COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

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

COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Natural Language Understanding with Common Sense Reasoning

Dan Roth

Department of Computer Science

University of Illinois at Urbana-Champaign

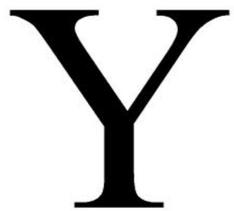
With thanks to:

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.

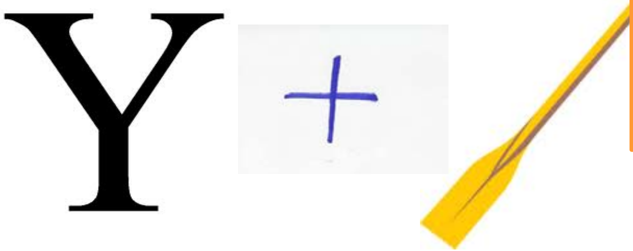
Please…




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



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.

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.

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?

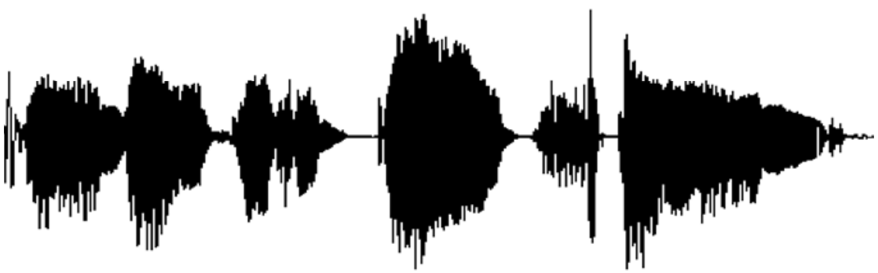
[Joint Research Program with Developmental Psycholinguist Cindy Fisher]

- Topid rivvo den marplox.

The Language-World Mapping Problem

“the world”

“the language”



[Topid rivvo den marplox.]

Observe how Words are Distributed Across Situations

Smur! Rivvo della frowler. Scene 1

Topid rivvo den marplox.

Blert dor marplox, arno. Scene 3

Marplox dorinda blicket. Scene n

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]

[Johanna rivvo den sheep.]

Nouns identified

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.]

Nouns identified

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.]

Nouns identified

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

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

Transitive: The boy is daxing the girl!

preferential looking paradigm

Strong Predictions [Gertner & Fisher, 2006]

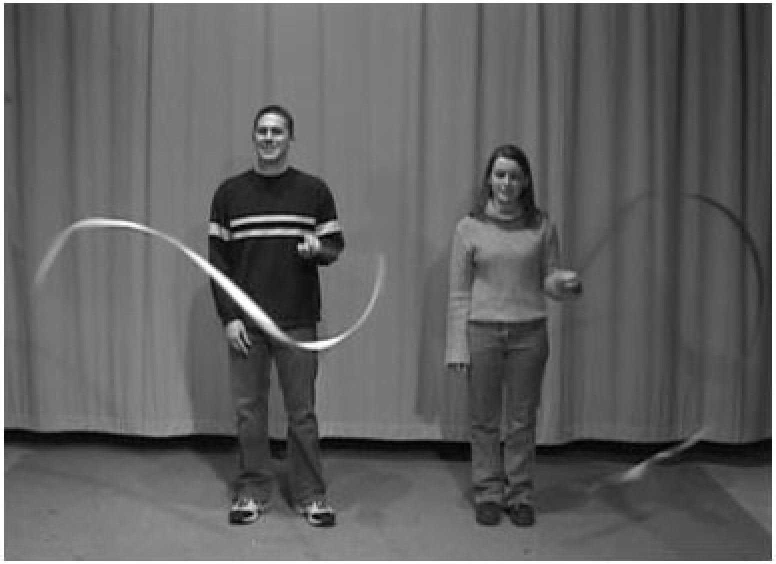
- Test 21 month olds on assigning arguments with novel verbs
- How order of nouns influences interpretation: Transitive & Intransitive

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!

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!

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.

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

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

- Interpretation builds on expectations that rely on knowledge.

Natural Language Inferences

- 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 ….

