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 web
  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

<table>
<thead>
<tr>
<th>Instance</th>
<th>Iteration</th>
<th>date learned</th>
<th>confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>dilator muscle of pupil is a muscle</td>
<td>210</td>
<td>17-Feb-2011</td>
<td>100.0</td>
</tr>
<tr>
<td>boyden cave is a cave</td>
<td>211</td>
<td>18-Feb-2011</td>
<td>100.0</td>
</tr>
<tr>
<td>pondicherry is a state or a province</td>
<td>211</td>
<td>18-Feb-2011</td>
<td>100.0</td>
</tr>
<tr>
<td>vena brachialis is a vein</td>
<td>211</td>
<td>18-Feb-2011</td>
<td>97.5</td>
</tr>
<tr>
<td>scott rigell is a U.S. politician</td>
<td>210</td>
<td>17-Feb-2011</td>
<td>96.0</td>
</tr>
<tr>
<td>toronto is the home city of the sports team</td>
<td>210</td>
<td>17-Feb-2011</td>
<td>93.8</td>
</tr>
<tr>
<td>jrn mcmurry is the CEO of boeing</td>
<td>210</td>
<td>17-Feb-2011</td>
<td>96.0</td>
</tr>
<tr>
<td>microsoft is a company that produces windows vista</td>
<td>213</td>
<td>22-Feb-2011</td>
<td>100.0</td>
</tr>
<tr>
<td>frogs is an animal that is a kind of small animals</td>
<td>210</td>
<td>17-Feb-2011</td>
<td>93.8</td>
</tr>
<tr>
<td>kcr is a TV station in the city houston</td>
<td>210</td>
<td>17-Feb-2011</td>
<td>93.8</td>
</tr>
</tbody>
</table>
Semi-Supervised Bootstrap Learning

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

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

It's underconstrained!!

- mayor of arg1
- live in arg1
- arg1 is home of traits such as arg1

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 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 \text{athlete}(NP1), \text{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 $\rightarrow 10^{10}$ 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 retraining itself from the updated KB
- “context” methods create growing subsets of contexts
### NELL Architecture

![Diagram of NELL Architecture](image)

### Never-Ending Language Learning

**Predicate** | **Feature** | **Weight**
--- | --- | ---
mountain | LAST=peak | 1.791
mountain | LAST=mountain | 1.093
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=is | 1.282
visualArtMovement | PREFIX=jour | -0.234
visualArtMovement | PREFIX=budd | -0.253

<table>
<thead>
<tr>
<th>Predicate</th>
<th>Web URL</th>
<th>Extraction Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>academicField</td>
<td><a href="http://scholndowais.msu.edu/student/ScholSearch.Asp">http://scholndowais.msu.edu/student/ScholSearch.Asp</a></td>
<td> [X] =</td>
</tr>
<tr>
<td>bird</td>
<td><a href="http://www.michaelforsberg.com/stock.html">http://www.michaelforsberg.com/stock.html</a></td>
<td>&lt;/li&gt; &lt;li&gt;&lt;X&gt; by [Y] &amp;#8211;</td>
</tr>
<tr>
<td>bookAuthor</td>
<td><a href="http://lifebehindthecurve.com/">http://lifebehindthecurve.com/</a></td>
<td></td>
</tr>
</tbody>
</table>
Coupled Training Helps!

[Carlson et al., WSDM 2010]

Using only two views:
Text, HTML contexts.

<table>
<thead>
<tr>
<th>PRECISION</th>
<th>Text uncpl</th>
<th>HTML uncpl</th>
<th>Coupled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories</td>
<td>.41</td>
<td>.59</td>
<td>.90</td>
</tr>
<tr>
<td>Relations</td>
<td>.69</td>
<td>.91</td>
<td>.95</td>
</tr>
</tbody>
</table>

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

| EconomicSector | 23 | 10 | 77 |
| Food | 53 | 60 | 83 |
| Furniture | 70 | 80 | 100 |
| Hobby | 0 | 57 | 90 |
| KitchenItem | 3 | 13 | 100 |
| Mammal | 50 | 50 | 90 |
| Movie | 57 | 100 | 100 |
| NewspaperCompany | 60 | 97 | 100 |
| Politician | 60 | 57 | 100 |
| Product | 83 | 77 | 70 |
| ProductType | 63 | 63 | 80 |
| 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:

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

– 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
Example Learned Horn Clauses

0.95  \text{athletePlaysSport}(?x,\text{basketball}) \leftarrow \text{athleteInLeague}(?x,\text{NBA})

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

0.91  \text{teamPlaysInLeague}(?x,\text{NHL}) \leftarrow \text{teamWonTrophy}(?x,\text{Stanley\_Cup})

0.90  \text{athleteInLeague}(?x,?y) \leftarrow \text{athletePlaysForTeam}(?x,?z), \text{teamPlaysInLeague}(?z,?y)

0.88  \text{cityInState}(?x,?y) \leftarrow \text{cityCapitalOfState}(?x,?y), \text{cityInCountry}(?y,\text{USA})

0.62* \text{newspaperInCity}(?x,\text{New\_York}) \leftarrow \text{companyEconomicSector}(?x,\text{media}), \text{generalizations}(?x,\text{blog})

Some rejected learned rules

\text{teamPlaysInLeague}(?x,\text{nba}) \leftarrow \text{teamPlaysSport}(?x,\text{basketball})
0.94  \[ 35 0 35 \] [positive negative unlabeled]

