Never Ending Learning

Tom Mitchell

Machine Learning Department
Carnegie Mellon University

New paradigm for Machine Learning:

Never-ending learning agents

- Persistent software individual
- Learns many functions / knowledge types
- Learns easier things first, then more difficult
- The more it learns, the more it can learn next
- Learns from experience, and from advice

NELL: Never-Ending Language Learner

Inputs:

- initial ontology
- dozen 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 ontology
    2. learn to read (perform #1) better than yesterday

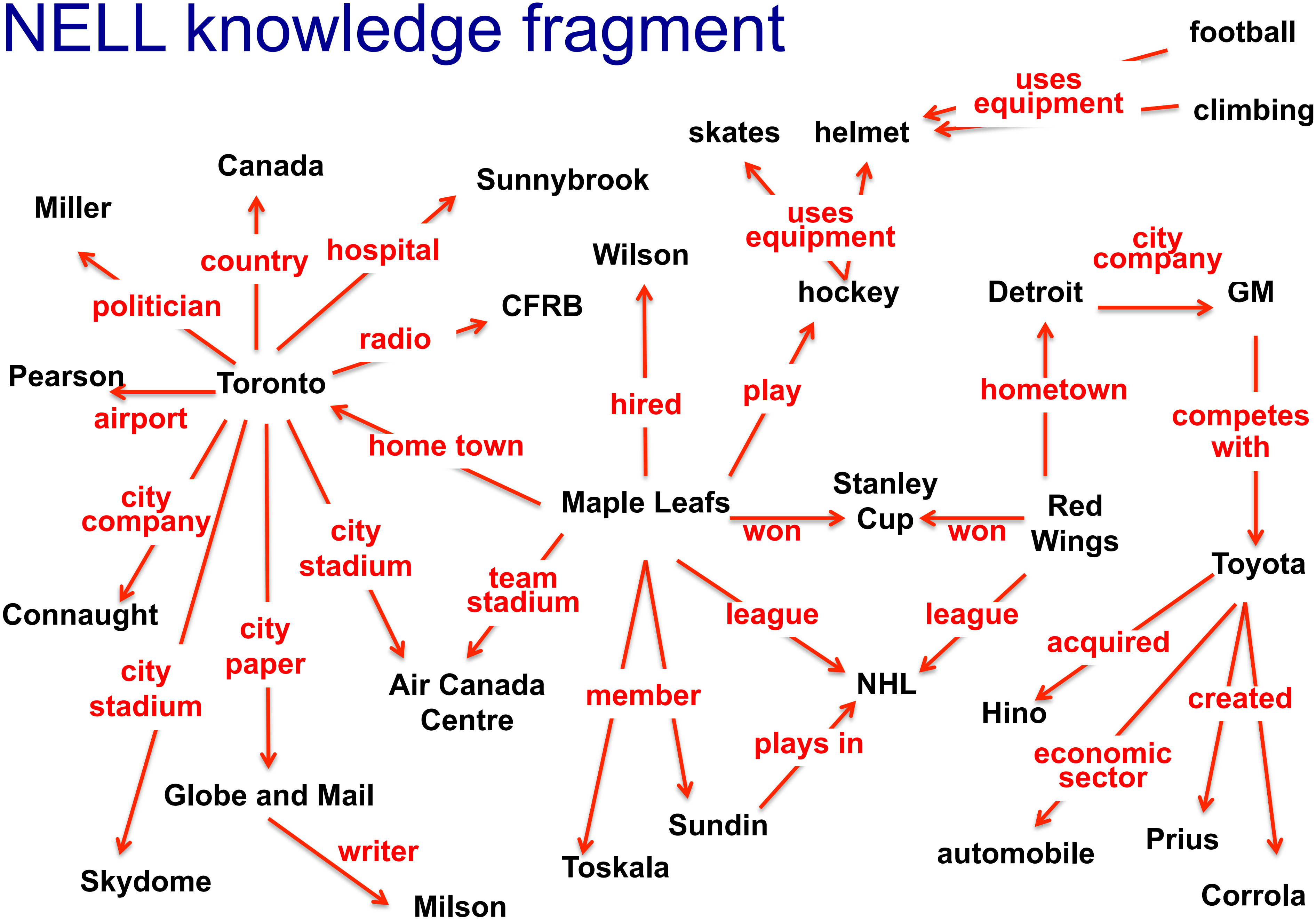
NELL today

Running 24x7, since January, 12, 2010

Result:

- KB with > 50 million candidate beliefs, growing daily
- learning to read better each day
- learning to reason, as well as read
- automatically extending its ontology

NELL knowledge fragment



NELL Today

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

Recently-Learned Facts

instance	iteration	date learned	conf
sacramento_convention_center is a stadium or event venue	737	04-jun-2013	
john_kenneth_macalister is a criminal	737	04-jun-2013	
birth_control_drugs is a personal care product	737	04-jun-2013	
almond_chocolate is a kind of candy	742	18-jun-2013	
garlic_shoots is an agricultural product	739	09-jun-2013	
hagar has husband abraham	742	18-jun-2013	
dave_murray is a musician who plays the guitar	739	09-jun-2013	
hart is a city located in the state or province georgia	742	18-jun-2013	
wood_prarie_farm is a farm in the state or province maine	742	18-jun-2013	
pepper is an agricultural product that is usually cooked with canola_oil	737	04-jun-2013	

How does NELL work?

