Never Ending Learning

Tom M. Mitchell

Justin Betteridge, Jamie Callan, Andy Carlson, William Cohen, Estevam Hruschka, Bryan Kisiel, Mahaveer Jain, Jayant Krishnamurthy, Edith Law, Thahir Mohamed, Mehdi Samadi, Burr Settles, Richard Wang, Derry Wijaya

Machine Learning Department
Carnegie Mellon University

October 2010

Humans learn many things, for years, and become better learners over time

Why not machines?

Never Ending Learning

Task: acquire a growing competence without asymptote

- over years
- multiple functions
- where learning one thing improves ability to learn the next
- acquiring data from humans, environment

Many candidate domains:

- Robots
- Softbots
- Game players

Years of Relevant AI/ML Research

- Architectures for problem solving/learning
  - SOAR [Newell, Laird, Rosenbloom 1986]
  - ICARUS [Langley], PRODIGY [Carbonell], …
- Large scale knowledge construction/extraction
  - Cyc [Lenat], KnowItAll, TextRunner [Etzioni et al 2004], WOE [Weld et al. 2009]
- Life long learning
  - Learning to learn [Thrun & Pratt, 1998], EBNN [Thrun & Mitchell 1993]
- Transfer learning
  - Multitask learning [Caruana 1995]
  - Transfer reinforcement learning [Parr & Russell 1998]
  - Learning with structured outputs [Taskar, 2009; Roth 2009]
- Active Learning
  - survey [Settles 2010]; Multi-task active learning [Harpale & Yang, 2010]
- Curriculum learning
  - [Bengio, et al., 2009; Krueger & Dayan, 2009; Ni & Ling, 2010]

NELL: Never-Ending Language Learner

Inputs:

- initial ontology
- handful of examples of each predicate in ontology
- 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: Never-Ending Language Learner

Goal:

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

Today…

Running 24 x 7, since January, 2010

Input:

- ontology defining ~500 categories and relations
- 10-20 seed examples of each
- 500 million web pages (ClueWeb – Jamie Callan)

Result:

- continuously growing KB with ~440,000 extracted beliefs

NELL Today

- http://rtw.ml.cmu.edu

Recently-Learned Facts

Refresh

instance	iteration	date learned	confidence
honours is an award, championship, or tournament trophy	158	12-oct-2010	93.8
game_learning is a cognitive action	158	12-oct-2010	97.1
el_capitan is a song	157	11-oct-2010	100.0
fondue_set is an item found in the kitchen	154	28-sep-2010	95.1
cheap_levitra is a drug	155	29-sep-2010	93.8
kstu_tv is a TV affiliate of the network fox	159	17-oct-2010	93.8
joe_satriani is a musician who plays the guitar	154	28-sep-2010	99.2
news competes with bbc	156	09-oct-2010	98.4
bankatlantic_center is the home venue for the sports team florida_panthers	158	12-oct-2010	96.9
museum_island is a tourist attraction in the city rome	156	09-oct-2010	100.0

Semi-Supervised Bootstrap Learning

it's underconstrained!!

Extract cities:

Paris
Pittsburgh
Seattle
Cupertino

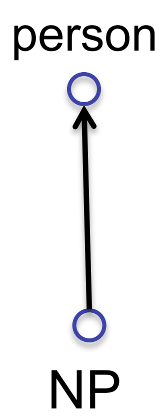
San Francisco
Austin
denial

anxiety
selfishness
Berlin

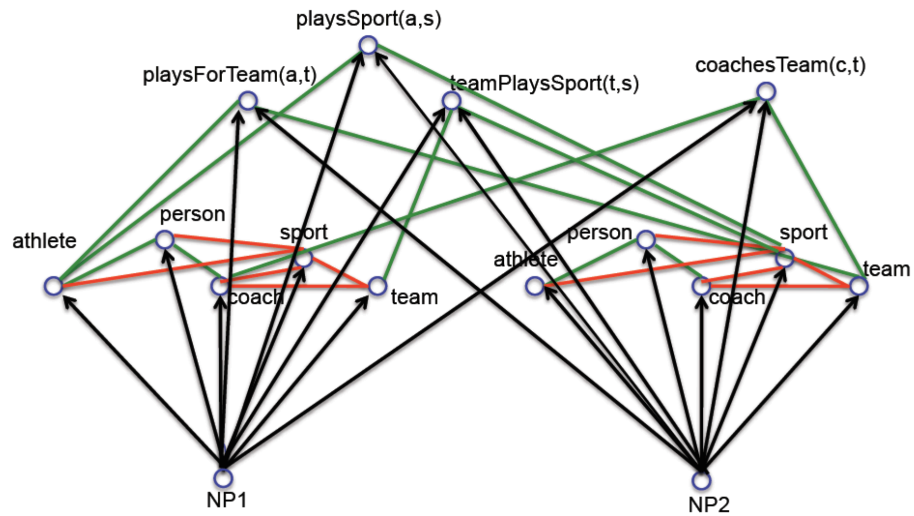
mayor of arg1
live in arg1

arg1 is home of
traits such as arg1

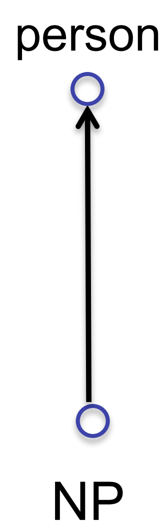
Key Idea 1: Coupled semi-supervised training of many functions



