Cognitive and application-driven ML for natural language

Christos Christodouloupoloulos
OSU - April 5, 2016
Natural Language (Processing) tasks

- Translation
- Sentiment Analysis
- Question Answering
- Following Instructions
- Human Interaction

Credit: Josh Lee (@wtrslid)
Natural Language (Processing) tasks

- Machine Translation
- Sentiment Analysis
- Question Answering
- Following Instructions
- Human Interaction

ASR/OCR
- Morphological Analysis
- PoS Tagging
- Syntactic Parsing
- Semantic Parsing
- Reasoning
- Generation

Credit: Josh Lee (@wtrsld)
Natural Language (Processing) tasks

- Machine Translation
- Sentiment Analysis
- Question Answering
- Following Instructions
- Human Interaction

Not really a pipeline!
Natural Language (Processing) tasks

Machine Translation
Sentiment Analysis
Question Answering
Following Instructions
Human Interaction

Arguments for less supervision

- Expensive (time/money)
- Not always reliable or “correct”
- Not how humans learn
Natural Language (Processing) tasks

Two questions

- Can we predict multiple NL phenomena jointly and efficiently?
- Can we do it an a cognitively inspired unsupervised way?

Arguments for less supervision

- Expensive (time/money)
- Not always reliable or “correct”
- Not how humans learn
Outline

- Application-Driven ML: **Extended SRL**
  - Modular semantic structures
  - Inference over independently-trained models

- Cognitive ML: **Baby SRL**
  - Modeling early stages of language acquisition
  - Testbed for psycholinguistic theories and unsupervised ML

- Unifying the two: **Unsupervised CCG induction**
  - Structured lexical-syntactic-semantic representation

- Future directions
Extended SRL (Joint work with Dan Roth)

- Semantic role labeling covers only specific predicate-argument relations
- Need for comprehensive semantic representations (SRL++)
  - Events, Entailment, Winograd schemas
- Difficult to produce hand-annotated resources (e.g. AMR, λ-calculus)
  - Especially for other languages/genres†
- Lots of independently annotated data for semantic tasks
  - Ontonotes, PDTB, Semeval/CoNLL/*SEM shared tasks

†AMR annotation of The Little Prince
http://amr.isi.edu/download/amr-bank-v1.6.txt
Extended SRL (Joint work with Dan Roth)

- Semantic role labeling covers only specific predicate-argument relations
- Need for comprehensive semantic representations (SRL++)
  - Events, entailment, Winograd schemas
- Difficult to produce hand-annotated resources (e.g. AMR, λ-calculus)
  - Especially for other languages/genres
- Lots of independently annotated data for semantic tasks
  - Ontonotes, PDTB, Semeval/CoNLL/*SEM shared tasks

It was a picture of a boa constrictor in the act of swallowing an animal.

(p / picture
 :domain (i / it)
 :topic (b2 / boa
 :mod (c2 / constrictor)
 :ARG0-of (s / swallow-01
 :ARG1 (a / animal))))

†AMR annotation of The Little Prince
http://amr.isi.edu/download/amr-bank-v1.6.txt
Extended SRL (Joint work with Dan Roth)

- Semantic role labeling covers only specific predicate-argument relations
- Need for comprehensive semantic representations (SRL++)
  - Events, Entailment, Winograd schemas
- Difficult to produce hand-annotated resources (e.g. AMR, \(\lambda\)-calculus)
  - Especially for other languages/genres
- \textbf{Lots} of independently annotated data for semantic tasks
  - Ontonotes, PDTB, Semeval/CoNLL/*SEM shared tasks

\[\dagger\] AMR annotation of The Little Prince

\[\text{http://amr.isi.edu/download/amr-bank-v1.6.txt}\]
Extended SRL

Modular semantic representation  Inference with independent models
Extended SRL

Modular semantic representation  Inference with independent models

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Five</th>
<th>prisoners</th>
<th>are</th>
<th>held</th>
<th>captive</th>
<th>in</th>
<th>Guantanamo</th>
</tr>
</thead>
</table>
Extended SRL

Modular semantic representation

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Five</th>
<th>prisoners</th>
<th>are</th>
<th>held</th>
<th>captive</th>
<th>in</th>
<th>Guantanamo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantities</td>
<td>[#5] [unit: prisoner]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Extended SRL