Semi-Supervised Bootstrap Learning

it's underconstrained!!

Find cities:

Paris
Pittsburgh
Seattle
Montpelier

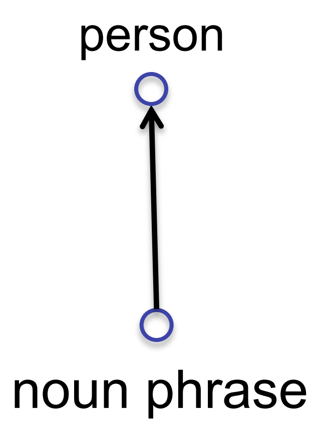
mayor of arg1
live in arg1

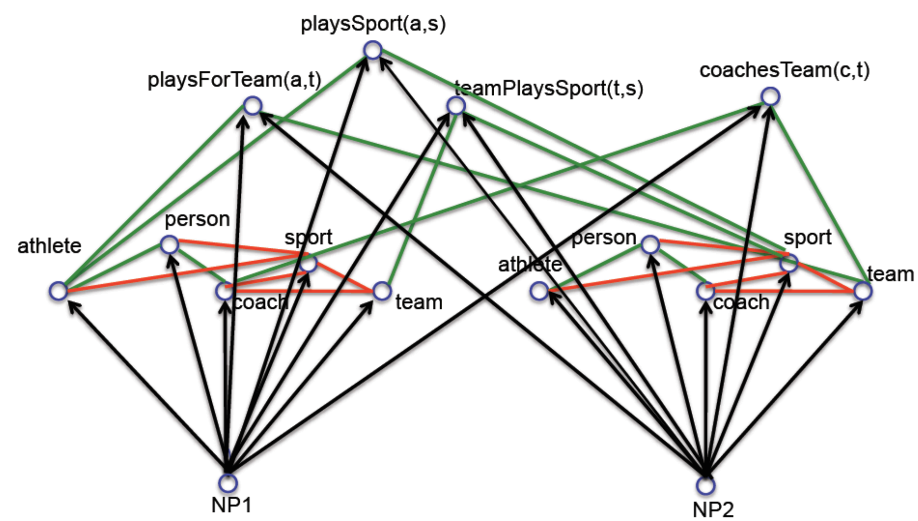
San Francisco
Berlin
denial

arg1 is home of
traits such as arg1

anxiety
selfishness
London

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

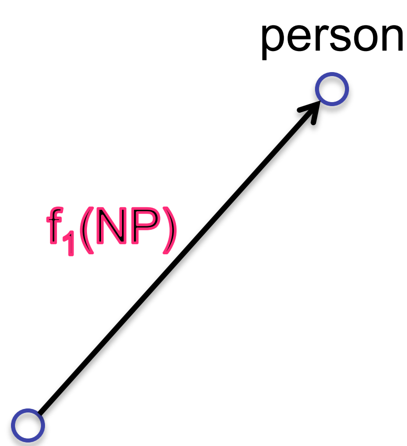




hard (underconstrained) semi-supervised learning problem

much easier (more constrained) semi-supervised learning problem

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



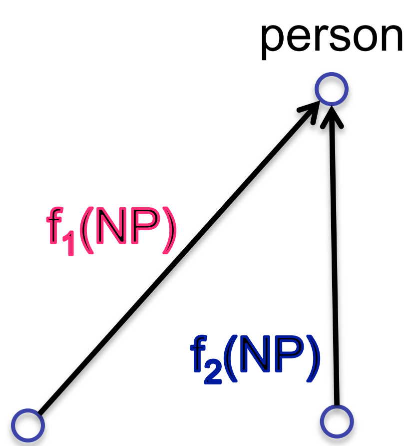
Supervised training of 1 function:

Minimize: $$\sum_{<np,person> \,\in\, labeled\ data} |f_1(np) - person|$$

NP: NP context distribution

__ is a friend
rang the __
...
__ walked in

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



NP: NP context distribution | NP morphology

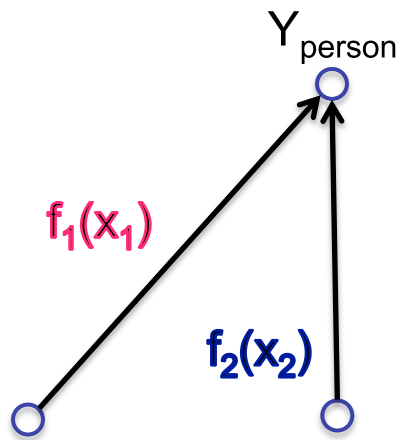
__ is a friend
rang the __
…
__ walked in

capitalized?
ends with '...ski'?
…
contains "univ."?