hard
(underconstrained)
semi-supervised
learning problem

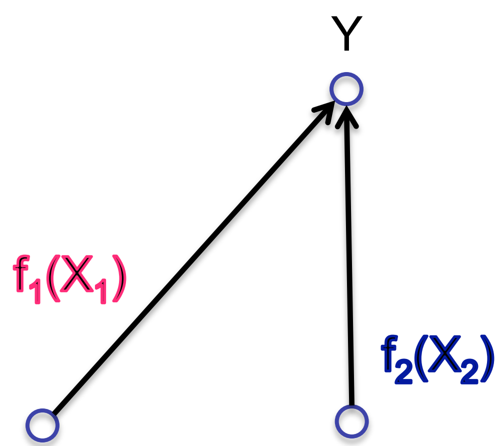


much easier (more constrained)
semi-supervised learning problem


person
NP

Coupled Training Type 1: Co-Training, Multiview, Co-regularization

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

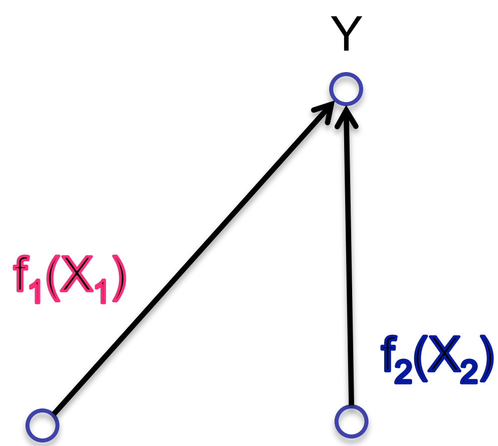


$\mathbf{X} = < X_1 \quad , \quad X_2 >$

Constraint: $f_1(x_1) = f_2(x_2)$

Coupled Training Type 1: Co-Training, Multiview, Co-regularization

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



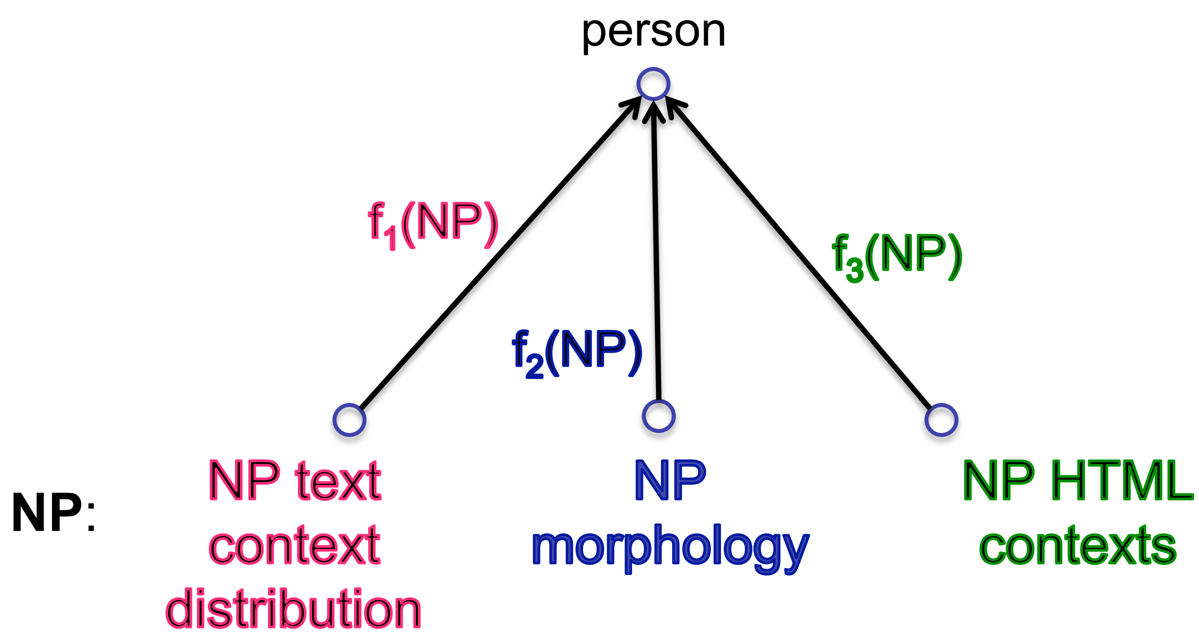
X = < X_1 , X_2 >

Constraint: $f_1(x_1) = f_2(x_2)$

If f_1, f_2 PAC learnable,
X_1, X_2 conditionally indep
Then PAC learnable from
<u>unlabeled</u> data and
weak initial learner

and disagreement between
f_1, f_2 bounds error of each

Type 1 Coupling Constraints in NELL

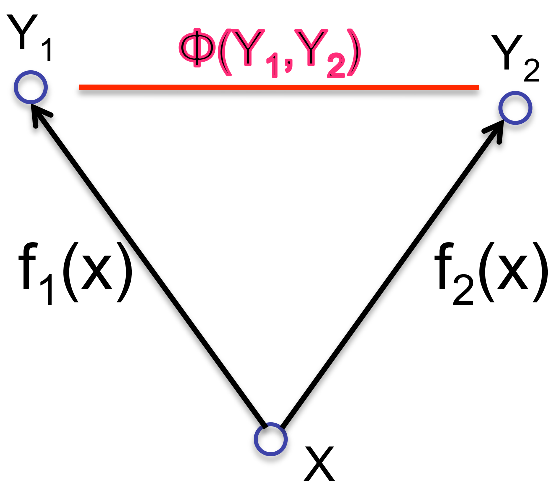


__ is a friend	***capitalized?***	***www.celebrities.com:***
rang the __	***ends with '...ski'?***	*** __ ***
...	**...**	
__ walked in	***contains "univ."?***	**...**

Coupled training type 2
Structured Outputs, Multitask, Posterior Regularization, Multilabel

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

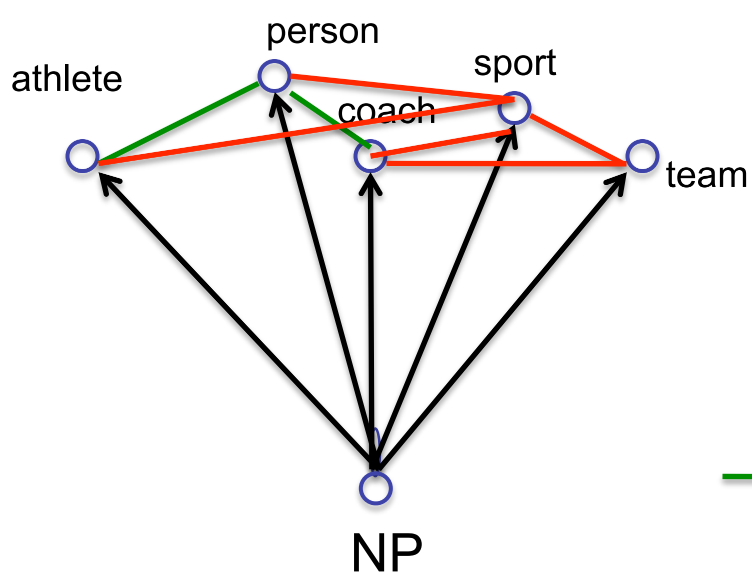
Learn functions with same input, different outputs, where we know some constraint $\Phi(Y_1,Y_2)$



Constraint: $\Phi(f_1(x), f_2(x))$

Effectiveness ~ probability that $\Phi(Y_1,Y_2)$ will be violated by incorrect f_j and f_k

Type 2 Coupling Constraints in NELL

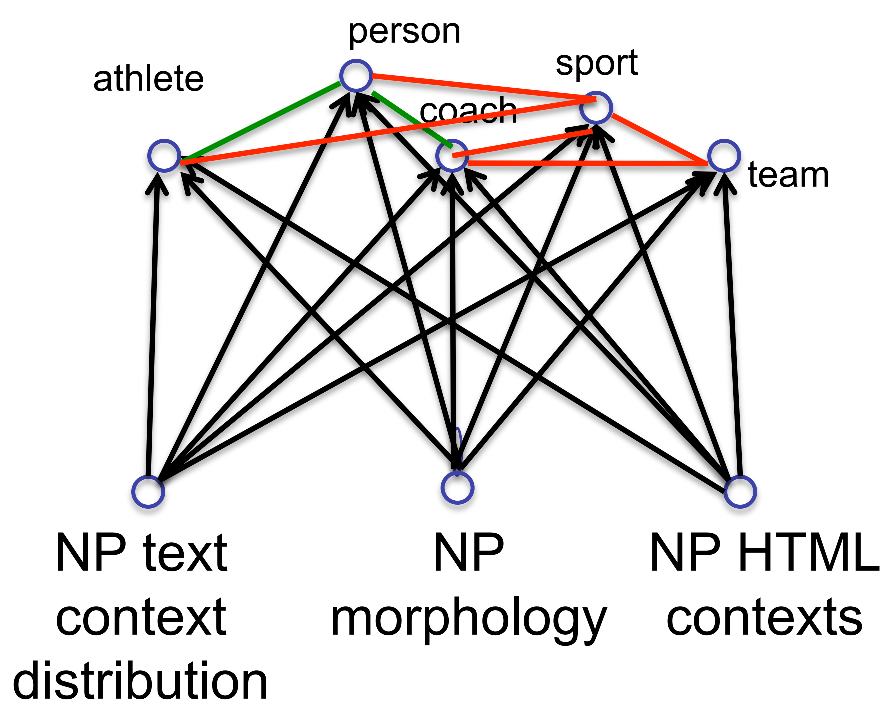


athlete(NP) → person(NP)

athlete(NP) → NOT sport(NP)

NOT athlete(NP) ← sport(NP)

