Natural Language Understanding with Common Sense Reasoning

Dan Roth
Department of Computer Science
University of Illinois at Urbana-Champaign

July 2015
IJCAI-15 Workshop on Neural-Symbolic Learning and Reasoning
Natural Language Understanding with Common Sense Reasoning

Dan Roth

Department of Computer Science
University of Illinois at Urbana-Champaign

With thanks to:

Collaborators: Kai-Wei Chang, Ming-Wei Chang, Xiao Chen, Cindy Fisher, Daniel Khashabi, Haoruo Peng, Lev Ratinov, Subhro Roy, ...

Funding: NSF; DHS; NIH; DARPA; IARPA, ARL, ONR
DASH Optimization (Xpress-MP); Gurobi.
Please...
Please...

- Identify units
- Consider multiple interpretations and representations
  - Pictures, text, layout, spelling, phonetics
- Put it all together: Determine “best” global interpretation
- Satisfy expectations
  - Slide; puzzle
(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don’t know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.
Comprehension

(ENGLAND, June, 1989) – Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don’t know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin’s dad was a magician.
4. Christopher Robin must be at least 65 now.
Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don’t know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.  
2. Winnie the Pooh is a title of a book.  
3. Christopher Robin’s dad was a magician.  
4. Christopher Robin must be at least 65 now.
(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don’t know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.  
2. Winnie the Pooh is a title of a book.  
3. Christopher Robin’s dad was a magician.  
4. Christopher Robin must be at least 65 now.
Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don’t know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin’s dad was a magician.
4. Christopher Robin must be at least 65 now.
Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don’t know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.  
2. Winnie the Pooh is a title of a book.  
3. Christopher Robin’s dad was a magician.  
4. Christopher Robin must be at least 65 now.
Comprehension

(ENGLAND, June, 1989) – Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don’t know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.  
2. Winnie the Pooh is a title of a book.  
3. Christopher Robin’s dad was a magician.  
4. Christopher Robin must be at least 65 now.

This is an Inference Problem
How do we Acquire Language?

[Joint Research Program with Developmental Psycholinguist Cindy Fisher]

- Topid rivvo den marplox.
The Language-World Mapping Problem

“the language”

[Topid rivvo den marplox.]

“the world”
Observe how Words are Distributed Across Situations

Smur! Rivvo della frowler.  

Topid rivvo den marplox.

Blert dor marplox, arno.

Marplox dorinda blicket.
Structure-Mapping: A proposed starting point for syntactic bootstrapping

- Children can learn the meanings of some nouns via cross-situational observation alone [Fisher 1996, Gillette, Gleitman, Gleitman, & Lederer, 1999; Snedeker & Gleitman, 2005]
Children can learn the meanings of some nouns via cross-situational observation alone [Fisher 1996, Gillette, Gleitman, Gleitman, & Lederer, 1999; Snedeker & Gleitman, 2005]

But how do they learn the meaning of verbs?

- Sentences comprehension is grounded by the acquisition of an initial set of concrete nouns
- These nouns yields a skeletal sentence structure — candidate arguments; cue to its semantic predicate—argument structure.
- Represent sentence in an abstract form that permits generalization to new verbs

[Joanna rivvo den sheep.]
Structure-Mapping: A proposed starting point for syntactic bootstrapping

- Children can learn the meanings of some nouns via cross-situational observation alone [Fisher 1996, Gillette, Gleitman, Gleitman, & Lederer, 1999; Snedeker & Gleitman, 2005]

- But how do they learn the meaning of verbs?
  - Sentences comprehension is grounded by the acquisition of an initial set of concrete nouns
  - These nouns yields a skeletal sentence structure — candidate arguments; cue to its semantic predicate—argument structure.
  - Represent sentence in an abstract form that permits generalization to new verbs

  [Johanna rivvo den sheep.]
Strong Predictions  [Gertner & Fisher, 2006]

- Test *21 month olds* on assigning arguments with *novel verbs*
- How order of nouns influences interpretation: Transitive & Intransitive
Strong Predictions  [Gertner & Fisher, 2006]

- Test 21 month olds on assigning arguments with novel verbs
- How order of nouns influences interpretation: Transitive & Intransitive
Strong Predictions  [Gertner & Fisher, 2006]

- Test 21 month olds on assigning arguments with novel verbs
- How order of nouns influences interpretation: Transitive & Intransitive

Transitive: The boy is daxing the girl!

preferential looking paradigm
Strong Predictions [Gertner & Fisher, 2006]

- Test 21 month olds on assigning arguments with novel verbs
- How order of nouns influences interpretation: Transitive & Intransitive

Transitive: The boy is daxing the girl!

preferential looking paradigm
Strong Predictions  [Gertner & Fisher, 2006]

- Test 21 month olds on assigning arguments with novel verbs
- How order of nouns influences interpretation: Transitive & Intransitive

Agent-first: The boy and the girl are daxing!
Agent-last: The girl and the boy are daxing!

preferential looking paradigm
Strong Predictions  [Gertner & Fisher, 2006]

- Test 21 month olds on assigning arguments with novel verbs
- How order of nouns influences interpretation: Transitive & Intransitive

Agent-first: The boy and the girl are daxing!
Agent-last: The girl and the boy are daxing!

preferential looking paradigm
Strong Predictions [Gertner & Fisher, 2006]

- Test 21 month olds on assigning arguments with novel verbs
- How order of nouns influences interpretation: Transitive & Intransitive

Agent-first: The boy and the girl are daxing!
Agent-last: The girl and the boy are daxing!
Strong Predictions [Gertner & Fisher, 2006]

- Test 21 month olds on assigning arguments with novel verbs
- How order of nouns influences interpretation: Transitive & Intransitive

Agent-first: The boy and the girl are daxing!
Agent-last: The girl and the boy are daxing!

Error disappears by 25 months
preferential looking paradigm
Current Project: BabySRL

- **Realistic Computational model** for Syntactic Bootstrapping via Structure Mapping:
  - Verbs meanings are learned via their syntactic argument-taking roles
  - Semantic feedback to improve syntactic & meaning representation
Current Project: BabySRL

- **Realistic Computational model** for Syntactic Bootstrapping via Structure Mapping:
  - Verbs meanings are learned via their syntactic argument-taking roles
  - Semantic feedback to improve **syntactic & meaning representation**

- Develop Semantic Role Labeling System (BabySRL) to experiment with theories of early language acquisition
  - SRL as minimal level language understanding
  - Determine who does what to whom.

