Relation Recognition

Fang Li

Dept. of Computer Science & Engineering

Contents

-- Distant supervised learning

-- Deep learning

Distant supervision method

"Distant supervision for relation extraction without labeled data"

- What means "**distant supervision**"?
- What are the **advantages** of the method?
- What are the **disadvantages** of the method?

Distant Supervision

--Mike Mintz, et al. Distant supervision for relation extraction without labeled data ACL2009

- Combing bootstrapping with supervised learning
- ✓ Instead of 5 seeds, use a large database to get huge number of seeds
- ✓ Create lots of features from all these examples
- ✓ Combine in a supervised classifier

Existing Knowledge Base

DBPedia or Freebase:

tens of thousands of examples of many relations; such as,

- *place-of-birth<Edwin Hubble, Marshfield>*
- *Place-of-birth<Albert Einstein,Ulm>*

…

Wikipedia:

Extract all sentences that have two named entities that match the tuple, like the following:

- ...*Hubble was born in Marshfield...*
- *...Einstein, born (1879), Ulm...*
- *...Hubble's birthplace in Marshfield...*

Collecting Training Data

Corpus text

Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, …
Bill Gates attended Harvard from…
Google was founded by Larry Page …

Training data

(Bill Gates, Microsoft)
Label: Founder
Feature: X founded Y

Freebase

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)

a freely available online database of structured semantic data. They use 1.8 million instances of 102 relations connecting 940,000 entities

The distant supervision algorithm for relation extraction

function DISTANT SUPERVISION(*Database D, Text T*) **returns** *relation classifier C*

foreach relation *R*
  foreach tuple (*e1,e2*) of entities with relation *R* in *D*
    sentences ← Sentences in *T* that contain *e1* and *e2*
    f ← Frequent features in *sentences*
    observations ← observations + new training tuple (*e1, e2, f, R*)
C ← Train supervised classifier on *observations*
return *C*

Distantly supervised learning of relation extraction patterns

1. For each relation

   Born-In

2. For each tuple in big database

   <Edwin Hubble, Marshfield>
   <Albert Einstein, Ulm>

3. Find sentences in large corpus with both entities

   Hubble was born in Marshfield
   Einstein, born (1879), Ulm
   Hubble's birthplace in Marshfield

4. Extract frequent features (parse, words, etc)

   PER was born in LOC
   PER, born (XXXX), LOC
   PER's birthplace in LOC

5. Train supervised classifier using thousands of patterns

   $P(\text{born-in} \mid f_1, f_2, f_3, \ldots, f_{70000})$

Distant supervision: Lexical Features:

- The sequence of words between the two entities
- The part-of-speech tags of these words
- A flag indicating which entity came first in the sentence
- A window of k words to the left of Entity 1 and their part-of-speech tags
- A window of k words to the right of Entity 2 and their part-of-speech tags

Distant supervision: Syntactic Features:

Use parser MINIPAR

- A dependency path between the two entities.
- For each entity, one 'window' node that is not part of the dependency path.

Distant supervision: Features

P(place_of_birth | $f_1, f_2, f_3, \ldots f_{7000}$)

词汇特征的
K=0,1,2

Feature type	Left window	NE1	Middle	NE2	Right window
Lexical	[]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[]
Lexical	[Astronomer]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[,]
Lexical	[#PAD#, Astronomer]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[, Missouri]
Syntactic	[]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[]
Syntactic	[Edwin Hubble $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[]
Syntactic	[Astronomer $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[]
Syntactic	[]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{lex-mod}$,]
Syntactic	[Edwin Hubble $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{lex-mod}$,]
Syntactic	[Astronomer $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{lex-mod}$,]
Syntactic	[]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{inside}$ Missouri]
Syntactic	[Edwin Hubble $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{inside}$ Missouri]
Syntactic	[Astronomer $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{inside}$ Missouri]

Table 3: Features for 'Astronomer Edwin Hubble was born in Marshfield, Missouri'.