Natural Language Inferences

- 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 Inferences

- 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 Inferences

- 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 Inferences

- 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 Inferences

- 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 Inferences

visitors

- 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 Inferences

visitors

- 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 Inferences

visitors

- 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 Inferences

visitors

- 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 Inferences

visitors

- 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 Inferences

visitors

- 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 Inferences

visitors

- 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 Inferences

visitors

- 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 Inferences

visitors

- 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.

This is an Inference Problem

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

Expectation is a knowledge intensive component

- 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

Expectation is a knowledge intensive component

- 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

Expectation is a knowledge intensive component

- 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

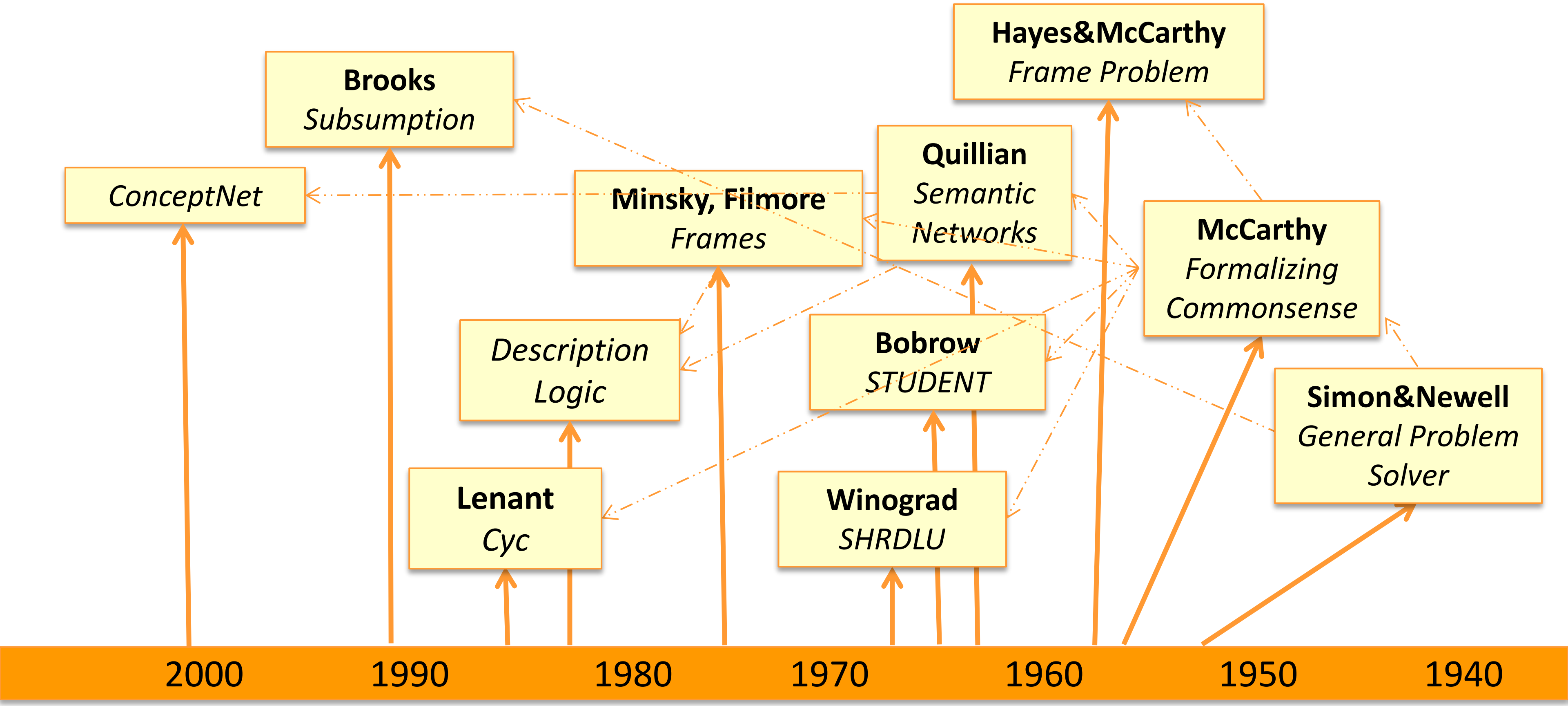
Expectation is a knowledge intensive component

- 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.