- Inputs and knowledge sources
  - Only those we can defend children have access to
BabySRL: Key Components

[Connor et. al.’13: Starting from Scratch in Semantic Role Labeling: Early Indirect Supervision]

- **Representation:**
  - Theoretically motivated representation of the input
  - Shallow, abstract, sentence representation consisting of
    - # of nouns in the sentence
    - Noun Patterns (1st of two nouns)
    - Relative position of nouns and predicates

- **Learning:**
  - Guided by knowledge kids have
    - **Classify words by** part-of-speech
    - **Identify arguments and predicates**
    - **Determine the role** arguments take
  - Minimal Supervision that is Defensible from psycholinguistic evidence
BabySRL: Key Components
[Connor et. al.’13: Starting from Scratch in Semantic Role Labeling: Early Indirect Supervision]

- **Representation:**
  - Theoretically motivated representation of the input
  - Shallow, abstract, sentence representation consisting of
    - # of nouns in the sentence
    - Noun Patterns (1st of two nouns)
    - Relative position of nouns and predicates

- **Learning:**
  - Guided by knowledge kids have
    - Classify words by part-of-speech
    - Identify arguments and predicates
    - Determine the role arguments take
  - Minimal Supervision that is Defensible from psycholinguistic evidence
BabySRL: Key Components

[Connor et. al.'13: Starting from Scratch in Semantic Role Labeling: Early Indirect Supervision]

- Representation:
  - Theoretically motivated representation of the input
  - Shallow, abstract, sentence representation consisting of
    - # of nouns in the sentence
    - Noun Patterns (1st of two nouns)
    - Relative position of nouns and predicates

- Learning:
  - Guided by knowledge kids have
    - Classify words by part-of-speech
    - Identify arguments and predicates
    - Determine the role arguments take
  - Minimal Supervision that is Defensible from psycholinguistic evidence

Some of the representations are abstract (non-lexical). Learning is guided by abstract expectations.
BabySRL: Key Components

[Connor et. al.’13: Starting from Scratch in Semantic Role Labeling: Early Indirect Supervision]

- **Representation:**
  - Theoretically motivated representation of the input
  - Shallow, abstract, sentence representation consisting of
    - # of nouns in the sentence
    - Noun Patterns (1st of two nouns)
    - Relative position of nouns and predicates

- **Learning:**
  - Guided by knowledge kids have
    - Classify words by part-of-speech
    - Identify arguments and predicates
    - Determine the role arguments take
  - Minimal Supervision that is Defensible from psycholinguistic evidence

Some of the representations are abstract (non-lexical). Learning is guided by abstract expectations.

Among other findings, our models reproduce mistakes kids make, and recover from them (with more learning).
Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don’t know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.  
2. Winnie the Pooh is a title of a book.  
3. Christopher Robin’s dad was a magician.  
4. Christopher Robin must be at least 65 now.

This is an Inference Problem
Comprehension

- Dan is flying to Philadelphia this weekend. Penn is organizing a workshop on the Penn Discourse Treebank.
  - ➔ Dan is attending the workshop
  - ➔ The Workshop is in Philadelphia
Comprehension

- Dan is flying to Philadelphia this weekend. Penn is organizing a workshop on the Penn Discourse Treebank.
  - Dan is attending the workshop
  - The Workshop is in Philadelphia

- Interpretation builds on expectations that rely on knowledge.
Comprehension

- Dan is flying to Philadelphia this weekend. Penn is organizing a workshop on the Penn Discourse Treebank.
  - ➔ Dan is attending the workshop
  - ➔ The Workshop is in Philadelphia

- Jan is a black Dutch man.
  - ➔ Jan is a black man.

- **Interpretation** builds on **expectations** that rely on knowledge.
Comprehension

- Dan is flying to Philadelphia this weekend. Penn is organizing a workshop on the Penn Discourse Treebank.
  - Dan is attending the workshop
  - The Workshop is in Philadelphia

- Jan is a black Dutch man.
  - Jan is a black man.

- Interpretation builds on expectations that rely on knowledge.
  - Jan is a short Dutch man.
  - Jan is a short man.
Comprehension

- Dan is flying to Philadelphia this weekend. Penn is organizing a workshop on the Penn Discourse Treebank.
  - □ ➔ Dan is attending the workshop
  - □ ➔ The Workshop is in Philadelphia

- Jan is a black Dutch man.
  - □ ➔ Jan is a black man.

- Jan is a short Dutch man.
  - □ ➔ Jan is a short man.

- Interpretation builds on expectations that rely on knowledge.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

- 49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

- 49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by....

- 49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

- visitors

49 people were hit by a suicide bomber in Akuressa.
At least 14 people have been killed in southern Sri Lanka, police say. The telecoms minister was among about 35 injured in the blast site at the town of Akuressa, 160km (100 miles) south of the capital, Colombo. Government officials were attending a function at a mosque to celebrate an Islamic holiday at the time. The defense ministry said the suicide attack was carried out by ....

- 49 people were hit by a suicide bomber in Akuressa.
Natural Language Understanding

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions so they satisfy our expectations
Natural Language Understanding

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions so they satisfy our expectations
Natural Language Understanding

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions so they satisfy our expectations
Natural Language Understanding

Natural language understanding decisions are global decisions that require

- Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
- The ability to put these predictions together coherently
- Knowledge, that guides the decisions so they satisfy our expectations

Natural Language Interpretation is a Common Sense driven Inference Process that is best thought of as a knowledge constrained optimization problem, done on top of multiple statistically learned models.
Natural Language Understanding

- Natural language understanding decisions are global decisions that require
  - Making (local) predictions driven by different models trained in different ways, at different times/conditions/scenarios
  - The ability to put these predictions together coherently
  - Knowledge, that guides the decisions so they satisfy our expectations

Natural Language Interpretation is a Common Sense driven Inference Process that is best thought of as a knowledge constrained optimization problem, done on top of multiple statistically learned models.

Many forms of Inference; a lot boil down to determining best assignment
Common Sense Reasoning was formulated traditionally as a “reasoning” process, irrespective of learning and the resulting knowledge representation.
Common Sense Reasoning was formulated traditionally as a “reasoning” process, irrespective of learning and the resulting knowledge representation.
What is Needed?
What is Needed?
What is Needed?

Training on the go!
What is Needed?

Training on the go!
What is Needed?

- A computational Framework
- Three Examples:
  - Pronoun Resolution
  - Quantitative Reasoning
  - Semantic Parsing
The Neuro-Symbolic Connection
The Neuro-Symbolic Connection

- We don’t need to discuss implementation
The Neuro-Symbolic Connection

- We don’t need to discuss implementation
  - But you may think about some of the common sense requirements that come up from the discussion that follows as “desiderata”.
The Neuro-Symbolic Connection

- We don’t need to discuss implementation
  - But you may think about some of the common sense requirements that come up from the discussion that follows as “desiderata”.