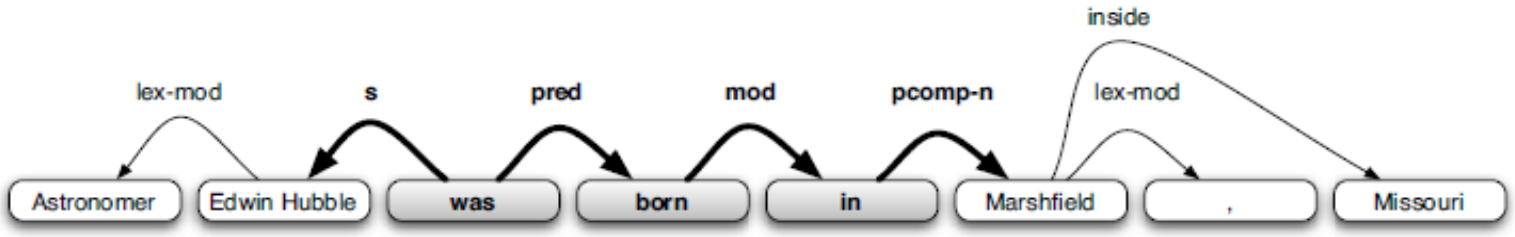


Figure 1: Dependency parse with dependency path from 'Edwin Hubble' to 'Marshfield' highlighted in boldface.

Examples of high-weight features for several relations

	SYN	is $\Uparrow_s$	ORG	$\Uparrow_s$ is $\Downarrow_{pred}$ band $\Downarrow_{mod}$ from $\Downarrow_{pcn}$	LOC	$\Uparrow_s$ is
/people/deceased_person/place_of_death	LEX		PER	died in	LOC	
	SYN	hanged $\Uparrow_s$	PER	$\Uparrow_s$ hanged $\Downarrow_{mod}$ in $\Downarrow_{pcn}$	LOC	$\Uparrow_s$ hanged
/people/person/nationality	LEX		PER	is a citizen of	LOC	
	SYN		PER	$\Downarrow_{mod}$ from $\Downarrow_{pcn}$	LOC	
/people/person/parents	LEX		PER	, son of	PER	
	SYN	father $\Uparrow_{gen}$	PER	$\Uparrow_{gen}$ father $\Downarrow_{person}$	PER	$\Uparrow_{gen}$ father
/people/person/place_of_birth	LEX$\curvearrowleft$		PER	is the birthplace of	PER	
	SYN		PER	$\Uparrow_s$ born $\Downarrow_{mod}$ in $\Downarrow_{pcn}$	LOC	
/people/person/religion	LEX		PER	embraced	LOC	
	SYN	convert $\Downarrow_{appo}$	PER	$\Downarrow_{appo}$ convert $\Downarrow_{mod}$ to $\Downarrow_{pcn}$	LOC	$\Downarrow_{appo}$ convert

Training and Testing

- Training: 900,000 Freebase relation instances, 800,000 Wikipedia articles
- Testing: 900,000 Freebase relation instances, 400,000 different articles
- Classifer: multi-class logistic regression classifier which returns a relation name and a confidence score.

Freebase Examples

Relation name	Size	Example
/people/person/nationality	281,107	John Dugard, South Africa
/location/location/contains	253,223	Belgium, Nijlen
/people/person/profession	208,888	Dusa McDuff, Mathematician
/people/person/place_of_birth	105,799	Edwin Hubble, Marshfield
/dining/restaurant/cuisine	86,213	MacAyo's Mexican Kitchen, Mexican
/business/business_chain/location	66,529	Apple Inc., Apple Inc., South Park, NC
/biology/organism_classification_rank	42,806	Scorpaeniformes, Order
/film/film/genre	40,658	Where the Sidewalk Ends, Film noir
/film/film/language	31,103	Enter the Phoenix, Cantonese
/biology/organism_higher_classification	30,052	Calopteryx, Calopterygidae
/film/film/country	27,217	Turtle Diary, United States
/film/writer/film	23,856	Irving Shulman, Rebel Without a Cause
/film/director/film	23,539	Michael Mann, Collateral
/film/producer/film	22,079	Diane Eskenazi, Aladdin
/people/deceased_person/place_of_death	18,814	John W. Kern, Asheville
/music/artist/origin	18,619	The Octopus Project, Austin
/people/person/religion	17,582	Joseph Chartrand, Catholicism
/book/author/works_written	17,278	Paul Auster, Travels in the Scriptorium
/soccer/football_position/players	17,244	Midfielder, Chen Tao
/people/deceased_person/cause_of_death	16,709	Richard Daintree, Tuberculosis
/book/book/genre	16,431	Pony Soldiers, Science fiction
/film/film/music	14,070	Stavisky, Stephen Sondheim
/business/company/industry	13,805	ATS Medical, Health care

