Never Ending Language Learning

Carnegie Mellon University

Tenet 1:

Understanding requires a <u>belief system</u>

We'll never produce natural language <u>understanding</u> systems until we have systems that react to arbitrary sentences by saying one of:

- I understand, and already knew that
- I understand, and didn't know, but accept it
- I understand, and disagree because …

Tenet 2:

We'll never really understand learning until we build machines that

- learn many different things,
- over years,
- and become better <u>learners</u> over time.

NELL: Never-Ending Language Learner

Inputs:

- initial ontology
- few examples of each ontology predicate
- the web
- occasional interaction with human trainers

The task:

- run 24x7, forever
- each day:
    1. extract more facts from the web to populate the initial ontology
    2. learn to read (perform #1) better than yesterday

NELL today

Running 24x7, since January, 12, 2010

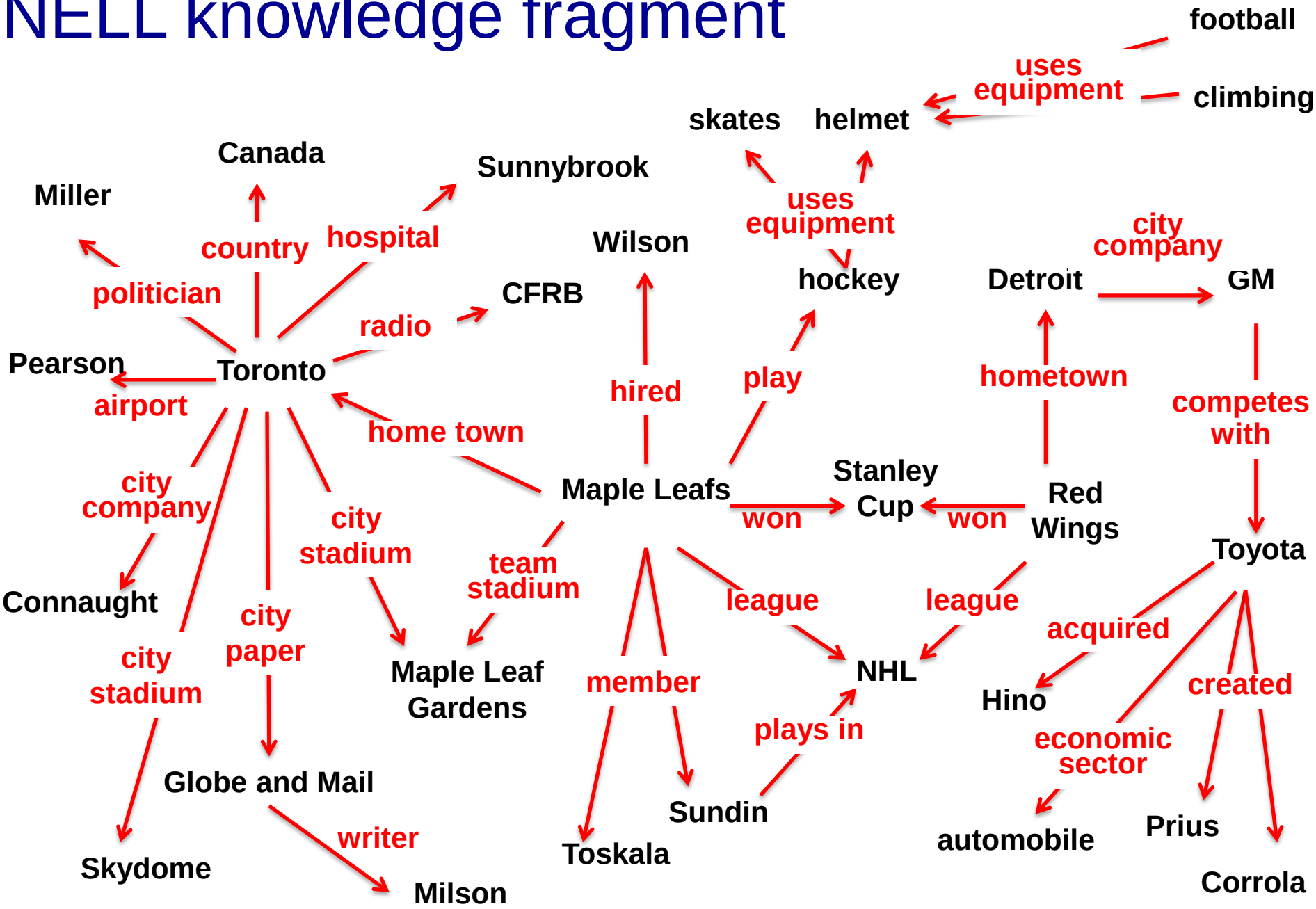
Inputs:

- ontology defining >600 categories and relations
- 10-20 seed examples of each
- 500 million web pages
- 100,000 web search queries per day
- ~ 5 minutes/day of human guidance

Result:

- KB with > 15 million candidate beliefs, growing daily
- learning to reason, as well as read
- automatically extending its ontology

NELL knowledge fragment



NELL Today

- http://rtw.ml.cmu.edu ← follow NELL here
- eg. “diabetes”, “Avandia”, ,“tea”, “IBM”, “love” “baseball” “BacteriaCausesCondition” …

Recently-Learned Facts

instance	iteration	date learned
banana_nut_chocolate_chip_bread is a baked good	446	03-nov-2011
nisha_ganatra is an African person	446	03-nov-2011
corky_miller is a male	445	01-nov-2011
two_marriages is a parlour game	443	29-oct-2011
tanner_police_department is a part of the government	446	03-nov-2011
georgia_aquarium is a tourist attraction in the city atlanta	448	05-nov-2011
florida_hospital is a hospital in the city orlando	446	03-nov-2011
adobe_systems_incorporated is a company also known as adobe	445	01-nov-2011
the sports team yankees was the winner of n1923_world_series	448	05-nov-2011
north_arm is a city located in the state or province ohio	446	03-nov-2011

NELL Today

- http://rtw.ml.cmu.edu ← follow NELL here
- eg. “diabetes”, “Avandia”, ,“tea”, “IBM”, “love” “baseball” “BacteriaCausesCondition” …

Recently-Learned Facts twitter

instance	iteration	date learned	con
the_rembrandts is a TV show	535	21-mar-2012	
alexander_memorial_coliseum is a building	535	21-mar-2012	
jason_bergmann is an athlete	535	21-mar-2012	
cnc_costume_national_shoes is a kind of clothing	535	21-mar-2012	
korinthia is a state or a province	535	21-mar-2012	
children is an animal that can develop problems	536	23-mar-2012	
burlington is a city that lies on the river skagit	536	23-mar-2012	
grains is a generalization of maize	536	23-mar-2012	
j_s__bach is a musician who plays the piano	539	26-mar-2012	
mary is the parent of god_the_son	538	25-mar-2012	

NELL Today

- http://rtw.ml.cmu.edu ← follow NELL here
- eg. "diabetes", "Avandia", ,"tea", "IBM", "love" "baseball" "BacteriaCausesCondition" …

Recently-Learned Facts twitter

instance	iteration	date learned
hartsfield is an airport	554	22-apr-2012
ted_marcoux is a commedian	554	22-apr-2012
graph_paper is an office supply	559	01-may-2012
san_antonio_museum_of_art is a museum	554	22-apr-2012
south_ossetia is a country	556	26-apr-2012
duke_ellington is a musician who plays the piano	559	01-may-2012
classic_books is a generalization of pride_and_prejudice	559	01-may-2012
state is a synonym for department	557	29-apr-2012
brookfield_zoo is a tourist attraction in the city chicago	559	01-may-2012
ford_motor is a company that produces falcon	559	01-may-2012

Semi-Supervised Bootstrap Learning

it's underconstrained!!

Extract cities:

Paris
Pittsburgh
Seattle
Cupertino

San Francisco
Austin
denial

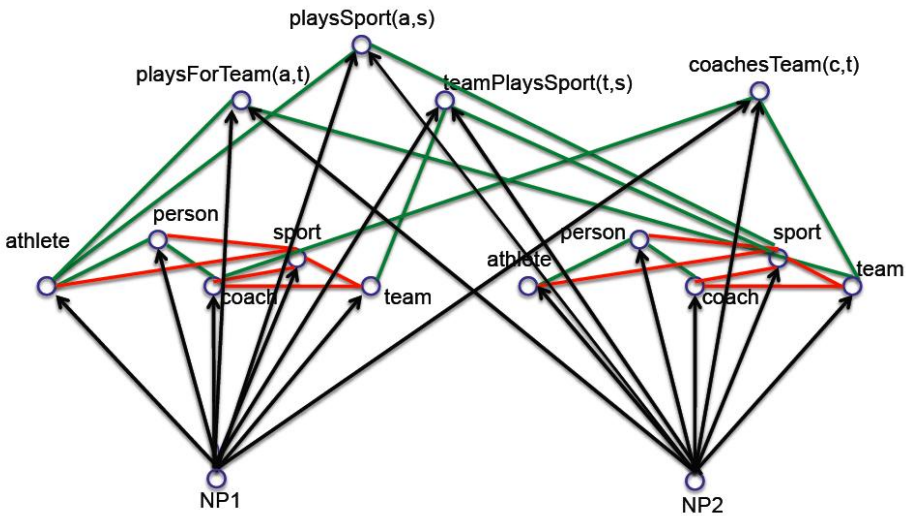
anxiety
selfishness
Berlin

mayor of arg1
live in arg1

arg1 is home of
traits such as arg1

Key Idea 1: Coupled semi-supervised training of many functions

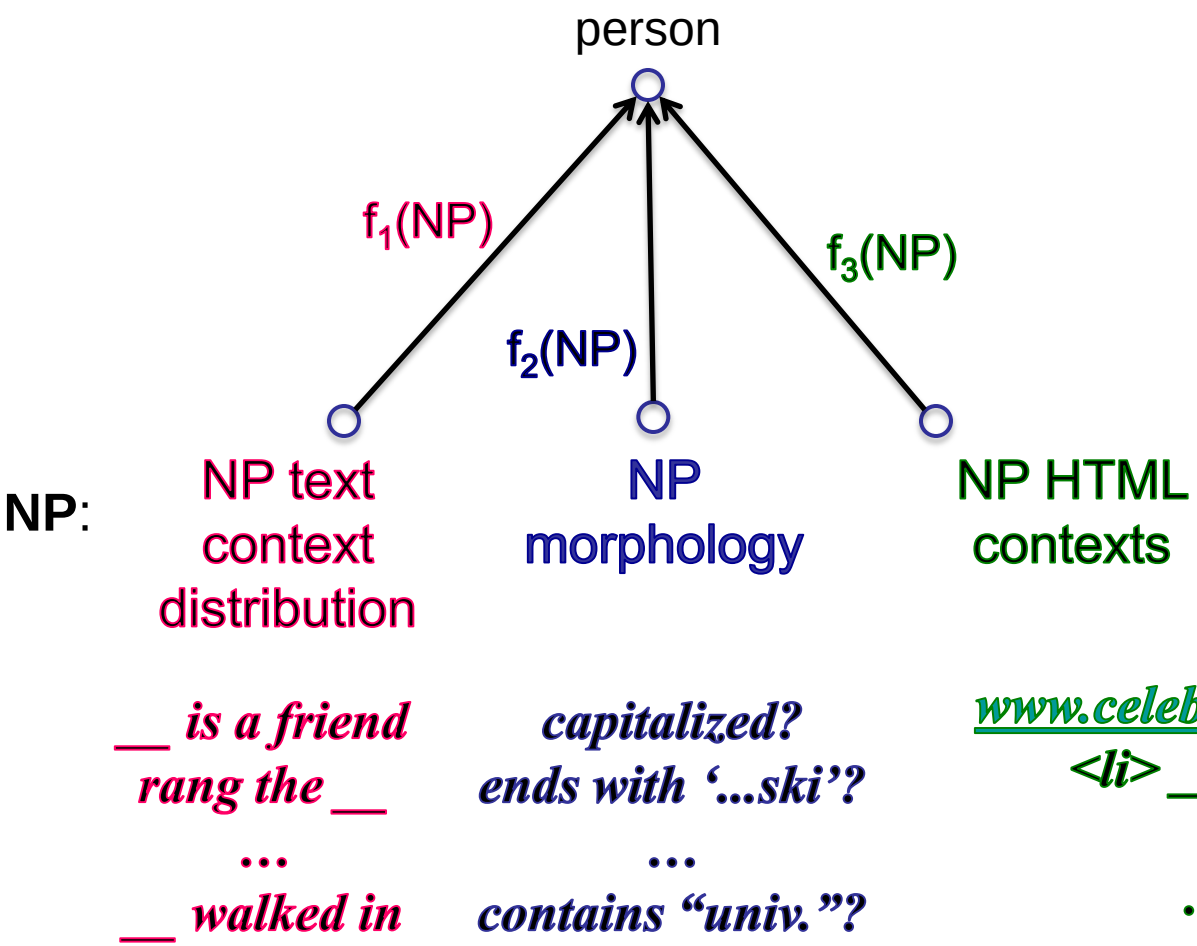
hard (underconstrained) semi-supervised learning problem



much easier (more constrained) semi-supervised learning problem

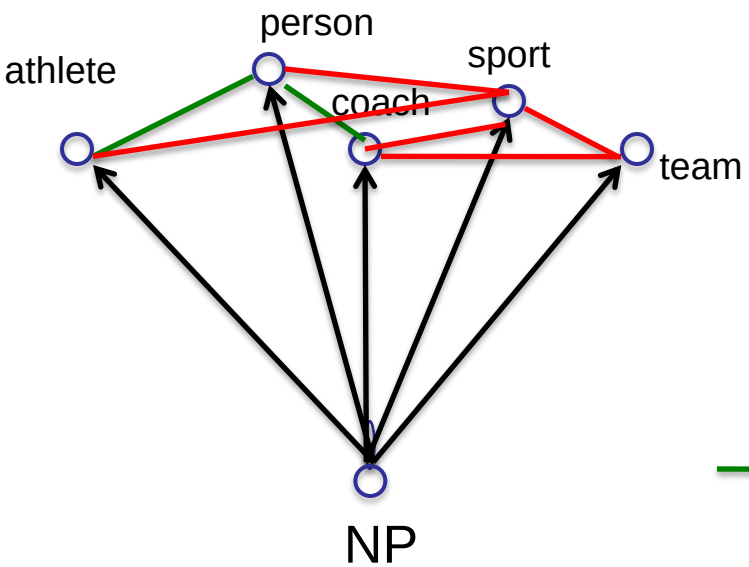
Type 1 Coupling: Co-Training, Multi-View Learning

[Blum & Mitchell; 98]
[Dasgupta et al; 01]
[Ganchev et al., 08]
[Sridharan & Kakade, 08]
[Wang & Zhou, ICML10]



Type 2 Coupling: Multi-task, Structured Outputs

[Daume, 2008]
[Bakhir et al., eds. 2007]
[Roth et al., 2008]
[Taskar et al., 2009]
[Carlson et al., 2009]

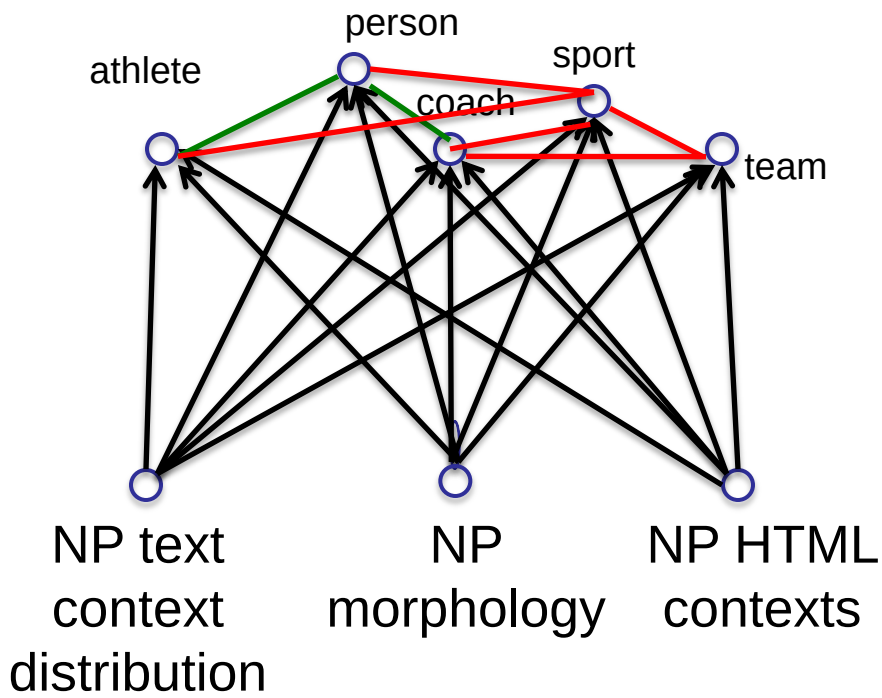


athlete(NP) → person(NP)

athlete(NP) → NOT sport(NP)

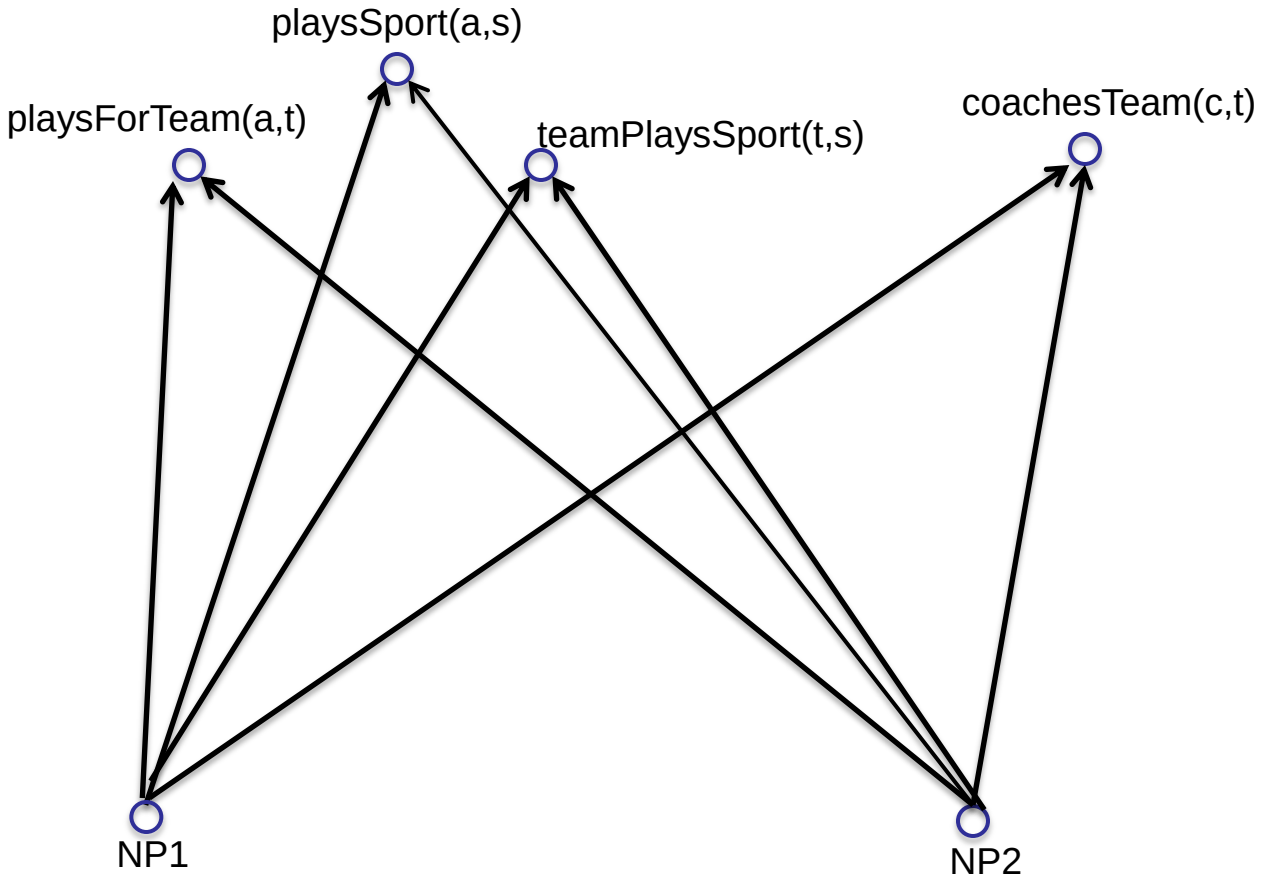
NOT athlete(NP) ← sport(NP)