- We care about plausible interaction models
The Neuro-Symbolic Connection

- We don’t need to discuss implementation
  - But you may think about some of the common sense requirements that come up from the discussion that follows as “desiderata”.

- We care about plausible interaction models
  - In terms of levels/types of supervision
The Neuro-Symbolic Connection

- We don’t need to discuss implementation
  - But you may think about some of the common sense requirements that come up from the discussion that follows as “desiderata”.

- We care about plausible interaction models
  - In terms of levels/types of supervision

- We care about general purpose inference mechanisms that are realistic (in terms of scale).
The Neuro-Symbolic Connection

- We don’t need to discuss implementation
  - But you may think about some of the common sense requirements that come up from the discussion that follows as “desiderata”.

- We care about plausible interaction models
  - In terms of levels/types of supervision

- We care about general purpose inference mechanisms that are realistic (in terms of scale).
  - Can these desiderata serve to motivate a concrete research program in computational neuroscience, with the goal of addressing these?
The Neuro-Symbolic Connection

- We don’t need to discuss implementation
  - But you may think about some of the *common sense requirements* that come up from the discussion that follows as “desiderata”.

- We care about plausible interaction models
  - In terms of levels/types of supervision

- We care about general purpose inference mechanisms that are realistic (in terms of scale).
  - Can these desiderata serve to motivate a concrete research program in computational neuroscience, with the goal of *addressing these*?
    [Credit Isaac Noble for a discussion that led to this bullet]
Recognizing Entities and Relations

Dole’s wife, Elizabeth, is a native of N.C.
Recognizing Entities and Relations

Dole’s wife, Elizabeth, is a native of N.C.

\[ \text{E}_1 \xrightarrow{R_{12}} \text{E}_2 \xrightarrow{R_{23}} \text{E}_3 \]
Joint Inference with General Constraint Structure [Roth&Yih’04,07,...]

Recognizing Entities and Relations

Dole’s wife, Elizabeth, is a native of N.C.
Joint Inference with General Constraint Structure [Roth&Yih’04,07,....]
Recognizing Entities and Relations

Dole ’s wife, Elizabeth , is a native of N.C.

**E1**  **E2**  **E3**

\[ R_{12} \quad R_{23} \]

<table>
<thead>
<tr>
<th>other</th>
<th>0.05</th>
<th>0.10</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>per</td>
<td>0.85</td>
<td>0.60</td>
<td>0.50</td>
</tr>
<tr>
<td>loc</td>
<td>0.10</td>
<td>0.30</td>
<td>0.45</td>
</tr>
</tbody>
</table>

| irrelevant  | 0.05 | 0.10 | 0.05 |
| spouse_of   | 0.45 | 0.05 | 0.50 |
| born_in     | 0.50 | 0.85 | 0.05 |
Recognizing Entities and Relations

Joint Inference with General Constraint Structure [Roth&Yih’04,07,....]

Dole’s wife, Elizabeth, is a native of N.C.

E₁  R₁₂  E₂  R₂₃  E₃

other  0.05  other  0.10  other  0.05
per  0.85  per  0.60  per  0.50
loc  0.10  loc  0.30  loc  0.45

irrelevant  0.05  irrelevant  0.10
spouse_of  0.45  spouse_of  0.05
born_in  0.50  born_in  0.85
 Joint Inference with General Constraint Structure [Roth&Yih’04,07,....]

Recognizing Entities and Relations

Dole’s wife, Elizabeth, is a native of N.C.

E1  R_{12}  E2  R_{23}  E3

<table>
<thead>
<tr>
<th></th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
</tr>
</thead>
<tbody>
<tr>
<td>irrelevant</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spouse_of</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>born_in</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
</tr>
</thead>
<tbody>
<tr>
<td>irrelevant</td>
<td></td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>spouse_of</td>
<td></td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>born_in</td>
<td></td>
<td>0.85</td>
<td></td>
</tr>
</tbody>
</table>
Joint Inference with General Constraint Structure [Roth&Yih’04,07,...]
Recognizing Entities and Relations

Dole’s wife, Elizabeth, is a native of N.C.

E1 → E2 → E3

\[ R_{12} \]
\[ R_{23} \]

<table>
<thead>
<tr>
<th>Other</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per</td>
<td>0.85</td>
</tr>
<tr>
<td>Loc</td>
<td>0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per</td>
<td>0.60</td>
</tr>
<tr>
<td>Loc</td>
<td>0.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per</td>
<td>0.50</td>
</tr>
<tr>
<td>Loc</td>
<td>0.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Irrelevant</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>spouse_of</td>
<td>0.45</td>
</tr>
<tr>
<td>born_in</td>
<td>0.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Irrelevant</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>spouse_of</td>
<td>0.05</td>
</tr>
<tr>
<td>born_in</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Recognizing Entities and Relations

Joint Inference with General Constraint Structure [Roth & Yih, 2004, 2007, …]

Dole’s wife, Elizabeth, is a native of N.C.

\[ E_1 \xrightarrow{R_{12}} E_2 \xrightarrow{R_{23}} E_3 \]

\begin{tabular}{|c|c|}
  \hline
  other & 0.05 \\
  \hline
  per & 0.85 \\
  \hline
  loc & 0.10 \\
  \hline
\end{tabular}

\begin{tabular}{|c|c|}
  \hline
  other & 0.10 \\
  \hline
  per & 0.60 \\
  \hline
  loc & 0.30 \\
  \hline
\end{tabular}

\begin{tabular}{|c|c|}
  \hline
  other & 0.05 \\
  \hline
  per & 0.50 \\
  \hline
  loc & 0.45 \\
  \hline
\end{tabular}

\begin{tabular}{|c|c|}
  \hline
  irrelevant & 0.05 \\
  \hline
  spouse_of & 0.45 \\
  \hline
  born_in & 0.50 \\
  \hline
\end{tabular}

\begin{tabular}{|c|c|}
  \hline
  irrelevant & 0.10 \\
  \hline
  spouse_of & 0.05 \\
  \hline
  born_in & 0.85 \\
  \hline
\end{tabular}

Joint inference gives good improvement.
Joint Inference with General Constraint Structure [Roth&Yih'04, 07, …]

Recognizing Entities and Relations

Dole’s wife, Elizabeth, is a native of N.C.

Key Questions:
How to guide the global inference?
How to learn the model(s)?