Human Evaluation Result

Relation name	100 instances			1000 instances		
	Syn	Lex	Both	Syn	Lex	Both
/film/director/film	**0.49**	0.43	0.44	**0.49**	0.41	0.46
/film/writer/film	**0.70**	0.60	0.65	**0.71**	0.61	0.69
/geography/river/basin_countries	0.65	0.64	**0.67**	**0.73**	0.71	0.64
/location/country/administrative_divisions	0.68	0.59	**0.70**	**0.72**	0.68	**0.72**
/location/location/contains	0.81	**0.89**	0.84	**0.85**	0.83	0.84
/location/us_county/county_seat	0.51	0.51	**0.53**	0.47	**0.57**	0.42
/music/artist/origin	0.64	0.66	**0.71**	0.61	**0.63**	0.60
/people/deceased_person/place_of_death	0.80	0.79	**0.81**	0.80	**0.81**	0.78
/people/person/nationality	0.61	0.70	**0.72**	0.56	0.61	**0.63**
/people/person/place_of_birth	**0.78**	0.77	**0.78**	0.88	0.85	**0.91**
Average	0.67	0.66	**0.69**	**0.68**	0.67	0.67

Table 5: Estimated precision on human-evaluation experiments of the highest-ranked 100 and 1000 results per relation, using stratified samples. 'Average' gives the mean precision of the 10 relations. Key: Syn = syntactic features only. Lex = lexical features only. We use stratified samples because of the overabundance of *location-contains* instances among our high-confidence results.

Examples of their results

Relation name	New instance
/location/location/contains	Paris, Montmartre
/location/location/contains	Ontario, Fort Erie
/music/artist/origin	Mighty Wagon, Cincinnati
/people/deceased_person/place_of_death	Fyodor Kamensky, Clearwater
/people/person/nationality	Marianne Yvonne Heemskerk, Netherlands
/people/person/place_of_birth	Wavell Wayne Hinds, Kingston
/book/author/works_written	Upton Sinclair, Lanny Budd
/business/company/founders	WWE, Vince McMahon
/people/person/profession	Thomas Mellon, judge

Ten relation instances extracted by our system that did not appear in Freebase.

Advantages

- Does not need human annotation
- Large training corpus

Disadvantages of the method

- Noises in training data, for example, founder(Bill Gates, Microsoft)

S1: Bill Gates is one of the founders of Microsoft Co.

S2: Bill Gates has founded the Microsoft Co.

S3: Bill Gates was CEO of the Microsoft Co. ×

S4: Bill Gates discussed with the CEO of the Microsoft Co. for his retirement. ×

- Relations are disjoined

Founded(Jobs, Apple), CEO-of(Jobs, Apple) can not be extracted both.

How to improve it ?

Improved method:

Knowledge-Based Weak Supervision for Information Extraction of Overlapping Relations

Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, Daniel S. Weld
Computer Science & Engineering
University of Washington
Seattle, WA 98195, USA

Relation	Freebase Matches		MULTIR	
	#sents	% true	$\tilde{P}$	$\tilde{R}$
/business/person/company	302	89.0	100.0	25.8
/people/person/place_lived	450	60.0	80.0	6.7
/location/location/contains	2793	51.0	100.0	56.0
/business/company/founders	95	48.4	71.4	10.9
/people/person/nationality	723	41.0	85.7	15.0
/location/neighborhood/neighborhood_of	68	39.7	100.0	11.1
/people/person/children	30	80.0	100.0	8.3
/people/deceased_person/place_of_death	68	22.1	100.0	20.0
/people/person/place_of_birth	162	12.0	100.0	33.0
/location/country/administrative_divisions	424	0.2	N/A	0.0

