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

March 2011

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
- Tweeters

NELL: Never-Ending Language Learner

Inputs:

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

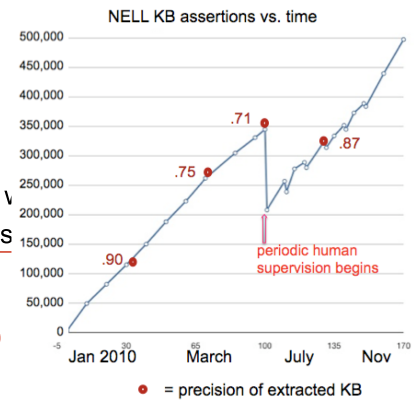
The task:

- 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 v
    2. learn to read better than yes



Today…

Running 24x7, since January, 12, 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 >525,000 extracted beliefs

NELL Today

- http://rtw.ml.cmu.edu
- eg., “Disney”, “Mets”, “IBM”, “Pittsburgh” …

Recently-Learned Facts twitter Refresh

instance	iteration	date learned	confidence
dilator muscle of pupil is a muscle	210	17-feb-2011	100.0
boyden cave is a cave	211	18-feb-2011	100.0
pondicherry is a state or a province	211	18-feb-2011	100.0
vena brachialis is a vein	211	18-feb-2011	97.5
scott rigell is a U.S. politician	210	17-feb-2011	96.9
toronto is the home city of the sports team ryerson	210	17-feb-2011	93.8
jim mcnerney is the CEO of boeing	210	17-feb-2011	96.9
microsoft is a company that produces windows vista	213	22-feb-2011	100.0
frogs is an animal that is a kind of small animals	210	17-feb-2011	93.8
kprc is a TV station in the city houston	210	17-feb-2011	93.8

Semi-Supervised Bootstrap Learning

it's underconstrained!!

Extract cities:

Paris
Pittsburgh
Seattle
Cupertino

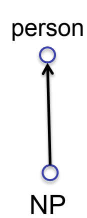
San Francisco
Austin
denial

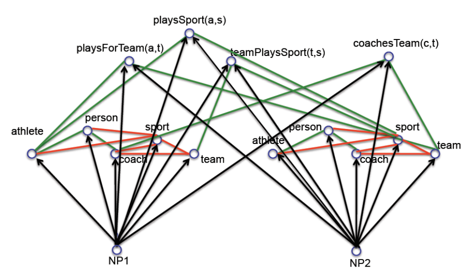
anxiety
selfishness
Berlin

mayor of arg1
live in arg1

arg1 is home of
traits such as arg1

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



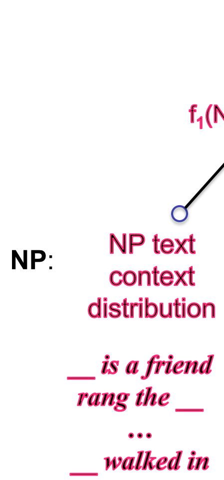


hard (underconstrained) semi-supervised learning problem

much easier (more constrained) semi-supervised learning problem

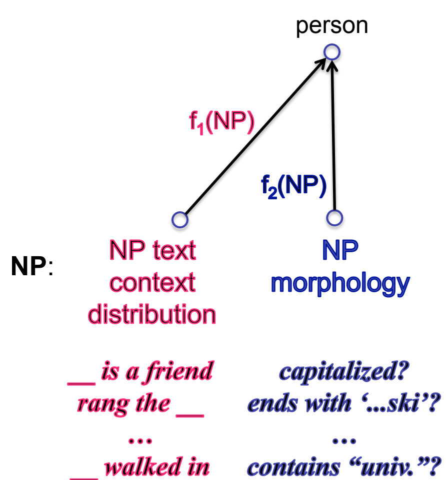
Type 1 Coupling: Co-Training, Multi-View Learning

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



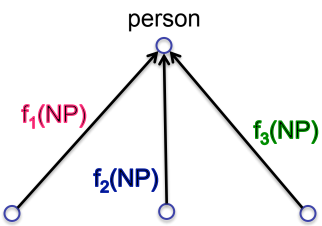
Type 1 Coupling: Co-Training, Multi-View Learning

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



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

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

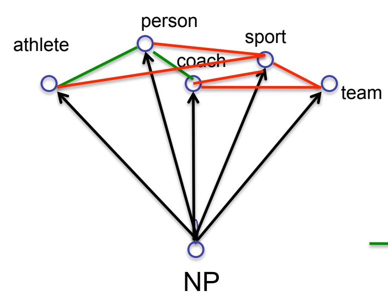


NP:

NP text context distribution	NP morphology	NP HTML contexts
__ is a friend	*capitalized?*	*www.celebrities.com:*
rang the __	*ends with '...ski'?*	* __ *
...	*...*	
__ walked in	*contains "univ."?*	...

