Never Ending Learning

Tom Mitchell

Machine Learning Department
Carnegie Mellon University
New paradigm for Machine Learning:

Never-ending learning agents

- Persistent software individual
- Learns many functions / knowledge types
- Learns easier things first, then more difficult
- The more it learns, the more it can learn next
- Learns from experience, and from advice
NELL: Never-Ending Language Learner

Inputs:
• initial ontology
• dozen 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 ontology
  2. learn to read (perform #1) better than yesterday
NELL today

Running 24x7, since January, 12, 2010

Result:

- KB with > 50 million candidate beliefs, growing daily
- learning to read better each day
- learning to reason, as well as read
- automatically extending its ontology
NELL Today

- [http://rtw.ml.cmu.edu](http://rtw.ml.cmu.edu) ➔ follow NELL here
- eg. “diabetes”, “Avandia”, “tea”, “IBM”, “love” “baseball” “BacteriaCausesCondition” “kitchenItem” “ClothingGoesWithClothing” ...

Recently-Learned Facts

<table>
<thead>
<tr>
<th>instance</th>
<th>iteration</th>
<th>date learned</th>
<th>content</th>
</tr>
</thead>
<tbody>
<tr>
<td>sacramento_convention_center is a stadium or event venue</td>
<td>737</td>
<td>04-jun-2013</td>
<td></td>
</tr>
<tr>
<td>john_kenneth_macalister is a criminal</td>
<td>737</td>
<td>04-jun-2013</td>
<td></td>
</tr>
<tr>
<td>birth_control_drugs is a personal care product</td>
<td>737</td>
<td>04-jun-2013</td>
<td></td>
</tr>
<tr>
<td>almond_chocolate is a kind of candy</td>
<td>739</td>
<td>09-jun-2013</td>
<td></td>
</tr>
<tr>
<td>garlic_shoots is an agricultural product</td>
<td>742</td>
<td>18-jun-2013</td>
<td></td>
</tr>
<tr>
<td>hagar has husband abraham</td>
<td>742</td>
<td>18-jun-2013</td>
<td></td>
</tr>
<tr>
<td>dave_murray is a musician who plays the guitar</td>
<td>739</td>
<td>09-jun-2013</td>
<td></td>
</tr>
<tr>
<td>hart is a city located in the state or province georgia</td>
<td>742</td>
<td>18-jun-2013</td>
<td></td>
</tr>
<tr>
<td>wood_prairie_farm is a farm in the state or province maine</td>
<td>742</td>
<td>18-jun-2013</td>
<td></td>
</tr>
<tr>
<td>pepper is an agricultural product that is usually cooked with canola oil</td>
<td>737</td>
<td>04-jun-2013</td>
<td></td>
</tr>
</tbody>
</table>
How does NELL work?
Semi-Supervised Bootstrap Learning

Find cities:

Paris
Pittsburgh
Seattle
Montpelier

San Francisco
Berlin
denial

anxiety
selfishness
London

mayor of arg1
live in arg1

arg1 is home of
traits such as arg1

it’s underconstrained!!

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

Supervised training of 1 function:

Minimize: \[ \sum_{<np,\text{person}> \in \text{labeled data}} |f_1(np) - person| \]
Type 1 Coupling: Co-Training, Multi-View Learning

Coupled training of 2 functions:

Minimize: \[ \sum_{<np,person> \in \text{labeled data}} |f_1(np) - person| + \sum_{<np,person> \in \text{labeled data}} |f_2(np) - person| + \sum_{np \in \text{unlabeled data}} |f_1(np) - f_2(np)| \]

NP:
- NP context distribution
- NP morphology

_ is a friend
rang the ___
... ends with ‘...ski’?
__ walked in
... contains “univ.”?
Type 1 Coupling: Co-Training, Multi-View Learning

Theorem (Blum & Mitchell, 1998):

If $f_1$ and $f_2$ are PAC learnable from noisy labeled data, and $X_1, X_2$ are conditionally independent given $Y$, then $f_1, f_2$ are PAC learnable from polynomial unlabeled data plus a weak initial predictor.

NP:

- \_ is a friend
- rang the \_
- ... ends with ‘...ski’?
- ... walked in
- contains “univ.”?
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:
- context
distribution
- morphology

is a friend
rang the
walked in
capitalized?
ends with ‘...ski’?
contains “univ.”?
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

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

Graph with nodes labeled athlete, person, sport, coach, and team, connected by edges to show relationships.
Type 3 Coupling: Learning Relations
Type 3 Coupling: Argument Types
Type 3 Coupling: Argument Types

\[ \text{playsSport}(\text{NP1}, \text{NP2}) \rightarrow \text{athlete}(\text{NP1}), \text{sport}(\text{NP2}) \]

\[ \text{playsSport}(a, s) \rightarrow \text{playsForTeam}(a, t), \text{teamPlaysSport}(t, s) \]

over 2500 coupled functions in NELL
NELL: Learned reading strategies

Plays_Sport(arg1, arg2):