Performance of Relation Classification (SemiEval-2010 task 8 dataset)

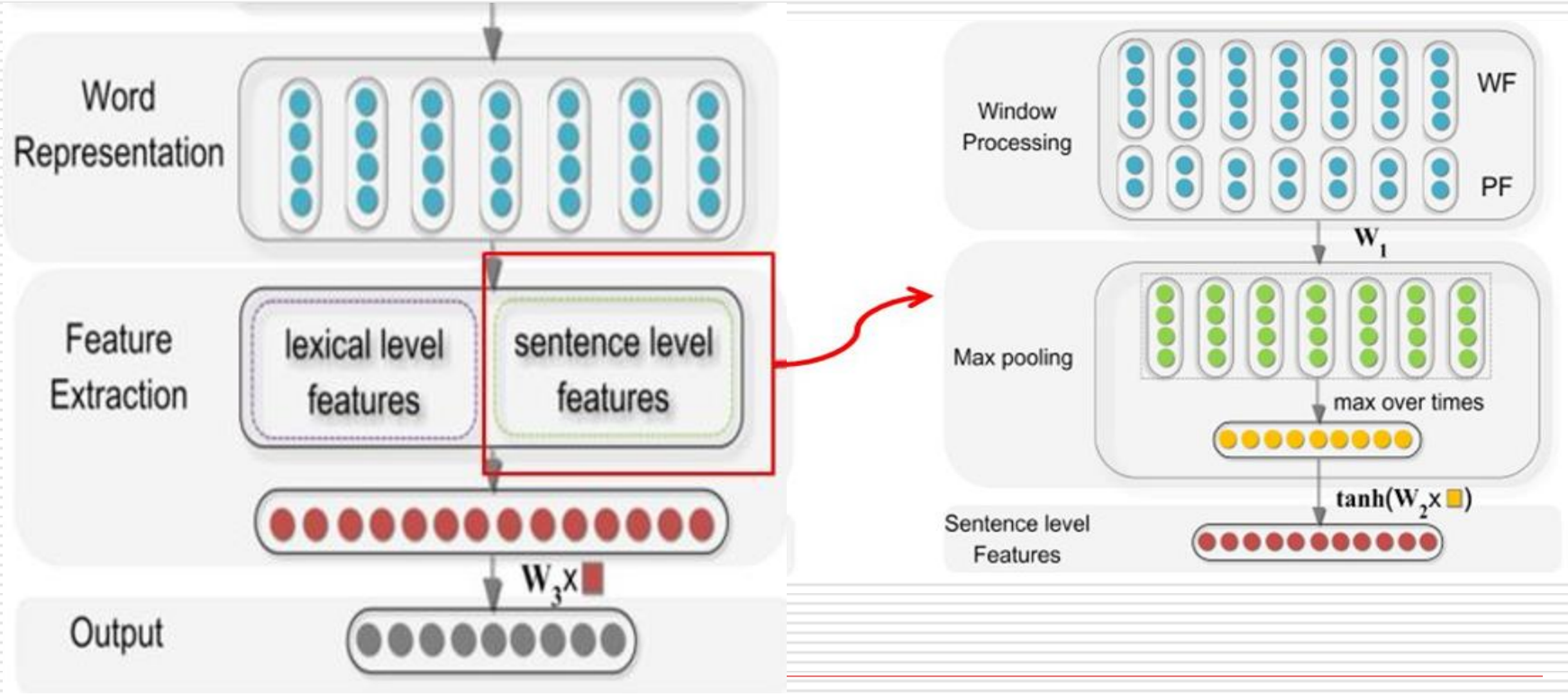
- From Xianpei Han's tutorial in 2016,10,21 "Knowledge graph based sematic relation Extraction"