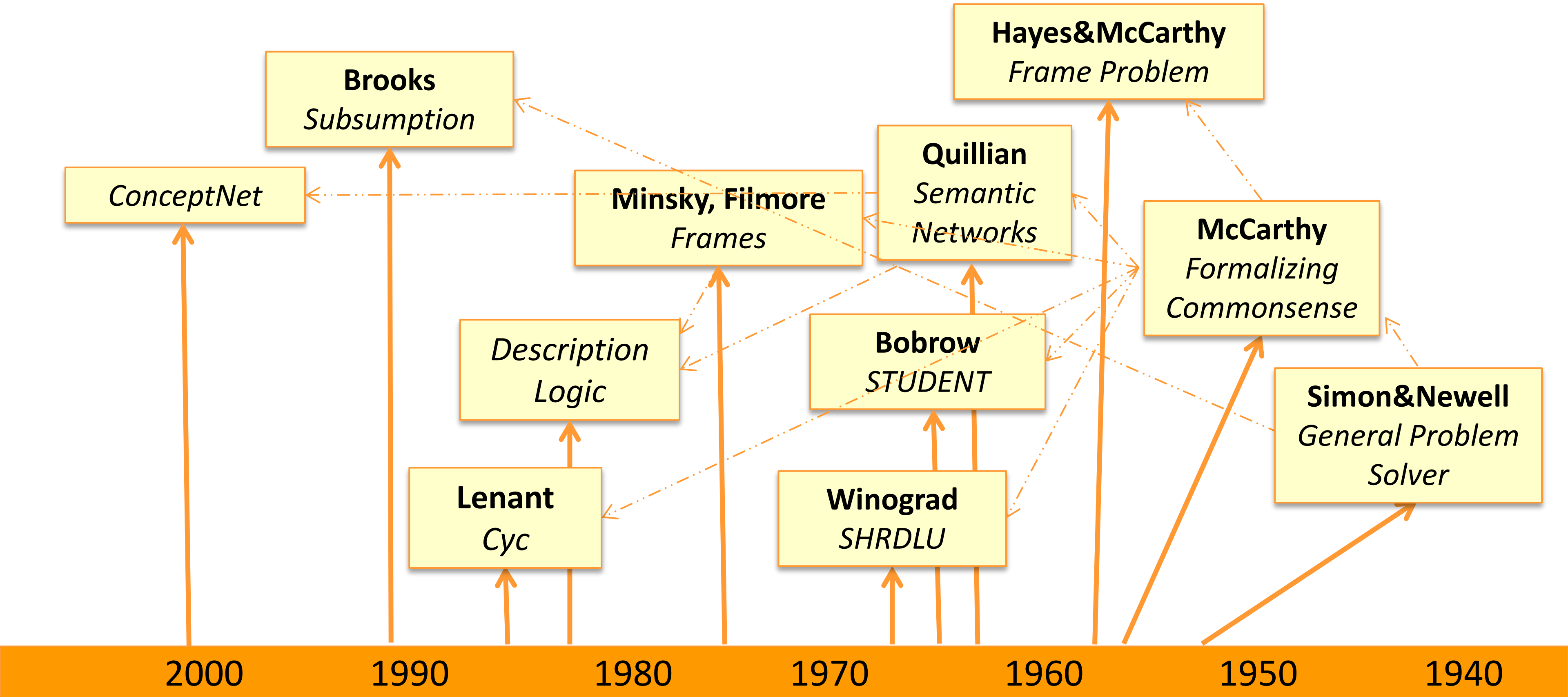
Many forms of Inference; a lot boil down to determining best assignment

A Biased View of Common Sense Reasoning



Common Sense Reasoning was formulated traditionally as a "reasoning" process, irrespective of learning and the resulting knowledge representation.