Joint inference gives good improvement

<table>
<thead>
<tr>
<th></th>
<th>other</th>
<th>per</th>
<th>loc</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>0.05</td>
<td>0.85</td>
<td>0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>other</th>
<th>per</th>
<th>loc</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2</td>
<td>0.10</td>
<td>0.60</td>
<td>0.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>other</th>
<th>per</th>
<th>loc</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_{12}</td>
<td>0.05</td>
<td>0.50</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>other</th>
<th>per</th>
<th>loc</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_{23}</td>
<td>0.05</td>
<td>0.45</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>irrelevant</th>
<th>spouse_of</th>
<th>born_in</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>0.05</td>
<td>0.45</td>
<td>0.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>irrelevant</th>
<th>spouse_of</th>
<th>born_in</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2</td>
<td>0.10</td>
<td>0.05</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Joint Inference with General Constraint Structure

Recognizing Entities and Relations

Models could be learned separately/jointly; constraints may come up only at decision time.

Dole's wife, Elizabeth, is a native of N.C.

Key Questions:
- How to guide the global inference?
- How to learn the model(s)?

Joint inference gives good improvement.
### Joint Inference with General Constraint Structure

#### Recognizing Entities and Relations

<table>
<thead>
<tr>
<th></th>
<th>other</th>
<th>per</th>
<th>loc</th>
</tr>
</thead>
<tbody>
<tr>
<td>irrelevant</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spouse_of</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>born_in</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spouse_of</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>born_in</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>per</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>loc</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>per</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>loc</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>per</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>loc</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Models could be learned separately/jointly; constraints may come up only at decision time.

An **Objective function that incorporates learned knowledge** (output constraints)

A **Constrained Conditional Model**

Key Questions:
- How to guide the global inference?
- How to learn the model(s)?

Joint inference gives good improvement
Constrained Conditional Models

\[ y = \arg\max_{y \in \mathcal{Y}} w^T \phi(x, y) \]
Constrained Conditional Models

\[ y = \arg\max_{y \in \mathcal{Y}} w^T \phi(x, y) \]
Constrained Conditional Models

\[ y = \arg\max_{y \in \mathcal{Y}} \ w^T \phi(x, y) \]
Constrained Conditional Models

\[ y = \arg\max_{y \in Y} w^T \phi(x, y) \]

Weight Vector for “local” models

Features, classifiers; log-linear models (HMM, CRF) or a combination
Constrained Conditional Models

\[ y = \arg\max_{y \in \mathcal{Y}} w^T \phi(x, y) + u^T C(x, y) \]

Knowledge component: (Soft) constraints

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination
Constrained Conditional Models

\[ y = \arg\max_{y \in \mathcal{Y}} w^T \phi(x, y) + u^T C(x, y) \]

- Weight Vector for "local" models
- Features, classifiers; log-linear models (HMM, CRF) or a combination
- Knowledge component: (Soft) constraints
- How far \( y \) is from a "legal/expected" assignment
- Penalty for violating the constraint.
Constrained Conditional Models

\[ y = \arg\max_{y \in Y} w^T \phi(x, y) + u^T C(x, y) \]

- **Training**: learning the objective function \((w, u)\)
  - Decouple? Decompose? Force \(u\) to model hard constraints?

- Penalty for violating the constraint.
- Knowledge component: (Soft) constraints
- Weight Vector for “local” models
- How far \(y\) is from a “legal/expected” assignment

Features, classifiers; log-linear models (HMM, CRF) or a combination
Constrained Conditional Models

\[ y = \arg\max_{y \in Y} w^T \phi(x, y) + u^T C(x, y) \]

- **Training**: learning the objective function \((w, u)\)
  - Decouple? Decompose? Force \(u\) to model hard constraints?

- A way to push the learned model to **satisfy our output expectations** (or expectations from a latent representation)
  - [CoDL, Chang et. al (07, 12); Posterior Regularization, Ganchev et. al (10); Unified EM (Samdani et. al (12))]
Constrained Conditional Models

- **Training:** learning the objective function \((w, u)\)
  - Decouple? Decompose? Force \(u\) to model hard constraints?

- A way to push the learned model to **satisfy our output expectations** (or expectations from a latent representation)
  - [CoDL, Chang et. al (07, 12); Posterior Regularization, Ganchev et. al (10); Unified EM (Samdani et. al (12))]

\[
y = \arg\max_y \sum \phi(x, y) w_{x,y} \quad \text{subject to Constraints } C(x, y)
\]

- Weight Vector for “local” models
- Features, classifiers; log-linear models (HMM, CRF) or a combination
- Knowledge component: (Soft) constraints
- Penalty for violating the constraint.
- How far \(y\) is from a “legal/expected” assignment
Constrained Conditional Models

- Training: learning the objective function \((w, u)\)
  - Decouple? Decompose? Force \(u\) to model hard constraints?
- A way to push the learned model to **satisfy our output expectations** (or expectations from a latent representation)
  - [CoDL, Chang et al (07, 12); Posterior Regularization, Ganchev et al (10); Unified EM (Samdani et al (12))]

\[
y = \arg\max_y \sum_{\phi(x,y)} w_{x,y} \text{ subject to Constraints } C(x,y)
\]
Constrained Conditional Models

Training: learning the objective function \((w, u)\)

- Decouple? Decompose? Force \(u\) to model hard constraints?

A way to push the learned model to **satisfy our output expectations** (or expectations from a latent representation)

- [CoDL, Chang et. al (07, 12); Posterior Regularization, Ganchev et. al (10); Unified EM (Samdani et. al (12)]
Examples: CCM Formulations

\[ y = \arg\max_{y \in \mathcal{Y}} \ w^T \phi(x, y) + u^T \mathcal{C}(x, y) \]
Examples: CCM Formulations

\[ y = \arg\max_{y \in Y} w^T \phi(x, y) + u^T C(x, y) \]

While \( \phi(x, y) \) and \( C(x, y) \) could be the same; we want \( C(x, y) \) to express high level declarative knowledge over the statistical models.
Examples: CCM Formulations

\[ y = \arg\max_{y \in Y} w^T \phi(x, y) + u^T C(x, y) \]

While \( \phi(x, y) \) and \( C(x, y) \) could be the same; we want \( C(x, y) \) to express high level declarative knowledge over the statistical models.

Formulate NLP Problems as ILP problems (inference may be done otherwise)
1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)
Examples: CCM Formulations

\[ y = \arg\max_{y \in Y} w^T \phi(x, y) + u^T C(x, y) \]

While \( \phi(x, y) \) and \( C(x, y) \) could be the same; we want \( C(x, y) \) to express high level declarative knowledge over the statistical models.

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Sequential Prediction

HMM/CRF based:
\[ \text{Argmax} \sum \lambda_{ij} x_{ij} \]

Knowledge/Linguistics Constraints

Cannot have both A states and B states in an output sequence.
Examples: CCM Formulations

\[ y = \operatorname{argmax}_{y \in \mathcal{Y}} \quad w^T \phi(x, y) + u^T C(x, y) \]

While \( \phi(x, y) \) and \( C(x, y) \) could be the same; we want \( C(x, y) \) to express high level declarative knowledge over the statistical models.