Multi-view, Multi-Task Coupling

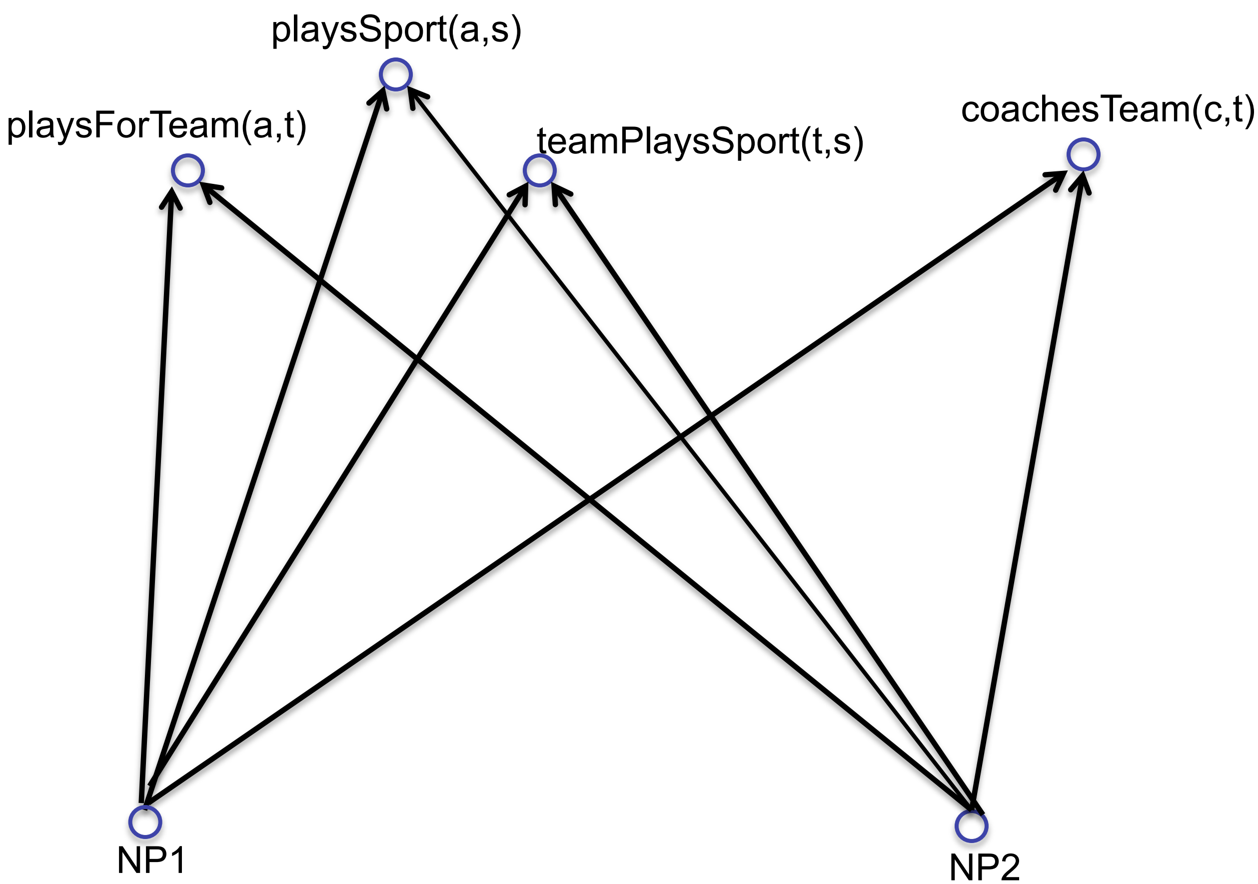


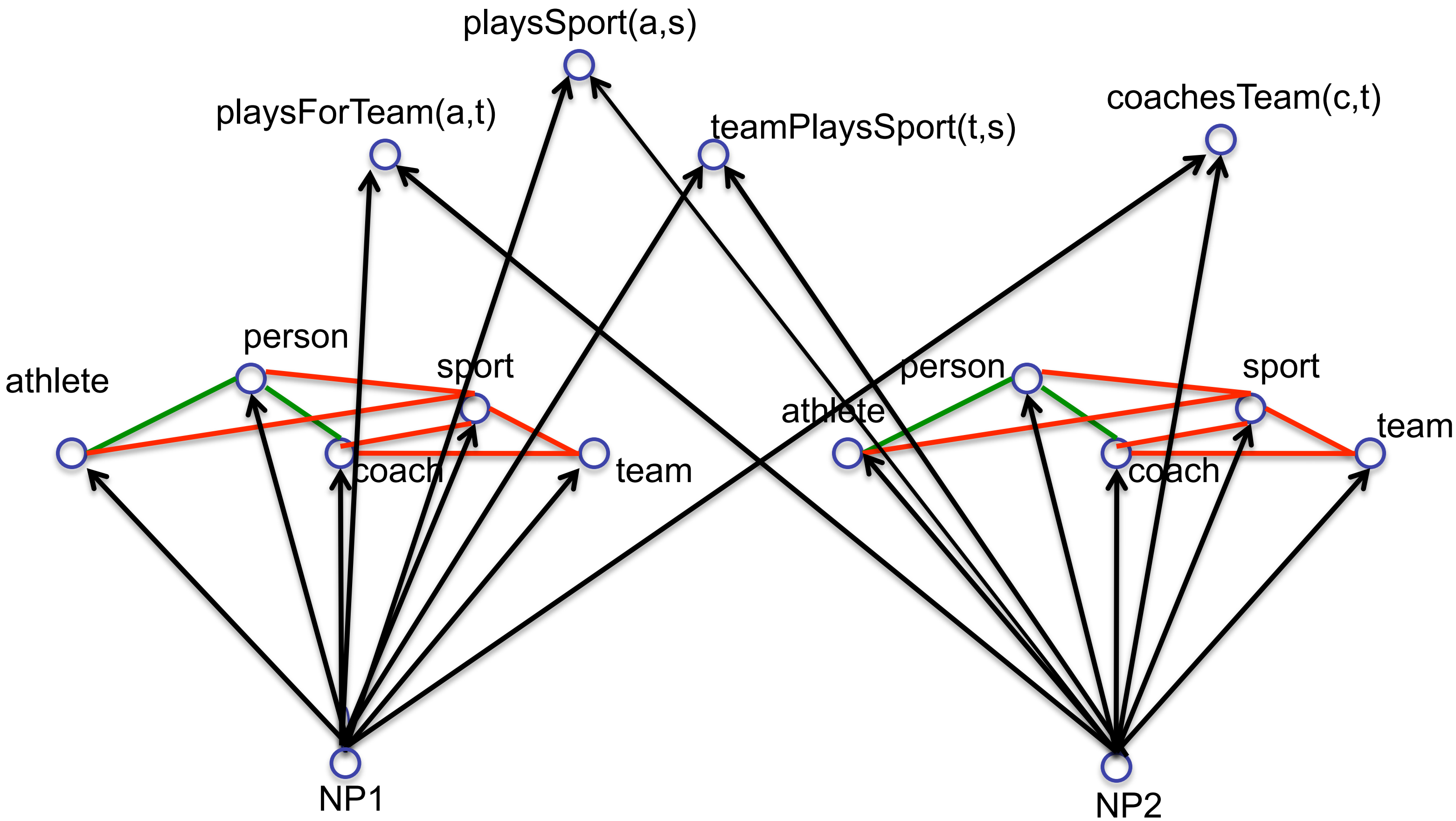
NP:

C categories, V views, CV ≈ 250*3=750 coupled functions

pairwise constraints on functions ≈ 10^5

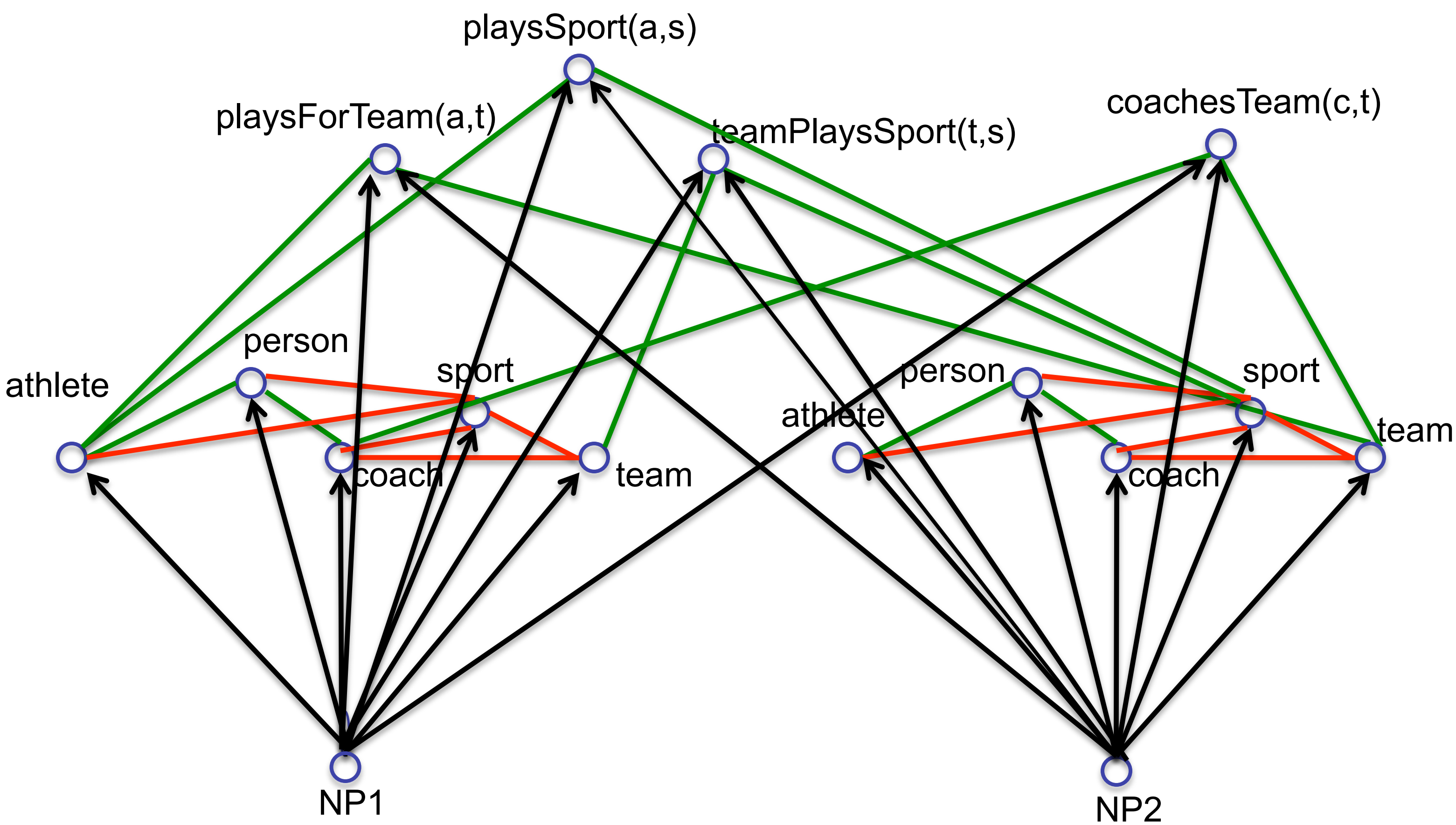
Learning Relations between NP's




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

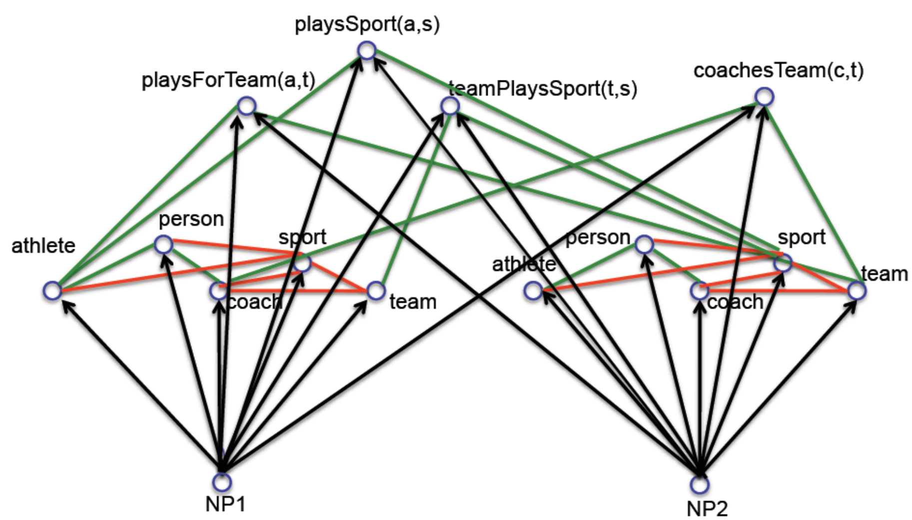
Type 3 Coupling: Argument Types

Constraint: f3(x1,x2) → (f1(x1) AND f2(x2))



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

Pure EM Approach to Coupled Training



E: jointly estimate latent labels for each function of each unlabeled example

M: retrain all functions, based on these probabilistic labels

Scaling problem:

- **E** step: 20M NP's, 10^{14} NP pairs to label
- **M** step: 50M text contexts to consider for each function → 10^{10} parameters to retrain
- even more URL-HTML contexts…

NELL's Approximation to EM

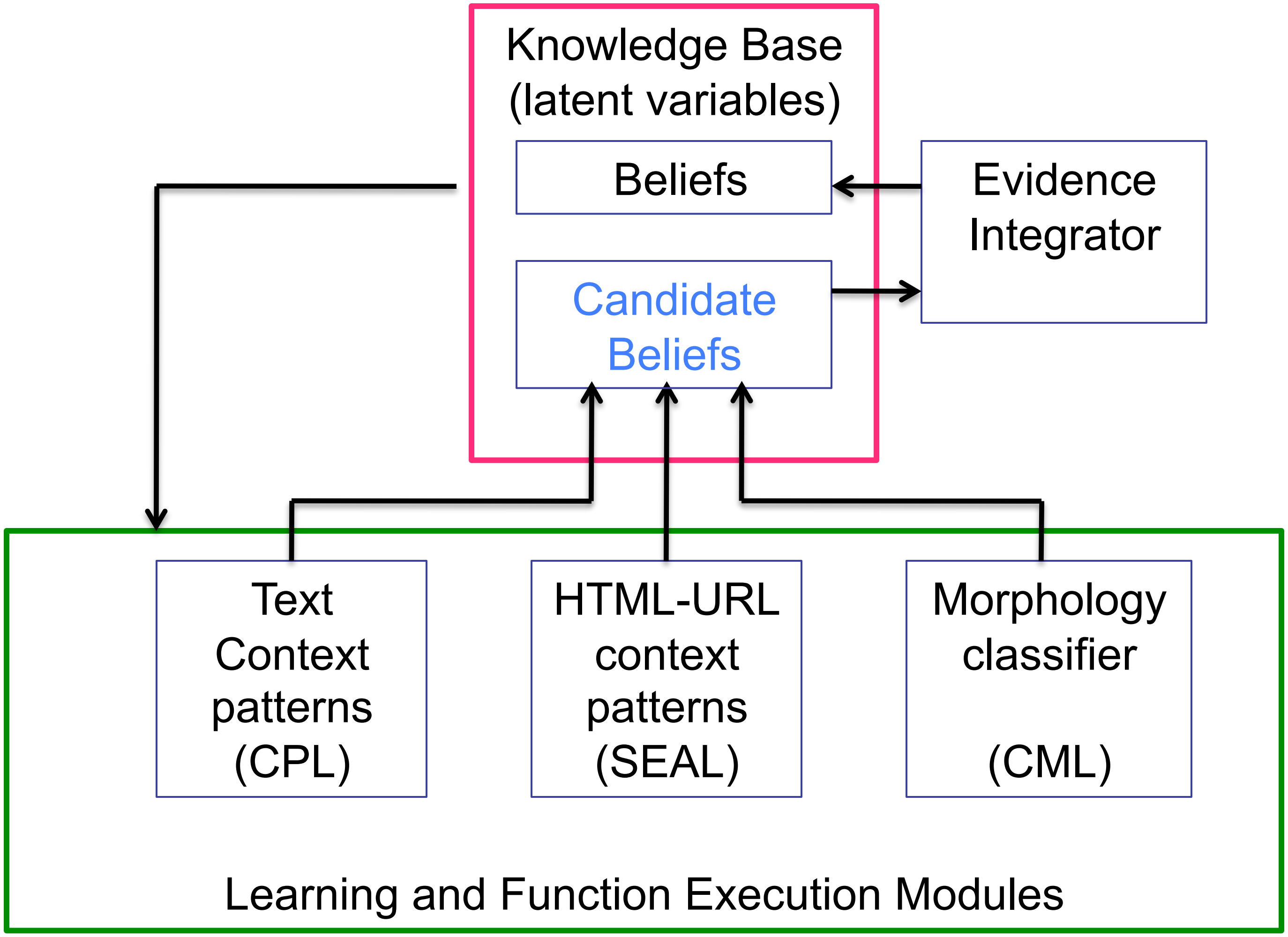
E' step:

- Consider only a growing subset of the latent variable assignments
  - category variables: up to 250 NP's per category per iteration
  - relation variables: add only if confident and args of correct type
  - <u>this set of explicit latent assignments ***IS*** the knowledge base</u>

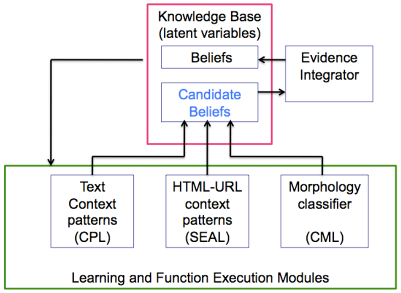
M' step:

- Each view-based learner retrains itself from the updated KB
- "context" methods create growing subsets of contexts

NELL Architecture



Never-Ending Language Learning



arg1_was_playing_arg2 arg2_megastar_arg1 arg2_icons_arg1
arg2_player_named_arg1 arg2_prodigy_arg1
arg1_is_the_tiger_woods_of_arg2 arg2_career_of_arg1
arg2_greats_as_arg1 arg1_plays_arg2 arg2_player_is_arg1
arg2_legends_arg1 arg1_announced_his_retirement_from_arg2
arg2_operations_chief_arg1 arg2_player_like_arg1
arg2_and_golfing_personalities_including_arg1 arg2_players_like_arg1
arg2_greats_like_arg1 arg2_players_are_steffi_graf_and_arg1
arg2_great_arg1 arg2_champ_arg1 arg2_greats_such_as_arg1
arg2_professionals_such_as_arg1 arg2_hit_by_arg1 arg2_greats_arg1
arg2_icon_arg1 arg2_stars_like_arg1 arg2_pros_like_arg1
arg1_retires_from_arg2 arg2_phenom_arg1 arg2_lesson_from_arg1
arg2_architects_robert_trent_jones_and_arg1 arg2_sensation_arg1
arg2_pros_arg1 arg2_stars_venus_and_arg1 arg2_hall_of_famer_arg1
arg2_superstar_arg1 arg2_legend_arg1 arg2_legends_such_as_arg1
arg2_players_is_arg1 arg2_pro_arg1 arg2_player_was_arg1
arg2_god_arg1 arg2_idol_arg1 arg1_was_born_to_play_arg2
arg2_star_arg1 arg2_hero_arg1 arg2_players_are_arg1
arg1_retired_from_professional_arg2 arg2_legends_as_arg1
arg2_autographed_by_arg1 arg2_champion_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
visualArtMovement	PREFIX=journ	-0.234
visualArtMovement	PREFIX=budd	-0.253

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] –

Coupled Training Helps!

[Carlson et al., WSDM 2010]

Using only two views:
Text, HTML contexts.

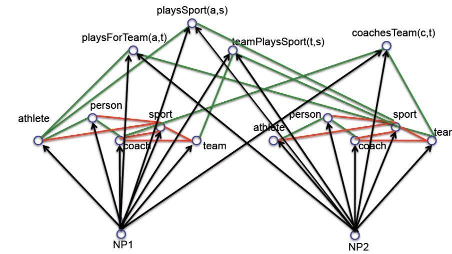
PRECISION	Text uncpl	HTML uncpl	Coupled
Categories	.41	.59	.90
Relations	.69	.91	.95

10 iterations,
200 M web pages
44 categories, 27 relations
199 extractions per category

	text	HTML	Coupled
EconomicSector	23	10	77
Emotion	53	60	83
Food	70	80	100
Furniture	0	57	90
Hobby	33	50	90
KitchenItem	3	13	100
Mammal	50	50	90
Movie	57	100	100
NewspaperCompany	60	97	100
Politician	60	37	100
Product	83	77	70
ProductType	63	63	50
Profession	53	57	93
ProfessionalOrganization	63	77	87
Reptile	3	27	100
Room	0	7	100
Scientist	30	17	100
Shape	7	7	85
Sport	13	83	73
SportsEquipment	10	23	23
SportsLeague	7	27	86
SportsTeam	30	87	87
Stadium	57	63	90
StateOrProvince	63	93	77
Tool	13	90	97
Trait	40	47	97
University	97	90	93
Vehicle	30	13	77

If coupled learning is the key idea, 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

Discover New Coupling Constraints

For each relation:

seek probabilistic first order Horn Clauses

- Positive examples: extracted beliefs in the KB
- Negative examples: ???

Ontology to the rescue:

numberOfValues(teamPlaysSport) = 1

numberOfValues(competesWith) = any

can infer negative examples from positive for this, but not for this

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]

Rule Learning Summary

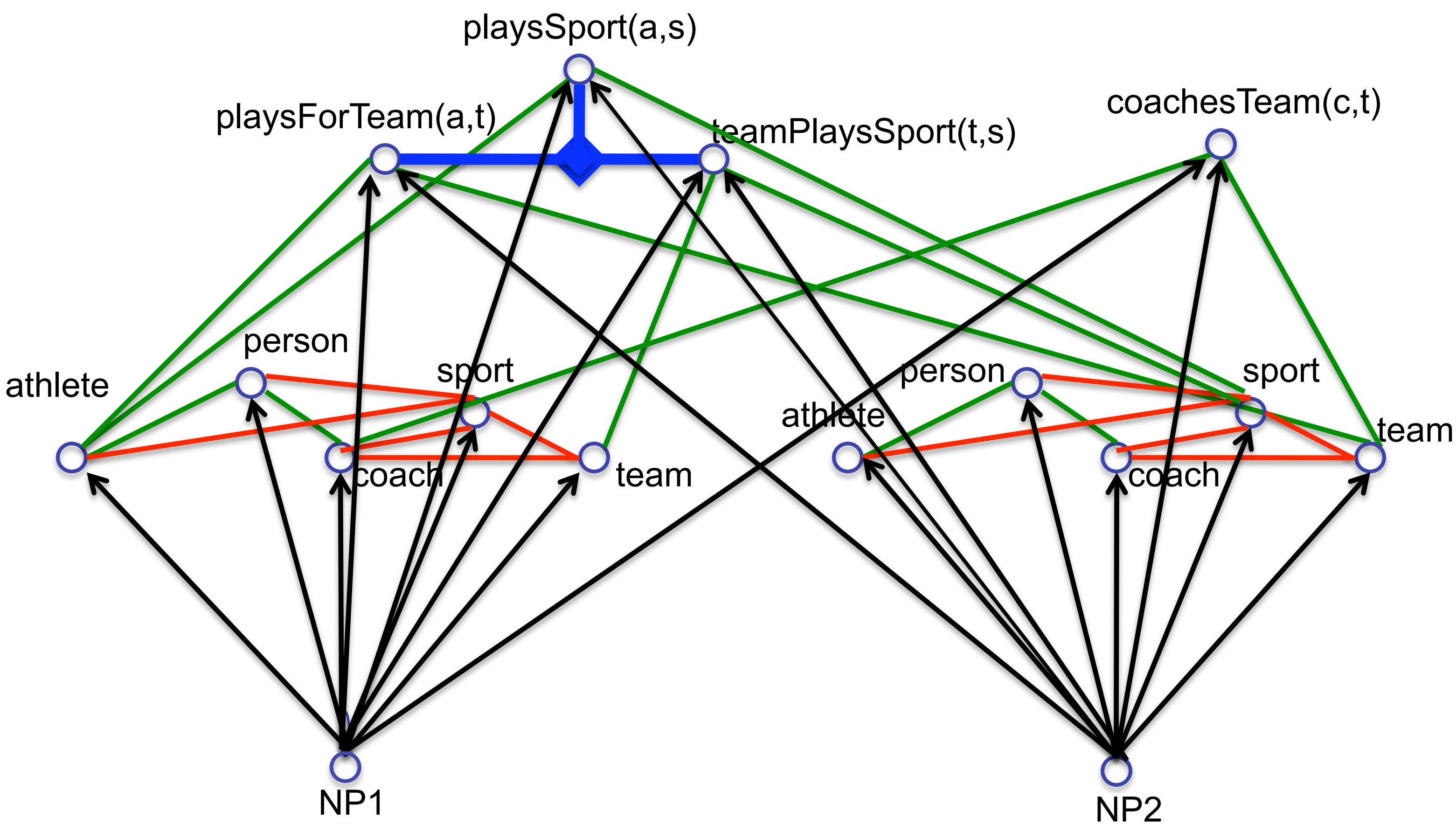
- Rule learner run every 10 iterations
- Manual filtering of rules