Modular semantic representation

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Five</th>
<th>prisoners</th>
<th>are</th>
<th>held</th>
<th>captive</th>
<th>in</th>
<th>Guantanamo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantities</td>
<td>[#5] [unit: prisoner]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NER</td>
<td>Card.</td>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Inference with independent models
Extended SRL

Modular semantic representation

Inference with independent models

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Five</th>
<th>prisoners</th>
<th>are</th>
<th>held</th>
<th>captive</th>
<th>in</th>
<th>Guantanamo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantities</td>
<td>[#5] [unit: prisoner]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NER</td>
<td>Card.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Location</td>
</tr>
<tr>
<td>VerbSRL</td>
<td>Arg1</td>
<td>hold</td>
<td></td>
<td>Arg2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NomSRL</td>
<td>Arg1</td>
<td>captive</td>
<td></td>
<td>ArgM-LOC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrepSRL</td>
<td>Gov</td>
<td></td>
<td>Loc</td>
<td>Obj</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Extended SRL

Modular semantic representation

| Tokens | Five | prisoners | are | held | captive | in | Guantanamo |
| Quantities | [#5] [unit: prisoner] |
| NER | Card. | Location |
| VerbSRL | Arg1 | hold | Arg2 |
| NomSRL | Arg1 | captive | ArgM-LOC |
| PrepSRL | Gov | Loc | Obj |
| LVC | Light Verb |
| FrameNet | Being_in_captivity |
| Wiki | Guantanamo_Bay detention camp |
**Extended SRL**

**Modular semantic representation**

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Five</th>
<th>prisoners</th>
<th>are</th>
<th>held</th>
<th>captive</th>
<th>in</th>
<th>Guantanamo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantities</td>
<td>[#5] [unit: prisoner]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NER</td>
<td>Card.</td>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VerbSRL</td>
<td>Arg1</td>
<td>Guantanamo_Bay_ detention_camp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NomSRL</td>
<td>ArgM-LOC</td>
<td>captive</td>
<td>ArgM-LOC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrepSRL</td>
<td>Gov</td>
<td>Loc</td>
<td>Obj</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVC</td>
<td>Light Verb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FrameNet</td>
<td>Being_in_ captivity</td>
<td>Guantanamo_Bay_ detention_camp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*How can we join predictions across these phenomena?*
### Extended SRL

**Modular semantic representation**

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Five</th>
<th>prisoners</th>
<th>are</th>
<th>held</th>
<th>captive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantities</td>
<td>[#5] [unit: prisoner]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NER</td>
<td>Card.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VerbSRL</td>
<td>Arg1</td>
<td>hold</td>
<td>Arg2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NomSRL</td>
<td>Arg1</td>
<td>captive</td>
<td>ArgM-LOC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FrameNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wiki</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Inference with independent models**

\[ \text{NomSRL}(\text{arg})::\text{ArgM-LOC} \rightarrow \neg \text{PrepSRL}(\text{prep})::\text{Instrument} \]

- **NER** (phrase)::Cardinal → Quant(phrase)::Number
- **Light Verb**
- **Location**
- **Gov**
- **Loc**
- **Obj**
- **Guantanamo_Bay detention camp**
Constraint-driven inference

● Constrained Conditional Models [Chang et al., 2012]
  ○ First-order constraints
  ○ ILP inference

● Multi-view model combination [Burkett et al., 2010]
Constraint-driven inference

- Constrained Conditional Models [Chang et al., 2012]
  - First-order constraints
  - ILP inference
- Multi-view model combination [Burkett et al., 2010]
Constraint-driven inference

- Constrained Conditional Models
  - First-order constraints
  - ILP inference
- Multi-view model combination

Algorithm 1: Weight update algorithm

1. for $m_i \in m$ do
2.   $\eta_i \leftarrow 0.1$;
3.   for $(x_j, y_j) \in D_i$ do
4.     if $\hat{y}_{m_i}(x_j) \neq y_j \land \sum_{k=1}^{m} C_k(x, y^*) > 0$
5.       then
6.         $\lambda_i \leftarrow \lambda_i + \eta_i / \sqrt{j}$
7.     end
8.   end
9. end
Extended SRL: Systems

- Clauses
- *Comma SRL
- *Coreference resolution
- FrameNet
- Light-verb constructions
- Metonymy
- Multi-word expressions
- NER
- *Nominal SRL
- Phrasal-verb constructions
- PP attachment