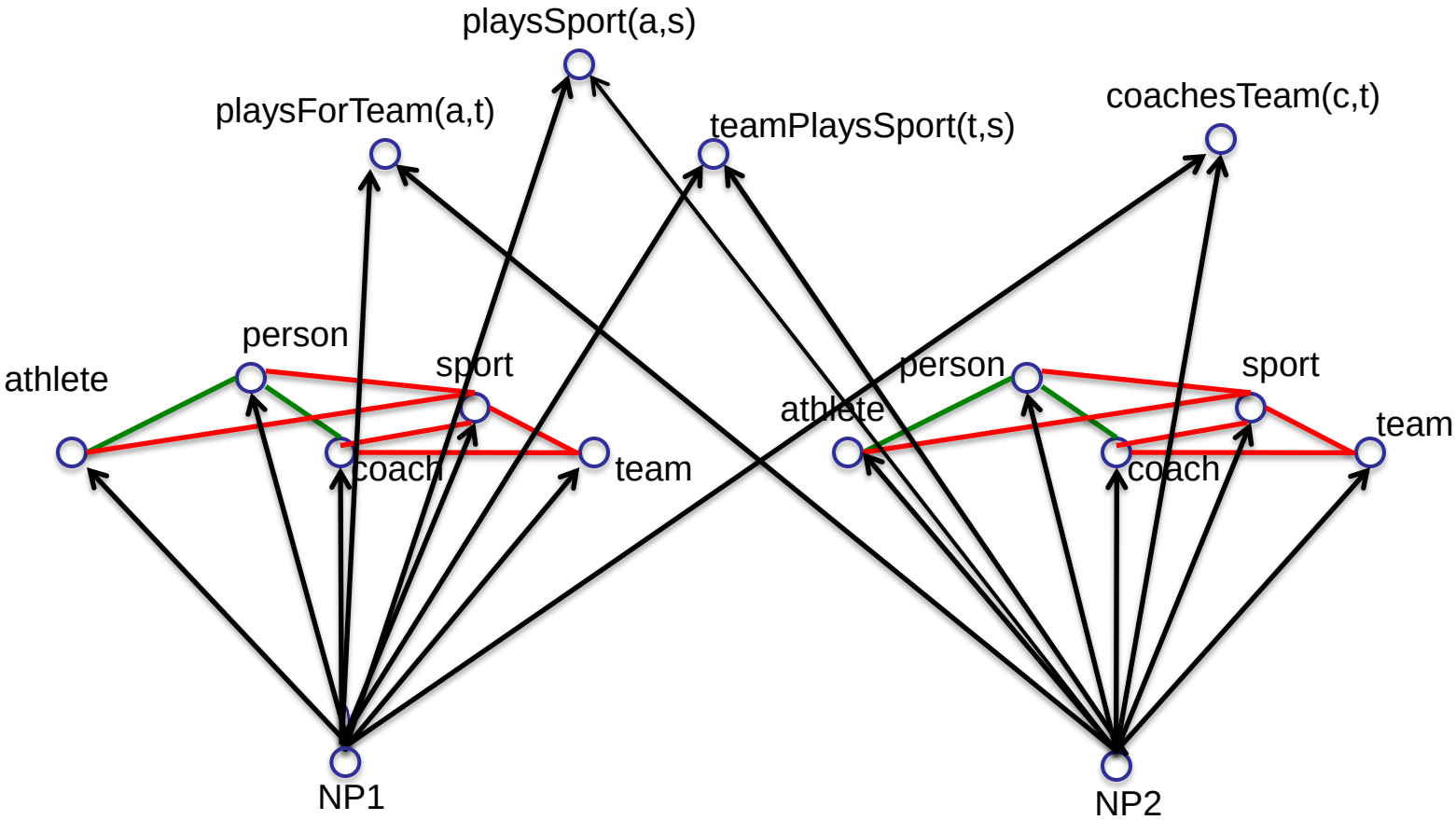
Multi-view, Multi-Task Coupling



NP:

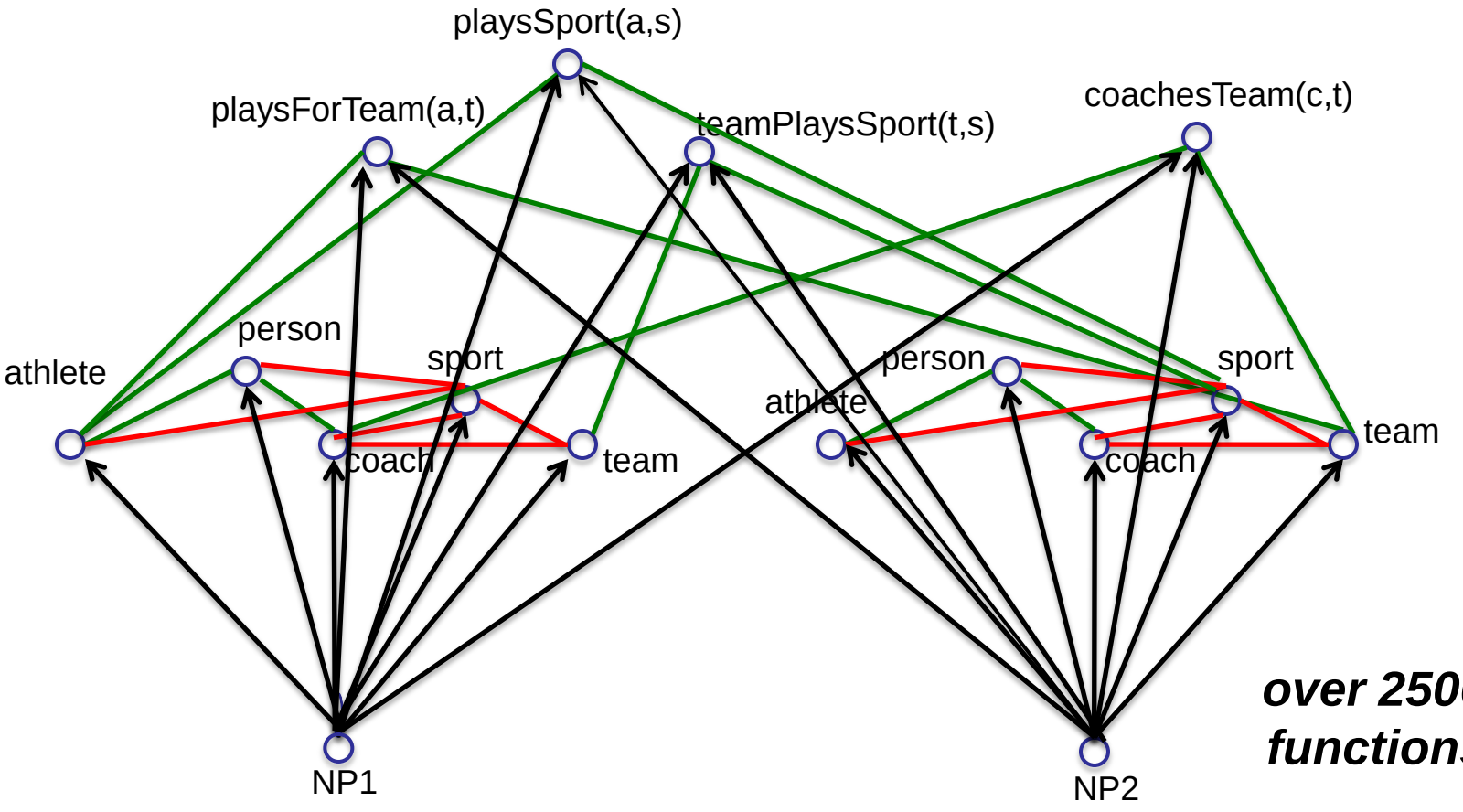
Learning Relations between NP's




playsSport(a,s)
playsForTeam(a,t)
teamPlaysSport(t,s)
coachesTeam(c,t)
person
athlete
sport
coach
team
person
sport
athlete
team
coach
NP1
NP2

