Never Ending Language Learning

Carnegie Mellon University
Tenet 1:
Understanding requires a belief system

We’ll never produce natural language understanding 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 learners 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
## Recently-Learned Facts

<table>
<thead>
<tr>
<th>Instance</th>
<th>Iteration</th>
<th>Date Learned</th>
</tr>
</thead>
<tbody>
<tr>
<td>banana_nut_chocolate_chip_bread is a baked good</td>
<td>446</td>
<td>03-nov-2011</td>
</tr>
<tr>
<td>nisha ganatra is an African person</td>
<td>446</td>
<td>03-nov-2011</td>
</tr>
<tr>
<td>corky miller is a male</td>
<td>445</td>
<td>01-nov-2011</td>
</tr>
<tr>
<td>two marriages is a parlour game</td>
<td>443</td>
<td>29-oct-2011</td>
</tr>
<tr>
<td>tanner police department is a part of the government</td>
<td>446</td>
<td>03-nov-2011</td>
</tr>
<tr>
<td>georgia aquarium is a tourist attraction in the city atlanta</td>
<td>448</td>
<td>05-nov-2011</td>
</tr>
<tr>
<td>florida hospital is a hospital in the city orlando</td>
<td>446</td>
<td>03-nov-2011</td>
</tr>
<tr>
<td>adobe systems incorporated is a company also known as adobe</td>
<td>445</td>
<td>01-nov-2011</td>
</tr>
<tr>
<td>the sports team yankees was the winner of n1923 world series</td>
<td>448</td>
<td>05-nov-2011</td>
</tr>
<tr>
<td>north arm is a city located in the state or province ohio</td>
<td>446</td>
<td>03-nov-2011</td>
</tr>
</tbody>
</table>
NELL Today

- [http://rtw.ml.cmu.edu](http://rtw.ml.cmu.edu) ← follow NELL here

### Recently-Learned Facts

<table>
<thead>
<tr>
<th>Instance</th>
<th>Iteration</th>
<th>Date Learned</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>the_rembrandts</td>
<td>535</td>
<td>21-mar-2012</td>
<td>is a TV show</td>
</tr>
<tr>
<td>alexander_memorial_coliseum</td>
<td>535</td>
<td>21-mar-2012</td>
<td>is a building</td>
</tr>
<tr>
<td>jason_bergmann</td>
<td>535</td>
<td>21-mar-2012</td>
<td>is an athlete</td>
</tr>
<tr>
<td>cnc_costume_national_shoes</td>
<td>535</td>
<td>21-mar-2012</td>
<td>is a kind of clothing</td>
</tr>
<tr>
<td>korinthia</td>
<td>535</td>
<td>21-mar-2012</td>
<td>is a state or a province</td>
</tr>
<tr>
<td>children</td>
<td>536</td>
<td>23-mar-2012</td>
<td>is an animal that can develop problems</td>
</tr>
<tr>
<td>burlington</td>
<td>536</td>
<td>23-mar-2012</td>
<td>is a city that lies on the river skagit</td>
</tr>
<tr>
<td>grains</td>
<td>536</td>
<td>23-mar-2012</td>
<td>is a generalization of maize</td>
</tr>
<tr>
<td>j_s_bach</td>
<td>539</td>
<td>26-mar-2012</td>
<td>is a musician who plays the piano</td>
</tr>
<tr>
<td>mary</td>
<td>538</td>
<td>25-mar-2012</td>
<td>is the parent of god_the_son</td>
</tr>
</tbody>
</table>
**NELL Today**

- [http://rtw.ml.cmu.edu](http://rtw.ml.cmu.edu) ➜ follow NELL here

### Recently-Learned Facts

<table>
<thead>
<tr>
<th>instance</th>
<th>iteration</th>
<th>date learned</th>
</tr>
</thead>
<tbody>
<tr>
<td>hartsfield is an airport</td>
<td>554</td>
<td>22-apr-2012</td>
</tr>
<tr>
<td>ted marcoux is a commedian</td>
<td>554</td>
<td>22-apr-2012</td>
</tr>
<tr>
<td>graph paper is an office supply</td>
<td>559</td>
<td>01-may-2012</td>
</tr>
<tr>
<td>san antonio museum of art is a museum</td>
<td>554</td>
<td>22-apr-2012</td>
</tr>
<tr>
<td>south ossetia is a country</td>
<td>556</td>
<td>26-apr-2012</td>
</tr>
<tr>
<td>duke ellington is a musician who plays the piano</td>
<td>559</td>
<td>01-may-2012</td>
</tr>
<tr>
<td>classic books is a generalization of pride and prejudice</td>
<td>559</td>
<td>01-may-2012</td>
</tr>
<tr>
<td>state is a synonym for department</td>
<td>557</td>
<td>29-apr-2012</td>
</tr>
<tr>
<td>brookfield zoo is a tourist attraction in the city chicago</td>
<td>559</td>
<td>01-may-2012</td>
</tr>
<tr>
<td>ford motor is a company that produces falcon</td>
<td>559</td>
<td>01-may-2012</td>
</tr>
</tbody>
</table>
Semi-Supervised Bootstrap Learning