A Biased View of Common Sense Reasoning



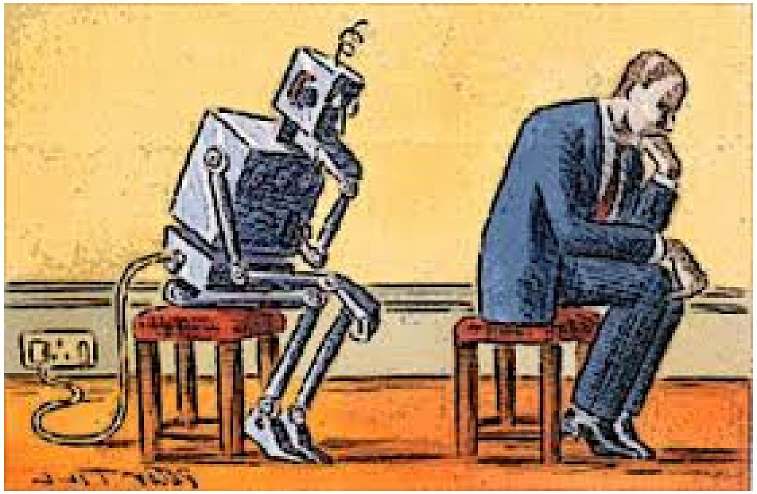
Common Sense Reasoning was formulated traditionally as a "reasoning" process, irrespective of learning and the resulting knowledge representation.

What is Needed?

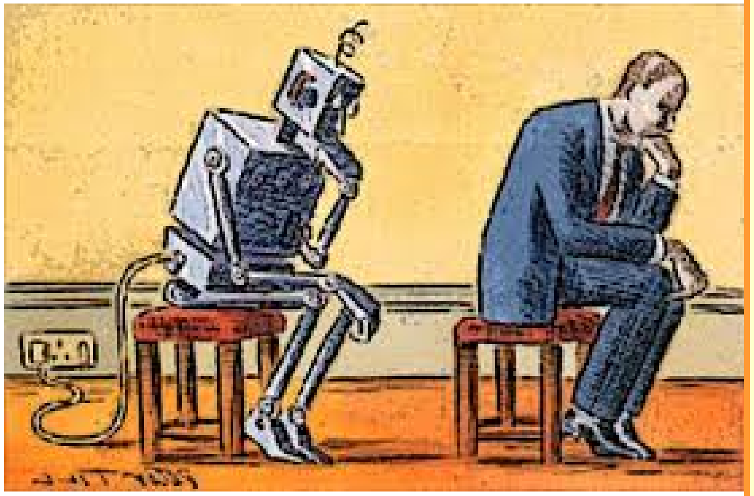
What is Needed?

What is Needed?

What is Needed?



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]

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

Recognizing Entities and Relations

$\boxed{\text{Dole}}$'s wife, $\boxed{\text{Elizabeth}}$, is a native of $\boxed{\text{N.C.}}$

$E_1 \qquad E_2 \qquad E_3$

$R_{12} \qquad R_{23}$

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

Recognizing Entities and Relations

other	0.05
per	0.85
loc	0.10

other	0.10
per	0.60
loc	0.30

other	0.05
per	0.50
loc	0.45

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

E1 E2 E3

R_{12} R_{23}

irrelevant	0.05
spouse_of	0.45
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	0.85

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

Recognizing Entities and Relations

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	**0.50**
loc	0.45

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

E_1 E_2 E_3

R_{12} R_{23}

irrelevant	0.05
spouse_of	0.45
born_in	**0.50**

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

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

Recognizing Entities and Relations

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	**0.50**
loc	0.45

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

E1 E2 E3

R_{12} R_{23}

irrelevant	0.05
spouse_of	0.45
born_in	**0.50**

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

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

Recognizing Entities and Relations

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	0.50
loc	**0.45**

$\boxed{\text{Dole}}$'s wife, $\boxed{\text{Elizabeth}}$, is a native of $\boxed{\text{N.C.}}$

$E_1 \xrightarrow{R_{12}} E_2 \xrightarrow{R_{23}} E_3$

irrelevant	0.05
spouse_of	0.45
born_in	**0.50**

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

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

Recognizing Entities and Relations

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	0.50
loc	**0.45**

$\boxed{\text{Dole}}$'s wife, $\boxed{\text{Elizabeth}}$, is a native of $\boxed{\text{N.C.}}$

$E1 \xrightarrow{R_{12}} E2 \xrightarrow{R_{23}} E3$

irrelevant	0.05
spouse_of	0.45
born_in	**0.50**

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

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

Recognizing Entities and Relations

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	0.50
loc	**0.45**

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

E1 $\xrightarrow{R_{12}}$ E2 $\xrightarrow{R_{23}}$ E3

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Joint Inference with General Constraint Structure [R

Recognizing Entities and Relations

Joint inference gives good improvement

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

othe	0.05
p	0.50
loc	**0.45**

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

$E1 \xrightarrow{R_{12}} E2 \xrightarrow{R_{23}} E3$

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Joint Inference with General Constraint Structure [R

Recognizing Entities and Relations

Joint inference gives good improvement

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

othe	0.05
p	0.50
loc	**0.45**

Dole 's wife, Elizabeth , is a na

E1 E2

R_{12} R_{23}

Key Questions:

How to guide the global inference?

