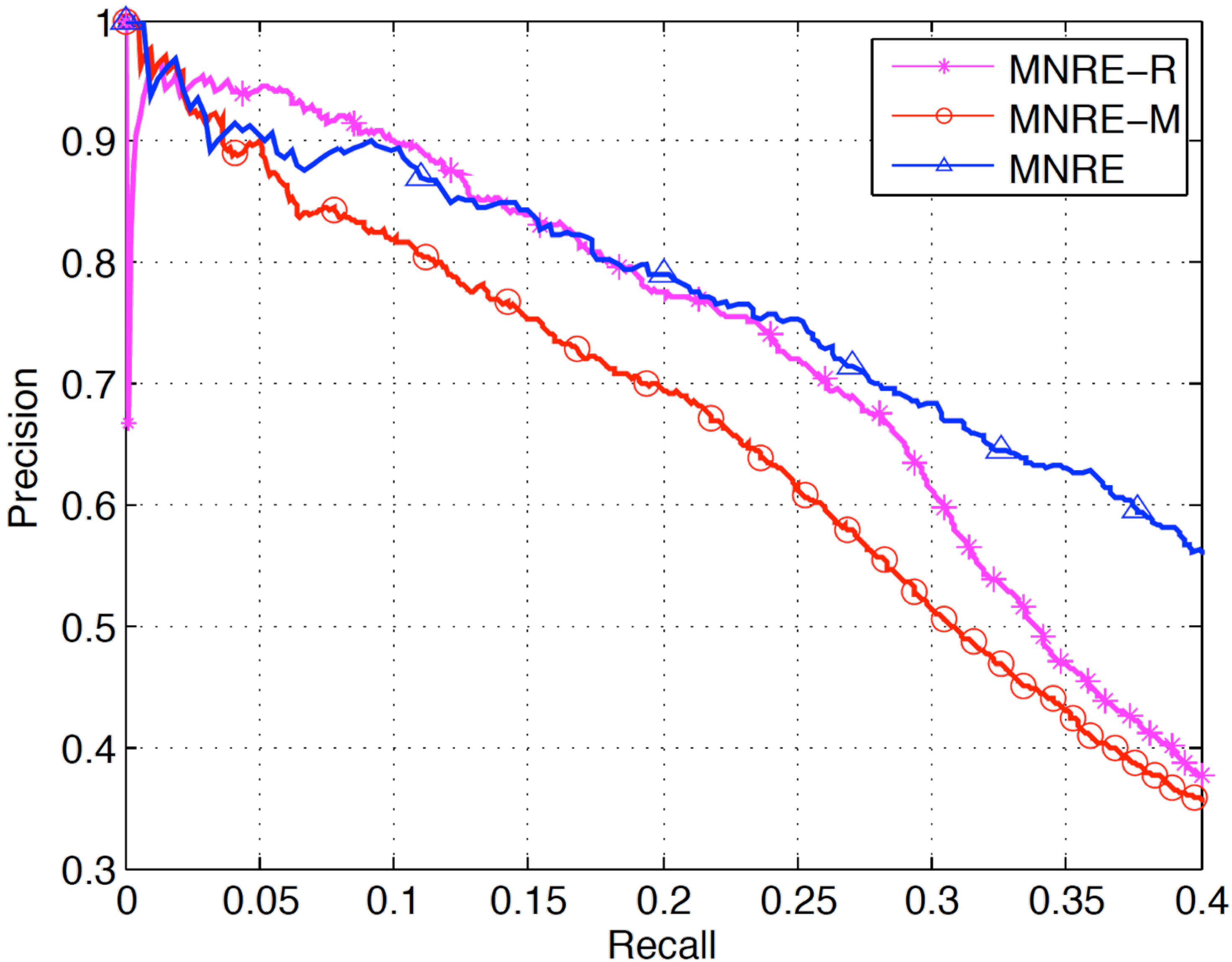
Effectiveness of Complementarity



Effectiveness of Complementarity

Relation	\#Sent-En	\#Sent-Zh	CNN-En	CNN-Zh	MNRE-En	MNRE-Zh
Contains	993	6,984	17.95	69.87	73.72	75.00
HeadquartersLocation	1,949	210	43.04	0	41.77	50.63
Father	1,833	983	64.71	77.12	86.27	83.01
CountryOfCitizenship	25,322	15,805	95.22	93.23	98.41	98.21

Comparison of Relation Matrix



Other Challenges

- One(Zero)-shot Relation Extraction
- Open Information Extraction
- Utilize Document Information

Open Source Tool

- Knowledge Representation
  - https://github.com/thunlp/KB2E
  - https://github.com/thunlp/Fast-TransX
  - https://github.com/thunlp/TensorFlow-TransX

- Knowledge Acquisition
  - https://github.com/thunlp/NRE
  - https://github.com/thunlp/TensorFlow-NRE
  - https://github.com/thunlp/MNRE

Q&A

Thanks