- arg1_was_playing_arg2
- arg2_megastar_arg1
- arg2_prodigy_arg1
- 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_and_golfing_personalities_including_arg1
- arg2_greats_like_arg1
- arg2_players_like_arg1
- arg2_icon_arg1
- arg1_retrires_from_arg2
- arg2_professionals_such_as_arg1
- arg2_architects_robert_trent_jones_and_arg1
- arg2_superstar_arg1
- arg2_god_arg1
- arg1_retired_from_professional_arg2
- arg1_was_born_to_play_arg2
- arg2_star_arg1
- arg2_hero_arg1
- arg2_god_arg1
- arg2_idol_arg1
- arg2_is_arg1
- arg2_pro_arg1
- arg2_player_was_arg1
- arg2_greats_arg1
- arg2_champ_arg1
- arg2_greats_such_as_arg1
- arg2_icon_arg1
- arg2_stars_like_arg1
- arg2_pros_arg1
- arg2_bins_venus_and_arg1
- arg2_hall_of_famer_arg1
- arg2_legends_arg1
- arg2_legends_such_as_arg1
- arg2_greats_arg1
- arg2_champ_arg1
- arg2_greats_another_arg1
- arg2_sensation_arg1
- arg2_considered_arg1

### Predicate Table

<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>newspaper</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>1.982</td>
</tr>
</tbody>
</table>

### Web URLs

- academicField: [http://scholendow.ais.msu.edu/student/ScholSearch.Asp](http://scholendow.ais.msu.edu/student/ScholSearch.Asp)

### Extraction Template

```html
&lt;script&gt;&lt;/script&gt;

```
Initial NELL Architecture

Knowledge Base (latent variables)

Beliefs

Candidate Beliefs

Evidence Integrator

Text Context patterns (CPL)

HTML-URL context patterns (SEAL)

Morphology classifier (CML)

Human advice

Continually Learning Extractors
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:

\[
\begin{align*}
0.93 & \quad \text{athletePlaysSport}(?x,?y) \leftarrow \text{athletePlaysForTeam}(?x,?z) \\
& \quad \text{teamPlaysSport}(?z,?y)
\end{align*}
\]

- connects previously uncoupled relation predicates
- infers new beliefs for KB
- modified version of FOIL [Quinlan]
- restricted rule language: form connected KB subgraphs
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 \( \text{team Plays In League}(?x, \text{NHL}) \leftarrow \text{team Won Trophy}(?x, \text{Stanley Cup}) \)

0.90 \( \text{athlete In League}(?x, ?y) \leftarrow \text{athlete Plays For Team}(?x, ?z), \text{team Plays In League}(?z, ?y) \)

0.88 \( \text{city In State}(?x, ?y) \leftarrow \text{city Capital Of State}(?x, ?y), \text{city In Country}(?y, \text{USA}) \)

0.62* \( \text{newspaper In City}(?x, \text{New York}) \leftarrow \text{company Economic Sector}(?x, \text{media}), \text{generalizations}(?x, \text{blog}) \)
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     | 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             | ARG1 is due to ARG2  
ARG1 is caused by ARG2 | pinched nerve, herniated disk  
tennis elbow, tendonitis  
blepharospasm, dystonia | IsDueTo |
| CellType            | ARG1 that release ARG2  
ARG2 releasing ARG1 | epithelial cells, surfactant  
nurons, serotonin  
mast cells, histomine | 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: sample of 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
Ontology Extension (2)  

[Burr Settles]

Goal:
- Add new subcategories

Approach:
- For each category $C$,
  - train NELL to read the relation
    $\text{SubsetOf}_C : C \rightarrow C$

*no new software here*
NELL: example self-discovered subcategories

Animal:
  • Pets
    – Hamsters, Ferrets, Birds, Dog, Cats, Rabbits, Snakes, Parrots, Kittens, …
  • Predator
    – Bears, Foxes, Wolves, Coyotes, Snakes, Racoons, Eagles, Lions, Leopards, Hawks, Humans, …

Learned reading patterns for Subset(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 mammals and arg2"  "arg1 and other Ice Age arg2"  "arg1 or other biting arg2"  "arg1 and other marsh arg2"  "arg1 and other migrant arg2"  "arg1 and other monogastic arg2"  "arg1 and other mythical arg2"  "arg1 and other nesting arg2"
NELL: example self-discovered subcategories

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

Chemical:
- Fossil fuels
  - Carbon, Natural gas, Coal, Diesel, Monoxide, Gases, …
- Gases
  - Helium, Carbon dioxide, Methane, Oxygen, Propane, Ozone, Radon…

Learned reading patterns:
"arg1 and other medium sized arg2"  "arg1 and other jungle arg2"  "arg1 and other magnificent arg2"  "arg1 and other pesky arg2"  "arg1 and other mammals and arg2"  "arg1 and other Ice Age arg2"  "arg1 or other biting arg2"  "arg1 and other marsh arg2"  "arg1 and other migrant arg2"  "arg1 and other monogastric arg2"  "arg1 and other mythical arg2"  "arg1 and other nesting arg2"

"arg1 and other hydrocarbon arg2"  "arg1 and other aqueous arg2"  "arg1 and other hazardous air arg2"  "arg1 and oxygen are arg2"  "arg1 and such synthetic arg2"  "arg1 as a lifting arg2"  "arg1 as a tracer arg2"  "arg1 as the carrier arg2"  "arg1 as the inert arg2"  "arg1 as the primary cleaning arg2"  "arg1 and other noxious arg2"  "arg1 and other trace arg2"  "arg1 as the reagent arg2"  "arg1 as the tracer"
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 latent 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
thank you

and thanks to:
Darpa, Google, NSF, Intel, Yahoo!, Microsoft, Fulbright