How to learn the model(s)?

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Joint Inference with General Constraint Structure [R

Recognizing Entities and Relations

Joint inference gives good improvement

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

othe	0.05
[illegible]	0.50
loc	**0.45**

Dole 's wife, Elizabeth , is a na

E1 E2

R_{12} R_{23}

Key Questions:

How to guide the global inference?

How to learn the model(s)?

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

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

Joint Inference with General Constraint Structure [R

Recognizing Entities and Relations

Joint inference gives good improvement

other	0.05
per	**0.85**
[illegible]	0.10

other	0.10
per	**0.60**
loc	0.30

othe	0.05
p	0.50
loc	**0.45**

Key Questions:

global inference?

An Objective function that incorporates learned models with knowledge (output constraints)

A Constrained Conditional Model

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

i	[illegible]
spouse_of	0.05
born_in	**0.85**

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

Constrained Conditional Models

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}} \; \mathbf{w}^{\mathbf{T}} \phi(\mathbf{x}, \mathbf{y})$$

Constrained Conditional Models

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^{\mathbf{T}} \phi(\mathbf{x}, \mathbf{y})$$

Constrained Conditional Models

$$\mathbf{y = argmax_{y \in \mathcal{Y}}\ w^T \phi(x, y)}$$

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

Constrained Conditional Models

$$\mathbf{y = argmax_{y \in \mathcal{Y}}\ w^T \phi(x, y)}$$

Weight Vector for "local" models

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

Constrained Conditional Models

$$\mathbf{y = argmax_{y \in \mathcal{Y}}\ w^T \phi(x, y) + u^T \mathcal{C}(x, y)}$$

Knowledge component: (Soft) constraints

Weight Vector for "local" models

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

Constrained Conditional Models

Penalty for violating the constraint.

$$\mathbf{y = argmax_{y \in \mathcal{Y}}\ w^T \phi(x, y) + u^T \mathcal{C}(x, y)}$$

Knowledge component: (Soft) constraints

Weight Vector for "local" models

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

How far y is from a "legal/expected" assignment

Constrained Conditional Models

Penalty for violating the constraint.

$$\mathbf{y = argmax_{y \in \mathcal{Y}}\ w^T \phi(x, y) + u^T \mathcal{C}(x, y)}$$

Knowledge component: (Soft) constraints

Weight Vector for “local” models

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

How far y is from a “legal/expected” assignment

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

Constrained Conditional Models

Penalty for violating the constraint.

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}} \; \mathbf{w}^{\mathbf{T}} \phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^{\mathbf{T}} \mathcal{C}(\mathbf{x}, \mathbf{y})$$

Knowledge component: (Soft) constraints

Weight Vector for "local" models

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

How far y is from a "legal/expected" assignment

- 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

Penalty for violating the constraint.

$$y = \text{argmax}_y \sum \mathbf{1}_{\phi(x,y)}\, w_{x,y} \quad \text{subject to Constraints } C(x,y)$$

Knowledge component: (Soft) constraints

Weight Vector for "local" models

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

How far y is from a "legal/expected" assignment

- 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

Any MAP problem w.r.t. any probabilistic model, can be formulated as an ILP [Roth+ 04, Taskar 04]

$y = \mathrm{argmax}_y \sum \mathbf{1}_{\phi(x,y)} w_{x,y}$ subject to Constraints $C(x,y)$

Knowledge component: (Soft) constraints

Weight Vector for "local" models

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

How far y is from a "legal/expected" assignment

- 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

Any MAP problem w.r.t. any probabilistic model, can be formulated as an ILP [Roth+ 04, Taskar 04]

Variables are models

$$y = \text{argmax}_y \sum \mathbf{1}_{\phi(x,y)}\, w_{x,y} \quad \text{subject to Constraints } C(x,y)$$

Knowledge component: (Soft) constraints

Weight Vector for "local" models

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

How far y is from a "legal/expected" assignment

- 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

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}} \ \mathbf{w}^T \phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^T \mathcal{C}(\mathbf{x}, \mathbf{y})$$

Examples: CCM Formulations

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}} \; \mathbf{w}^T \phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^T \mathcal{C}(\mathbf{x}, \mathbf{y})$$

While ϕ**(x, y) and** $\mathcal{C}$**(x, y)** could be the same; we want C(x, y) to express high level declarative knowledge over the statistical models.

