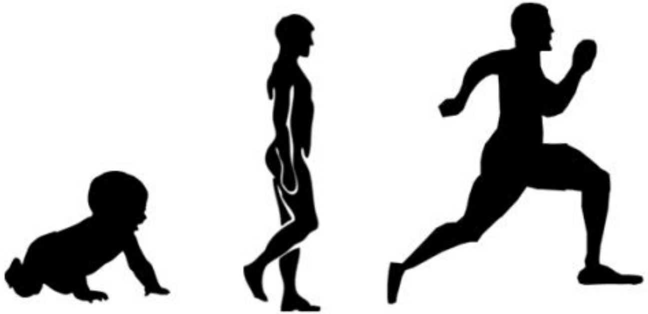
Never Ending Language Learning

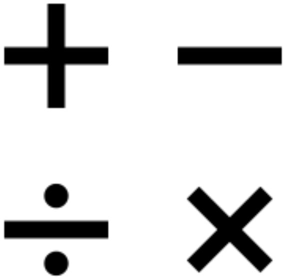
T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, J. Welling

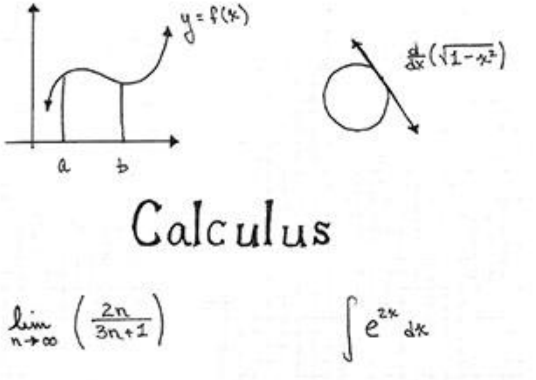
Slides borrowed from Tom M. Mitchell and Andrew Carlson

Human Learning

- Curricular
- Diverse, Multi-task
- **Never Ending**





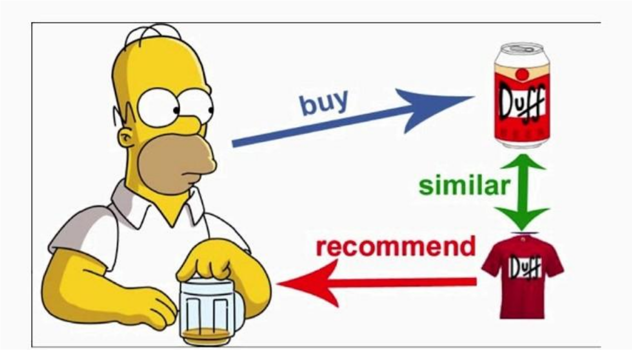


Pratyush, Soumya

Machines and Humans learn in fundamentally different ways

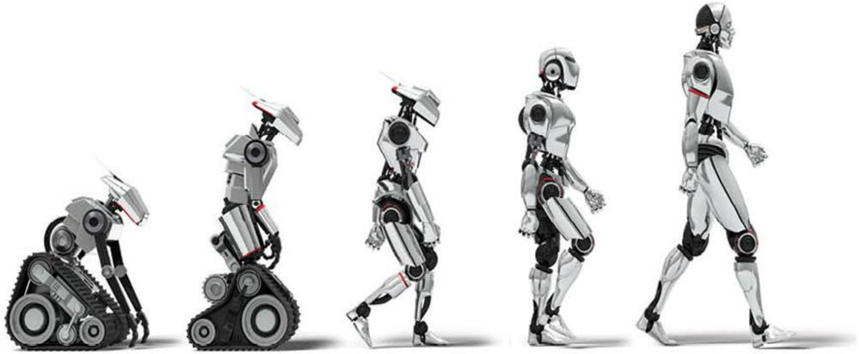
Typical Machine Learning

- Supervised
- Single-Task
- Performance plateaus
- **Not never-ending**



Never Ending Machine Learning

- Robotics
- Role Playing games



NELL - Never-Ending Language Learner

- Semi-supervised Learning
- Bootstrapped Learning
- Multi-Task Learning
- Active Learning
- Curriculum Learning

All this leads to...

- **Never-Ending Learning**

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 - **How will we know?**

NELL is a Knowledge Base

Knowledge Base is a **belief system.**

Knowledge Base **reduces redundancy on the web**.