Extract cities:

- Paris
- Pittsburgh
- Seattle
- Cupertino
- San Francisco
- Austin
- denial
- anxiety
- selfishness
- Berlin

It’s underconstrained!!

- 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]
Multi-view, Multi-Task Coupling
Learning Relations between NP’s
Type 3 Coupling: Argument Types

playsSport(NP1, NP2) $\rightarrow$ athlete(NP1), sport(NP2)

over 2500 coupled functions in NELL
Basic NELL Architecture

Continually Learning Extractors

Knowledge Base (latent variables)
- Beliefs
- Candidate Beliefs

Evidence Integrator

Text Context patterns (CPL)
HTML-URL context patterns (SEAL)
Morphology classifier (CML)
NELL: Learned reading strategies

Plays_Sport(arg1,arg2):

arg1_was_playing_arg2 arg2_megastar
arg2_player_named_arg1 arg2_prodigy
arg1_is_the_tiger_woods_of_arg2
arg2_greats_as_arg1 arg1_plays_arg2
arg2_legends_arg1 arg1_announced_his_retirement_from_arg2
arg2_operations_chief_arg1 arg2_plays_arg2
arg2_and_golfing_personalities_including_arg1
arg2_greats_like_arg1 arg2_players
arg2_icon_arg1 arg2_stars_like_arg1
arg1_retires_from_arg2 arg2_phenom
arg2_architects_robert_trent_jones_arg1
arg2_professionals_such_as_arg1 arg2_sensation_arg1
arg2_icon.Arg1 arg2_stars_venus_and_arg1
arg2_greats_arg1 arg2_champ_arg1
arg2_greats_such_as_arg1
arg2_professionals_such_as_arg1 arg2_hit_by_arg2
arg2_icon_arg1 arg2_stars_like_arg1
arg1_retires_from_arg2 arg2_legends_arg1
arg2_autographed_by_arg1

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

<table>
<thead>
<tr>
<th>Predicate</th>
<th>Web URL</th>
<th>Extraction Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>bird</td>
<td><a href="http://www.michaelforsberg.com/stock.html">http://www.michaelforsberg.com/stock.html</a></td>
<td></td>
</tr>
<tr>
<td>bookAuthor</td>
<td><a href="http://lifebehindthecurve.com/">http://lifebehindthecurve.com/</a></td>
<td></td>
</tr>
</tbody>
</table>
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:

\[
0.93 \ \text{athletePlaysSport}(?x,?y) \leftarrow \text{athletePlaysForTeam}(?x,?z) \\
\text{teamPlaysSport}(?z,?y)
\]

- 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,?x,?x)
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 \( \text{playsSport}(?x,?y) \leftarrow \text{playsForTeam}(?x,?z), \text{teamPlaysSport}(?z,?y) \)
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*]

<table>
<thead>
<tr>
<th>Category Pair</th>
<th>Text contexts</th>
<th>Extracted Instances</th>
<th>Suggested Name</th>
</tr>
</thead>
</table>
| 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, histamine  | 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
Inference by KB Random Walks

If: \( x_1 \) competes with \((x_1, x_2)\) then: economic sector \((x_2, x_3)\)

Then: economic sector \((x_1, x_3)\)
Inference by KB Random Walks

[Laot et al, EMNLP 2011]

KB:

Random walk path type:

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]
CityLocatedInCountry(Pittsburgh) = ?

[Feature = Typed Path]
CityInState, CityInstate$^{-1}$, CityLocatedInCountry

Feature Value: 0.8
Weight: 0.32

[Logistic Regression]

[Lao et al, EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

[Feature = Typed Path]

CityInState, CityInstate\(^{-1}\), CityLocatedInCountry

Feature Value

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CityLocatedInCountry(Pittsburgh)</td>
<td>0.8</td>
</tr>
</tbody>
</table>

[Logistic Regression]

Weight

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.32</td>
</tr>
</tbody>
</table>
Feature = Typed Path
CityInState, CityInstate⁻¹, CityLocatedInCountry

Feature Value
CityLocatedInCountry(Pittsburgh) = ?