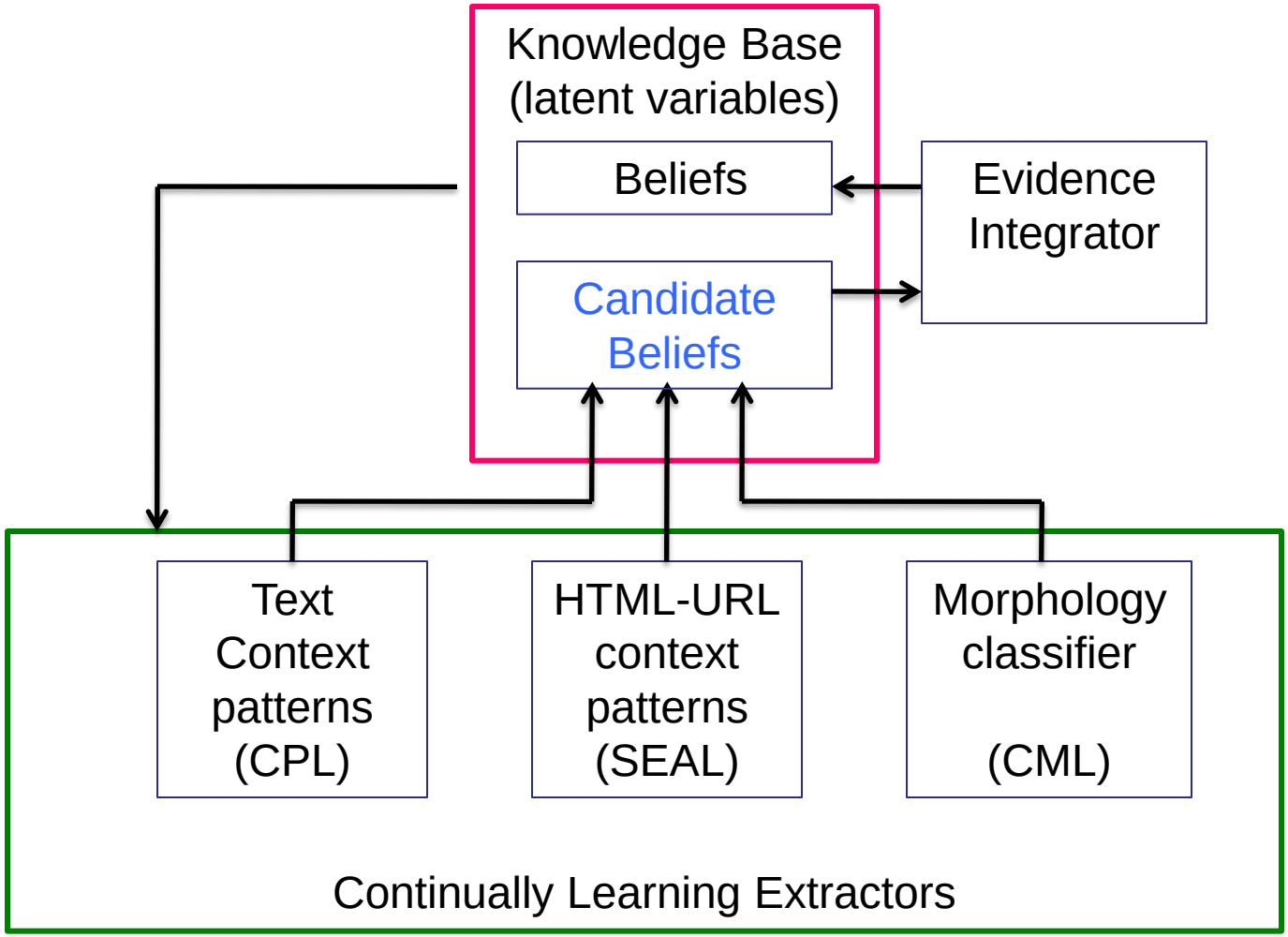
Type 3 Coupling: Argument Types

playsSport(NP1,NP2) → athlete(NP1), sport(NP2)



over 2500 coupled functions in NELL

Basic NELL Architecture



NELL: Learned reading strategies

Plays_Sport(arg1,arg2):

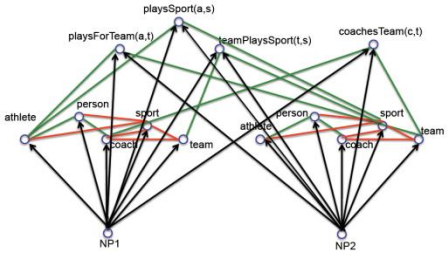
arg1_was_playing_arg2 arg2_megas
arg2_player_named_arg1 arg2_prod
arg1_is_the_tiger_woods_of_arg2 ar
arg2_greats_as_arg1 arg1_plays_arg
arg2_legends_arg1 arg1_announced
arg2_operations_chief_arg1 arg2_pla
arg2_and_golfing_personalities_includ
arg2_greats_like_arg1 arg2_players_
arg2_great_arg1 arg2_champ_arg1
arg2_professionals_such_as_arg1 arg
arg2_icon_arg1 arg2_stars_like_arg1
arg1_retires_from_arg2 arg2_phenom
arg2_architects_robert_trent_jones_ar
arg2_pros_arg1 arg2_stars_venus_a
arg2_superstar_arg1 arg2_legend_a
arg2_players_is_arg1 arg2_pro_arg1
arg2_god_arg1 arg2_idol_arg1 arg1

Predicate	Feature	Weight
mountain	LAST=peak	1.791
mountain	LAST=mountain	1.093
mountain	FIRST=mountain	-0.875
musicArtist	LAST=band	1.853
musicArtist	POS=DT_NNS	1.412
musicArtist	POS=DT_JJ_NN	-0.807
newspaper	LAST=sun	1.330
newspaper	LAST=university	-0.318
newspaper	POS=NN_NNS	-0.798
university	LAST=college	2.076
university	PREFIX=uc	1.999
university	LAST=state	1.992
university	LAST=university	1.745
university	FIRST=college	-1.381
visualArtMovement	SUFFIX=ism	1.282

Predicate	Web URL	Extraction Template
academicField	http://scholendow.ais.msu.edu/student/ScholSearch.Asp	[X] -
athlete	http://www.quotes-search.com/d_occupation.aspx?o=+athlete	<a href='d_author.aspx?a=[X]'>-
bird	http://www.michaelforsberg.com/stock.html	<option>[X]</option>
bookAuthor	http://lifebehindthecurve.com/	[X] by [Y] –

If coupled learning is the key,
how can we get new coupling constraints?

Key Idea 2:



Discover New Coupling Constraints

- first order, probabilistic horn clause constraints:

0.93 athletePlaysSport(?x,?y) ← athletePlaysForTeam(?x,?z)
teamPlaysSport(?z,?y)

  – connects previously uncoupled relation predicates

  – infers new beliefs for KB

Example Learned Horn Clauses

0.95 athletePlaysSport(?x,basketball) ← athleteInLeague(?x,NBA)

0.93 athletePlaysSport(?x,?y) ← athletePlaysForTeam(?x,?z)
teamPlaysSport(?z,?y)

0.91 teamPlaysInLeague(?x,NHL) ← teamWonTrophy(?x,Stanley_Cup)

0.90 athleteInLeague(?x,?y) ←athletePlaysForTeam(?x,?z),
teamPlaysInLeague(?z,?y)

0.88 cityInState(?x,?y) ← cityCapitalOfState(?x,?y), cityInCountry(?y,USA)

0.62* newspaperInCity(?x,New_York) ← companyEconomicSector(?x,media)
generalizations(?x,blog)

Some rejected learned rules

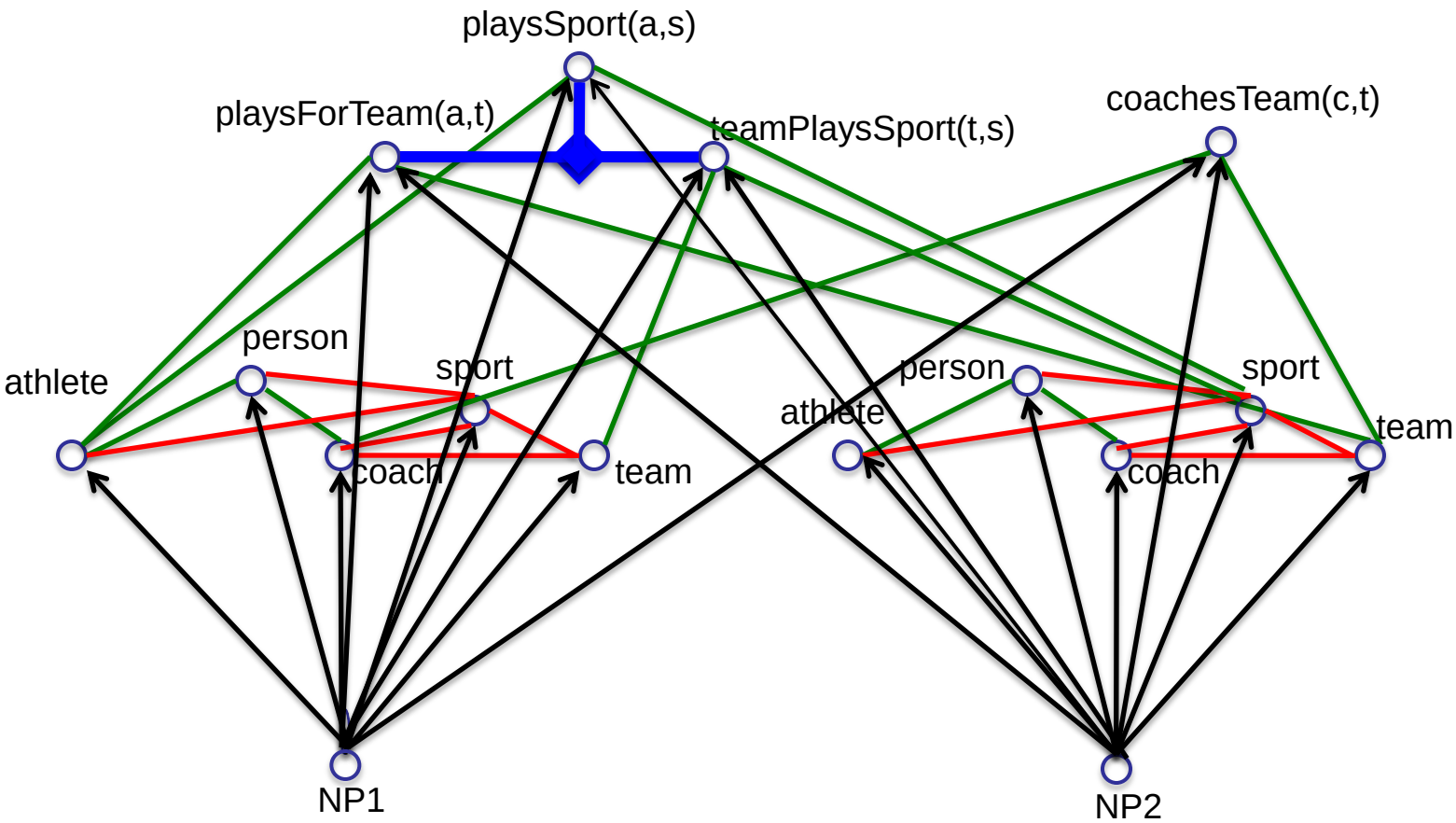
teamPlaysInLeague{?x nba} ← teamPlaysSport{?x basketball}
0.94 [35 0 35] [positive negative unlabeled]

cityCapitalOfState{?x ?y} ← cityLocatedInState{?x ?y}, teamPlaysInLeague{?y nba}
0.80 [16 2 23]

teamplayssport{?x, basketball} ← generalizations{?x, university}
0.61 [246 124 3063]