- Collection of tuples - (subject, relation, object)
- Open vs Closed
- Typed vs Untyped

NELL is a Typed, Closed KB

Text → Facts → Knowledge Base → Applications

Demo

[Tea]

[Diabetes]

[Pakistan People's Party]

Lovish

Learning Task 1 : Category Classification of Noun Phrases

Semi-Supervised Bootstrap Learning

Semantic drift

<u>Extract cities:</u>

Paris
Pittsburgh
Seattle
Cupertino

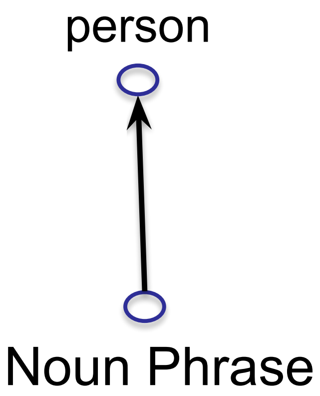
mayor of arg1
live in arg1

San Francisco
Austin
denial

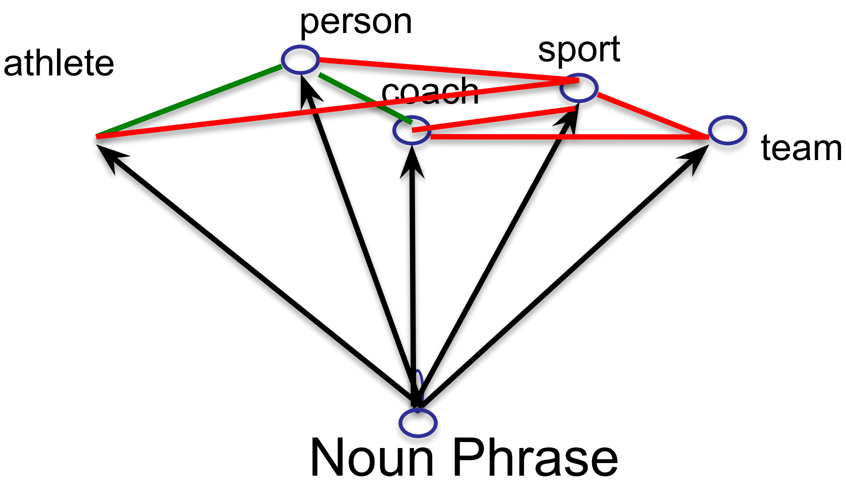
arg1 is home of
traits such as arg1

anxiety
selfishness
Berlin

Solution : Coupled Training using Constraints

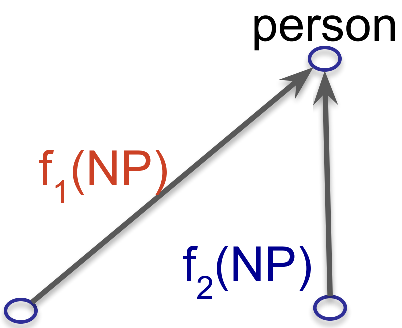


hard
(underconstrained)



much easier (more constrained)

Example : Coupled Training using Constraints



NP: NP context distribution | NP morphology

__ is a friend | *capitalized?*

rang the __ | *ends with '...ski'?*

... | *...*

__ walked in | *contains "univ."?*

Coupled training of 2 functions:

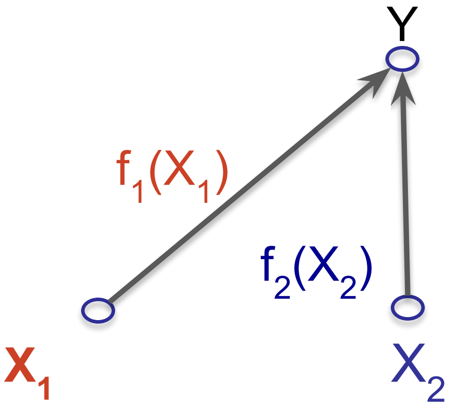
$$\text{Minimize: } \sum_{<np,person> \,\in\, labeled\ data} |f_1(np) - person|$$

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

$$+ \sum_{np \,\in\, unlabeled\ data} |f_1(np) - f_2(np)|$$

Consistency ≡ Accuracy ??

Example : Coupled Training using Constraints



If f_1, f_2 PAC learnable, X_1, X_2 conditionally independent given Y, disagreement between f_1 and f_2 bounds the error of each.

NP:

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

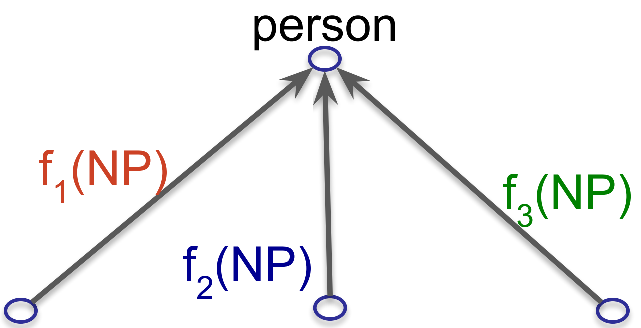
Consistency ≡ Accuracy ??

Never-Ending Learning Design Principle 1

“To achieve successful semi-supervised learning, couple the training of many different learning tasks.”