Classifier	Feature set	F_1
SVM	POS, WordNet, prefixes and other morphological features, depdency parse, Levin classes, PropBank, FanmeNet, NomLex-Plus, Google n-gram, paraphrases, TextRunner	82.2
RNN	Word embeddings	74.8
	Word embeddings, POS, NER, WordNet	77.6
MVRNN	Word embeddings	79.1
	Word embeddings, POS, NER, WordNet	82.4
CNN	Word embeddings	69.7
	Word embeddings, word position embeddings, WordNet	82.7
Chain CNN	Word embeddings, POS, NER, WordNet	82.7
FCM	Word embeddings	80.6
	Word embeddings, depedency parsing, NER	83.0
CR-CNN	Word embeddings	82.8†
	Word embeddings, position embeddings	82.7
	Word embeddings, position embeddings	**84.1**†
SDP-LSTM	Word embeddings	82.4
	Word embeddings, POS embeddings, WordNet embeddings, grammar relation embeddings	**83.7**

New Trends: Deep Learning

- Training data → Word Embedding

Acme Inc. hired Mr Smith as their new CEO, replacing Mr Bloggs.



Zeng, Relation classification via convolutional deep neural network Coling 2014

Extract relations using new method

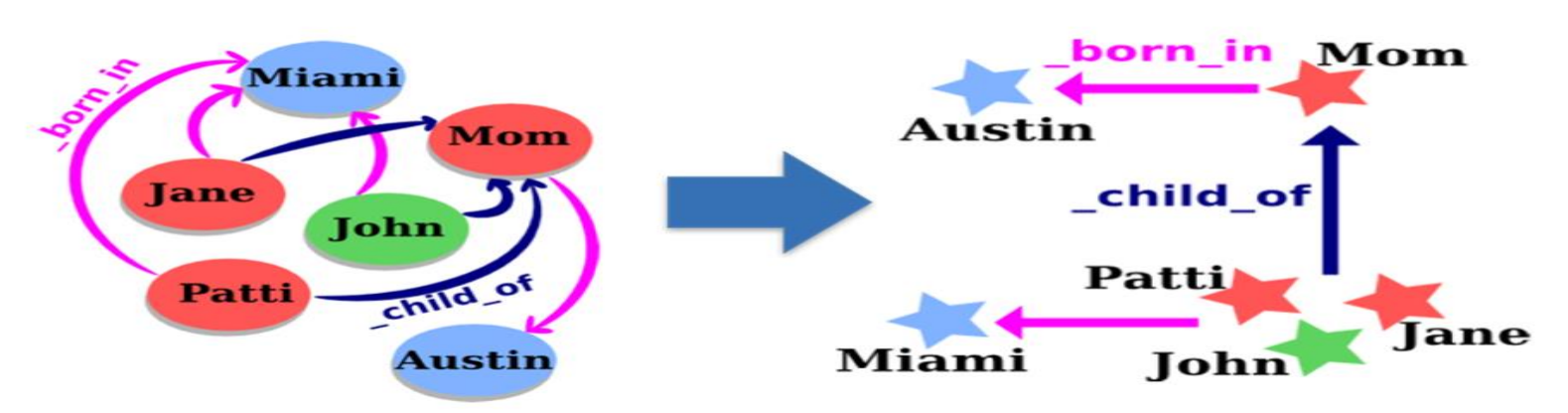
Translating Embeddings for Modeling Multi-relational Data

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán
Université de Technologie de Compiègne – CNRS
Heudiasyc UMR 7253
Compiègne, France
{bordesan, nusunier, agarciad}@utc.fr

Jason Weston, Oksana Yakhnenko
Google
111 8th avenue
New York, NY, USA
{jweston, oksana}@google.com

Relation Representation (Triplets)