Logistic Regression
Weight
0.8
0.32

[Feature Value]

CityLocatedInCountry(Pittsburgh) = ? [Lao et al, EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

Pr(U.S. | Pittsburgh, TypedPath)

Feature = Typed Path
CityInState, CityInstate\(^{-1}\), CityLocatedInCountry

Feature Value
0.8

Logistic Regression
Weight
0.32

[Lao et al, EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

[Feature = Typed Path]
CityInState, CityInState$^{-1}$, CityLocatedInCountry
AtLocation$^{-1}$, AtLocation, CityLocatedInCountry

Feature Value
CityLocatedInCountry(Pittsburgh) = 0.8

Logistic Regression
Weight
0.32
0.20

[Lao et al, EMNLP 2011]
Feature = Typed Path
CityInState, CityInState\(^{-1}\), CityLocatedInCountry
AtLocation\(^{-1}\), AtLocation, CityLocatedInCountry

Feature Value
CityLocatedInCountry(Pittsburgh) = ?

Feature Value
0.8
0.32
0.20

[Logistic Regression]

[Source: Lao et al, EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

Feature = Typed Path
CityInState, CityInState^{-1}, CityLocatedInCountry
AtLocation^{-1}, AtLocation, CityLocatedInCountry

Feature Value
CityInState 0.8
CityInState^{-1} 0.32
CityLocatedInCountry 0.20

Logistic Regression Weight

[Chao et al., EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

Feature = Typed Path
CityInState, CityInState\(^{-1}\), CityLocatedInCountry
AtLocation\(^{-1}\), AtLocation, CityLocatedInCountry

Feature Value
0.8
0.6

Logistic Regression
Weight
0.32
0.20

[Lao et al, EMNLP 2011]
Feature = Typed Path
CityInState, CityInstate$^{-1}$, CityLocatedInCountry
AtLocation$^{-1}$, AtLocation, CityLocatedInCountry
...

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

[Logistic Regression]

CityLocatedInCountry(Pittsburgh) = ?

[Laos et al, EMNLP 2011]
Random walk inference: learned path types

CityLocatedInCountry\((city, country)\):

8.04 cityliesonriver, cityliesonriver\(^{-1}\), citylocatedincounty
5.42 hasofficeincity\(^{-1}\), hasofficeincity, citylocatedincounty
4.98 cityalsoknownnas, cityalsoknownnas, citylocatedincounty
2.85 citycapitalofcountry, citylocatedincounty\(^{-1}\), citylocatedincounty
2.29 agentactsinlocation\(^{-1}\), agentactsinlocation, citylocatedincounty
1.22 statehascapital\(^{-1}\), statelocatedincounty
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):

<table>
<thead>
<tr>
<th>NELL/PRA ranking</th>
<th>Human Ranking (9 subjs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>Apple</td>
</tr>
<tr>
<td>Oracle</td>
<td>Google</td>
</tr>
<tr>
<td>IBM</td>
<td>Yahoo</td>
</tr>
<tr>
<td>Apple</td>
<td>IBM</td>
</tr>
<tr>
<td>SAP</td>
<td>Redhat</td>
</tr>
<tr>
<td>Yahoo</td>
<td>Oracle</td>
</tr>
<tr>
<td>Facebook</td>
<td>Facebook</td>
</tr>
<tr>
<td>Redhat</td>
<td>SAP</td>
</tr>
<tr>
<td>Lenovo</td>
<td>SAP</td>
</tr>
<tr>
<td>FedEx</td>
<td>Lenovo</td>
</tr>
<tr>
<td>SAS</td>
<td>FedEx</td>
</tr>
<tr>
<td>Boeing</td>
<td>Boeing</td>
</tr>
<tr>
<td>Honda</td>
<td>HP</td>
</tr>
<tr>
<td>Dupont</td>
<td>Du Pont</td>
</tr>
<tr>
<td>Lufthansa</td>
<td>Lufthansa</td>
</tr>
<tr>
<td>Exxon</td>
<td>Exxon</td>
</tr>
<tr>
<td>Pfizer</td>
<td>Pfizer</td>
</tr>
</tbody>
</table>

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

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