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]

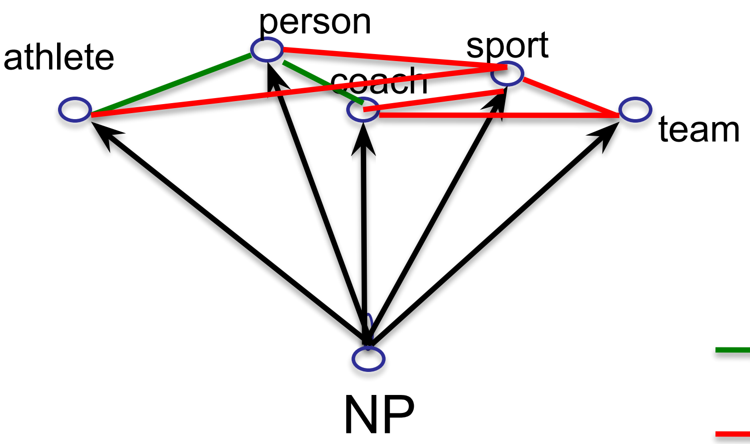


NP:

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: Subset/Superset

Type 3 Coupling: Multi Label Mutual Exclusion



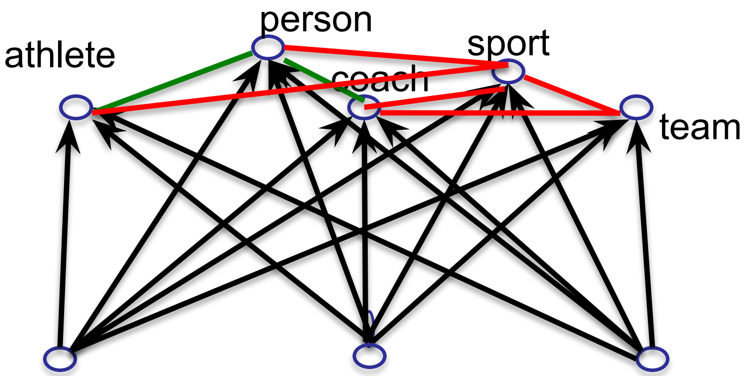
[Daume, 2008]
[Bakhir et al., eds. 2007]
[Roth et al., 2008]
[Taskar et al., 2009]
[Carlson et al., 2009]

athlete(NP) → person(NP)

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

Type 2 Coupling: Subset/Superset

Type 3 Coupling: Multi Label Mutual Exclusion



NP:

NP text context distribution

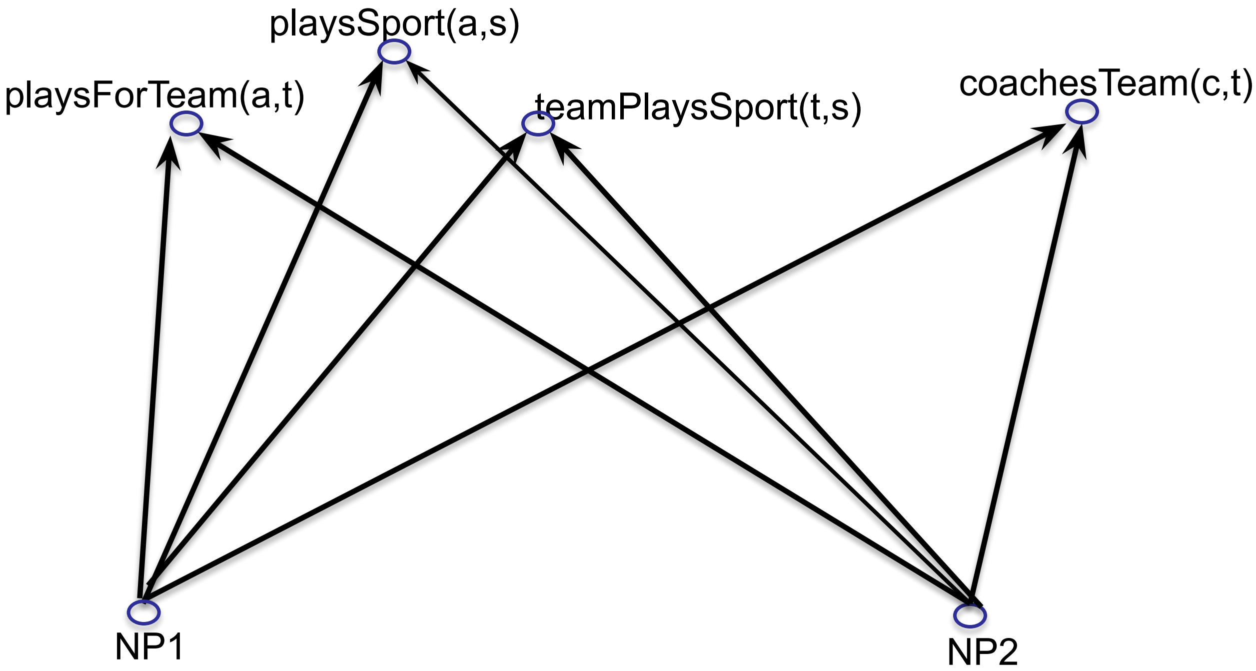
NP morphology

NP HTML contexts

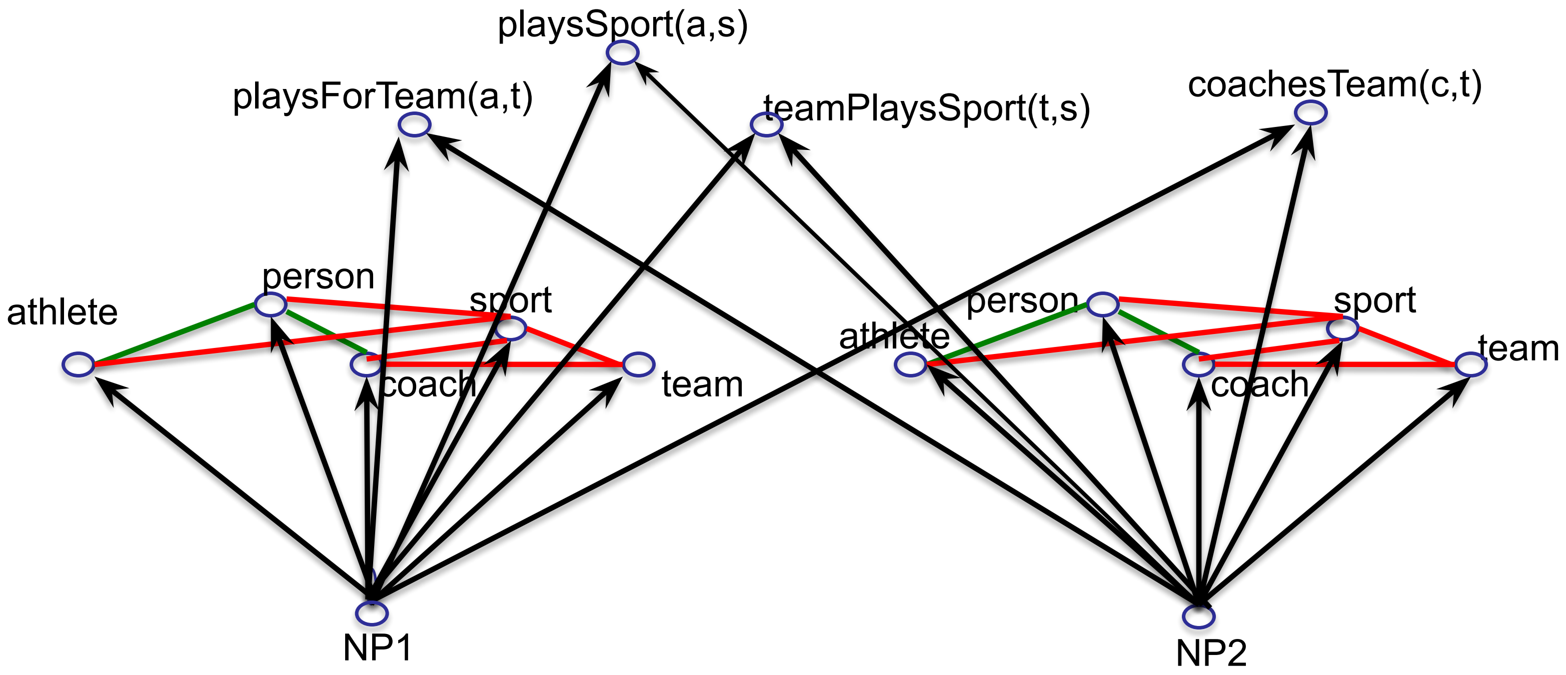
Atishya?

