Never-Ending Language Learning

Tom Mitchell, William Cohen, and Many Collaborators
Carnegie Mellon University
We will never really understand learning until we build machines that
• learn many different things,
• from years of diverse experience,
• in a staged, curricular fashion,
• and become better learners over time.
Tenet 2:
Natural language understanding requires a belief system

A natural language understanding system should react to text by saying either:
• I understand, and already knew that
• I understand, and didn’t know, but accept it
• I understand, and disagree because …
NELL: Never-Ending Language Learner

Inputs:
• initial ontology (categories and relations)
• 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:
- knowledge base with 90 million candidate beliefs
- learning to read
- learning to reason
- extending ontology
NELL Is Improving Over Time (Jan 2010 to Nov 2014)

number of NELL beliefs vs. time

reading accuracy vs. time (average over 31 predicates)

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


Recently-Learned Facts

<table>
<thead>
<tr>
<th>instance</th>
<th>iteration</th>
<th>date learned</th>
</tr>
</thead>
<tbody>
<tr>
<td>jerry_patton is a chef</td>
<td>881</td>
<td>24-oct-2014</td>
</tr>
<tr>
<td>flavones is a video game</td>
<td>883</td>
<td>02-nov-2014</td>
</tr>
<tr>
<td>smith_s_rose_bellied_lizard is a reptile</td>
<td>883</td>
<td>02-nov-2014</td>
</tr>
<tr>
<td>gray_flycatcher is a bird</td>
<td>883</td>
<td>02-nov-2014</td>
</tr>
<tr>
<td>basalt_plains can be a part of a landscape</td>
<td>883</td>
<td>02-nov-2014</td>
</tr>
<tr>
<td>louisiana is a state or province located in the geopolitical location u_s</td>
<td>886</td>
<td>21-nov-2014</td>
</tr>
<tr>
<td>kansas_city_chiefs is a sports team that won the super_bowl</td>
<td>886</td>
<td>21-nov-2014</td>
</tr>
<tr>
<td>miller has been charged with contempt</td>
<td>886</td>
<td>21-nov-2014</td>
</tr>
<tr>
<td>the companies new_york and london_sunday_times compete with each other</td>
<td>884</td>
<td>08-nov-2014</td>
</tr>
<tr>
<td>aberdeen is a city located in the geopolitical location the_united_kingdom</td>
<td>886</td>
<td>21-nov-2014</td>
</tr>
</tbody>
</table>
Portuguese NELL

[Estevam Hruschka, 2014]

conflitomilitar
(category)

See learned instances of conflitomilitar as a list or on a map

Metadata

- allLearnedPatterns
  - "a armada durante _" "a causa diplomática _" "a ci on _" "a armamentista durante _" "a declaração de capitulão _" "a disputa tecnológica _" "a fronteira interalem_" "a iminência _" "a guerra _o francesa durante _" "a independência _" "a P.Y.S.B.E. na _" "a P.Y.S.B.E. _" "a ponte resquícios _" "promover _" "acabaram a produção _" "acusa _o crime _" "agudiza _o no _" "antecederam os conflitos da _" "antigas _" "arquimimigos na _" "As decadas do cabaré _" "As origens do conflito A _" "as razões teóricas para _" "iraquianas durante _" "aviões de luta e _" "bacilos durante _" "batalha da propaganda durante _" "bimotor na _" "catrefações _" "cidades do leste durante _" "combates de avião _" "combate _" "conflito militar apelidado de _" "conflito militar assim chamado de _" "conflito militar tal como _" "conflitos militares assim como _"
How does NELL work?
Semi-Supervised Bootstrap Learning

Learn which noun phrases are 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: \( \sum_{<np,person> \in \text{labeled data}} |f_1(np) - person| \)
Type 1 Coupling: Co-Training, Multi-View Learning

Coupled training of 2 functions:

Minimize: \( \sum_{\langle np, person \rangle \in \text{labeled data}} |f_1(np) - person| + \sum_{\langle np, person \rangle \in \text{labeled data}} |f_2(np) - person| + \sum_{np \in \text{unlabeled data}} |f_1(np) - f_2(np)| \)

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

[Blum & Mitchell; 98]
[Dasgupta et al; 01 ]
[Ganchev et al., 08]
[Sridharan & Kakade, 08]
[Wang & Zhou, ICML10]
## NELL: Learned reading strategies