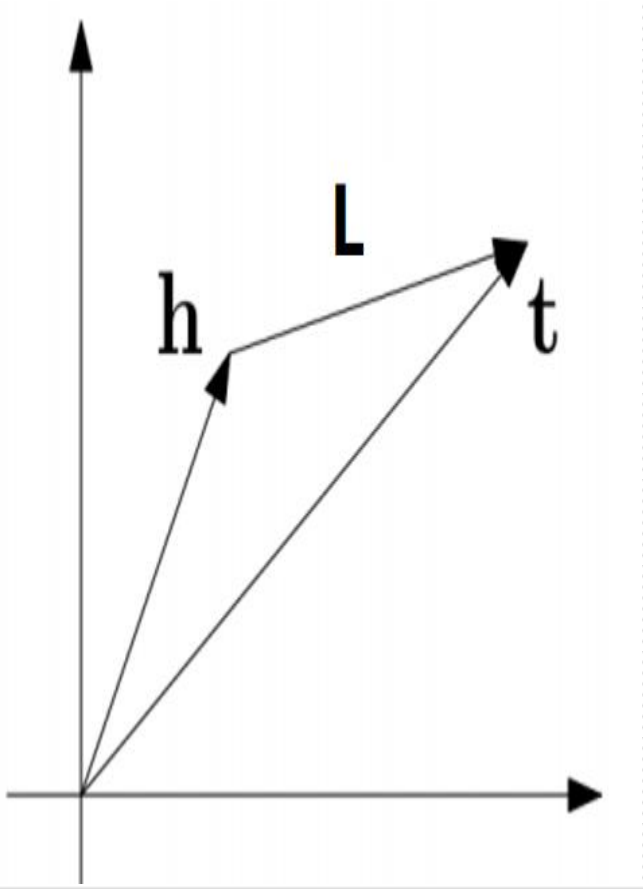
- (head, label, tail) i.e (h,l,t): there exists a **relationship** of name label between the entities head and tail.



- (Mom born_in Austin)

TransE:

- If (h,l,t) holds, then the **embedding** of the tail entity *t* should be close to the **embedding** of the head entity *h* plus some **vector** that depends on the relationship *l*



Example

- Aim: H+L=T

T should be a nearest neighbor of H+L, while H+L should be far away from T otherwise.

(Yaoming born_in Shanghai) √

(Yaoming born_in Beijing) X

d(yaoming+born_in,Beijing) >> d(yaoming+born_in, Shanghai)

d: distance as dissimilarity function

Learn TransE

- Minimize a margin-based ranking criterion over the training set:

$$\mathcal{L} = \sum_{(h,\ell,t)\in S} \sum_{(h',\ell,t')\in S'_{(h,\ell,t)}} \left[\gamma + d(\boldsymbol{h}+\boldsymbol{\ell},\boldsymbol{t}) - d(\boldsymbol{h'}+\boldsymbol{\ell},\boldsymbol{t'})\right]_+$$

Positive examples

Negative examples

- Corrupted triplets:

$$S'_{(h,\ell,t)} = \{(h',\ell,t)|h' \in E\} \cup \{(h,\ell,t')|t' \in E\}$$

The optimization is carried out by stochastic gradient descent.
Additional constraint: the L2-norm of the embeddings of entities is 1

Experimental Results

INPUT (HEAD AND LABEL)	PREDICTED TAILS
J. K. Rowling `influenced by`	*G. K. Chesterton*, J. R. R. Tolkien, *C. S. Lewis*, **Lloyd Alexander**, Terry Pratchett, Roald Dahl, Jorge Luis Borges, *Stephen King*, Ian Fleming
Anthony LaPaglia `performed in`	*Lantana*, *Summer of Sam*, *Happy Feet*, *The House of Mirth*, Unfaithful, **Legend of the Guardians**, Naked Lunch, X-Men, The Namesake
Camden County `adjoins`	**Burlington County**, *Atlantic County*, *Gloucester County*, Union County, Essex County, New Jersey, Passaic County, Ocean County, Bucks County
The 40-Year-Old Virgin `nominated for`	*MTV Movie Award for Best Comedic Performance*, *BFCA Critics' Choice Award for Best Comedy*, *MTV Movie Award for Best On-Screen Duo*, MTV Movie Award for Best Breakthrough Performance, **MTV Movie Award for Best Movie**, MTV Movie Award for Best Kiss, D. F. Zanuck Producer of the Year Award in Theatrical Motion Pictures, Screen Actors Guild Award for Best Actor - Motion Picture
Costa Rica football team `has position`	*Forward*, *Defender*, *Midfielder*, **Goalkeepers**, Pitchers, Infielder, Outfielder, Center, Defenseman
Lil Wayne `born in`	**New Orleans**, Atlanta, Austin, St. Louis, Toronto, New York City, Wellington, Dallas, Puerto Rico
WALL-E `has the genre`	Animations, Computer Animation, *Comedy film*, *Adventure film*, *Science Fiction*, **Fantasy**, Stop motion, *Satire*, Drama