Learning Task 2 : Relation Classification

Learning Relations between Noun Phrases

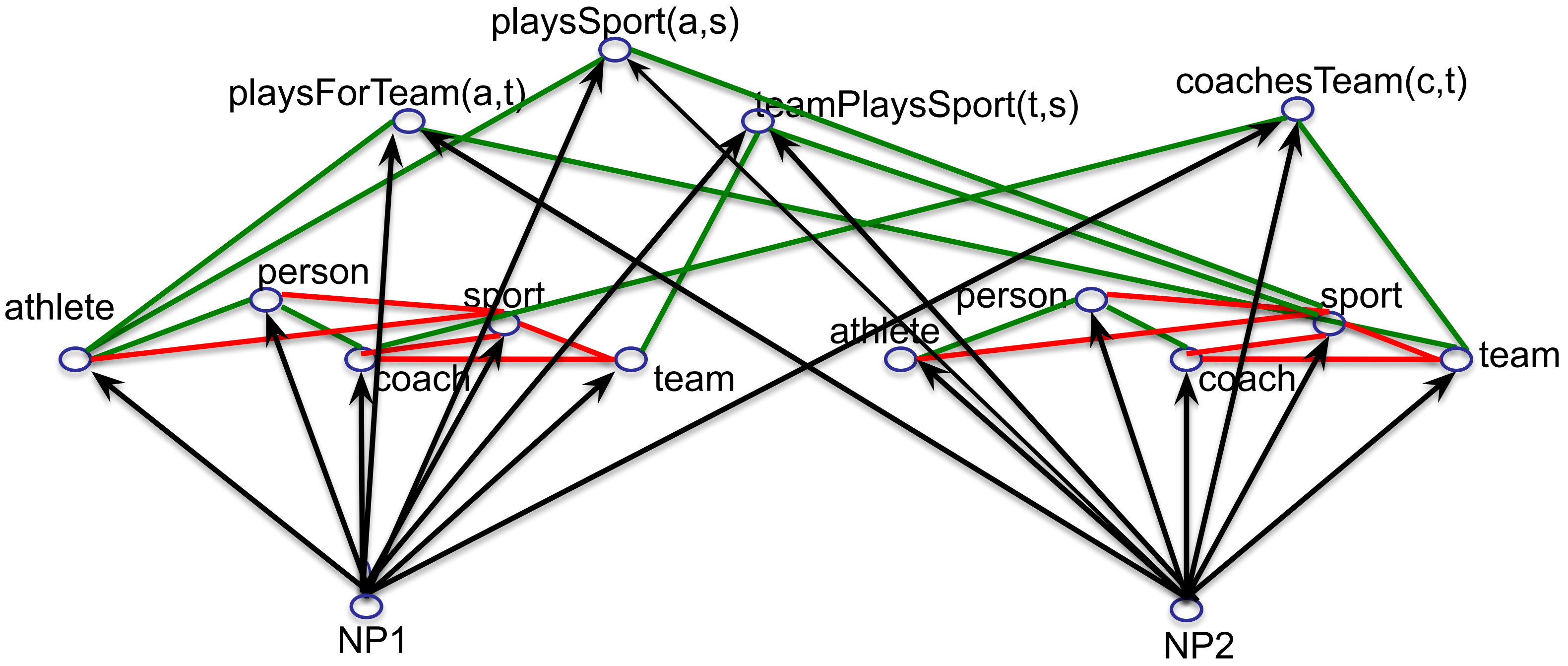


Learning Relations between Noun Phrases



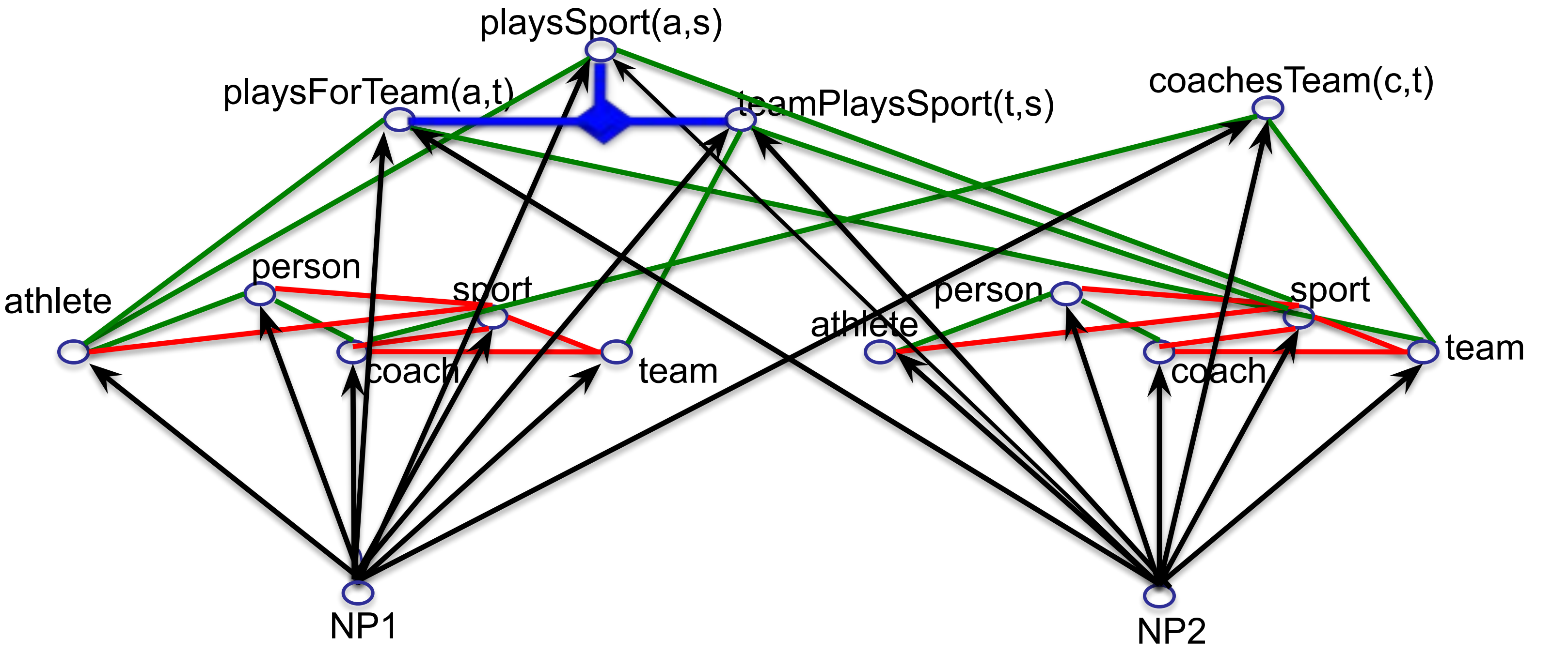
Type 4 Coupling: Argument Types

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



Type 5 Coupling: Horn Clauses

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



How did we get Horn Clauses?