Coupled training of 2 functions:

Minimize:
$$\sum_{<np,person> \in \, labeled\ data} |f_1(np) - person|$$
$$+ \sum_{<np,person> \in \, labeled\ data} |f_2(np) - person|$$
$$+ \sum_{np \in \, unlabeled\ data} |f_1(np) - f_2(np)|$$

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



NP: X_1 = context distribution | X_2 = morphology

X_1 = context distribution	X_2 = morphology
__ is a friend	***capitalized?***
rang the __	***ends with '...ski'?***
…	***…***
__ walked in	***contains "univ."?***

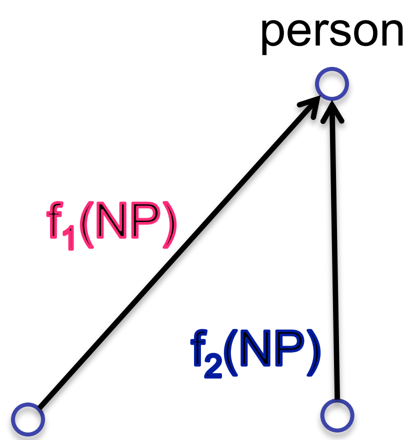
Theorem (Blum & Mitchell, 1998):

If f_1,and f_2 are PAC learnable from noisy *labeled* data, and X_1, X_2 are conditionally independent given Y,

Then f_1, f_2 are PAC learnable from polynomial *unlabeled* data plus a weak initial predictor

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]

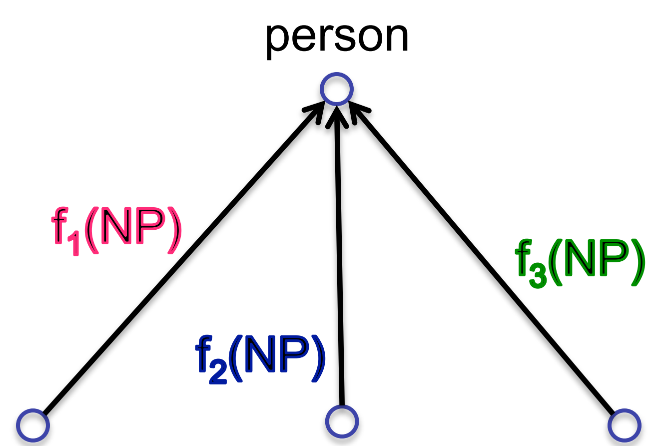


NP:

NP text context distribution	NP morphology
__ is a friend	*capitalized?*
rang the __	*ends with '...ski'?*
…	*…*
__ walked in	*contains "univ."?*

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]



NP:

NP text context distribution

NP morphology

NP HTML contexts

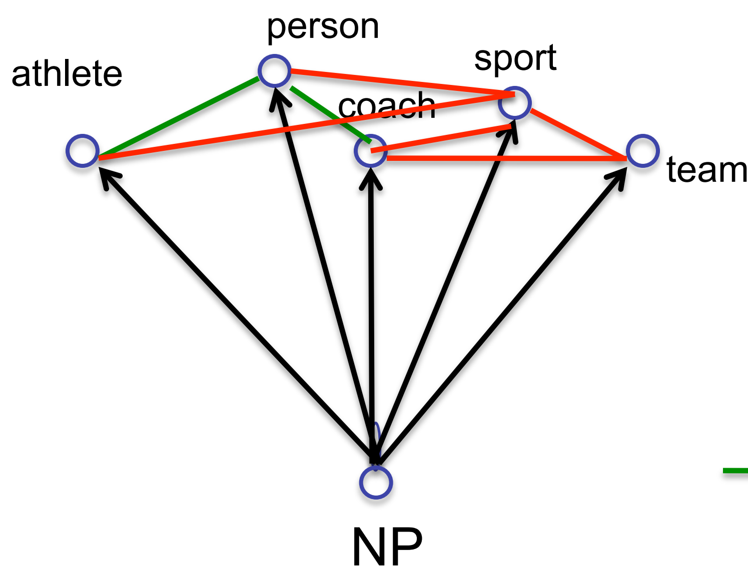
__ is a friend
rang the __
…
__ walked in

capitalized?
ends with '…ski'?
…
contains "univ."?

www.celebrities.com:
*** __ ***
…

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]