- After 120 iterations
  - 565 learned rules
  - 486 (86%) survived manual filter

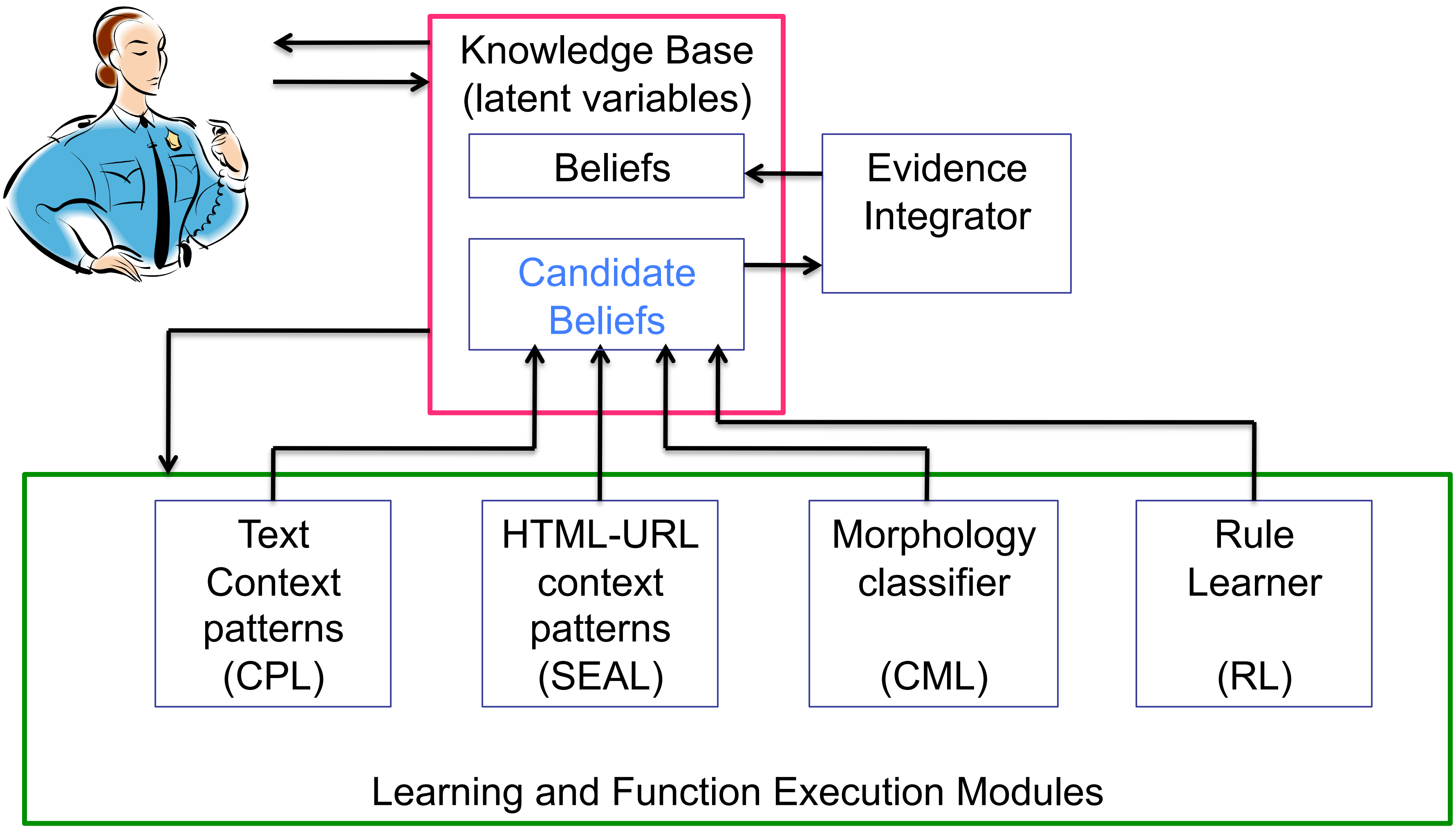
  - 3948 new beliefs inferred by these rules

Learned Probabilistic Horn Clause Rules

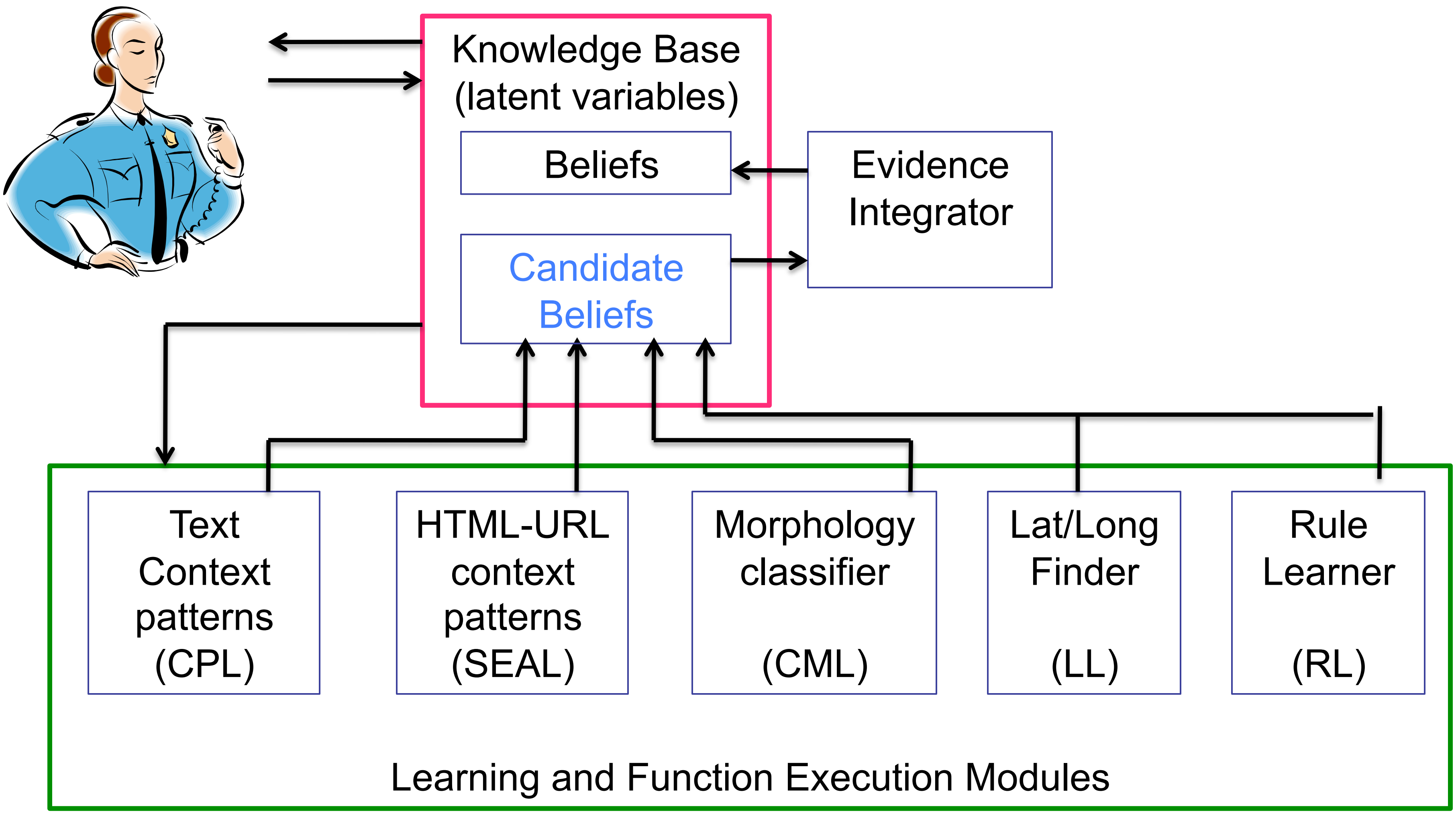
0.93 playsSport(?x,?y) ← playsForTeam(?x,?z), teamPlaysSport(?z,?y)