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Sentence Compression/Summarization:

Language Model based:

\[ \operatorname{Argmax} \sum \lambda_{ijk} x_{ijk} \]

Knowledge/Linguistics Constraints

If a modifier chosen, include its head
If verb is chosen, include its arguments
Examples: CCM Formulations

\[ y = \arg\max_{y \in Y} w^T \phi(x, y) + u^T C(x, y) \]

While \( \phi(x, y) \) and \( C(x, y) \) could be the same; we want \( C(x, y) \) to express high level declarative knowledge over the statistical models.

Formulate NLP Problems as ILP problems

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Sentence Compression/Summarization:
Language Model based:
\[ \text{Argmax} \sum \lambda_{ijk} x_{ijk} \]

Knowledge/Linguistics Constraints
- If a modifier chosen, include its head
- If verb is chosen, include its arguments
Examples: CCM Formulations

\[ y = \arg\max_{y \in Y} w^T \phi(x, y) + u^T C(x, y) \]

While \( \phi(x, y) \) and \( C(x, y) \) could be the same; we want \( C(x, y) \) to express high level declarative knowledge over the statistical models.

Formulate NLP Problems as ILP problems (inference may be done otherwise)
1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Constrained Conditional Models Allow:
- Decouple complexity of the learned model from that of the desired output
- Learn a simple model (multiple; pipelines); reason with a complex one.
- Accomplished by incorporating constraints to bias/re-rank global decisions to satisfy (minimally violate) expectations.
Semantic Role Labeling (SRL)

I left my pearls to my daughter in my will.

\[
I_{A0} \text{ left } [\text{my pearls}]_{A1} \text{ [to my daughter]}_{A2} \text{ [in my will]}_{AM-LOC}.
\]

- **A0** Leaver
- **A1** Things left
- **A2** Benefactor
- **AM-LOC** Location

I left my pearls to my daughter in my will.
Semantic Role Labeling (SRL)

I left my pearls to my daughter in my will.

\[ I_A^0 \text{ left } [\text{my pearls}]_{A1} \text{ to my daughter} ]_{A2} \text{ in my will } ]_{AM-LOC} .

- **A0** Leaver
- **A1** Things left
- **A2** Benefactor
- **AM-LOC** Location

I left my pearls to my daughter in my will.
Algorithmic Approach

- **Identify** argument candidates
  - Pruning [Xue & Palmer, EMNLP'04]
  - Argument Identifier
    - Binary classification

- **Classify** argument candidates
  - Argument Classifier
    - Multi-class classification

- **Inference**
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output
Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP’04]
  - Argument Identifier
    - Binary classification

- Classify argument candidates
  - Argument Classifier
    - Multi-class classification

- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output
Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP’04]
  - Argument Identifier
    - Binary classification

- Classify argument candidates
  - Argument Classifier
    - Multi-class classification

- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output
Algorithmic Approach

- **Identify** argument candidates
  - Pruning [Xue&Palmer, EMNLP’04]
  - Argument Identifier
    - Binary classification

- **Classify** argument candidates
  - Argument Classifier
    - Multi-class classification

- **Inference**
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output
Algorithmic Approach

- **Identify** argument candidates
  - Pruning [Xue & Palmer, EMNLP’04]
  - Argument Identifier
    - Binary classification

- **Classify** argument candidates
  - Argument Classifier
    - Multi-class classification

- **Inference**
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output
    - One inference problem for each verb predicate.
Algorithmic Approach

- **Identify** argument candidates
  - Pruning [Xue & Palmer, EMNLP’04]
  - Argument Identifier
    - Binary classification
- **Classify** argument candidates
  - Argument Classifier
    - Multi-class classification
- **Inference**
  
  \[
  \arg\max \sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}
  \]

  Subject to:
  - One label per argument: \( \sum_t y^{a,t} = 1 \)
  - No overlapping or embedding
  - Relations between verbs and arguments,....
Algorithmic Approach

- **Identify** argument candidates
  - Pruning [Xue&Palmer, EMNLP’04]
  - Argument Identifier
    - Binary classification

- **Classify** argument candidates
  - Argument Classifier
    - Multi-class classification

- **Inference**
  
  \[
  \arg\max_{a,t} \sum_{a,t} y_{a,t} c_{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}
  \]

  Subject to:
  - One label per argument: \( \sum_{t} y_{a,t} = 1 \)
  - No overlapping or embedding
  - Relations between verbs and arguments,....
Algorithmic Approach

- **Identify** argument candidates
  - Pruning [Xue&Palmer, EMNLP’04]
  - Argument Identifier
    - Binary classification

- **Classify** argument candidates
  - Argument Classifier
    - Multi-class classification

- **Inference**

\[
\text{argmax } \sum_{a,t} y_{a,t} c_{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}
\]

Subject to:
- One label per argument: \(\sum_t y_{a,t} = 1\)
- No overlapping or embedding
- Relations between verbs and arguments,....

Variable \(y_{a,t}\) indicates whether candidate argument \(a\) is assigned a label \(t\).
\(c_{a,t}\) is the corresponding model score.

Abstract representation of expectations/knowledge.
Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - Binary classification
- Classify argument candidates
  - Argument Classifier
    - Multi-class classification
- Inference
  
  \[ \text{argmax} \sum_{a,t} y_{a,t} c_{a,t} = \sum_{a,t} 1_{a=t} c_{a=t} \]

  Subject to:
  - One label per argument: \( \sum_t y_{a,t} = 1 \)
  - No overlapping or embedding
  - Relations between verbs and arguments,....

No duplicate argument classes
\[ \forall i, \sum_{y \in \mathcal{Y}} 1\{y_i=y\} = 1 \]

Unique labels
\[ \forall y \in \mathcal{Y}, \sum_{i=0}^{n-1} 1\{y_i=y\} \leq 1 \]

\[ \forall y \in \mathcal{Y_R}, \sum_{i=0}^{n-1} 1\{y_i=y=\text{“R-Ax”}\} \leq \sum_{i=0}^{n-1} 1\{y_i=\text{“Ax”}\} \]

\[ \forall j, y \in \mathcal{Y_C}, 1\{y_j=y=\text{“C-Ax”}\} \leq \sum_{i=0}^{j} 1\{y_i=\text{“Ax”}\} \]

I left my nice pearls to her

Abstract representation of expectations/knowledge
Algorithmic Approach

- **Identify** argument candidates
  - Pruning [Xue&Palmer, EMNLP’04]
  - Argument Identifier
    - Binary classification
- **Classify** argument candidates
  - Argument Classifier
    - Multi-class classification
- **Inference**

\[
\arg\max \sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}
\]

Subject to:
- One label per argument: \( \sum_t y^{a,t} = 1 \)
- No overlapping or embedding
- Relations between verbs and arguments,....