Learning Task 3 : Inference Rules among Belief Triples

Learning Horn Clauses

How :

- Data mining empirical evidence
- Path Ranking Algorithm (PRA)

Why :

- Infer new beliefs
- **Get more constraints !!**

Never-Ending Learning Design Principle 2

“To achieve successful semi-supervised learning, couple the training of many different learning tasks.”

“Allow the agent to learn additional coupling constraints.”

Examples of Learnt 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)

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

Due to "macro-reading"

Requires human supervision ~5 minutes a day

Are we done?
Will NELL learn forever now?

Learning Task 3 : New Relations and Sub-categories

Ontology Extension - Relation

[Mohamed et al., *EMNLP* 2011]

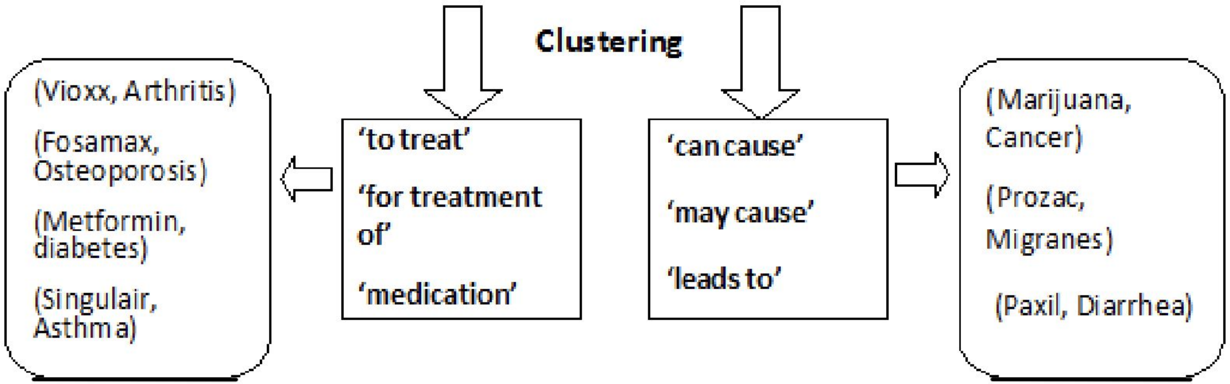
Key Idea -

- **Redundancy of information** in web data - the **same relational fact** is often stated **multiple times** in large text corpora, using **different context patterns**.

Approach :-

- For each pair of categories C1, C2
    - Build a Context by Context co-occurrence matrix.
    - Apply K-means clustering to get candidate relations.
    - Rank and get top 50 instance pairs as seed instances.

Contexts/ Contexts	may cause	can cause	can lead to	to treat	for treatment of	medication
may cause	0.176	0.074	0.030	0.015	0.011	0.000
can cause	0.051	0.150	0.039	0.018	0.013	0.010
can lead to	0.034	0.064	0.189	0.019	0.021	0.018
to treat	0.006	0.011	0.007	0.109	0.043	0.015
for treatment of	0.005	0.008	0.008	0.045	0.086	0.023
medication	0.000	0.011	0.009	0.030	0.036	0.111



Ontology Extension - Relation (Errors)

Table 2. Examples of Incorrect category instances.

name(category1 -main context- category2)	Relation Contexts	Seed Instances
SportsGame -Beating- Country	'beating'	"tournament,Sri Lanka" "champions, France" "match, canada"
Animal -will eat- Condiment	'will eat' 'eating'	"wolf, sheep" "fox, rabbit" "lion, lamb"

Table 5. Examples of relations representing facts that are not concrete.

Name	Relation Contexts	Seed Instances
Emotion -of living in- StateOrProvince	'of living in'	"joy, california" "excitement, colorado" "fear, iowa"
BodyPart -to keep- BodyPart	'to keep' 'guard'	"hand, eye" "nose, throat" "eye, brain" "elbow, hand"

Keshav, Rajas

Source of error - NELL Itself !

Solution : Classifier → Human supervision

Ontology Extension - Sub-category

[Burr Settles]

Key Idea -

- Formulate the problem as **finding a new relation**.

Approach :-

- For each category C
  - Train NELL to **read** the relation **SubsetOfC: C→C**

NELL : Self-Discovered Sub-categories

Sankalan, Shubham

Animal:

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

Learning categories?

Learned reading patterns for AnimalSubset(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 migrant arg2"

"arg1 and other Ice Age arg2"
"arg1 or other biting arg2"
"arg1 and other mammals and arg2"
"arg1 and other marsh arg2"
"arg1 and other monogastric arg2"

NELL : Self-Discovered Sub-categories

Sankalan, Shubham

Animal:

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

everypromotedthing

Learned reading patterns for AnimalSubset(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 migrant arg2”

"arg1 and other Ice Age arg2"

"arg1 or other biting arg2"

"arg1 and other mammals and arg2"

"arg1 and other marsh arg2"

"arg1 and other monogastric arg2"

Never-Ending Learning Design Principle 3

“To achieve successful semi-supervised learning, couple the training of many different learning tasks.”