Type 2 Coupling: Multi-task, Structured Outputs

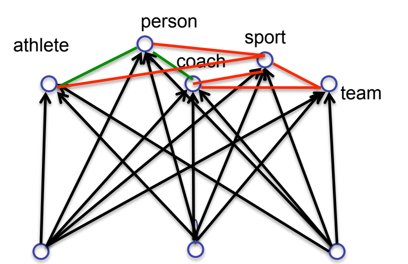
[Daume, 2008]
[Bakhir et al., eds. 2007]
[Roth et al., 2008]
[Taskar et al., 2009]
[Carlson et al., 2009]



athlete(NP) → person(NP)

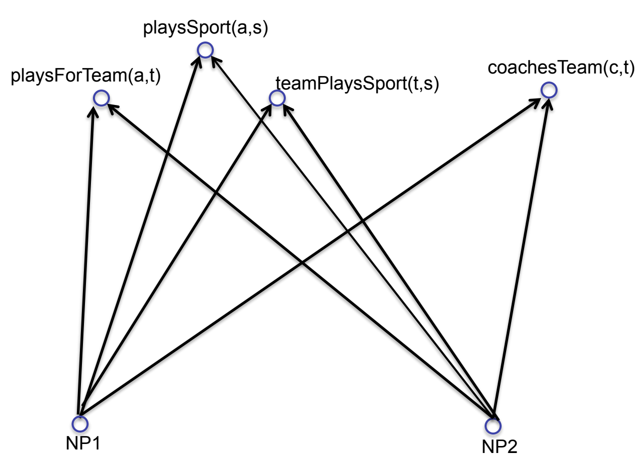
athlete(NP) → NOT sport(NP)
NOT athlete(NP) ← sport(NP)