\text{cityCapitalOfState}(?x,?y) \leftarrow \text{cityLocatedInState}(?x,?y), \text{teamPlaysInLeague}(?y,\text{nba})
0.80  \[ 16 2 23 \]

\text{teamplayssport}(?x,\text{basketball}) \leftarrow \text{generalizations}(?x,\text{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)
Learning and Function Execution Modules

NELL Architecture

Knowledge Base (latent variables)
Beliefs
Candidate Beliefs
Evidence Integrator

Text Context patterns (CPL)
HTML-URL context patterns (SEAL)
Morphology classifier (CML)
Rule Learner (RL)

Learning and Function Execution Modules

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 KB assertions vs. time

Jan 2010 March July Nov

.90 .75 .71 .87

periodic human supervision begins

= precision of extracted KB
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

<table>
<thead>
<tr>
<th>Category Pair</th>
<th>Name</th>
<th>Text contexts</th>
<th>Extracted Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>MusicInstrument</td>
<td>Master</td>
<td>ARG1 master ARG2</td>
<td>sitar, George Harrison</td>
</tr>
<tr>
<td>Master</td>
<td></td>
<td>ARG1 virtuoso ARG2</td>
<td>tenor sax, Stan Getz</td>
</tr>
<tr>
<td>Master</td>
<td></td>
<td>ARG1 legend ARG2</td>
<td>trombone, Tommy Dorsey</td>
</tr>
<tr>
<td>MusicInstrument</td>
<td>Master</td>
<td>ARG2 plays ARG1</td>
<td>vibes, Lionel Hampton</td>
</tr>
<tr>
<td>Disease</td>
<td>IsDueTo</td>
<td>ARG1 is due to ARG2</td>
<td>pinched nerve, herniated disk</td>
</tr>
<tr>
<td>Disease</td>
<td>IsDueTo</td>
<td>ARG1 is caused by ARG2</td>
<td>tennis elbow, tendonitis</td>
</tr>
<tr>
<td>Disease</td>
<td>IsDueTo</td>
<td></td>
<td>blepharospasm, dystonia</td>
</tr>
<tr>
<td>Chemical</td>
<td>ThatRelease</td>
<td>ARG1 that release ARG2</td>
<td>epithelial cells, surfactant</td>
</tr>
<tr>
<td>Chemical</td>
<td>ThatRelease</td>
<td>ARG2 releasing ARG1</td>
<td>neurons, serotonin</td>
</tr>
<tr>
<td>Chemical</td>
<td>ThatRelease</td>
<td></td>
<td>mast cells, histamine</td>
</tr>
<tr>
<td>Mammals</td>
<td>Eat</td>
<td>ARG1 eat ARG2</td>
<td>koala bears, eucalyptus</td>
</tr>
<tr>
<td>Plant</td>
<td>Eat</td>
<td>ARG2 eating ARG1</td>
<td>grasses, goats, saplings</td>
</tr>
</tbody>
</table>

### Ontology Extension (2)

- NELL sometimes extracts subclasses instead of instances:
  - chemicals: carbon_dioxide, amonia, gas.

- Idea: have NELL learn to read the “Is_A” relation

- Result: NELL currently learns (reads about) new subcategories and their members
Results: Ontology extension by reading

<table>
<thead>
<tr>
<th>Original Category</th>
<th>SubType discovered by reading</th>
<th>Extracted Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical</td>
<td>Gases</td>
<td>ammonia, carbon_dioxide, carbon_monoxide, methane, sulphur, oxides, nitrous_oxides, water_vapor, ozone, nitrogen</td>
</tr>
<tr>
<td>Animal</td>
<td>LiveStock</td>
<td>chickens, cows, sheep, goats, pigs</td>
</tr>
<tr>
<td>Profession</td>
<td>Professionals</td>
<td>surgeons, chiropractors, dentists, engineers, medical staff, midwives, professors, scientists, specialists, technologists, aides</td>
</tr>
</tbody>
</table>

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]
Which Noun Phrases refer to Which Concepts?

[Krishnamurthy & Mitchell, 2011]

**Observed NP’s**

- Apple_theNP
- AppleInc_theNP

**Unobserved Concepts**

- Apple_theFruit
- Apple_theCompany

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

---

**Evaluated Precision/Recall of Pairwise Coreference Decisions:**

<table>
<thead>
<tr>
<th>Category</th>
<th>Precision</th>
<th>Recall</th>
<th>Freebase concepts per NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>athlete</td>
<td>0.95</td>
<td>0.56</td>
<td>1.8</td>
</tr>
<tr>
<td>city</td>
<td>0.97</td>
<td>0.25</td>
<td>3.9</td>
</tr>
<tr>
<td>coach</td>
<td>0.86</td>
<td>0.94</td>
<td>1.1</td>
</tr>
<tr>
<td>company</td>
<td>0.85</td>
<td>0.41</td>
<td>2.4</td>
</tr>
<tr>
<td>country</td>
<td>0.74</td>
<td>0.56</td>
<td>1.8</td>
</tr>
<tr>
<td>sportsteam</td>
<td>0.89</td>
<td>0.30</td>
<td>3.3</td>
</tr>
<tr>
<td>stadium</td>
<td>0.83</td>
<td>0.61</td>
<td>1.6</td>
</tr>
</tbody>
</table>

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_at__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**

---

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