- *Preposition SRL
- Quantities
- Sentiment analysis (aspect-based)
- Sentence specificity
- Temporal extraction
- *Verb SRL
- VP ellipsis
- *Wikification

* denotes previously implemented system
Detour: CogComp software

- **LBJava** [Rizzolo & Roth, 2010]
  - Rapid development of ML software

- **EDISON + TextAnnotation** [Sammons et al., 2016]
  - Data structures and feature extraction
Detour: CogComp software

- **LBJava** [Rizzolo & Roth, 2010]
  - Rapid development of ML software

- **EDISON + TextAnnotation** [Sammons et al., 2016]
  - Data structures and feature extraction

**NER system definition:**

```java
import edu.illinois.cs.cogcomp.esrl.core.features.*;
import edu.illinois.cs.cogcomp.core.datastructures.textannotation.Constituent;

discrete NERLabel(Constituent phrase) <- { return phrase.getLabel(); };

discrete NERClassifier(Constituent phrase) <-
  learn NERLabel
  using Capitalization, WordBigrams, POSBigrams, WordContextBigrams,
  POSContextBigrams, ChunkContextBigrams

with SparseNetworkLearner {
  SparseAveragedPerceptron.Parameters p =
    new SparseAveragedPerceptron.Parameters();
  p.learningRate = .1;
  p.thickness = 2;
  baseLTU = new SparseAveragedPerceptron(p);
}
```

- **LBJava** [Rizzolo & Roth, 2010]
  - Rapid development of ML software

- **EDISON + TextAnnotation** [Sammons et al., 2016]
  - Data structures and feature extraction
Detour: CogComp software

- **LBJava** [Rizzolo & Roth, 2010]
  - Rapid development of ML software

- **EDISON + TextAnnotation** [Sammons et al., 2016]
  - Data structures and feature extraction

**Output label**

**Classifier definition**

**NER system definition:**

```java
import edu.illinois.cs.cogcomp.esrl.core.features.*;
import edu.illinois.cs.cogcomp.core.datastructures.textannotation.Constituent;

discrete NERLabel(Constituent phrase) <- { return phrase.getLabel(); }

discrete NERClassifier(Constituent phrase) <-
    learn NERLabel using Capitalization, WordBigrams, POSBigrams, WordContextBigrams,
    POSContextBigrams, ChunkContextBigrams
    with SparseNetworkLearner {
        SparseAveragedPerceptron.Parameters p =
            new SparseAveragedPerceptron.Parameters();
        p.learningRate = .1;
        p.thickness = 2;
        baseLTU = new SparseAveragedPerceptron(p);
    }
```

- LBJava
- EDISON + TextAnnotation
Extended SRL: Constraints

- \([\text{LVC}(\text{phrase})::\text{true} \rightarrow \neg\text{PVC}(\text{phrase})::\text{true}] \land \text{[PVC}(\text{phrase})::\text{true} \rightarrow \text{phrase contains } \{\text{IN}|\text{PRT}|\text{RB}\}]\)
  - LVC accuracy: 81.2 → 82.2
  - Same PVC accuracy (89.9)
  - Candidate selection prevented constraint violation
Extended SRL: Constraints

- \([LVC(\text{phrase})::\text{true} \rightarrow \neg PVC(\text{phrase})::\text{true}] \land [PVC(\text{phrase})::\text{true} \rightarrow \text{phrase contains \{IN|PRT|RB\}}]\)
  - LVC accuracy: 81.2 → 82.2
  - Same PVC accuracy (89.9)
  - Candidate selection prevented constraint violation

- Quantities(phrase)::Date → Temporal(phrase)::Date
  - Quantities overall F1: 78.7 → 78.9 (Date acc: 79.2 → 85.5)
  - Temp overall F1: 75.7 → 75.9
  - Double implication doesn’t work
Extended SRL: Constraints

- \( \text{NER(phrase)}::\text{Cardinal} \rightarrow \text{Quant(phrase)}::\text{Number} \)
- \( \text{NER(phrase)}::\text{Date} \rightarrow \text{Temp(phrase)}::\text{Date} \)
- \( \text{NER(phrase)}::\text{Date} \rightarrow \text{Temp(phrase)}::\text{Date} \land \neg \text{Quant(phrase)}::\text{Number} \)
- \( \text{VerbSRL(arg)}::\text{ArgM-LOC} \rightarrow \neg \text{PrepSRL(prep)}::\text{Instrument} \)
- \( \text{Metonymy}(x)::\text{true} \land \text{NER}(x)::y \rightarrow \exists z \ \forall x' \ \text{Met}(x')::\text{false} \land \text{NER}(x')::z \land z \neq y \)
Extended SRL: Summary