Learned Probabilistic Horn Clause Rules

0.93 playsSport(?x,?y) ← playsForTeam(?x,?z), teamPlaysSport(?z,?y)



Key Idea 3:

Automatically extend ontology

Ontology Extension (1)

[Mohamed et al., *EMNLP* 2011]

Goal:

- Add new relations to ontology

Approach:

- For each pair of categories C1, C2,
  - co-cluster pairs of known instances, and text contexts that connect them

Example Discovered Relations

[Mohamed et al. *EMNLP* 2011]

Category Pair	Text contexts	Extracted Instances	Suggested Name
MusicInstrument Musician	ARG1 master ARG2 ARG1 virtuoso ARG2 ARG1 legend ARG2 ARG2 plays ARG1	sitar , George Harrison tenor sax, Stan Getz trombone, Tommy Dorsey vibes, Lionel Hampton	Master
Disease Disease	ARG1 is due to ARG2 ARG1 is caused by ARG2	pinched nerve, herniated disk tennis elbow, tendonitis blepharospasm, dystonia	IsDueTo
CellType Chemical	ARG1 that release ARG2 ARG2 releasing ARG1	epithelial cells, surfactant neurons, serotonin mast cells, histomine	ThatRelease
Mammals Plant	ARG1 eat ARG2 ARG2 eating ARG1	koala bears, eucalyptus sheep, grasses goats, saplings	Eat
River City	ARG1 in heart of ARG2 ARG1 which flows through ARG2	Seine, Paris Nile, Cairo Tiber river, Rome	InHeartOf

NELL: recently self-added relations

- athleteWonAward
- animalEatsFood
- languageTaughtInCity
- clothingMadeFromPlant
- beverageServedWithFood
- fishServedWithFood
- athleteBeatAthlete
- athleteInjuredBodyPart
- arthropodFeedsOnInsect
- animalEatsVegetable
- plantRepresentsEmotion
- foodDecreasesRiskOfDisease
- clothingGoesWithClothing
- bacteriaCausesPhysCondition
- buildingMadeOfMaterial
- emotionAssociatedWithDisease
- foodCanCauseDisease
- agriculturalProductAttractsInsect
- arteryArisesFromArtery
- countryHasSportsFans
- bakedGoodServedWithBeverage
- beverageContainsProtein
- animalCanDevelopDisease
- beverageMadeFromBeverage

Key Idea 4: Cumulative, Staged Learning

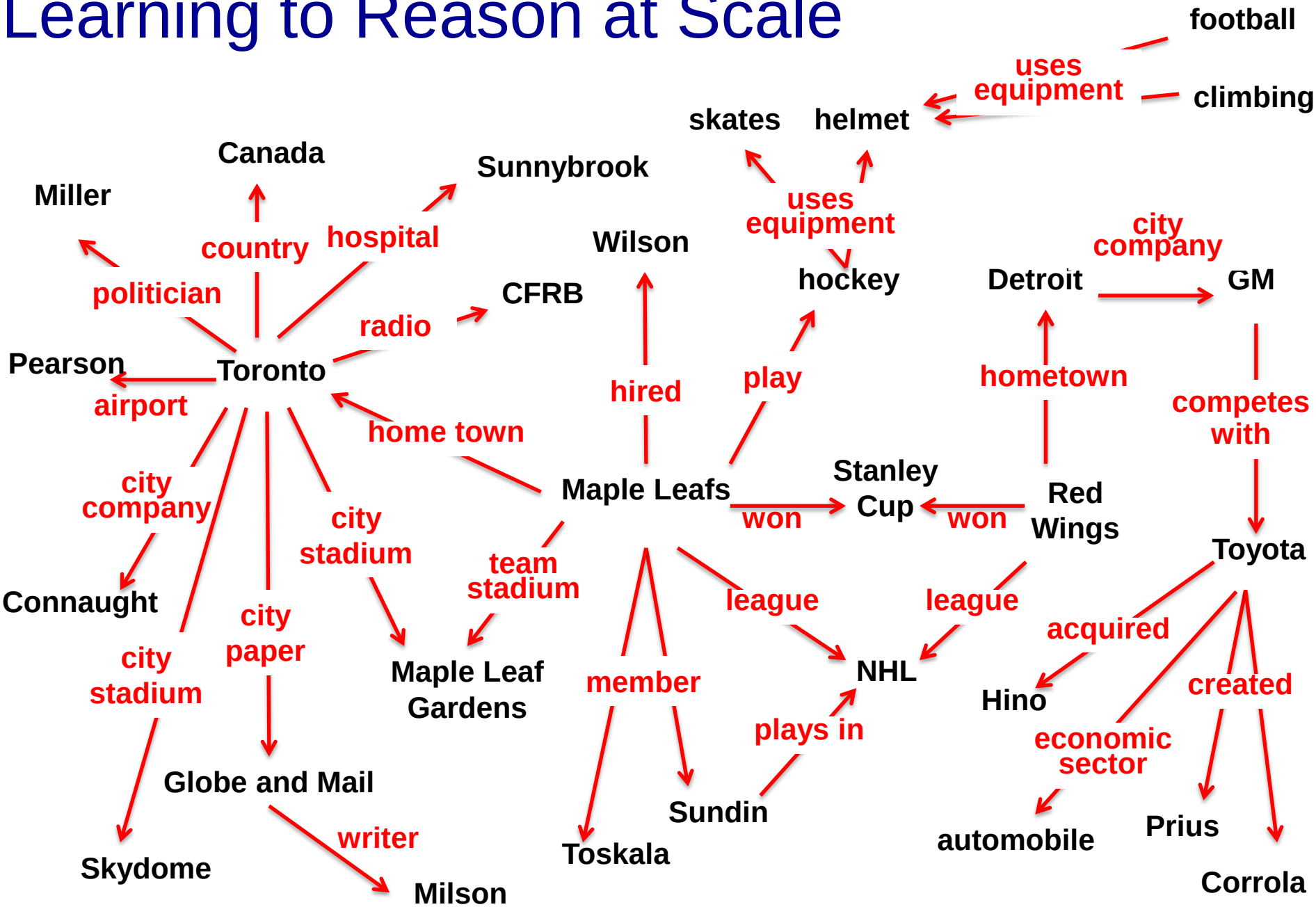
Learning X improves ability to learn Y

1. Classify noun phrases (NP's) by category
2. Classify NP pairs by relation
3. Discover rules to predict new relation instances
4. Learn which NP's (co)refer to which concepts
5. Discover new relations to extend ontology
6. Learn to infer relation instances via targeted random walks
7. Learn to assign temporal scope to beliefs
8. Learn to microread single sentences

---

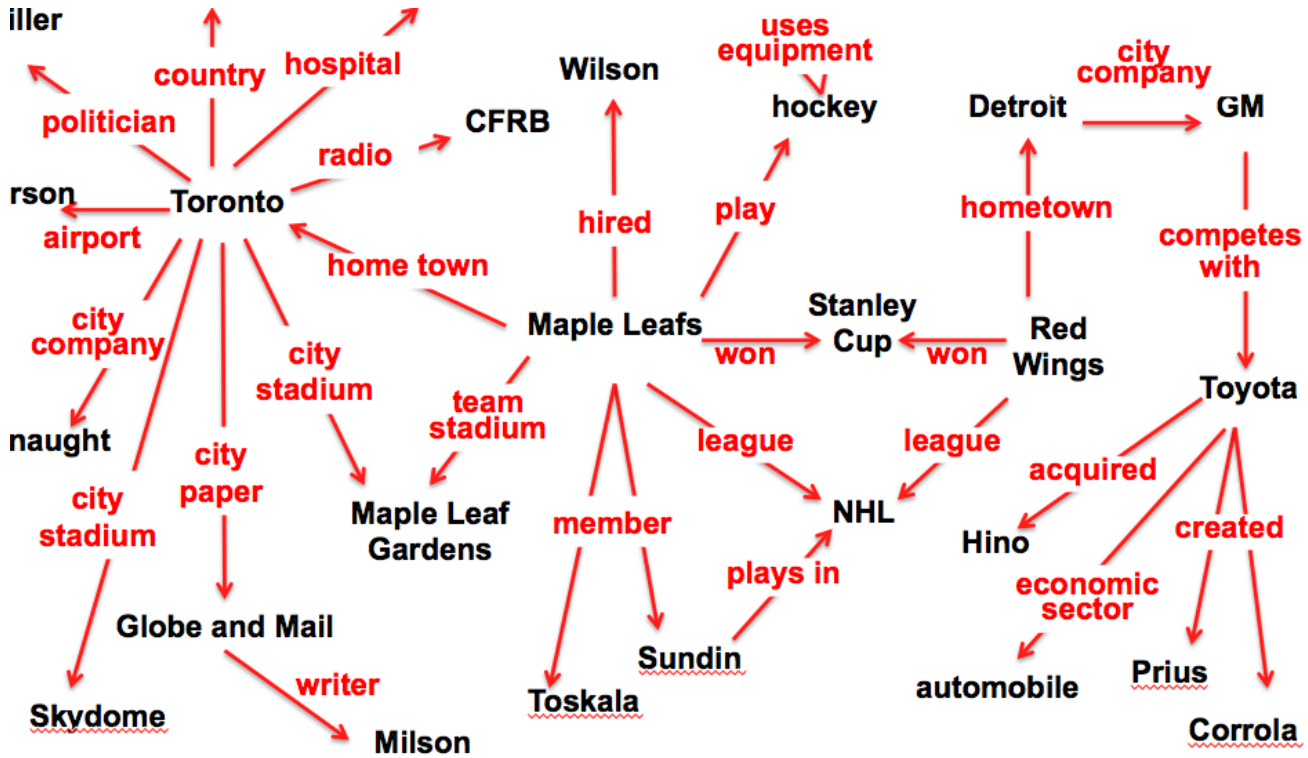
9. Vision: co-train text and visual object recognition
10. Goal-driven reading: predict, then read to corroborate/correct
11. Make NELL a conversational agent on Twitter