TransE's Result

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

Comparisons with several methods developed from TransE

Head	**University_of_Cambridge**		
Relation	**/education/education/student**		
Model	TransE	TransH	TransR
1	John_Cleese	Stephen_Fry	David_Attenborough
2	Samuel_Beckett	David_Attenborough	Stephen_Fry
3	Harold_Pinter	Ralph_Vaughan_Williams	Stephen_Hawking
4	Virginia_Woolf	Alan_Bennett	Ralph_Vaughan_Williams
5	Graham_Chapman	Francis_Bacon	Alan_Bennett
6	Philip_Pullman	Julian_Fellowes	Julian_Fellowes
7	Ian_McEwan	Hugh_Bonneville	Ernest_Rutherford
8	Douglas_Adams	Graham_Chapman	Jonathan_Lynn
9	Terry_Gilliam	Miriam_Margolyes	Tom_Hollander
10	Richard_Dawkins	Stephen_Hawking	Chris_Weitz

What are the disadvantages of TransE?

- Can not extract 1 to N, N to 1, N to N relations.

(USA president Obama)

(USA president Bush)

(USA president Trump)

Source Codes

- KB2E: https://github.com/thunlp/KB2E

TransE, TransH, TransR,…

References for Relation Extraction using deep learning methods

- Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. Relation classification via convolutional deep neural network. In Proceedings of COLING 2014.
- Daojian Zeng,Kang Liu,Yubo Chen,and Jun Zhao. Distant supervision for relation extraction via piecewise convolutional neural networks. In Proceedings of EMNLP
- Lin, et al. (2016). Neural Relation Extraction with Selective Attention over Instances. ACL
- Zeng, et al. (2015). Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks. EMNLP.

Summarization

- *Semi-supervised methods*
- *Distant supervision method*
- *Deep learning method*

Discussion topic: How to identify the CEO of the company?

- 1993年7月,深圳市政府任命**夏斌**为**深交所**总经理。（+）
- **任汇川，中国平安集团**总经理，生于1969年，1992毕业于哈尔滨工程大学，同年加入中国平安集团，是平安集团迄今最年轻的本土高层管理人员，也是该集团年轻管理团队的典范之一。（+）
- **湖北日报传媒集团**总经理**张勤耘**涉严重违纪被调查.（+）
- 2014年6月7日 - 中国移动2010年5月31日确认,**李跃**出任**中国移动**总经理,原中国移动总裁王建宙任中国移动党组书记兼任集团公司董事长。（+）
- 开业仅两年的前海人寿或将迎来第二波高管团队换血潮,正式获批上任不到半年的总经理傅杰也传闻将出走。（-）
- 公司实行总经理负责制,总经理是公司的法定代表人。（-）

About the Project (1)

- Task: Employment relation extraction
- Training corpus:本报北京１２月３０日讯新华社记者胡晓梦、本报记者吴亚明报道：新年将至，国务院侨务办公室主任郭东坡今天通过新闻媒介，向海外同胞和国内归侨、侨眷、侨务工作者发表新年贺词。

(胡晓梦,新华社)

(吴亚明,新民晚报)

(郭东坡,国务院侨务办)

About the Project (2)

- Methods:
  - Pattern-based
  - Supervised method or semi-supervised or unsupervised methods

Training corpus are put online.

- Evaluation:

Use test corpus with human annotated results to evaluate your algorithm.