**Learning Based Java:** allows a developer to encode constraints in First Order Logic; these are compiled into linear inequalities automatically.

Abstract representation of expectations/knowledge

Variable \( y^{a,t} \) indicates whether candidate argument \( a \) is assigned a label \( t \).

\( c^{a,t} \) is the corresponding model score
Algorithmic Approach

- **Identify** argument candidates
  - Pruning [Xue&Palmer, EMNLP’04]
  - Argument Identifier
    - Binary classification

- **Classify** argument candidates
  - Argument Classifier
    - Multi-class classification

- **Inference**

  \[
  \text{argmax} \sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}
  \]

  Subject to:
  - One label per argument: \(\sum_t y^{a,t} = 1\)
  - No overlapping or embedding
  - Relations between verbs and arguments,....

Use the **pipeline architecture’s simplicity** while **maintaining uncertainty**: keep probability distributions over decisions & use global inference at decision time.

**Learning Based Java**: allows a developer to encode constraints in First Order Logic; these are compiled into linear inequalities automatically.

Variable \(y^{a,t}\) indicates whether candidate argument \(a\) is assigned a label \(t\). \(c^{a,t}\) is the corresponding model score.

Abstract representation of expectations/knowledge
The Computational Process

- The computational process used in each of these examples is very similar to the one used in the babySRL

  - Models are induced via some interactive learning process
    - Feedback goes back to improve earlier learned models

  - Relatively abstract knowledge, is used
    - "Output expectations", or "constraints" on what can be represented guide learning and prediction (inference)

  - Knowledge impacts both latent representations and predictions

- Today, the key difference between the babySRL and our other models is in the level of supervision
  - And consequently, the type of text we can deal with.
Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.
ENGLAND, June, 1989 - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.
I. Coreference Resolution

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.
Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

- Big Problem; essential to text understanding; hard.
- Requires: good learning and inference models & knowledge
Recent Advances in Co-reference [Chang, Peng, Samdani, Khashabi]

- Latent Left-linking Model (L3M) model [ICML 14]

- Joint mention identification & co-reference resolution [CoNLL’15]

- Hard Co-reference Problems [NAACL’15]
Recent Advances in Co-reference \cite{Chang2014}

- **Latent Left-linking Model (L3M) model** \cite{ICML2014}
  
  A latent variable structured prediction model for discriminative supervised clustering. **Jointly** learns a similarity function and performs inference, assuming a latent left linking forest of mentions.

- **Joint mention identification & co-reference resolution** \cite{CoNLL2015}

- **Hard Co-reference Problems** \cite{NAACL2015}
Recent Advances in Co-reference [Chang, Peng, Samdani, Khashabi]

- Latent Left-linking Model (L3M) model [ICML 14]
  - A latent variable structured prediction model for discriminative supervised clustering. Jointly learns a similarity function and performs inference, assuming a latent left linking forest of mentions.

- Joint mention identification & co-reference resolution [CoNLL’15]
  - Augment the ILP based Inference formulation with “a legitimate mention” variable, to jointly determine if the mention is legitimate and what to co-ref it with

- Hard Co-reference Problems [NAACL’15]
Recent Advances in Co-reference

- Latent Left-linking Model (L3M) model [ICML 14]
  - A latent variable structured prediction model for discriminative supervised clustering. *Jointly* learns a similarity function and performs inference, assuming a *latent left linking forest* of mentions.

- Joint mention identification & co-reference resolution [CoNLL’15]
  - Augment the ILP based Inference formulation with “*a legitimate mention*” variable, to *jointly* determine if the mention is legitimate and what to co-ref it with

- Hard Co-reference Problems [NAACL’15]

All together, the outcome is the best end-to-end coreference results on CoNLL data and on ACE [CoNLL’15]
Recent Advances in Co-reference

- Latent Left-linking Model (L3M) model [ICML 14]
  - A latent variable structured prediction model for discriminative supervised clustering. *Jointly* learns a similarity function and performs inference, assuming a *latent left linking forest* of mentions.

- Joint mention identification & co-reference resolution [CoNLL’15]
  - Augment the ILP based Inference formulation with “*a legitimate mention*” variable, to *jointly* determine if the mention is legitimate and what to co-ref it with

*Hard Co-reference Problems [NAACL’15]*
Pronoun Resolution can be Really Hard

- When Tina pressed Joan to the floor she was punished.
- When Tina pressed Joan to the floor she was hurt.
- When Tina pressed charges against Joan she was jailed.
Pronoun Resolution can be Really Hard

- When Tina pressed Joan to the floor she was punished.
- When Tina pressed Joan to the floor she was hurt.
- When Tina pressed charges against Joan she was jailed.
Pronoun Resolution can be Really Hard

- When Tina pressed Joan to the floor she was punished.
- When Tina pressed Joan to the floor she was hurt.
- When Tina pressed charges against Joan she was jailed.

- Requires, among other things, thinking about the structure of the sentence – who does what to whom
Hard Co-reference Problems

- Requires knowledge Acquisition
Hard Co-reference Problems

- Requires knowledge Acquisition
  - The bee landed on the flower because it had/wanted pollen.
Hard Co-reference Problems

- Requires knowledge Acquisition
  - The bee landed on the flower because it had/wanted pollen.
    - Lexical knowledge
Hard Co-reference Problems

- Requires knowledge Acquisition
  - The bee landed on the flower because it had/wanted pollen.
    - Lexical knowledge
  - John Doe robbed Jim Roy. He was arrested by the police.
Hard Co-reference Problems

- Requires knowledge Acquisition
  - The bee landed on the flower because it had/wanted pollen.
    - Lexical knowledge
  - John Doe robbed Jim Roy. He was arrested by the police.