NELL Architecture



NELL Architecture, October 2010



NELL as of Oct 18, 2010

440K beliefs in 160 iterations

210 categories, 280 relations

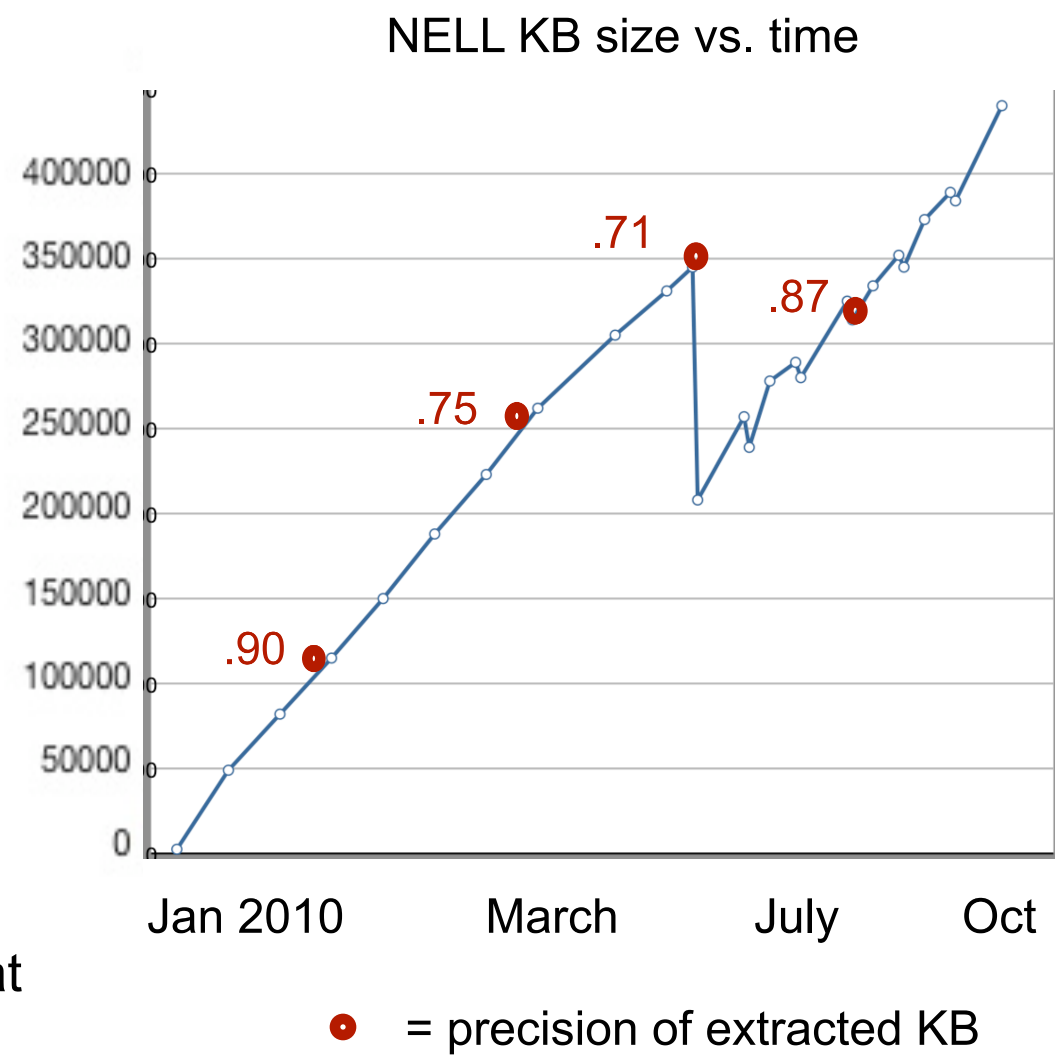
1470 coupled functions

> 40K text extraction patterns

> 548 accepted learned rules
leading to > 6000 new beliefs

65-75% of predicates currently being read well, remainder are receiving significant correction

Human check/clean KB
every 10 iterations, beginning at iteration 100



NELL – Human Feedback

beginning at iteration 100, human feedback every 10 iterations. 5 minutes per predicate

at iteration 100: 182 predicates in ontology

- 75% of predicates received minor or no correction
  - estimated precision 0.9-1.0
- 25% (45/182) received major corrections
  - estimated precision over recent iterations <<0.9
  - quick feedback: delete all extractions beyond iteration k
  - label some negative examples

NELL: “emotions”

shame
guilt
regret
embarrassment
stress
pity
empathy
resentment
awe
sympathy
laughter
despair
sorrow
concern
lust
loneliness
grief
disappointment

envy
gratitude
rage
pride
compassion
elation
anguish
hurt
relief
ecstasy
angst
dread
hopelessness
longing
remorse
anxieties
melancholy
fright

← **Earliest extractions**

NELL: "emotions" (at 100 iterations)

shame	envy	**2,636 extracted emotions,**	profound dislike
guilt	gratitude		split_personality
regret	rage		themotivation
embarrassment	pride	**490 extraction patterns**	fierce_joy
stress	compassion		practical_assistance
pity	elation		fearand
empathy	anguish		interest_toall
resentment	hurt	**← Earliest extractions**	differentnature
awe	relief		approval
sympathy	ecstasy		overwhelming_wave
laughter	angst	**Most recent extractions →**	vengence
despair	dread		policy_relevance
sorrow	hopelessness		disavowal
concern	longing		manifestation
lust	remorse		change
loneliness	anxieties		mild_bitterness
grief	melancholy		unfounded_fears
disappointment	fright		full_support

NELL: “emotions” 490 extraction patterns

tears of _
feelings such as _
heart filled with _
heart was filled with _
heart is filled with _
heart was full of _
feelings , such as _
twinge of _
pang of _
emotion such as _
heart is full of _
intense feelings of _
overwhelming feelings of _
heart full of _
hearts full of _
Feelings of _
It is with great _

deep feelings of _
mixed feelings of _
I was overcome with _
emotions , from _
feelings of intense _
strong feelings of _
I am filled with _
hearts filled with _
feelings of deep _
feelings of extreme _
paroxysms of _
I'm filled with _
source of deep _
he was filled with _
feeling of intense _
overwhelming feeling of _
I was filled with _

← Earliest

Most recent →

I just burst into _
People fall in _
big vote of _
I have been following with _
world looked on in _
other countries have expressed _
I was falling in _
issue is of great _
matters of mutual _
sheer driving _
Majesty expressed _
Association have expressed _
browser with JavaScript _
Friday expressed _
concurrent resolution expressing _

NELL – Newer Directions

Ontology Extension (1)

[Mohamed & Hruschka]

Goal:

- Discover frequently stated relations among ontology categories

Approach:

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

* additional experiments with Etzioni & Soderland using TextRunner

Preliminary Results

[Thahir Mohamed & Estevam Hruschka]

Category Pair	Name	Text contexts	Extracted Instances
MusicInstrument Musician	Master	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
Disease Disease	IsDueTo	ARG1 is due to ARG2 ARG1 is caused by ARG2	pinched nerve, herniated disk tennis elbow, tendonitis blepharospasm, dystonia
CellType Chemical	ThatRelease	ARG1 that release ARG2 ARG2 releasing ARG1	epithelial cells, surfactant neurons, serotonin mast cells, histomine
Mammals Plant	Eat	ARG1 eat ARG2 ARG2 eating ARG1	koala bears, eucalyptus sheep, grasses goats, saplings
…			

Ontology Extension (2)

[Burr Settles]