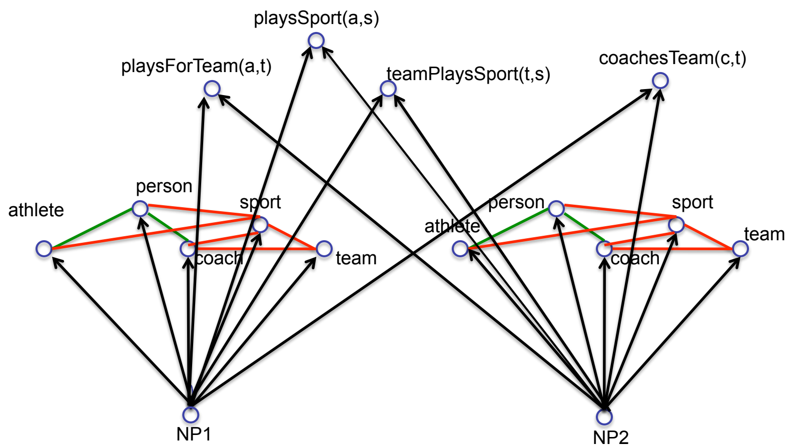
Multi-view, Multi-Task Coupling



NP: NP text context distribution | NP morphology | NP HTML contexts

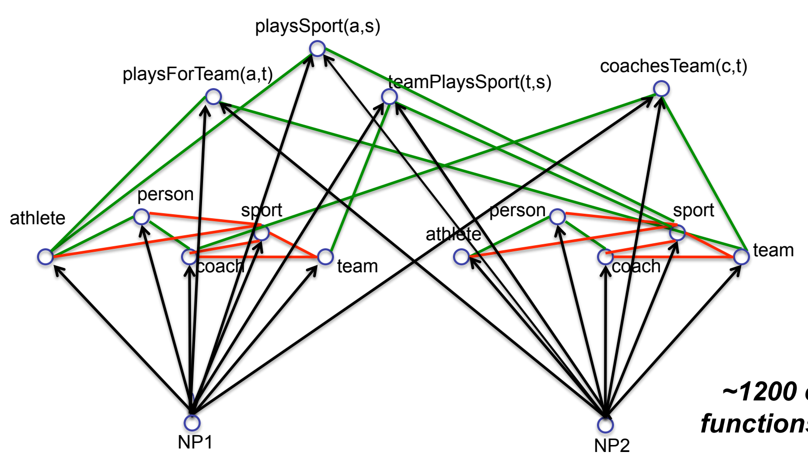
Learning Relations between NP's





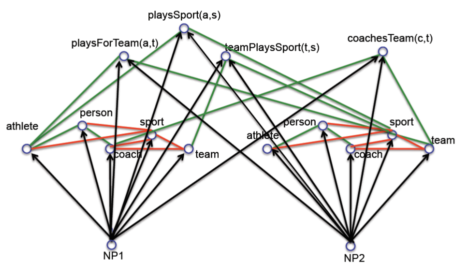
Type 3 Coupling: Argument Types

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



~1200 coupled functions in NELL

Pure EM Approach to Coupled Training



E: estimate labels for each function of each unlabeled example

M: retrain all functions, using 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

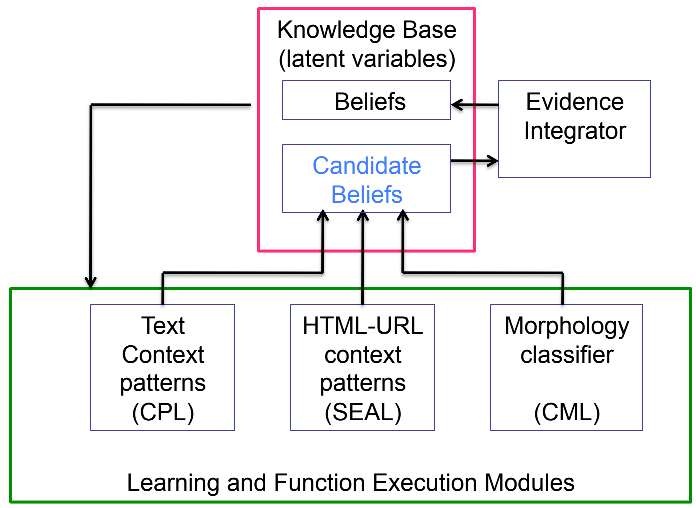
E' step:

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

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
arg1_plays_arg2   arg2_player_is_arg1   arg2_legends_arg1
arg1_announced_his_retirement_from_arg2   arg2_operations
arg2_player_like_arg1   arg2_and_golfing_personalities_includ
arg2_players_like_arg1   arg2_greats_like_arg1
arg2_players_are_steffi_graf_and_arg1   arg2_great_arg1   arg
arg2_greats_such_as_arg1   arg2_professionals_such_as_arg
arg2_hit_by_arg1 arg2_greats_arg1   arg2_icon_arg1   arg2_s
arg2_pros_like_arg1   arg1_retires_from_arg2   arg2_phenom
arg2_lesson_from_arg1   arg2_architects_robert_trent_jones_a
arg2_sensation_arg1 arg2_pros_arg1   arg2_stars_venus_and
arg2_hall_of_famer_arg1 arg2_superstar_arg1   arg2_legend
arg2_legends_such_as_arg1   arg2_players_is_arg1   arg2_pr
arg2_player_was_arg1   arg2_god_arg1   arg2_idol_arg1
arg1_was_born_to_play_arg2   arg2_star_arg1   arg2_hero_ar
arg2_players_are_arg1   arg1_retired_from_professional_arg2
arg2_legends_as_arg1   arg2_autographed_by_arg1   arg2_cha

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.