- The Subj of “rob” is more likely than the Obj of “rob” to be the Obj of “arrest”
Hard Co-reference Problems

- Requires knowledge Acquisition
  - The bee landed on the flower because it had/wanted pollen.
    - Lexical knowledge
  - John Doe robbed Jim Roy. He was arrested by the police.
    - The Subj of “rob” is more likely than the Obj of “rob” to be the Obj of “arrest”
Hard Co-reference Problems

- Requires knowledge Acquisition
  - The bee landed on the flower because it had/wanted pollen.
    - Lexical knowledge
  - John Doe robbed Jim Roy. He was arrested by the police.
    - The Subj of “rob” is more likely than the Obj of “rob” to be the Obj of “arrest”

- Requires an inference framework that can make use of this knowledge
ILP Formulation of Coreference Resolution

\[ y = \arg \max_y \sum_{uv} w_{uv} \cdot y_{uv} \]

s.t. \( \sum_{u < v} y_{uv} \leq 1, \forall v \)

\( y_{uv} \in \{0,1\} \)
ILP Formulation of Coreference Resolution

\[
y = \arg \max_y \sum_{uv} w_{uv} \cdot y_{uv} \\
\text{s.t} \sum_{u < v} y_{uv} \leq 1, \forall v \\
y_{uv} \in \{0,1\}
\]
ILP Formulation of Coreference Resolution

\[ y = \arg \max_y \sum_{uv} w_{uv} \cdot y_{uv} \]

s.t \( \sum_{u < v} y_{uv} \leq 1, \forall v \)

\( y_{uv} \in \{0,1\} \)
ILP Formulation of Coreference Resolution

\[ y = \arg \max_y \sum_{uv} w_{uv} \cdot y_{uv} \]

s.t. \[ \sum_u < v y_{uv} \leq 1, \forall v \]

\[ y_{uv} \in \{0,1\} \]

Variable \( y_{uv} \) indicates a coreference link \( u \to v \)
ILP Formulation of Coreference Resolution

\[ y = \arg \max_y \sum_{uv} w_{uv} \cdot y_{uv} \]
\[ \text{s.t } \sum_u < v y_{uv} \leq 1, \ \forall v \]
\[ y_{uv} \in \{0,1\} \]

**Best Link Approach:** only one of the antecedents \( u \) is linked to \( v \)

Variable \( y_{uv} \) indicates a coreference link \( u \rightarrow v \)
ILP Formulation of Coreference Resolution

\[ y = \arg \max_y \sum_{uv} w_{uv} \cdot y_{uv} \]

s.t \[ \sum_u < v y_{uv} < 1, \quad \forall v \]

\[ y_{uv} \in \{0, 1\} \]
ILP Formulation of Coreference Resolution

\[ y = \arg \max_y \sum_{uv} w_{uv} \cdot y_{uv} \]

s.t \( \sum_u < v \ y_{uv} < 1, \ \forall v \)

\[ y_{uv} \in \{0,1\} \]

\[
\begin{cases}
\text{if } s_i(u, v) \geq \alpha_i s_i(w, v) \Rightarrow y_{u,v} \geq y_{w,v}, \\
\text{if } s_i(u, v) \geq s_i(w, v) + \beta_i \Rightarrow y_{u,v} \geq y_{w,v}
\end{cases}
\]
ILP Formulation of Coreference Resolution

\[ y = \arg \max_y \sum_{uv} w_{uv} \cdot y_{uv} \]

s.t \[ \sum_{u < v} y_{uv} < 1, \ \forall v \]
\[ y_{uv} \in \{0,1\} \]

predicate schemas

\[
\begin{cases}
\text{if } s_i(u,v) \geq \alpha_i s_i(w,v) \Rightarrow y_{u,v} \geq y_{w,v}, \\
\text{if } s_i(u,v) \geq s_i(w,v) + \beta_i \Rightarrow y_{u,v} \geq y_{w,v}
\end{cases}
\]
ILP Formulation of Coreference Resolution

- \( y = \arg \max_y \sum_{uv} w_{uv} \cdot y_{uv} \)
  
  \[ \text{s.t} \sum_{u < v} y_{uv} < 1, \forall v \]
  \( y_{uv} \in \{0,1\} \)

- Acquire knowledge; formulated via “Predicate Schemas”.
  
  - Constraints over predicate schemas are instantiated given a new instance (document) and are incorporated “on-the-fly” into the ILP-based inference formulation to support preferred interpretations.
ILP Formulation of Coreference Resolution

\[ y = \text{arg max}_y \sum_{uv} w_{uv} \cdot y_{uv} \]
\[ \text{s.t} \sum_u < v y_{uv} < 1, \ \forall v \]
\[ y_{uv} \in \{0,1\} \]

Results in a state-of-the-art coreference that at the same time also handles hard instances at close to 90% Precision.

Predicate schemas

\[
\begin{cases}
\text{if } s_i(u, v) \geq \alpha_i s_i(w, v) \Rightarrow y_{u,v} \geq y_{w,v}, \\
\text{if } s_i(u, v) \geq s_i(w, v) + \beta_i \Rightarrow y_{u,v} \geq y_{w,v}
\end{cases}
\]

Acquire knowledge; formulated via “Predicate Schemas”.

Constraints over predicate schemas are instantiated given a new instance (document) and are incorporated “on-the-fly” into the ILP-based inference formulation to support preferred interpretations.
II. Quantities & Quantitative Reasoning

- A crucially important natural language understanding task.
- Election results; Stock Market; Casualties,...
II. Quantities & Quantitative Reasoning

- A crucially important natural language understanding task.
- Election results; Stock Market; Casualties,...

  The Emmanuel campaign funding totaled three times that of all his opponents put together.
II. Quantities & Quantitative Reasoning

- A crucially important natural language understanding task.
- Election results; Stock Market; Casualties,…
  
  The Emmanuel campaign funding totaled three times that of all his opponents put together.

- Understanding implies mapping the text to an arithmetic expression, or an equation: \[ E = 3 \sum_i o_i \]
II. Quantities & Quantitative Reasoning

- A crucially important natural language understanding task.
- Election results; Stock Market; Casualties,…
  
  The Emmanuel campaign funding totaled three times that of all his opponents put together.
- Understanding implies mapping the text to an arithmetic expression, or an equation: \( E = \sim 3 \sum \alpha \)
II. Quantities & Quantitative Reasoning

- A crucially important natural language understanding task.
- Election results; Stock Market; Casualties,...

  The Emmanuel campaign funding totaled three times that of all his opponents put together.

- Understanding implies mapping the text to an arithmetic expression, or an equation:  
  \[ E = \sim \sum_1^3 o_i \]

  John had 6 books; he wanted to give it to two of his friends. How many will each one get?
II. Quantities & Quantitative Reasoning

- A crucially important natural language understanding task.
- Election results; Stock Market; Casualties,...
  
The Emmanuel campaign funding totaled three times that of all his opponents put together.
- Understanding implies mapping the text to an arithmetic expression, or an equation: \( E = \sim 3 \sum_i o_i \)

  share it with
  
  John had 6 books; he wanted to give it to two of his friends. How many will each one get?
II. Quantities & Quantitative Reasoning

- A crucially important natural language understanding task.
- Election results; Stock Market; Casualties,...

The Emmanuel campaign funding totaled three times that of all his opponents put together.