### Mountain:
- "volcanic crater of _"  
- "volcanic eruptions like _"  
- "volcanic region of _"  
- "volcano , called _"  
- "volcano known as _"  
- "volcano Mt _"  
- "volcanoes , like _"  
- "volcanoes including _"  
- "volcano is called _"  
- "volcano known as _"  
- "volcano Mt _"  
- "volcano named _"  
- "volcanoes , including _"  
- "volcanoes such as _"  
- "We 've climbed _"  
- "weather atop _"  
- "weather station atop _"  
- "week hiking in _"  
- "weekend trip through _"  
- "West face of _"  
- "West ridge of _"  
- "white ledge in _"  
- "white summit of _"  
- "wilderness areas and _"  
- "winter ascents in _"  
- "winter ascent through _"  
- "world famous view of _"  
- "popping by _"  
- "you 've just climbed _"  
- " ' crater"  
- "' eruption"  
- "' foothills of _"  
- "'s drug guide"  
- "'s east ridge"  
- "'s Face"  
- "'s North Peak"  
- "'s North Ridge"  
- "'s southeast ridge"  
- "'s summit caldera"  
- "'s west ridge"  
- "(D,DDD ft)"  
- "consult el diablo"  
- "cooking planks"

### 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>1.282</td>
</tr>
<tr>
<td>visualArtMovement</td>
<td>PREFIX=journal</td>
<td>0.234</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;a href='d_author.aspx?a=[X]'&gt;&lt;li&gt;&lt;[X]&gt;&lt;/li&gt;&lt;/option&gt;&lt;option&gt;&lt;[X]&gt;&lt;/option&gt;&lt;/li&gt;&lt;[X] by [Y] &amp;#8211;&lt;/a&gt;
```
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]
Multi-view, Multi-Task Coupling

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

[Blum & Mitchell; 98]
[Dasgupta et al; 01]
[Ganchev et al., 08]
[Sridharan & Kakade, 08]
[Wang & Zhou, ICML10]
[Taskar et al., 2009]
[Carlson et al., 2009]
Type 3 Coupling: Relation Argument Types

- playsSport(a,s)
- playsForTeam(a,t)
- teamPlaysSport(t,s)
- coachesTeam(c,t)

NP1

NP2
Type 3 Coupling: Relation Argument Types

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

over 2500 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: 25M NPs, $10^{14}$ 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:
• Re-estimate the knowledge base:
  – but 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
Initial NELL Architecture

Continually Learning Reading Components

- Knowledge Base (latent variables)
  - Beliefs
  - Candidate Beliefs

- Text Context patterns (CPL)
- HTML-URL context patterns (SEAL)
- Morphology classifier (CML)
- Human advice

Knowledge Integrator
If coupled learning is the key, how can we get new coupling constraints?
Key Idea 2:

Discover New Coupling Constraints

- learn horn clause rules/constraints:

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

- learned by data mining the knowledge base
- connect previously uncoupled relation predicates
- infer new unread beliefs
- modified version of FOIL [Quinlan]
Learned Probabilistic Horn Clause Rules

0.93 \text{playsSport}(\text{?x,}\text{?y}) \leftarrow \text{playsForTeam}(\text{?x,}\text{?z}), \text{teamPlaysSport}(\text{?z,}\text{?y})
If: \( x_1 \) competes with \((x_1, x_2)\) and \( x_2 \) economic sector \((x_2, x_3)\)

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

Then: economic sector $(x_1, x_3)$

PRA:
1. restrict precondition to a chain.
2. inference by random walks

PRA: [Lao, Mitchell, Cohen, EMNLP 2011]
Inference by KB Random Walks

[KB, Lao, Mitchell, Cohen, EMNLP 2011]

KB:

Random walk path type:

Pr( R(x,y) ):

logistic function for R(x,y)

where \( i^{th} \) feature = probability of arriving at node y starting at node x, and taking a random walk along path of type i
CityLocatedInCountry(Pittsburgh) = ?

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

Feature Value

Logistic Regression Weight

0.32

Lao, Mitchell, Cohen, EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

[La, Mitchell, Cohen, *EMNLP 2011*]

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

**Feature Value**

**Logistic Regression Weight**
0.32
CityLocatedInCountry(Pittsburgh) = ?

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

[Feature Value]
Logistic Regression Weight 0.32

[Lao, Mitchell, Cohen, EMNLP 2011]
Feature = Typed Path
CityInState, CityInState⁻¹, CityLocatedInCountry

Feature Value

Logistic Regression
Weight
0.32

CityLocatedInCountry(Pittsburgh) = ?

[La, Mitchell, Cohen, EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

Pr(U.S. | Pittsburgh, TypedPath)

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