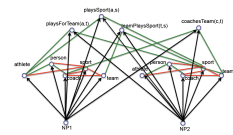
PRECISION	Text uncpl	HTML uncpl	Coupled
Categories	.41	.59	.90
Relations	.69	.91	.95

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

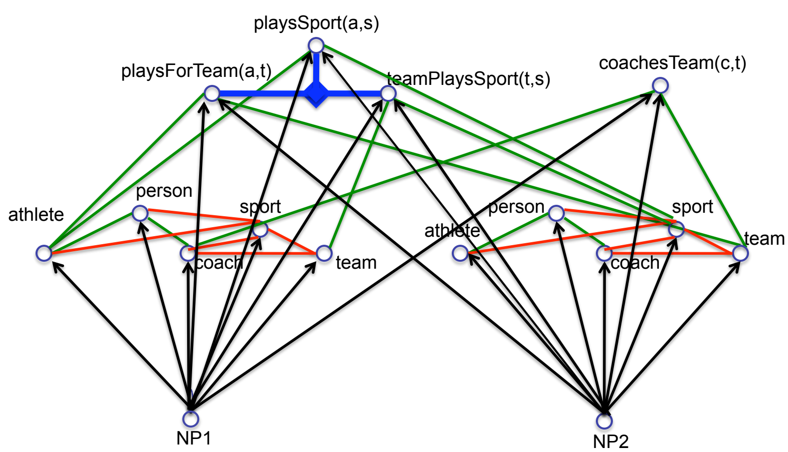
- Rule learner run every 10 iterations
- Manual filtering of rules

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

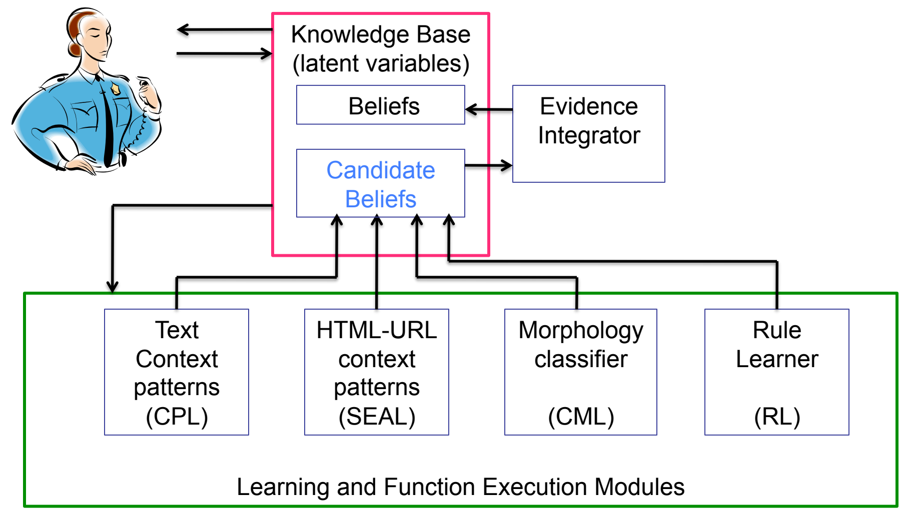
  - 3948 new beliefs inferred by these rules

Learned Probabilistic Horn Clause Rules

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



NELL Architecture



NELL as of March 6, 2011

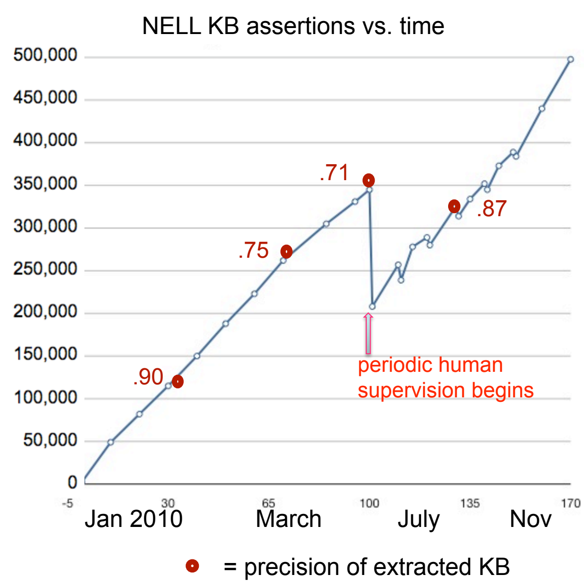
533K beliefs in 216 iterations
approx 85% correct
252 categories, 292 relations
1470 coupled functions