- Understanding implies mapping the text to an arithmetic expression, or an equation: \( E = \sum_{i} 3 o_i \)

John had 6 books; he wanted to give it to two of his friends. How many will each one get?
Gwen was organizing her book case making sure each of the shelves had exactly 9 books on it. She has 2 types of books – mystery books and picture books. If she had 3 shelves of mystery books and 5 shelves of picture books, how many books did she have total?
Gwen was organizing her book case making sure each of the shelves had exactly 9 books on it. She has 2 types of books – mystery books and picture books. If she had 3 shelves of mystery books and 5 shelves of picture books, how many books did she have total?

[Roy & Roth’15] suggests a solution that involves “parsing” the problem into an expression tree.
Gwen was organizing her book case making sure each of the shelves had exactly 9 books on it. She has 2 types of books – mystery books and picture books. If she had 3 shelves of mystery books and 5 shelves of picture books, how many books did she have total?

[Roy & Roth’15] suggests a solution that involves “parsing” the problem into an expression tree.
Gwen was organizing her book case making sure each of the shelves had exactly 9 books on it. She has 2 types of books – mystery books and picture books. If she had 3 shelves of mystery books and 5 shelves of picture books, how many books did she have total?

[Roy & Roth’15] suggests a solution that involves “parsing” the problem into an expression tree.
Inferring the Best Expression Tree

- **Decomposition**: Uniqueness properties of the $T(E)$ implies that it is determined by the unique $T$–operation between pairs of relevant quantities.
Inferring the Best Expression Tree

- **Decomposition:** Uniqueness properties of the $T(E)$ implies that it is determined by the unique $T$-operation between pairs of relevant quantities.

\[ E^* = \text{argmax} \sum_q R(q) \mathbf{1}_q + \square_{(q, q')} \text{Pair}(q, q', \odot(q, q')) \mathbf{1}_{q, q'} \]
Inferring the Best Expression Tree

- **Decomposition:** Uniqueness properties of the $T(E)$ implies that it is determined by the unique $T$–operation between pairs of relevant quantities.

$$E^* = \arg\max \sum_q R(q) \ 1_q + \bigwedge_{(q, q')} \text{Pair}(q, q', \odot(q, q')) \ 1_{q, q'}$$

Score of q being irrelevant to E
Inferring the Best Expression Tree

- **Decomposition:** Uniqueness properties of the $T(E)$ implies that it is determined by the unique $T$–operation between pairs of relevant quantities.

\[ E^* = \arg\max \sum_q R(q) \mathbf{1}_q + \Box_{(q, q')} \text{Pair}(q, q', \circ(q, q')) \mathbf{1}_{q, q'} \]

- Score of $q$ being irrelevant to $E$
- Score of $\circ$ being the unique operation between $(q_i, q_j)$
Inferring the Best Expression Tree

- **Decomposition**: Uniqueness properties of the $T(E)$ implies that it is determined by the unique $T$–operation between pairs of relevant quantities.

$$E^* = \arg\max \sum_q R(q) \mathbf{1}_q + \bigoplus_{(q, q')} \text{Pair}(q, q', \oplus(q, q')) \mathbf{1}_{q,q'}$$

- Subject to commonsense constraints.
  - Legitimacy
  - Positive Answer; Integral Answer; Range,...
Inferring the Best Expression Tree

- **Decomposition**: Uniqueness properties of the $T(E)$ implies that it is determined by the unique $T$–operation between pairs of relevant quantities.

$$E^* = \arg\max \sum_q R(q) \mathbf{1}_q + \bigotimes_{(q, q')} \text{Pair}(q, q', \odot(q, q')) \mathbf{1}_{q, q'}$$

- Subject to commonsense constraints.
  - Legitimacy
  - Positive Answer; Integral Answer; Range,...

Abstract Expectations developed given a text snippet
Inferring the Best Expression Tree

- **Decomposition**: Uniqueness properties of the $T(E)$ implies that it is determined by the unique $T$–operation between pairs of relevant quantities.

$$E^* = \arg\max \sum_q R(q) \mathbb{1}_q + \bigoplus_{(q, q')} \text{Pair}(q, q', \bigodot(q, q')) \mathbb{1}_{q,q'}$$

- Subject to *commonsense constraints*.
  - Legitimacy
  - Positive Answer; Integral Answer; Range,...

Results in a state-of-the-art results on multiple types of arithmetic word problems.
More Examples

- A lot of our natural language understanding work addresses similar issues and makes use of similar principles

  - Temporal Reasoning
    - We have expectations of transitivity, for example

  - Discourse Processing
    - We have expectations on “coherency” is conveying ideas

  - Knowledge Acquisition
    - We have expectations dictated by our prior knowledge

- See references for our work on various semantic processing tasks
Conclusion

- Natural Language Understanding is a Common Sense Inference problem.

- We would gain by thinking in a unified way on Learning, Knowledge (Representation and Acquisition) and Reasoning.

- Provided some recent samples from a research program that addresses
  - Learning, Inference and Knowledge via a unified approach
  - A constrained optimization framework that guides “best assignment” inference, with (declarative) expectations on the output.
Conclusion

- Natural Language Understanding is a Common Sense Inference problem.

- We would gain by thinking in a unified way on Learning, Knowledge (Representation and Acquisition) and Reasoning.

- Provided some recent samples from a research program that addresses
  - Learning, Inference and Knowledge via a unified approach
  - A constrained optimization framework that guides “best assignment” inference, with (declarative) expectations on the output.

- One can think about whether and how, we should constrain our models for addressing these “cognitive” problems, and our forms of interaction with data, in ways that are informed by neuroscience research.
Conclusion

- Natural Language Understanding is a Common Sense Inference problem.

- We would gain by thinking in a unified way on Learning, Knowledge (Representation and Acquisition) and Reasoning.

- Provided some recent samples from a research program that addresses
  - Learning, Inference and Knowledge via a unified approach
  - A constrained optimization framework that guides “best assignment” inference, with (declarative) expectations on the output.

- One can think about whether and how, we should constrain our models for addressing these “cognitive” problems, and our forms of interaction with data, in ways that are informed by neuroscience research.

Check out our CCM tutorial Software tools, demos, ...
Conclusion

- Natural Language Understanding is a Common Sense Inference problem.

- We would gain by thinking in a unified way on Learning, Knowledge (Representation and Acquisition) and Reasoning.

- Provided some recent samples from a research program that addresses
  - Learning, Inference and Knowledge via a unified approach
  - A constrained optimization framework that guides “best assignment” inference, with (declarative) expectations on the output.

- One can think about whether and how, we should constrain our models for addressing these “cognitive” problems, and our forms of interaction with data, in ways that are informed by neuroscience research.

Check out our CCM tutorial Software tools, demos, ...