- Combination of multiple phenomena
  - No need for joint annotations

- Joint inference via first-order constraints
  - Offer linguistic insights

- Flexible interface
  - Only requirement is list of $k$-best predictions
Extended SRL: Summary

- Combination of multiple phenomena
  - No need for joint annotations
- Joint inference via first-order constraints
  - Offer linguistic insights
- Flexible interface
  - Only requirement is list of $k$-best predictions

But can we do it with less (direct) supervision?

- Other languages
- “Better-than-gold” performance [e.g. Spitkovsky et al., 2011]

One solution: Look into human language acquisition

- Cognitive insights for ML models
- Testbed for cognitive theories
A (simple) model of language acquisition

“The girl *krads* the boy”

“The boy *krads*”

*krad* = RUN ??
*krad* = CHASE ??
Baby SRL (Joint work with Cindy Fisher and Dan Roth)

- **Syntactic bootstrapping**
  - Using the structure of the utterances to predict the semantic

- **An account of how syntactic bootstrapping can begin**
  - [Connor et al. (2010)]
  - Fisher et al. (2010)
  - Gutman et al. (2014)
  - van Schijndel & Elsner (2014)

- **Framed as an SRL problem**
  - Learn Agent/Patient roles for novel-verb utterances
Experiment 1: Supervised model

Given veridical feedback (“mind reading”), do low-level syntactic features capture anything useful about semantic roles/verb preferences?
Experiment 1: Supervised model

BabySRL corpus
[Connor et al. 2008]

Online Perceptron
$\alpha=0.1$, $\theta=4.0$

**Prediction**
- **Arg0**
  - \text{lex}: mary, like
  - \text{ncount}: 2
  - \text{npad}: 1of2
  - \text{vpos}: before
- **Arg1**
  - \text{lex}: flowers, like
  - \text{ncount}: 2
  - \text{npad}: 2of2
  - \text{vpos}: after

**Features**

**Feedback**
- Mary
  - \text{Arg0}
- flowers
  - \text{Arg1}
- Mary = \text{Arg0}
- flowers = \text{Arg1}
Results on novel-verb sentences

Structure-Mapping hypothesis [Fisher, 1996]
Results on novel-verb sentences

A *krads* B

A and B *krad*

Predicted error in 21mo
[Gertner & Fisher, 2006]
Experiment 1: Supervised model

Given **veridical feedback** ("mind reading"), do **low-level** syntactic features capture anything useful about semantic roles/verb preferences?

YES, but verb knowledge is crucial
Experiment 2: Removing bottom-up supervision

Can we predict nouns/verbs using distributional clusters and a few seed nouns?
Experiment 2: Predicting nouns

- HMM (80 states) + Variational Inference
  - Trained on 2.2M tokens of CDS
  - List of function words to separate clusters

- List of seed nouns
  - MacArthur-Bates CDI production norms [Dale & Fenson, 1996]
  - 75 nouns+pronouns (cutoff 50% at 21-mo)

Noun identification:
state $x$ contains $> k$ seed nouns

where $k = 4$ [or dynamic]
Step 1: Argument histograms
for each sentence:
  count the number of nouns
  collect histograms for each non-noun state

38: have, like, do, has, did, ought, remembers,…
…
47: little, other, new, good, big, next,…, yellow, nice, merry,…
…
59: baby, horsie, dog, bath,…, flowers, song, secret, cat,…
60: now, here, today, too, again, later, yesterday, first,…
61: it, you, me, her, nana, them,…, daddy, Mary,…
Step 2: Aggregate verb predictions
store histogram-based predictions

for each sentence:
   pick the HMM state most freq. appearing as verb

38: have, like, do, has, did, ought, remembers,…
…
47: little, other, new, good, big, next,…, yellow, nice, merry,…
…
59: baby, horsie, dog, bath,…, flowers, song, secret, cat,…
60: now, here, today, too, again, later, yesterday, first,…
61: it, you, me, her, nana, them,…, daddy, Mary,…
Results of verb/noun heuristics