Examples: CCM Formulations

$$\mathbf{y = argmax_{y \in \mathcal{Y}}\ w^T\phi(x, y) + u^T\mathcal{C}(x, y)}$$

While ϕ**(x, y) and** $\mathcal{C}$**(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

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^{\mathrm{T}}\mathcal{C}(\mathbf{x}, \mathbf{y})$$

While ϕ**(x, y) and** $\mathcal{C}$**(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:

Argmax $\sum \lambda_{ij}$ x_{ij}

Knowledge/Linguistics Constraints

Cannot have both A states and B states in an output sequence.

Examples: CCM Formulations

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^{\mathsf{T}}\phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^{\mathsf{T}}C(\mathbf{x}, \mathbf{y})$$

While ϕ**(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:

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

$$\mathbf{y = argmax_{y \in \mathcal{Y}}\ w^T\phi(x, y) + u^T\mathcal{C}(x, y)}$$

While ϕ**(x, y) and** $\mathcal{C}$**(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:

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

$$\mathbf{y} = \mathbf{argmax}_{\mathbf{y} \in \mathcal{Y}}\ \mathbf{w}^T\phi(\mathbf{x}, \mathbf{y}) + \mathbf{u}^T\mathcal{C}(\mathbf{x}, \mathbf{y})$$

While ϕ**(x, y) and** $\mathcal{C}$**(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]$_{A0}$ *left* [my pearls]$_{A1}$ [to my daughter]$_{A2}$ [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)

Archetypical Information Extraction Problem: E.g., Concept Identification and Typing, Event Identification, etc.

I *left* my pearls to my daughter in my will .

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

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

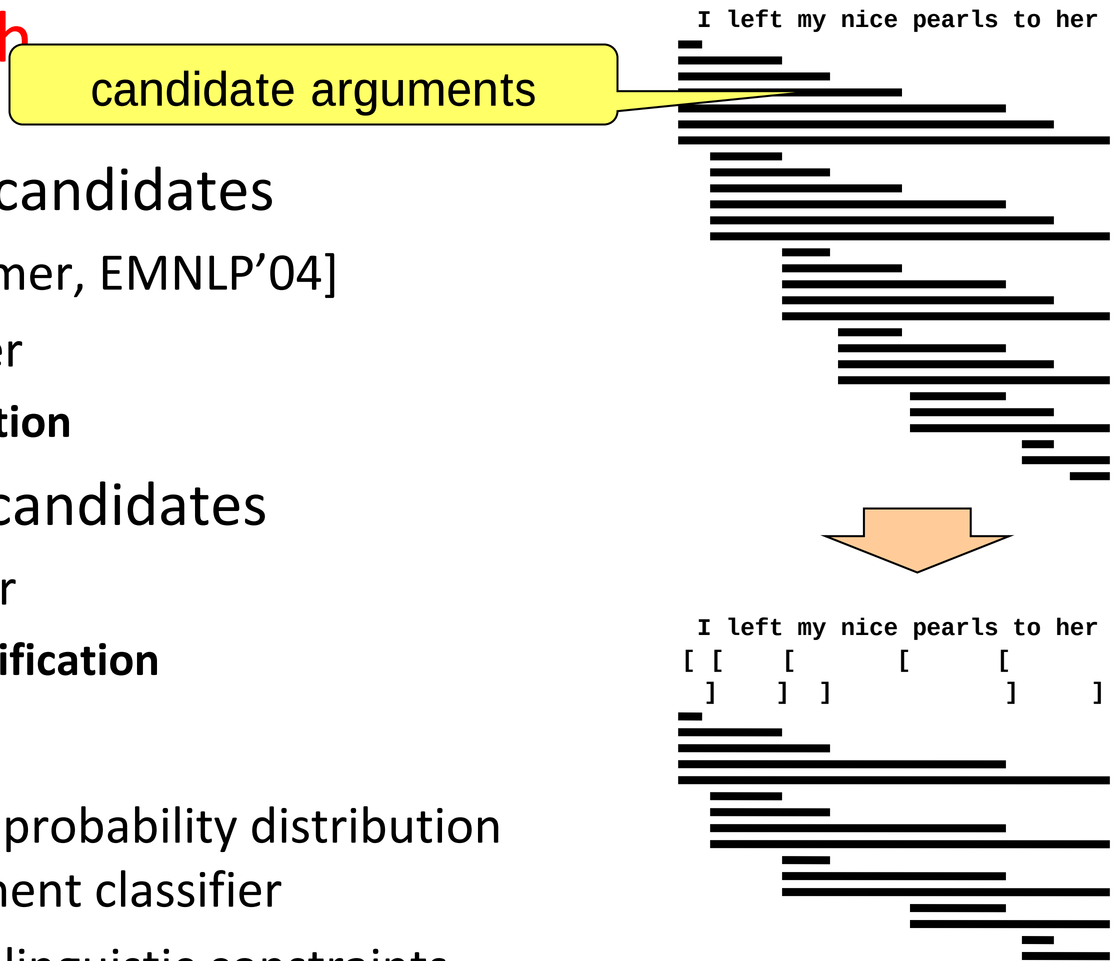
I *left* my pearls to my daughter in my will .