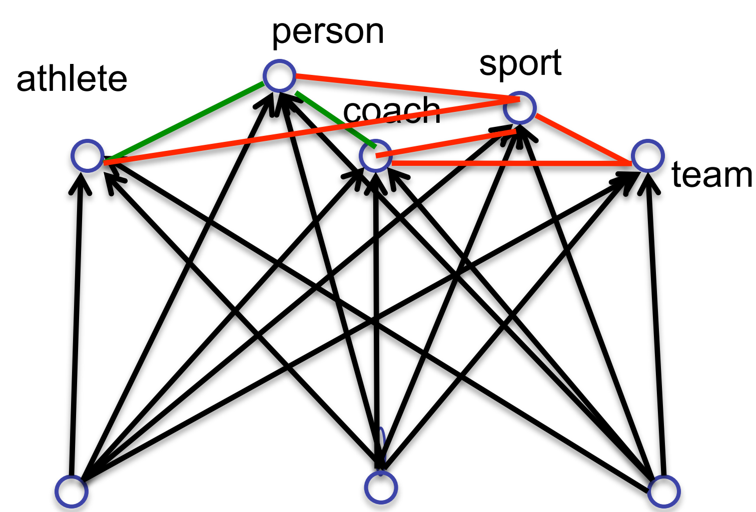


athlete(NP) → person(NP)

athlete(NP) → NOT sport(NP)

NOT athlete(NP) ← sport(NP)

Multi-view, Multi-Task Coupling



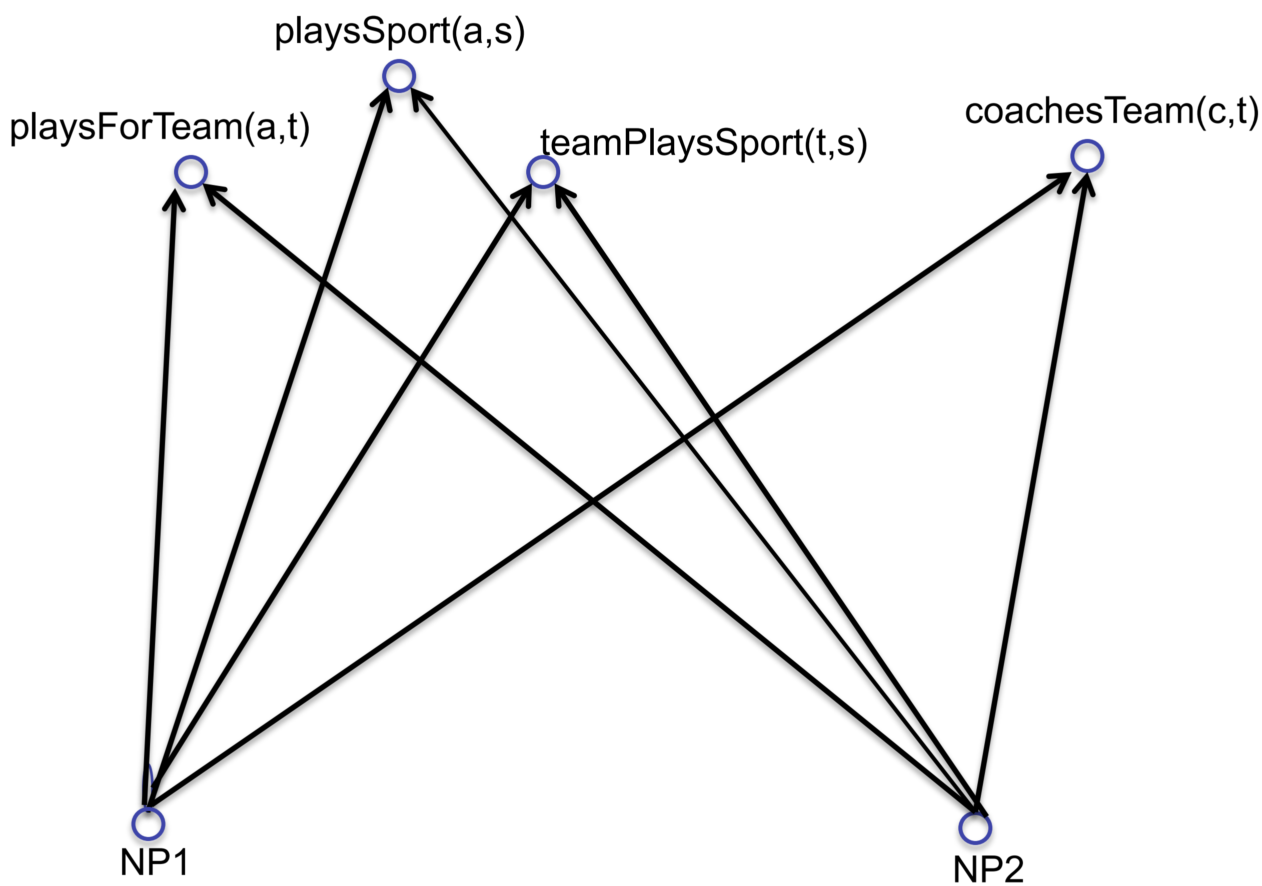
NP:

NP text context distribution

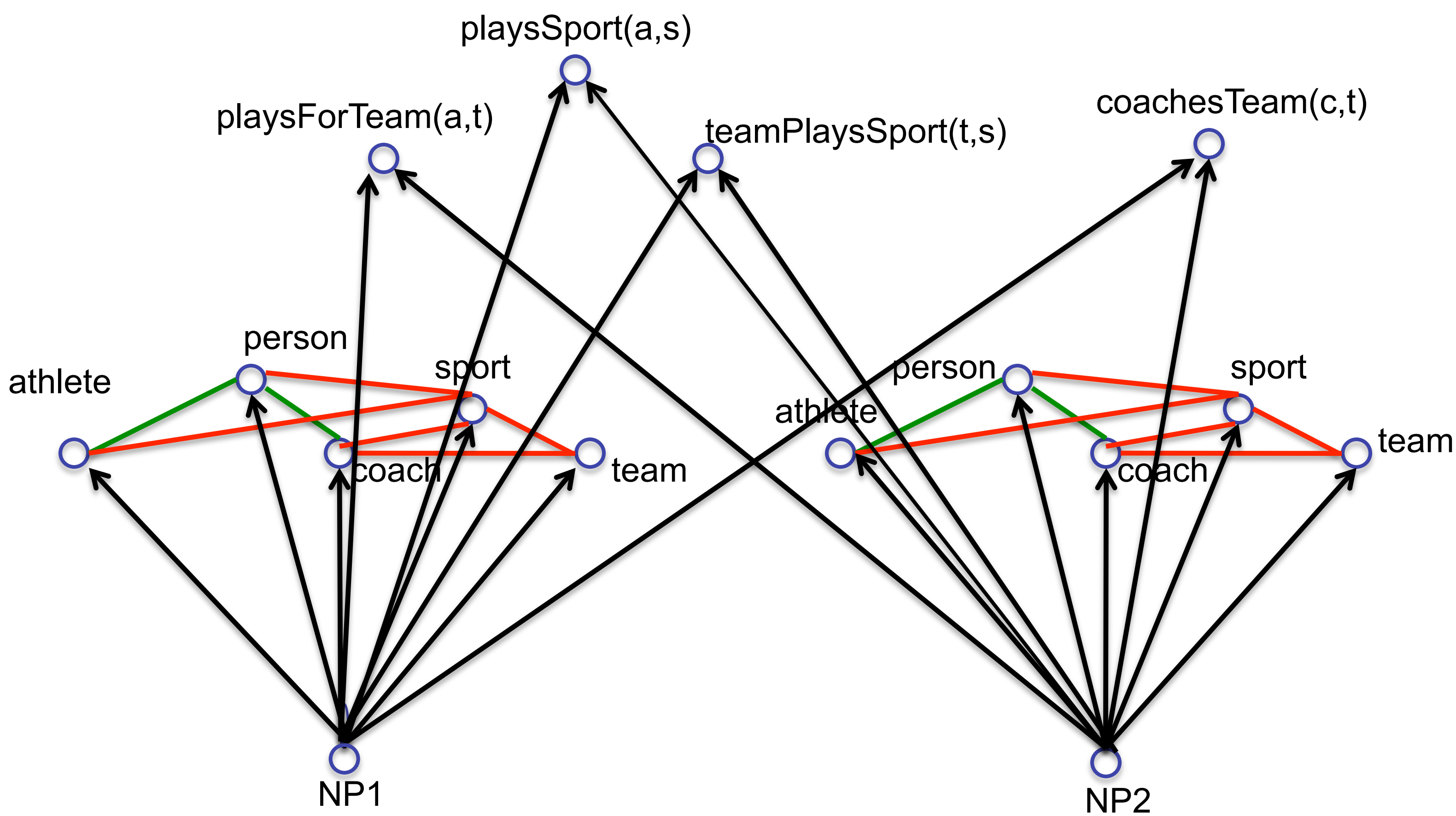
NP morphology

NP HTML contexts

Type 3 Coupling: Learning Relations

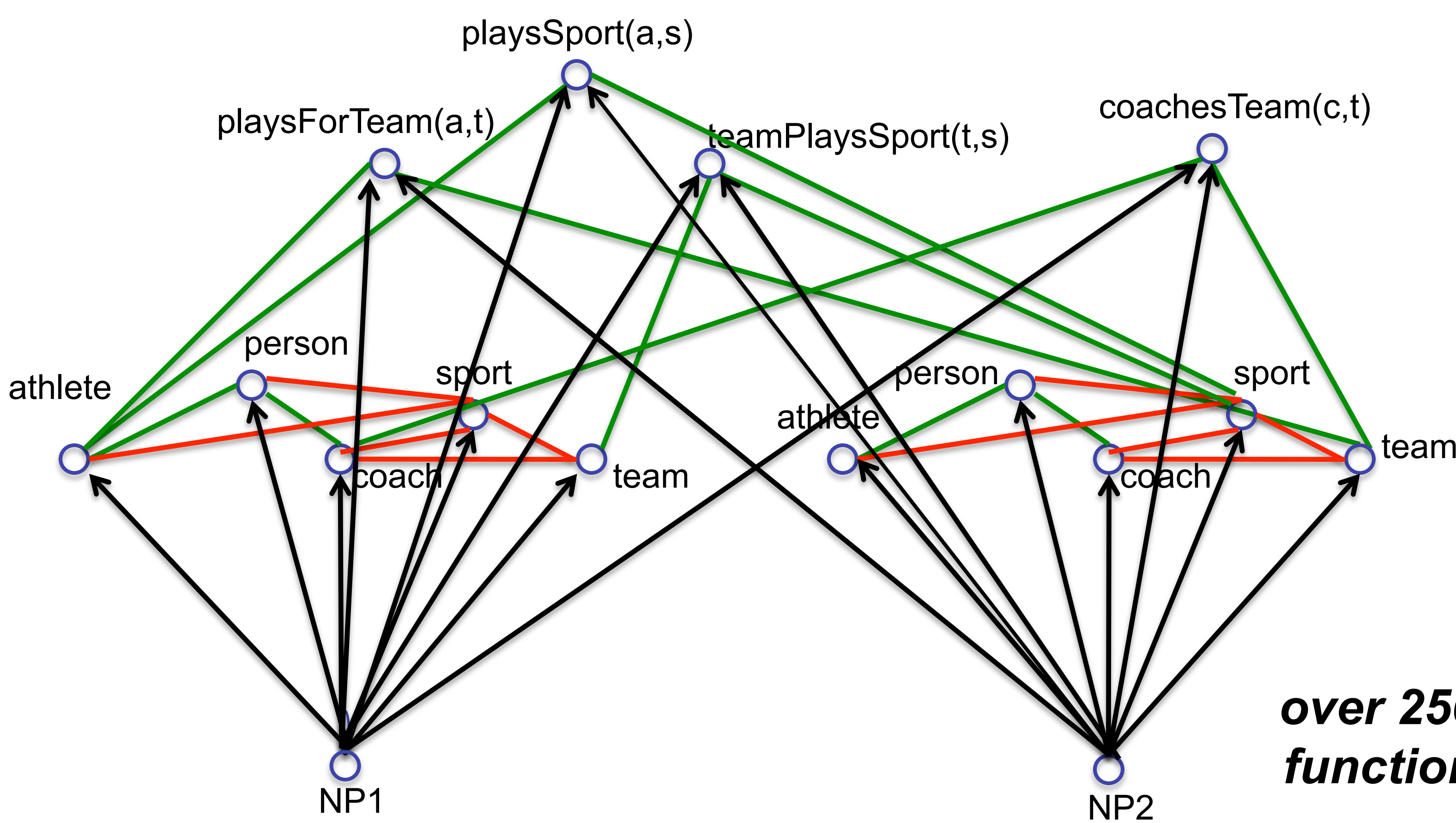


Type 3 Coupling: Argument Types