Feature Value
0.8

Weight
0.32

[Lao, Mitchell, Cohen, 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
- Weight: 0.20

[Source]

[Lao, Mitchell, Cohen, EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

[Lao, Mitchell, Cohen, EMNLP 2011]

Feature = Typed Path
CityInState, CityInState⁻¹, CityLocatedInCountry
AtLocation⁻¹, AtLocation, CityLocatedInCountry

<table>
<thead>
<tr>
<th>Feature Value</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.32</td>
</tr>
<tr>
<td>0.20</td>
<td></td>
</tr>
</tbody>
</table>
CityLocatedInCountry(Pittsburgh) = ?

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

<table>
<thead>
<tr>
<th>Feature Value</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.32</td>
</tr>
<tr>
<td>0.20</td>
<td></td>
</tr>
</tbody>
</table>

[Logistic Regression]

[Lao, Mitchell, Cohen, EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

[Lao, Mitchell, Cohen, EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

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

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

Logistic Regression
Weight
0.8 0.32
0.6 0.20
...
...

[Oto, Mitchell, Cohen, EMNLP 2011]
CityLocatedInCountry(Pittsburgh) = ?

[Lao, Mitchell, Cohen, EMNLP 2011]

1. Tractable (bounded length)
2. Anytime
3. Accuracy increases as KB grows
4. combines probabilities from different horn clauses

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

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

Weight
0.8 0.32
0.6 0.20
... ...
Random walk inference: learned rules

CityLocatedInCountry(city, country):

8.04 cityliesonriver, cityliesonriver\(^{-1}\), citylocatedincounty
5.42 hasofficeincity\(^{-1}\), hasofficeincity, citylocatedincounty
4.98 cityalsoknownas, cityalsoknownas, 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 learned rules for CityLocatedInCountry
Opportunity:

Can infer more if we start with more densely connected knowledge graph

→ as NELL learns, it will become more dense

→ augment knowledge graph with a second graph of corpus statistics:

<subject, verb, object> triples

[Gardner et al, 2014]
NELL: concepts and “noun phrases”

- **c:penguins** can refer to “Penguins” “Pens”
- **c:monongahela** can refer to “Monongahela” “Mon river”
- **c:river flows through** c:monongahela through Monongahela
- **c:hometown** c:pittsburgh hometown Pittsburgh
- [Gardner et al, 2014]
NELL: concepts and "noun phrases"

SVO triples from 500 M dependency parsed web pages (thank you Chris Re!)
**NELL: concepts and “noun phrases”**

- Circumvents NELL’s fixed vocabulary of relations!
- Sadly, adding these does not help: too sparse
- But clustering verb phrases based on latent embedding (NNMF), produces significant improvement
  - \{“lies on”, “runs through”, “flows through”, …\}
- Precision/recall over 15 NELL relations:
  - KB only: 0.80 / 0.33
  - KB + SVO_{latent}: 0.87 / 0.42

[SVO triples from 500 M dependency parsed web pages (thank you Chris Re!)](#)
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,
  • cluster pairs of known instances, in terms of text contexts that connect them
## Example Discovered Relations

[Mohamed et al. *EMNLP 2011*]

<table>
<thead>
<tr>
<th>Category Pair</th>
<th>Frequent Instance Pairs</th>
<th>Text Contexts</th>
<th>Suggested Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>MusicInstrument</td>
<td>sitar, George Harrison&lt;br&gt;tenor sax, Stan Getz&lt;br&gt;trombone, Tommy Dorsey&lt;br&gt;vibes, Lionel Hampton</td>
<td>ARG1 master ARG2&lt;br&gt;ARG1 virtuoso ARG2&lt;br&gt;ARG1 legend ARG2&lt;br&gt;ARG2 plays ARG1</td>
<td>Master</td>
</tr>
<tr>
<td>Musician</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease</td>
<td>pinched nerve, herniated disk&lt;br&gt;tennis elbow, tendonitis&lt;br&gt;blepharospasm, dystonia</td>
<td>ARG1 is due to ARG2&lt;br&gt;ARG1 is caused by ARG2</td>
<td>IsDueTo</td>
</tr>
<tr>
<td>Disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CellType Chemical</td>
<td>epithelial cells, surfactant neurons, serotonin&lt;br&gt;mast cells, histamine</td>
<td>ARG1 that release ARG2&lt;br&gt;ARG2 releasing ARG1</td>
<td>ThatRelease</td>
</tr>
<tr>
<td>Plant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mammals</td>
<td>koala bears, eucalyptus sheep, grasses&lt;br&gt;goats, saplings</td>
<td>ARG1 eat ARG2&lt;br&gt;ARG2 eating ARG1</td>
<td>Eat</td>
</tr>
<tr>