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

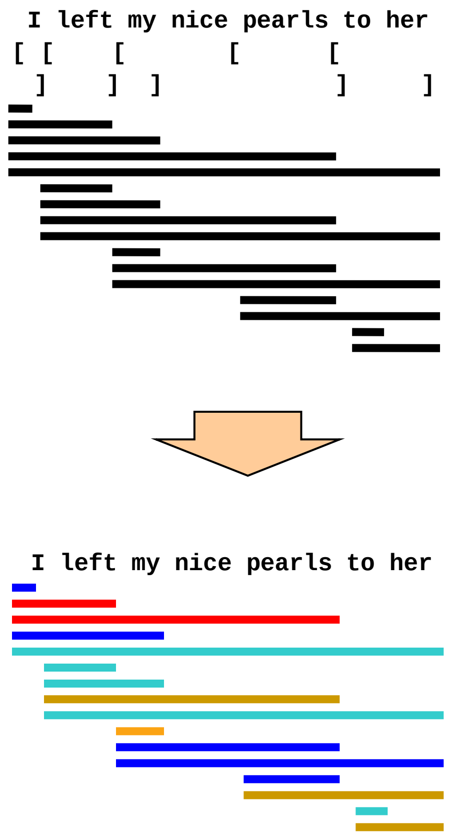
Algorithmic Approach

- 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



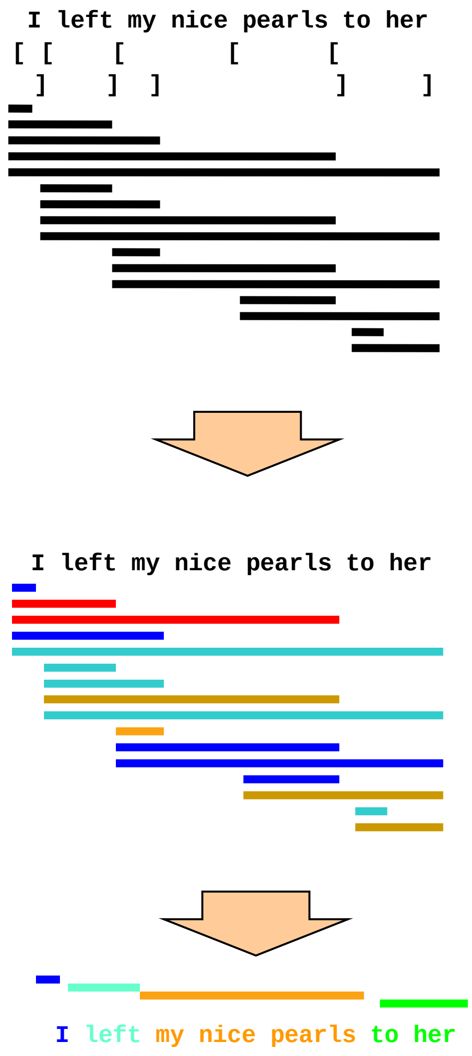
Algorithmic Approach

- 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

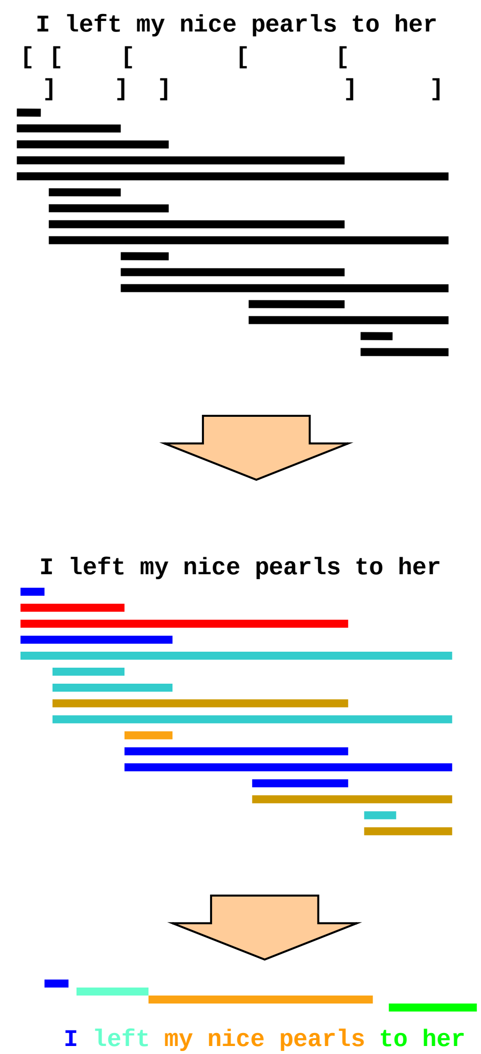
- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - 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 st ic constraints