“Allow the agent to learn additional coupling constraints.”

“Learn new representations that cover relevant phenomena beyond the initial representation.”

Never-Ending Learning Design Principle 4

What to do :

“To achieve successful semi-supervised learning, couple the training of many different learning tasks.”

“Allow the agent to learn additional coupling constraints.”

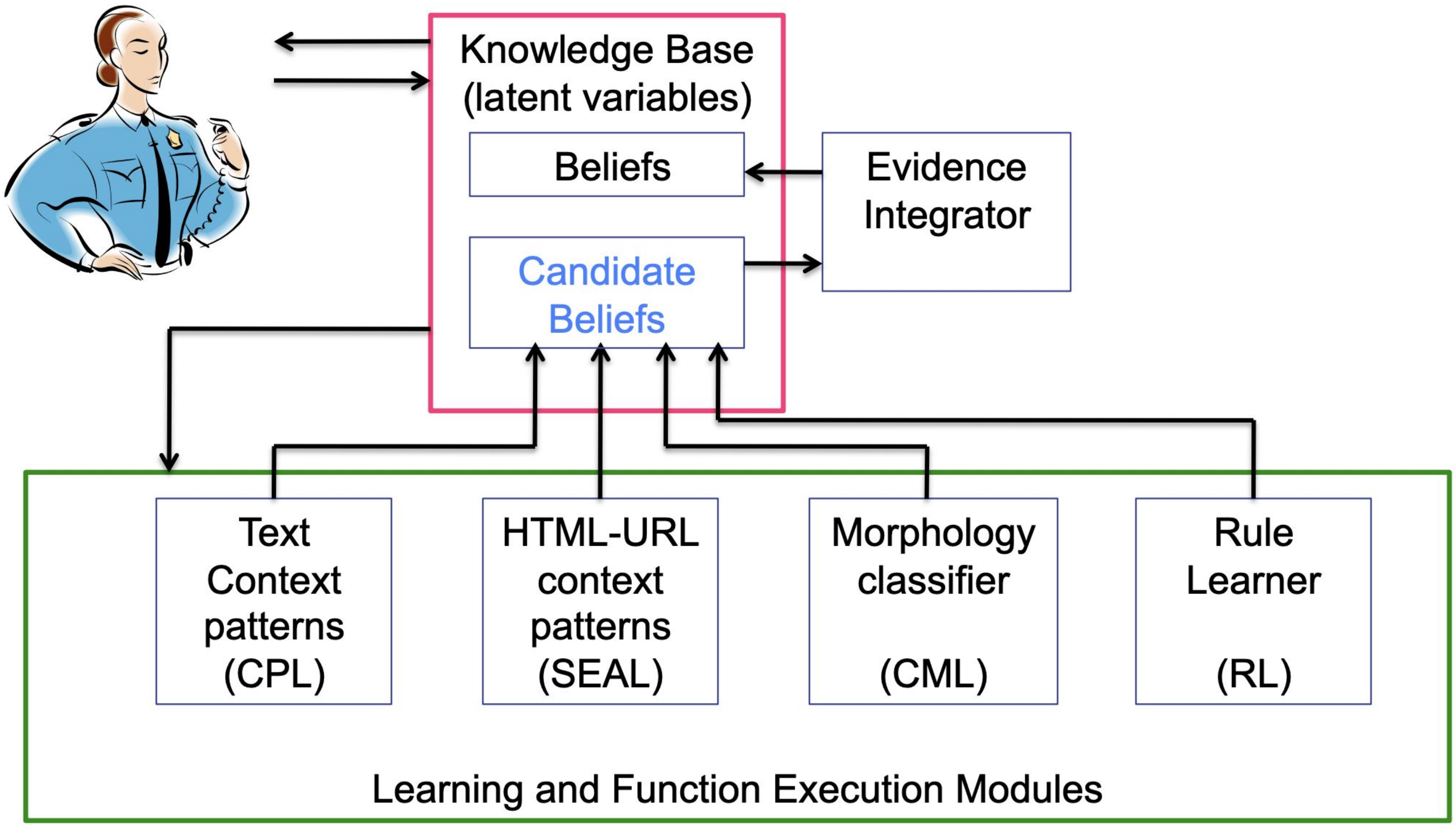
“Learn new representations that cover relevant phenomena beyond the initial representation.”

How to do:

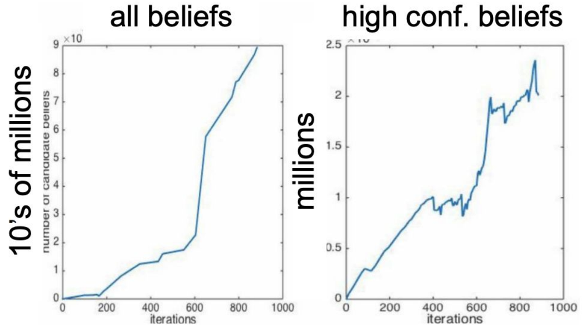
“Organize the set of learning tasks into an easy-to-increasingly-difficult curriculum.”