Learning to Reason at Scale



Inference by KB Random Walks

[Lao et al, *EMNLP* 2011]



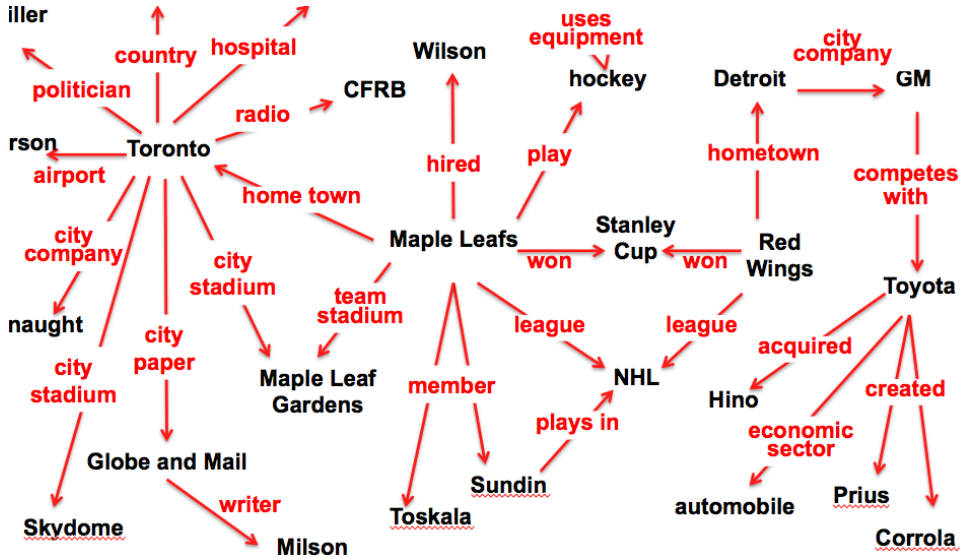
If: x1 — competes with (x1,x2) → x2 — economic sector (x2, x3) → x3

Then: economic sector (x1, x3)

Inference by KB Random Walks

[Lao et al, *EMNLP* 2011]

KB:



Random walk path type:

? —competes with→ ? —economic sector→ ?

Infer Pr(R(x,y)): Trained logistic function for R, where i^{th} feature is probability of arriving at node y when starting at node x, and taking a random walk along path type i

CityLocatedInCountry(Pittsburgh) = ?

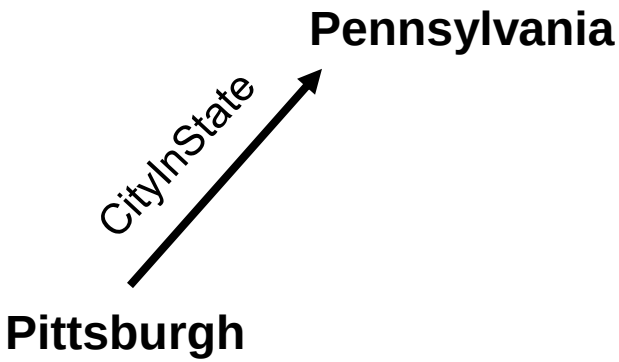
[Lao et al, *EMNLP* 2011]

Pittsburgh

Feature = Typed Path	**Feature Value**	**Logistic Regresssion Weight**
CityInState, CityInstate^{-1}, CityLocatedInCountry	0.8	0.32

CityLocatedInCountry(Pittsburgh) = ?

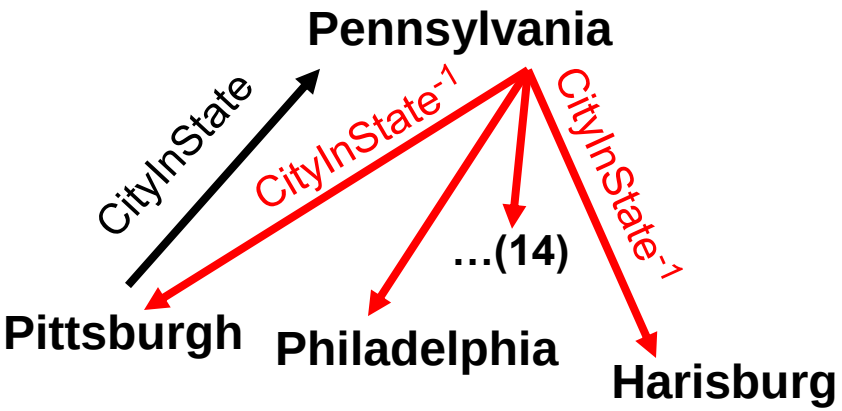
[Lao et al, *EMNLP* 2011]



Feature = Typed Path	**Feature Value**	**Logistic Regresssion Weight**
CityInState, CityInstate^{-1}, CityLocatedInCountry	0.8	0.32

CityLocatedInCountry(Pittsburgh) = ?

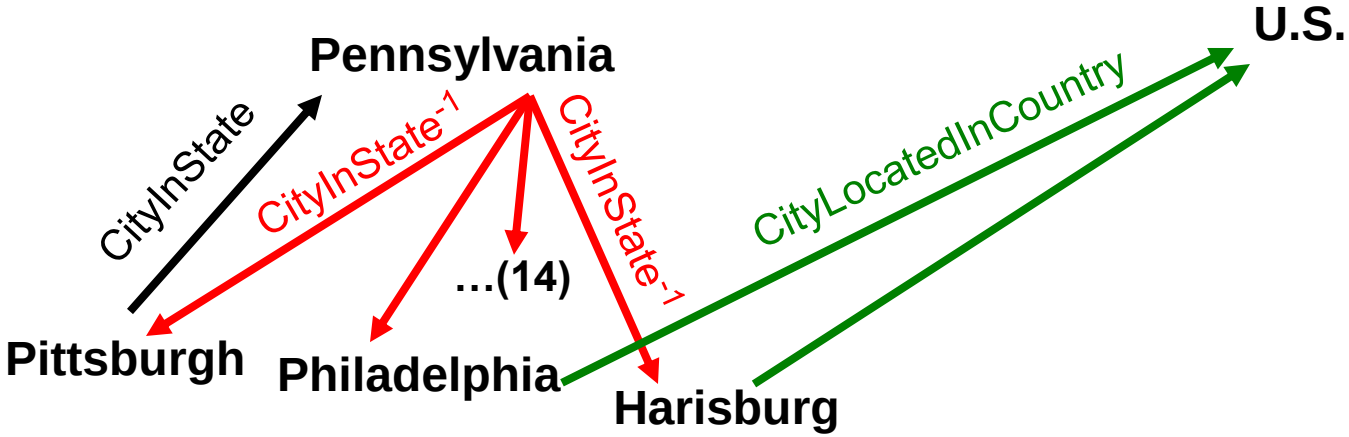
[Lao et al, *EMNLP* 2011]



Feature = Typed Path	**Feature Value**	**Logistic Regresssion Weight**
CityInState, CityInstate^{-1}, CityLocatedInCountry	0.8	0.32

CityLocatedInCountry(Pittsburgh) = ?

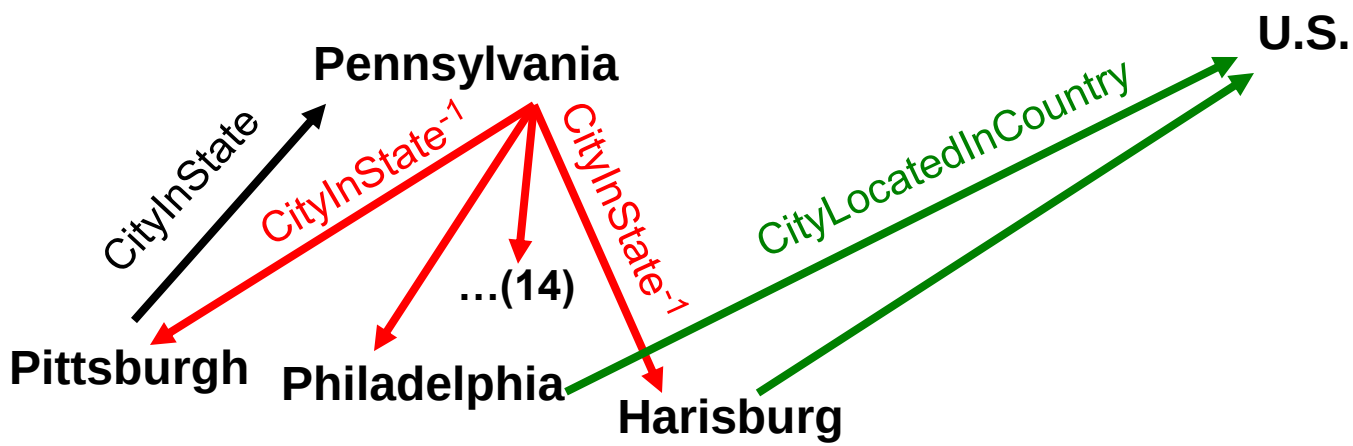
[Lao et al, *EMNLP* 2011]



Feature = Typed Path	**Feature Value**	**Logistic Regresssion Weight**
CityInState, CityInstate^{-1}, CityLocatedInCountry	0.8	0.32

CityLocatedInCountry(Pittsburgh) = ?