- NELL sometimes extracts subclasses instead of instances:
  - chemicals: carbon_dioxide, amonia, gas,
- So, add the relation "typeHasMember" to NELL's ontology
  - ChemicalType_Has_Chemical
  - AnimalType_Has_Animal
  - ProfessionType_Has_Profession
- NELL learns to read subcategory extensions to ontology

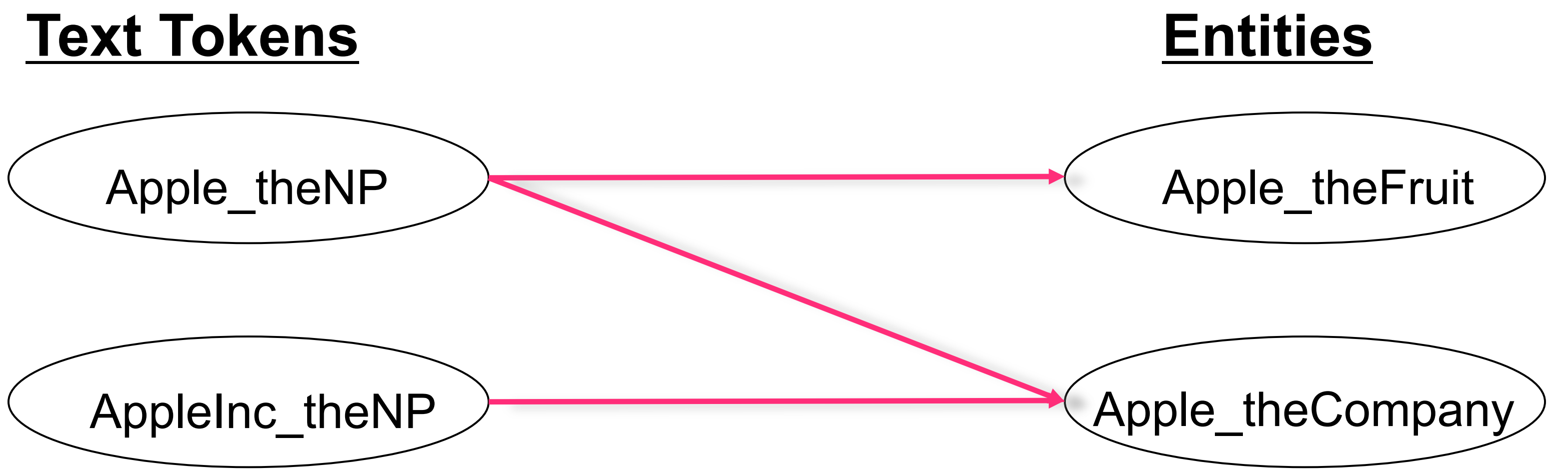
Results: Ontology extension by reading

Original Category	SubType discovered by reading	Extracted Instances
Chemical	Gases	amonia, carbon_dioxide, carbon_monoxide, methane, sulphur, oxides, nitrous_oxides, water_vapor, ozone, nitrogen
Animal	LiveStock	chickens, cows, sheep, goats, pigs
Profession	Professionals	surgeons, chiropractors, dentists, engineers, medical staff, midwives, professors, scientists, specialists, technologists, aides

Extraction patterns learned for populating AnimalType_Has_Animal
- arg2 like cows and arg1
- arg1 and other nonhuman arg2
- arg1 are mostly solitary arg2
- arg1 and other hoofed arg2
- …

Distinguishing Text Tokens from Entities

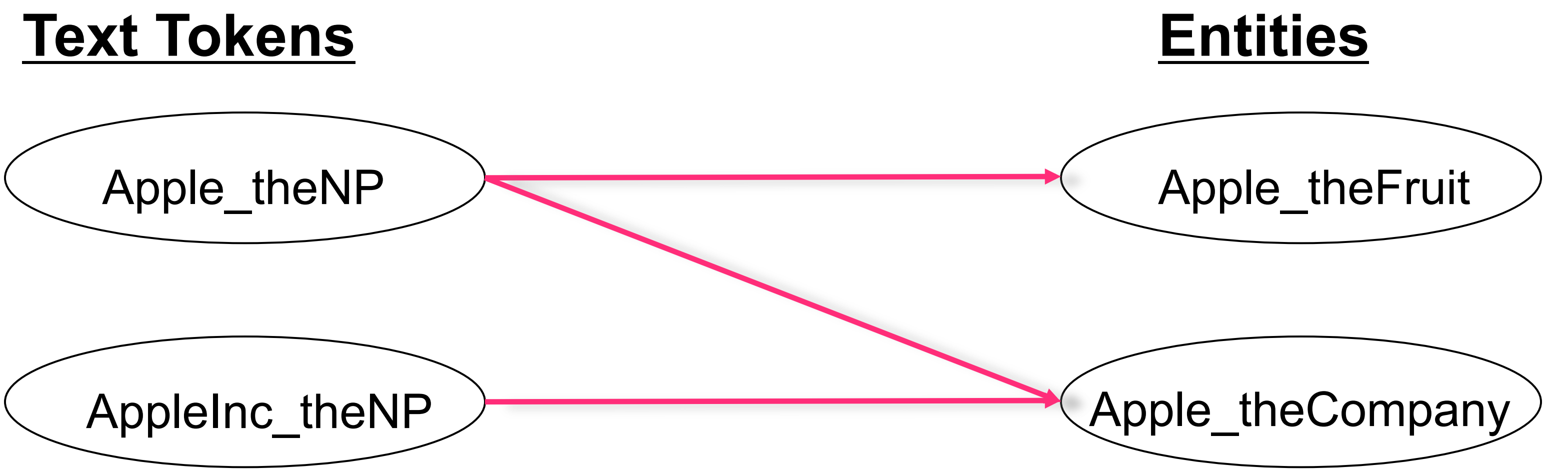
[Jayant Krishnamurthy]



Distinguish Text Tokens from Entities

COMING SOON…

[Jayant Krishnamurthy]



Coreference Resolution:

- Co-train classifier to predict coreference as f(string similarity, extracted beliefs)
- Small amount of supervision: ~10 labeled coreference decisions
- Cluster tokens using f as similarity measure

Preliminary Coreference Results

[Jayant Krishnamurthy]

Evaluated Precision/Recall of Pairwise Coreference Decisions:

Category	Precision	Recall
athlete	0.52	0.50
city	0.40	0.25
coach	0.76	0.76
company	0.80	0.63
country	0.86	0.15
sportsteam	0.88	0.21
stadium	0.70	0.18

Example “sportsteam” clusters:
st_louis_rams, louis_rams, st___louis_rams, rams, st__louis_rams
stanford_university, stanford_cardinals, stanford
pittsburgh_pirates, pirates, pittsburg_pirates
lakers, la_lakers, los_angeles_lakers
valdosta_blazers, valdosta_st__blazers, valdosta_state_blazers
illinois_state, illinois_state_university, illinois_university
...

Active Learning through CrowdSourcing

COMING SOON… [Edith Law, Burr Settles, Luis von Ahn]

- outsource actively-selected KB edits as a "human computation" trivia game: ***Polarity***

"positive" player

"negative" player

What will move forward research on Never Ending Learning?

Never Ending Learning: Thesis topics 1

Case study theses:

- office robot
- softbots
  - Web based research assistant
- game players
  - Why isn't there a never-ending chess learner?
- never-ending learners for sensors
  - intelligent street corner camera
  - intelligent traffic control light
  - intelligent traffic grid

Never Ending Learning: Thesis topics 2

- Scaling EM: billions of virtual(?) latent variables
  - convergence properties?
  - what properties of constraint graph predict success?

- How are correctness and self-consistency related?
  - disagreement bounds error when functions co-trained on conditionally independent features [Dasgupta, et al., 2003]

- Curriculum-based learning
  - what curriculum properties guarantee improved long term learning?

- Self-reflection:
  - what self-reflection and self-repairing capabilities assure “reachability” of target performance?

thank you!

and thanks to Yahoo! for M45 computing
and thanks to Google, NSF, Darpa for partial funding
and thanks to Microsoft for fellowship to Edith Law