- Noun discovery is very accurate (84.3%)
  - ~30 nouns needed
- Verb accuracy is 6.6% > guessing (75.5%)
  - Using only argument counting
- Verb recall is low (~36%)
  - Multiple-verb sentences
Experiment 2: SRL predictions

Mary
Arg0

flowers
Arg1

Mary = Arg0
flowers = Arg1

Features:
- **lex:** mary, like
- **ncount:** 2
- **npat:** 1
- **vpos:** before

- **lex:** flowers, like
- **ncount:** 2
- **npat:** 2
- **vpos:** after

N   V   A   N

Mary     likes      yellow    flowers
Experiment 2: SRL predictions

FEEDBACK

Mary

Arg0

flowers

Arg1

PREDICTION

Arg0

lex: mary, like
ncount: 2
npat: 1of2
vpos: before
...

Mary = Arg0

flowers = Arg1

Arg1

lex: flowers, like
ncount: 2
npat: 2of2
vpos: after
...

FEATURES

lex: mary, like
ncount: 2
npat: 1of2
vpos: before
...

38: have, like, do, has, did, ought, remembers, ...
...
47: little, other, new, good, big, next,..., yellow, nice, merry, ...

59: baby, horses, dog, bath,..., flowers, song, secret, cat,...
60: now, here, today, too, again, later, yesterday, first,...
61: at you, me, her, nana, them,..., daddy, Mary,...

Mary

likes

yellow

flowers

likes (HMM: 38)

yellow (HMM: 47)
Results on novel-verb sentences (transitive)

lex < npat < vpos
vpos-rand << vpos (~30%)
lex better than gold-PoS
Experiment 3: Reducing top-down supervision

```
Mary
 Arg0

flowers
 Arg1

Mary = Arg0
flowers = Arg1
```

**Features**
- **lex:** mary, like
- **ncount:** 2
- **npat:** 1of2
- **vpos:** before

**Prediction**
- **lex:** flowers, like
- **ncount:** 2
- **npat:** 2of2
- **vpos:** after

**Feedback**
- **Arg0**
  - lex: mary, like
  - npat: 1of2
  - vpos: before: 2.124
  - vpos: after: 1.271
  - ncount: 1 = 0.742

- **Arg1**
  - lex: flowers, like
  - npat: 2of2
  - vpos: after: 1.271
  - ncount: 1 = 0.742

```
38: have, like, do, has, did, ought, remembers,...
... little, other, new, good, big, next,..., yellow, nice, merry,...

59: baby, horsie, dog, bath,..., flowers, song, secret, cat,...
60: now, here, today, too, again, later,..., yesterday, first,...
61: it, you, me, her, nana, them,..., daddy, Mary,...

Mary likes yellow flowers
```

```
likes (HMM: 38) yellow (HMM: 47)
0 1 2 3 0 1 2 3
```
Experiment 3: Reducing top-down supervision

Mary = animate

Arg0

flowers = unknown

Arg1

Mary = Arg0

flowers = Arg1

lex: mary, like

ncount: 2

npat: 1of2

vpos: before

lex: flowers, like

ncount: 2

npat: 2of2

vpos: after

38: have, like, do, has, did, ought, remembers,...

... little, other, new, good, big, next,... yellow, nice, merry,...

59: baby, horrie dog, bath,... flowers, song, secret, cat,...

60: now, here, today, too, again, later, yesterday, first...

61: it you, me, her, nana, them,... daddy, Mary,...

Mary likes

yellow flowers
Experiment 3: Reducing top-down supervision

Input: BabySRL learner $L$
Data: CDS corpus $C$, animate nouns $N_a$, inanimate nouns $N_i$

foreach sentence $s \in C$ do

// Mark animate nouns as agents
if $\exists s_i \in N_a$ then
    SRL($s_i$) = Arg0;
    // Mark all other nouns as patients
    SRL($s_j$) = Arg1 $\forall s_j \in s, j \neq i$;
else
    // Mark inanimates as patients
    if $\exists s_i \in N_i$ then
        SRL($s_i$) = Arg1;
    else
        // Sentence has no animacy information
        skip $s$;
end

// If there are multiple animates
if $|SRL(s_i) = Arg0| > 1$ then
    // Let the learner decide
    SRL($\text{argmax}_{s_i} L(s_i = Arg0)) = Arg0$;
    // Mark all other nouns as agents
    SRL($s_j$) = Arg1 $\forall s_j \in s, j \neq i$;
end