Type 3 Coupling: Argument Types

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



over 2500 coupled functions in NELL

NELL: Learned reading strategies

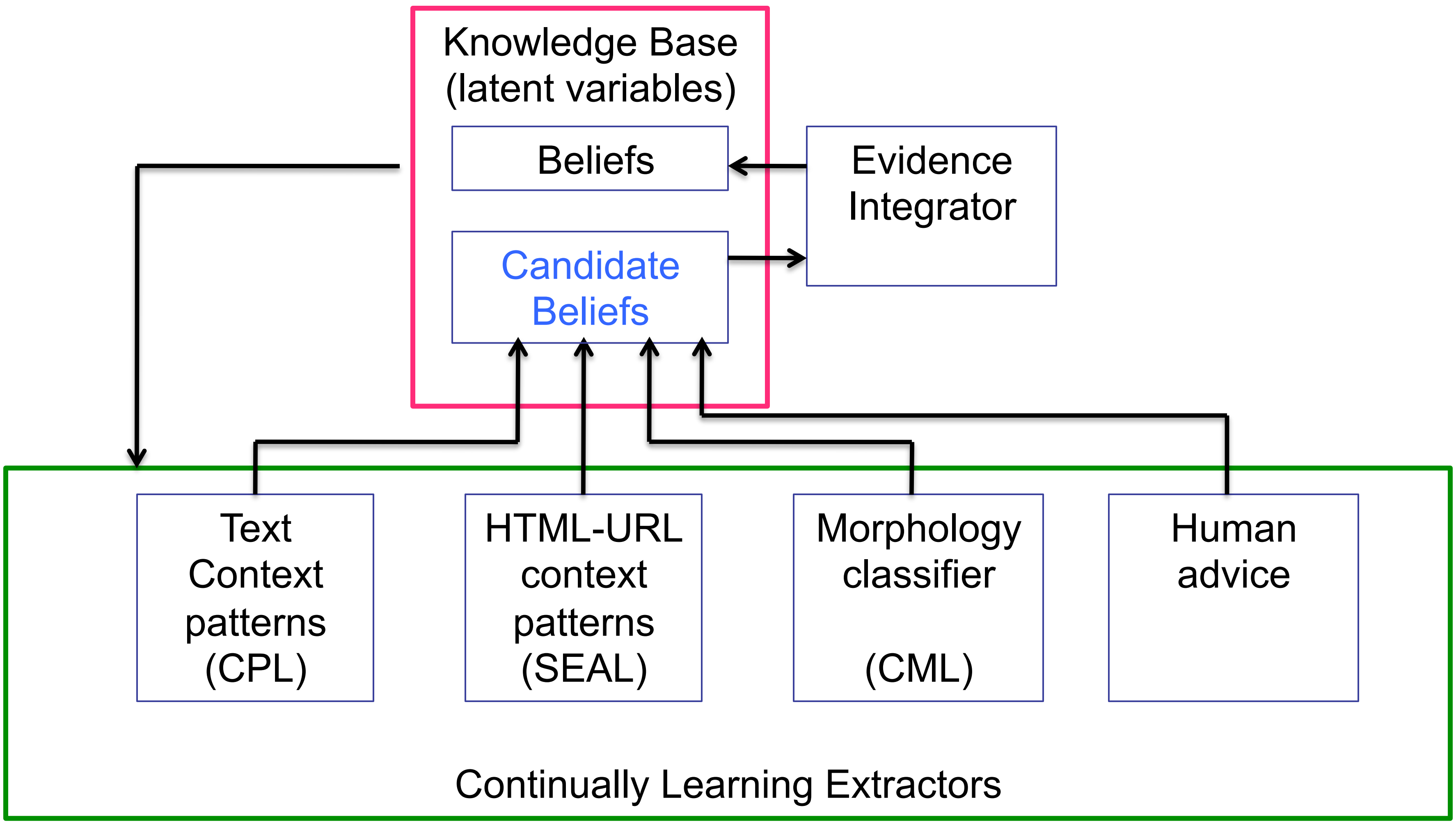
Plays_Sport(arg1,arg2):