[Lao et al, *EMNLP* 2011]

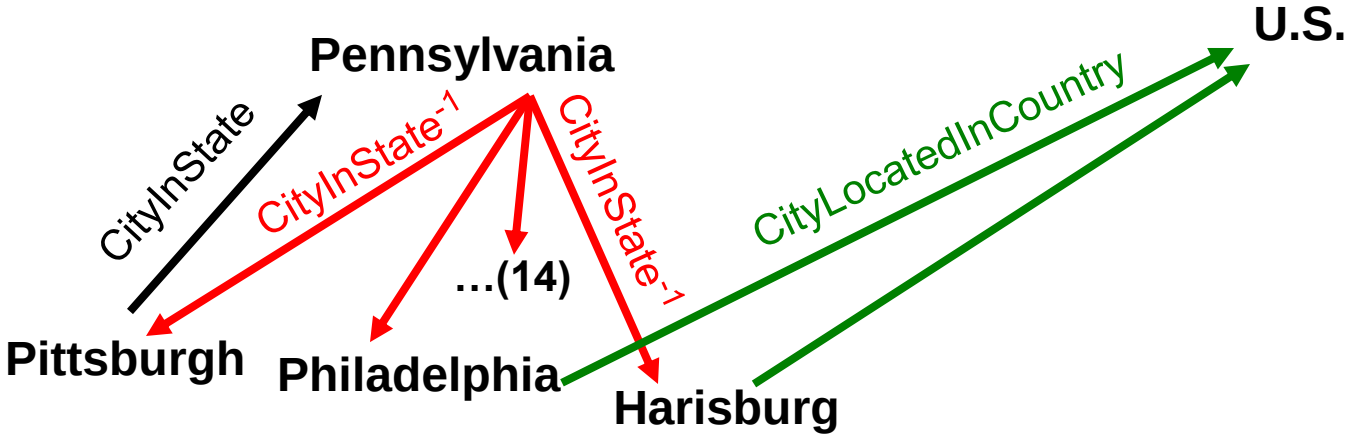


Pr(U.S. | Pittsburgh, TypedPath)

Feature = Typed Path	**Feature Value**	**Logistic Regresssion Weight**
CityInState, CityInstate^{-1}, CityLocatedInCountry	0.8	0.32

CityLocatedInCountry(Pittsburgh) = ?

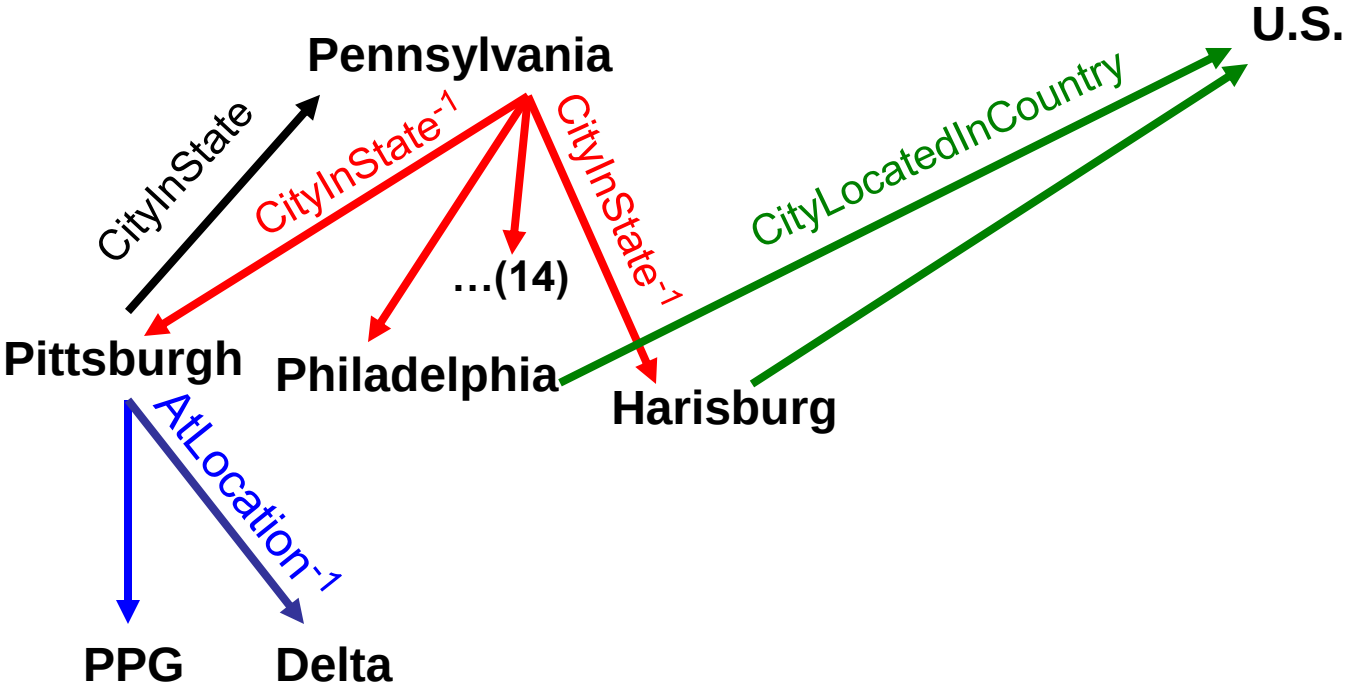
[Lao et al, *EMNLP* 2011]



Feature = Typed Path	**Feature Value**	**Logistic Regresssion Weight**
CityInState, CityInstate^{-1}, CityLocatedInCountry	0.8	0.32
AtLocation^{-1}, AtLocation, CityLocatedInCountry		0.20

CityLocatedInCountry(Pittsburgh) = ?

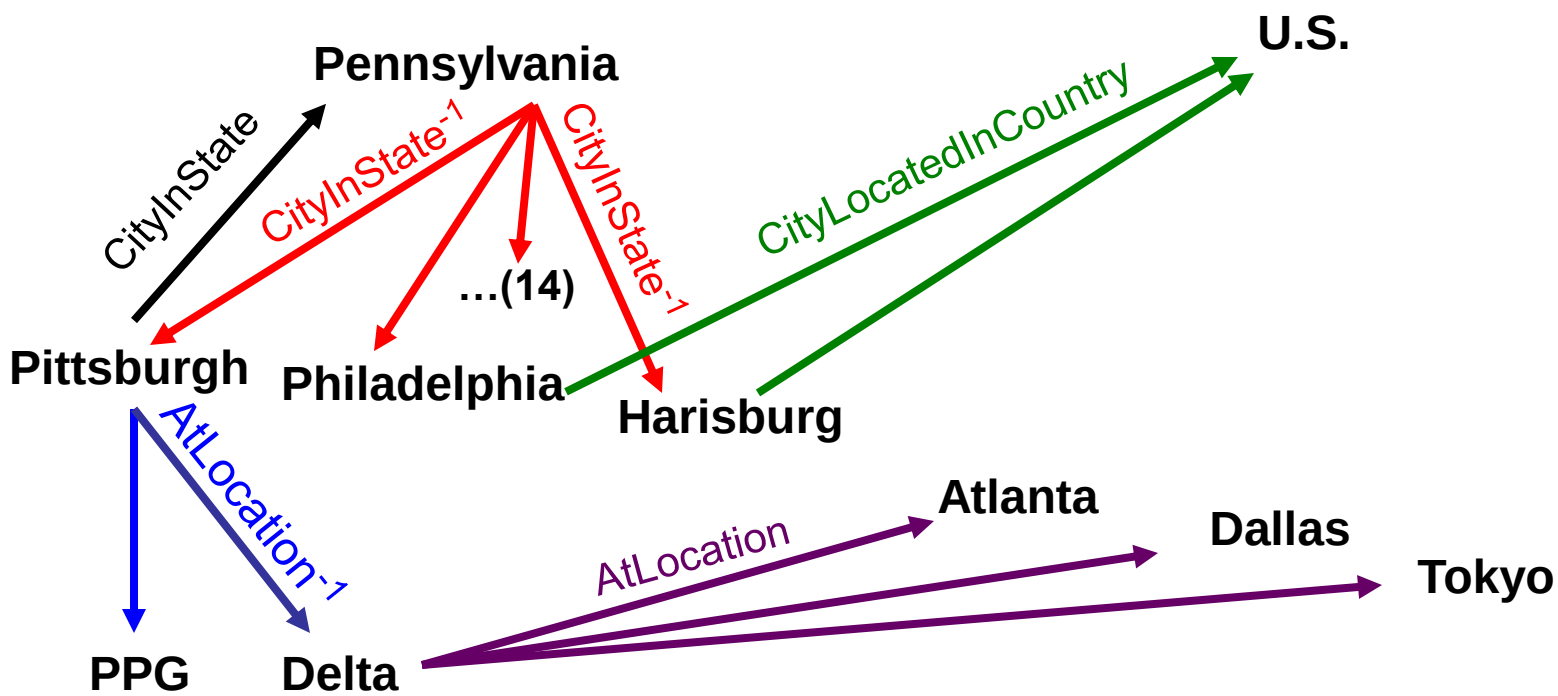
[Lao et al, *EMNLP* 2011]



Feature = Typed Path	**Feature Value**	**Logistic Regresssion Weight**
CityInState, CityInstate^{-1}, CityLocatedInCountry	0.8	0.32
AtLocation^{-1}, AtLocation, CityLocatedInCountry		0.20

CityLocatedInCountry(Pittsburgh) = ?