<td>Plant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>River</td>
<td>Seine, Paris&lt;br&gt;Nile, Cairo&lt;br&gt;Tiber river, Rome</td>
<td>ARG1 in heart of ARG2&lt;br&gt;ARG1 which flows through ARG2</td>
<td>InHeartOf</td>
</tr>
<tr>
<td>City</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
    SubsetOf\(_C\): C \rightarrow C

*no new software here, just add this relation to ontology*
NELL: subcategories discovered by reading

Animal:

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

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 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"
NELL: subcategories discovered by reading

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

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"

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 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 arg2"
NELL Architecture

Knowledge Base (latent variables)

Beliefs

Candidate Beliefs

Evidence Integrator

Text Context patterns (CPL)

Orthographic classifier (CML)

URL specific HTML patterns (SEAL)

Actively search for web text (OpenEval)

Infer new beliefs from old (PRA)

Image classifier (NEIL)

Human advice

Ontology extender (OntExt)
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. Vision: connect NELL and NEIL
8. Learn to microread single sentences
9. Learn to assign temporal scope to beliefs
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
Consistency
Correctness
Self reflection
The core problem:
• Agents can measure internal *consistency*, but not *correctness*

Challenge:
• Under what conditions does *consistency* → *correctness*?
The core problem:

• Agents can measure internal \textit{consistency}, but not \textit{correctness}

Challenge:

• Under what conditions does \textit{consistency} \rightarrow \textit{correctness}?
• Can an autonomous agent determine its accuracy from observed consistency?
Problem setting:

• have $N$ different estimates $f_1, \ldots, f_N$ of target function $f^*$

$$f_i : X \rightarrow Y; \quad Y \in \{0, 1\}$$

• agreement between $f_i, f_j$:

$$a_{ij} \equiv P_x(f_i(x) = f_j(x))$$
Problem setting:

- have N different estimates $f_1, \ldots, f_N$ of target function $f^*$
  
  $f_i : X \rightarrow Y; \ Y \in \{0, 1\}$

- agreement between $f_i, f_j$: $a_{ij} \equiv P_x(f_i(x) = f_j(x))$

Key insight: errors and agreement rates are related

$$a_{ij} = \Pr[\text{neither makes error}] + \Pr[\text{both make error}]$$

$$a_{ij} = 1 - e_i - e_j + 2e_{ij}$$

- prob. $f_i$ and $f_i$ agree
- prob. $f_i$ error
- prob. $f_j$ error
- prob. $f_i$ and $f_j$ both make error
Estimating Error from Unlabeled Data

1. IF $f_1, f_2, f_3$ make indep. errors, and accuracies > 0.5
   THEN
   \[ a_{ij} = 1 - e_i - e_j + 2e_i e_j \]
   \[ a_{ij} = 1 - e_i - e_j + 2e_i e_j \]

Measure errors from unlabeled data:
- use unlabeled data to estimate $a_{12}, a_{13}, a_{23}$
- solve three equations for three unknowns $e_1, e_2, e_3$
Estimating Error from Unlabeled Data

1. IF $f_1, f_2, f_3$ make indep. errors, accuracies $> 0.5$
   THEN $a_{ij} = 1 - e_i - e_j + 2e_{ij}$
   $\Rightarrow a_{ij} = 1 - e_i - e_j + 2e_i e_j$

2. but if errors not independent
Estimating Error from Unlabeled Data

1. **IF** \( f_1, f_2, f_3 \) make indep. errors, accuracies > 0.5
   **THEN** \( a_{ij} = 1 - e_i - e_j + 2e_{ij} \)
   \[ \rightarrow a_{ij} = 1 - e_i - e_j + 2e_i e_j \]

2. but if errors not independent

   \[
   \min \ (e_{ij} - e_i e_j)^2 \\
   \text{such that} \\
   (\forall i, j) \ a_{ij} = 1 - e_i - e_j + 2e_{ij}
   \]
True error (red), estimated error (blue)

NELL classifiers:

[Platanios, Blum, Mitchell, *UAI 2014*]
True error (red), estimated error (blue)

NELL classifiers:

Brain image fMRI classifiers:
Summary

1. Use coupled training for semi-supervised learning
2. Datamine the KB to learn probabilistic inference rules
3. Automatically extend ontology
4. Use staged learning curriculum

New directions:
• Self-reflection, self-estimates of accuracy (A. Platanios)
• Incorporate vision with NEIL (Abhinav Gupta)
• Microreading (Jayant Krishnamurthy, Ndapa Nakashole)
• Aggressive ontology expansion (Derry Wijaya)
• Portuguese NELL (Estevam Hrushka)
• never-ending learning phones? robots? traffic lights?
thank you

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

follow NELL on Twitter:  @CMUNELL
browse/download NELL’s KB at http://rtw.ml.cmu.edu