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_ar
arg2_legends_arg1 arg1_announced
arg2_operations_chief_arg1 arg2_pla
arg2_and_golfing_personalities_incluc
arg2_greats_like_arg1 arg2_players_
arg2_great_arg1 arg2_champ_arg1
arg2_professionals_such_as_arg1 ar
arg2_icon_arg1 arg2_stars_like_arg1
arg1_retires_from_arg2 arg2_phenor
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

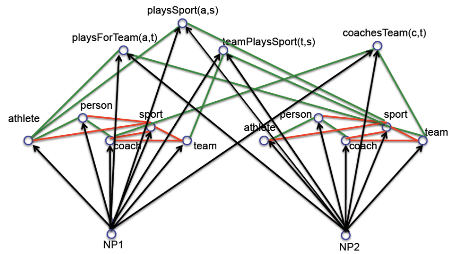
Predicate	Web URL	Extraction Template
academicField	http://scholendow.ais.msu.edu/student/ScholSearch.Asp	` [X] -`
athlete	http://www.quotes-search.com/d_occupation.aspx?o=+athlete	`<a href='d_author.aspx?a=[X]'>-`
bird	http://www.michaelforsberg.com/stock.html	`<option>[X]</option>`
bookAuthor	http://lifebehindthecurve.com/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

Initial NELL Architecture



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
  - modified version of FOIL [Quinlan]
  - restricted rule language: form connected KB subgraphs

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)