Mary = animate
Arg0

flowers = unknown
Arg1

Mary = Arg0
flowers = Arg1

.lex: mary, like
ncount: 2
npat: 1of2
vpos: before

lex: flowers, like
ncount: 2
npat: 2of2
vpos: after

Mary
likes
yellow
flowers
Results on novel-verb sentences (transitive)
Results on novel-verb sentences (transitive)

lex < npat ~ vpos
vpos-rand < vpos (~4%)
Baby SRL summary

- Modeling early language acquisition
  - Testbed for psycholinguistic theories
  - Replication of experimental results

- Structure-mapping for syntactic bootstrapping
  - Identifying verbs from noun structure
  - Predicting semantic roles using low-level syntactic features
Baby SRL summary

- Modeling early language acquisition
  - Testbed for psycholinguistic theories
  - Replication of experimental results

- Structure-mapping for syntactic bootstrapping
  - Identifying verbs from noun structure
  - Predicting semantic roles using low-level syntactic features

Getting closer to E-SRL

- Phrases instead of words
- Generalised predicate-argument structures
- Direct access to semantics
Bridging the gap: Inducing CCG structures
(Joint work with Yonatan Bisk and Julia Hockenmaier)

Combinatory Categorial Grammar [Steedman, 2000]

**CCGbank** [Hockenmaier & Steedman, 2007]
Bridging the gap: Inducing CCG structures
(Joint work with Yonatan Bisk and Julia Hockenmaier)

Can we induce a CCG from raw text?
Bisk & Hockenmaier (2013)
  Inducing CCG from raw text + PoS
Can we replace PoS with induced clusters?
  Bisk, *Christodoulopoulos*, & Hockenmaier, 2015
Unsupervised CCG induction

<table>
<thead>
<tr>
<th>Hertz</th>
<th>equipment</th>
<th>is a</th>
<th>major</th>
<th>supplier of rental equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labels</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R0</td>
<td>N/N</td>
<td>S</td>
<td>N</td>
<td>N/N</td>
</tr>
<tr>
<td>R1</td>
<td>N/N</td>
<td>S/S</td>
<td>S</td>
<td>S/S</td>
</tr>
<tr>
<td>R2</td>
<td>(S/S)/(S/S)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(S/S)/(S/S)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(S/S)/(S/S)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(S/S)/(S/S)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(S/S)/(S/S)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
<tr>
<td></td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
<td>(N/N)/(N/N)</td>
</tr>
</tbody>
</table>
Unsupervised CCG induction

Final round
Unsupervised CCG induction: Experiments

Directed Attachments on Dependency Treebanks

- This Work
- Bisk & Hockenmaier 2013 (Gold POS)
Unsupervised CCG induction: Summary

- Structured syntactic categories
  - Predicate-Argument structure directly from categories

- Transparent to semantics
  - Inverse of Kwiatkowski et al. (2012)
Overall Summary

- **Extended SRL**
  - Modular semantic structures with joint inference
  - Improved performance over independently-trained models
  - Constraint-based theories of language processing [McDonald et al., 1994; Ferreira et al., 2002]

- **Baby SRL**
  - Identifying verbs using only a few seed nouns
  - Low-level syntactic features can guide semantic learning
  - Empirical evidence for structure-mapping account of syntactic bootstrapping

- **Unsupervised CCG induction**
  - Structured lexical-syntactic-semantic representation from raw text
The road ahead

- **Extended SRL**
  - Mine for constraints
  - Entailment via latent structure alignment [Sammons et al., 2009]

- **Baby SRL**
  - Add internal structure (CCG)
  - More predicates, constructions (pro-drop, filler gap)

- **Unsupervised CCG induction**
  - Use weak semantic supervision
  - Add constraints from E-SRL
Thank you!

Dan Roth, Cindy Fisher, Mark Sammons, Vivek Srikumar, Parisa Kordjamshidi, Shyam Upadhyay, Daniel Khashabi, Stephen Mayhew, Mark Steedman, Yonatan Bisk, Julia Hockenmaier, Catriona Silvey

Source code:
https://gitlab-beta.engr.illinois.edu/cogcomp/illinois-esrl
https://gitlab-beta.engr.illinois.edu/babysrl-group/babysrl