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
12. Add a robot body to NELL

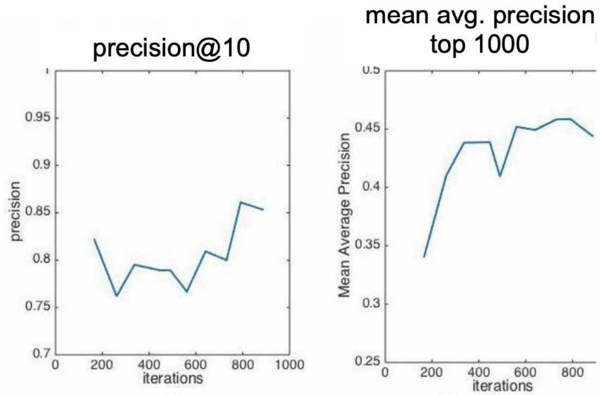
NELL Architecture (Simplified)



NELL Evaluation



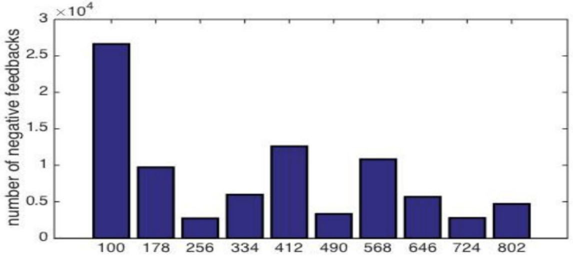
number of NELL beliefs vs. time



reading accuracy vs. time
(average over 31 predicates)

Sushant, Soumya

Methodology satisfactory?



human feedback vs. time
(average 2.4 feedbacks per predicate per month)

Limitations

- **Self-reflection** - How well am I doing?
- **“Macro-reading”** - Reliance on redundancy of web

Most frequent is read first, doesn’t know when to stop looking ...

NELL “emotions”

shame	envy
guilt	gratitude
regret	rage
embarrassment	pride
stress	compassion
pity	elation
empathy	anguish
resentment	hurt
awe	relief
sympathy	ecstasy
laughter	angst
despair	dread
sorrow	hopelessness
concern	longing
lust	remorse
loneliness	anxieties
grief	melancholy
disappointment	fright

← **Earliest extractions**

NELL "emotions" (at 100 iterations)

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

2,636 extracted emotions,

490 extraction patterns

← Earliest extractions

Most recent extractions →

profound dislike
split_personality
themotivation
fierce_joy
practical_assistance
fearand
interest_toall
differentnature
approval
overwhelming_wave
vengence
policy_relevance
disavowal
manifestation
change
mild_bitterness
unfounded_fears
full_support

Extensions

- Temporal Scoping
    - **Coupled Temporal Scoping of Relational Facts**. P.P. Talukdar, D.T. Wijaya and T.M. Mitchell. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM)*, 2012.
    - **Acquiring Temporal Constraints between Relations**. P.P. Talukdar, D.T. Wijaya and T.M. Mitchell. In *Proceedings of the Conference on Information and Knowledge Management (CIKM)*, 2012
    - **CTPs: Contextual Temporal Profiles for Time Scoping Facts via Entity State Change Detection**. D.T. Wijaya, N. Nakashole and T.M. Mitchell. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014
- Domain-Specific NELL
    - **Bootstrapping Biomedical Ontologies for Scientific Text using NELL**. Dana Movshovitz-Attias and William W. Cohen, , in *BioNLP-2012*, 2012
- Integration with other KBs
- “Micro-reading”

Limitations (Piazza)

- Temporal scoping [Almost everyone]
  - WasPrimeMinisterOf(), IsPrimeMinisterOf() [Atishya]
  - (e1,r,e2,timespan) Constraint Learning on timespans [Sushant]
  - Time histogram [Keshav]
- Dependent on human supervision [Deepanshu, Pratyush, Pawan, Shubham]
  - Is that a problem?
  - Adversarial human supervision? - think of “independent errors”
  - Can NELL work with absolutely no supervision?
- Missing implementation details [Sushant]
  - Mostly heuristics (Refer paper in additional material)

9 PhD Thesis in CMU! [Vipul]

Limitations (Piazza)

- Trustworthy sources on web? [Keshav, Lovish, Sankalan]
  - Fake news can fool “macro-reading”
  - Deal with fictitious context (Eg: Harry Potter) [Jigyasa]
- Negative feedback using low-scoring beliefs [Jigyasa]
  - What does a low score in macro-reading mean?
- Deletion of old facts [Many people]
  - Best guess - think how EM does it
- Knowledge graph embedding literature [Siddhant, Vipul]
- NELL will learn our biases - “character of the web” [Sankalan]
  - tomato

Discussion

- State of NELL today?
- Should NELL incorporate “what to read”?
    - How will it decide in an unsupervised setting?
- Where can we apply never-ending learning?
    - Any application in Computer Vision?
    - What will be the tasks and constraints?
- Have never-ending learning design principles changed?
    - Can we add something new?