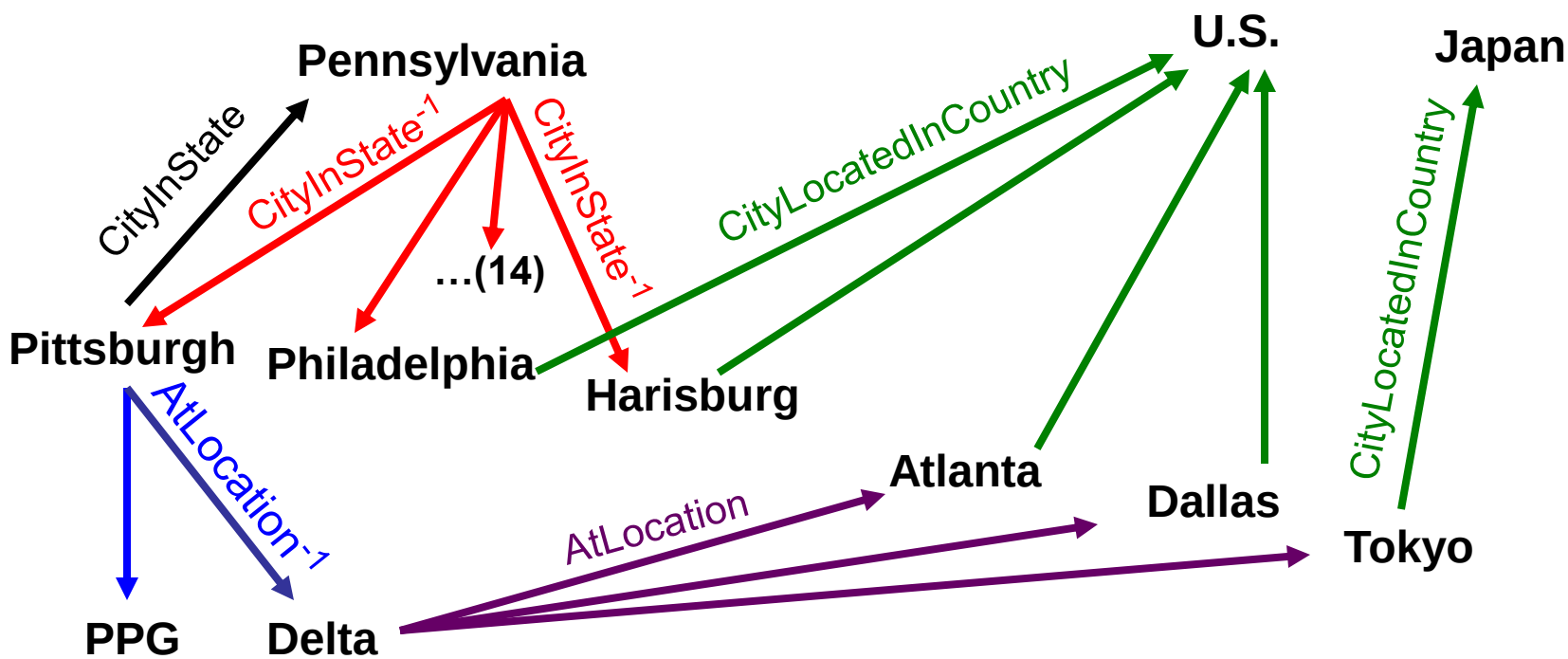
[Lao et al, *EMNLP* 2011]



Feature = Typed Path	**Feature Value**	**Logistic Regresssion Weight**
CityInState, CityInstate^{-1}, CityLocatedInCountry	0.8	0.32
AtLocation^{-1}, AtLocation, CityLocatedInCountry		0.20

CityLocatedInCountry(Pittsburgh) = ?

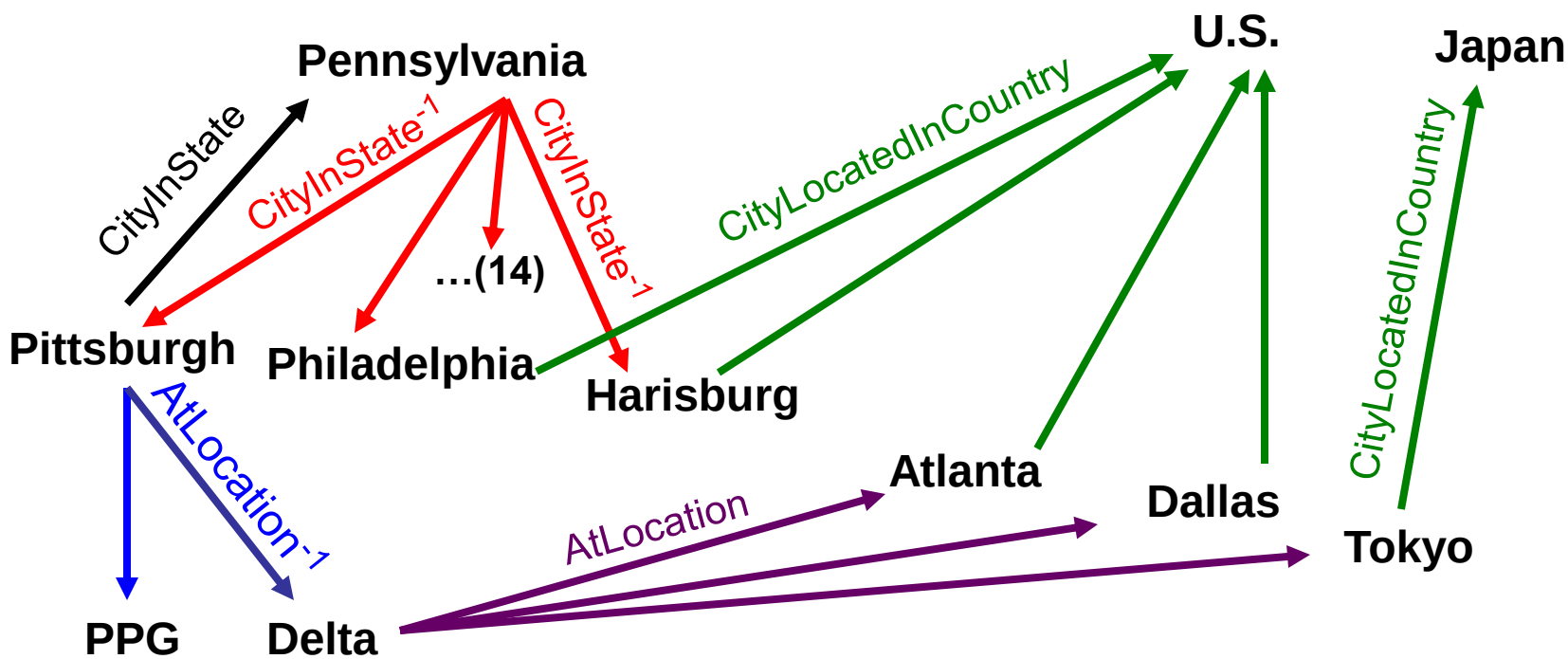
[Lao et al, *EMNLP* 2011]



Feature = Typed Path	Feature Value	Logistic Regresssion Weight
CityInState, CityInstate^{-1}, CityLocatedInCountry	0.8	0.32
AtLocation^{-1}, AtLocation, CityLocatedInCountry	0.6	0.20

CityLocatedInCountry(Pittsburgh) = ?

[Lao et al, *EMNLP* 2011]



Feature = Typed Path	**Feature Value**	**Logistic Regresssion Weight**
CityInState, CityInstate^{-1}, CityLocatedInCountry	0.8	0.32
AtLocation^{-1}, AtLocation, CityLocatedInCountry	0.6	0.20
…	…	…

CityLocatedInCountry(Pittsburgh) = U.S. p=0.58

Random walk inference: learned path types

CityLocatedInCountry(*city, country*):

8.04 cityliesonriver, cityliesonriver^{-1}, citylocatedincountry

5.42 hasofficeincity^{-1}, hasofficeincity, citylocatedincountry

4.98 cityalsoknownas, cityalsoknownas, citylocatedincountry

2.85 citycapitalofcountry,citylocatedincountry^{-1},citylocatedincountry

2.29 agentactsinlocation^{-1}, agentactsinlocation, citylocatedincountry

1.22 statehascapital^{-1}, statelocatedincountry

0.66 citycapitalofcountry

.
.
.

7 of the 2985 paths for inferring CityLocatedInCountry

Random Walk Inference: Example

Rank 17 companies by probability competesWith(MSFT,X):

NELL/PRA ranking

Google
Oracle
IBM
Apple
SAP
Yahoo
Facebook
Redhat
Lenovo
FedEx
SAS
Boeing
Honda
Dupont
Lufthansa
Exxon
Pfizer

Random Walk Inference: Example

Rank 17 companies by probability competesWith(MSFT,X):

NELL/PRA ranking	Human Ranking (9 subjs)
Google	Apple
Oracle	Google
IBM	Yahoo
Apple	IBM
SAP	Redhat
Yahoo	Oracle
Facebook	Facebook
Redhat	SAP
Lenovo	SAS
FedEx	Lenovo
SAS	Boeing
Boeing	Honda
Honda	FedEx
Dupont	Dupont
Lufthansa	Exxon
Exxon	Lufthansa
Pfizer	Pfizer

1. Tractable (bounded length)
2. Anytime
3. Accuracy increases as KB grows
4. Addresses question of how to combine probabilities from different horn clauses

demo

Random walk inference: learned path types

KB graph augmented by
Subj-Verb-Obj
corpus statistics

CompetesWith(*company, company*):

5.29 companyAlsoKnownAs, competesWith

2.12 companyAlsoKnownAs, producesProduct, agentInvolvedWith^{-1}

0.77 companyAlsoKnownAs, *subj_offer_obj*, *subj_offer_obj*$^{-1}$

0.65 companyEconomicSector, companyEconomicSector^{-1}

- 0.19 companyAlsoKnownAs
- 0.38 companyAlsoKnownAs, companyAlsoKnownAs

.
.
.

6 of the 7966 path types learned for CompetesWith

Summary

Key ideas:

- Coupled semi-supervised learning
- Learn new coupling constraints (Horn clauses)
- Automatically extend ontology
- Learn progressively more challenging types of K
- Scalable random walk probabilistic inference
  - Integrating symbolic extracted beliefs,
    + subsymbolic corpus statistics

thank you

and thanks to:

Darpa, Google, NSF, Intel, Yahoo!, Microsoft, Fullbright