> 85K learned text extraction patterns

> 548 accepted learned rules leading to > 6000 new beliefs

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

Human check/feedback, beginning at iteration 100



NELL – Newer Directions

Ontology Extension (1)

[Mohamed & Hruschka]

Goal:

- Automatically extend ontology with new relations

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,
- Idea: have NELL learn to real the “Is_A” relation
- Result: NELL currently learns (reads about) new subcategories and their members

Results: Ontology extension by reading

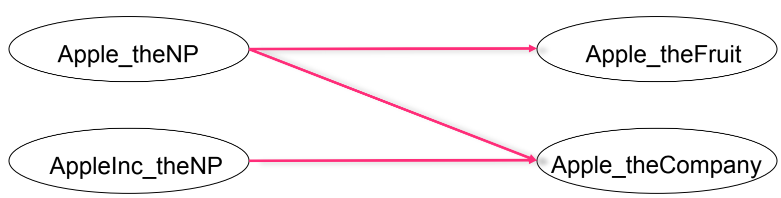
Original Category	SubType discovered by reading	Extracted Instances
Chemical	Gases	amonia, carbon_dioxide, carbon_monoxide, methane, sulphur, oxides, nitrous_oxides, water_vapor, ozone, nitrogen
Animal	LiveStock	chickens, cows, sheep, goats, pigs
Profession	Professionals	surgeons, chiropractors, dentists, 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
- …

Which Noun Phrases refer to Which Concepts?

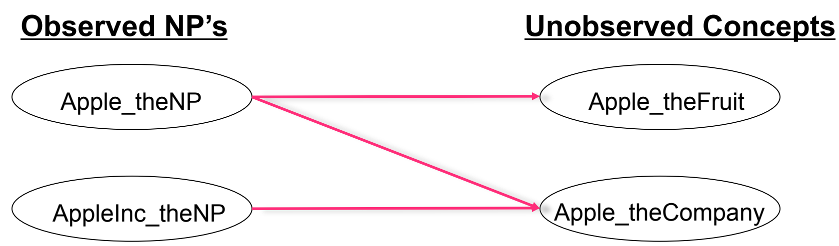
[Krishnamurthy & Mitchell, 2011]

Observed NP's **Unobserved Concepts**



Which Noun Phrases refer to Which Concepts?

[Krishnamurthy & Mitchell, 2011]



Coreference Resolution:

- Co-train classifier to predict NP coreference as f(string similarity, extracted beliefs)
- Small amount of supervision: ~10 labeled coreference decisions
- Cluster tokens using f as similarity measure
- Heuristic: one word sense per ontology category

Which Noun Phrases refer to Which Concepts?

[Krishnamurthy & Mitchell, 2011]

Evaluated Precision/Recall of Pairwise Coreference Decisions:

Category	Precision	Recall	Freebase concepts per NP
athlete	0.95	0.56	1.8
city	0.97	0.25	3.9
coach	0.86	0.94	1.1
company	0.85	041	2.4
country	0.74	0.56	1.8
sportsteam	0.89	0.30	3.3
stadium	0.83	0.61	1.6

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

Key Idea 3: Cumulative, Staged Learning

Learning X improves ability to learn Y

1. Classify 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. Read to find new subcategories for ontology
6. Cluster to discover new relations
7. Microread: NP types and relations within sentences
8. Microread: coreference within paragraphs
9. Microread: verb role labeling

Summary

- *Large scale coupled* semi-supervised training
- Automatically learn new coupling constraints/rules
- Cumulative learning

Many open research opportunities

- Role of self-reflection in never-ending learning
- Twitter dialogs with NELL
- Macro-reading to bootstrap microreading
- Give NELL a robot body
- Collaborate with other AI'ers across the web

Current NELL Team

Tom Mitchell
Professor

William Cohen
Professor

Estevam Hruschka, Jr.
Visiting Professor (USFCar)

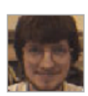
Burr Settles
Postdoctoral Fellow

Derry Wijaya
PhD Student (Language Technologies Institute)

Edith Law
PhD Student (Machine Learning Department)

Justin Betteridge
PhD Student (Language Technologies Institute)

Jayant Krishnamurthy
PhD Student (Computer Science Department)



Bryan Kisiel
Research Programmer

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