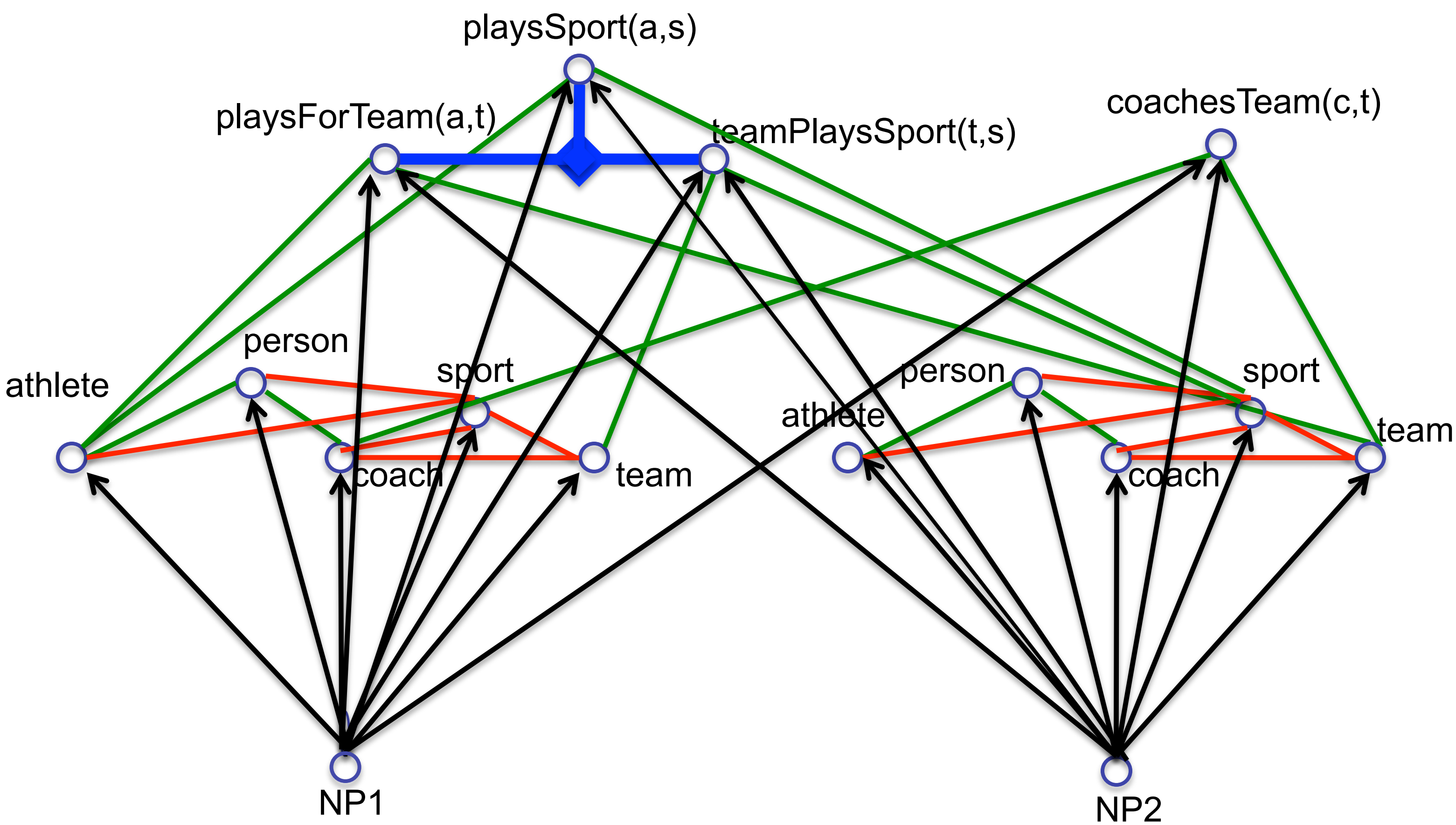
0.90 athleteInLeague(?x,?y) ←athletePlaysForTeam(?x,?z),
teamPlaysInLeague(?z,?y)

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

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

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: sample of 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

Ontology Extension (2)

[Burr Settles]

Goal:

- Add new subcategories

Approach:

- For each category C,
  - train NELL to read the relation
    $\text{SubsetOf}_C: C \rightarrow C$

*no new software here

NELL: example self-discovered subcategories

Animal:

- Pets
  - Hamsters, Ferrets, Birds, Dog, Cats, Rabbits, Snakes, Parrots, Kittens, …
- Predator
  - Bears, Foxes, Wolves, Coyotes, Snakes, Racoons, Eagles, Lions, Leopards, Hawks, Humans, …

Learned reading patterns for Subset(arg1,arg2)

"arg1 and other medium sized arg2" "arg1 and other jungle arg2" "arg1 and other magnificent arg2" "arg1 and other pesky arg2" "arg1 and other mammals and arg2" "arg1 and other Ice Age arg2" "arg1 or other biting arg2" "arg1 and other marsh arg2" "arg1 and other migrant arg2" "arg1 and other monogastric arg2" "arg1 and other mythical arg2" "arg1 and other nesting

NELL: example self-discovered subcategories

Animal:

- Pets
  - Hamsters, Ferrets, Birds, Dog, Cats, Rabbits, Snakes, Parrots, Kittens, …
- Predator
  - Bears, Foxes, Wolves, Coyotes, Snakes, Racoons, Eagles, Lions, Leopards, Hawks, Humans, …

Learned reading patterns:

"arg1 and other medium sized arg2" "arg1 and other jungle arg2" "arg1 and other magnificent arg2" "arg1 and other pesky arg2" "arg1 and other mammals and arg2" "arg1 and other Ice Age arg2" "arg1 or other biting arg2" "arg1 and other marsh arg2" "arg1 and other migrant arg2" "arg1 and other monogastric arg2" "arg1 and other mythical arg2" "arg1 and other nesting

Chemical:

- Fossil fuels
  - Carbon, Natural gas, Coal, Diesel, Monoxide, Gases, …
- Gases
  - Helium, Carbon dioxide, Methane, Oxygen, Propane, Ozone, Radon…

Learned reading patterns:

"arg1 and other hydrocarbon arg2" "arg1 and other aqueous arg2" "arg1 and other hazardous air arg2" "arg1 and oxygen are arg2" "arg1 and such synthetic arg2" "arg1 as a lifting arg2" "arg1 as a tracer arg2" "arg1 as the carrier arg2" "arg1 as the inert arg2" "arg1 as the primary cleaning arg2" "arg1 and other noxious arg2" "arg1 and other trace arg2" "arg1 as the reagent arg2" "arg1 as the tracer

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 latent 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
12. Add a robot body to NELL

thank you

and thanks to:

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