  - Infer t utput

One inference problem for each verb predicate.



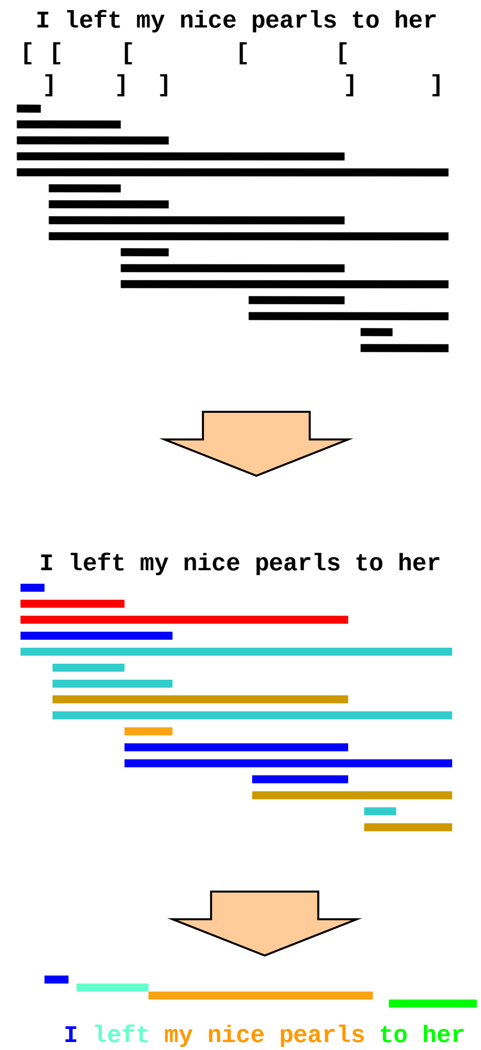
Algorithmic Approach

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

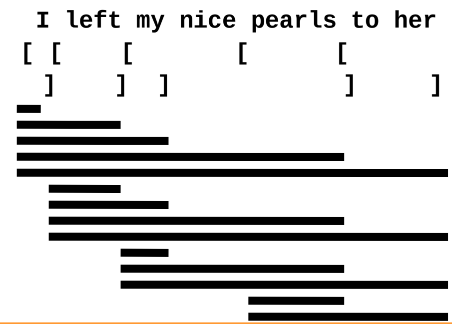
argmax $\sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$

Subject to:

- One label per argument: $\sum_{t} y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,….



Algorithmic Approach



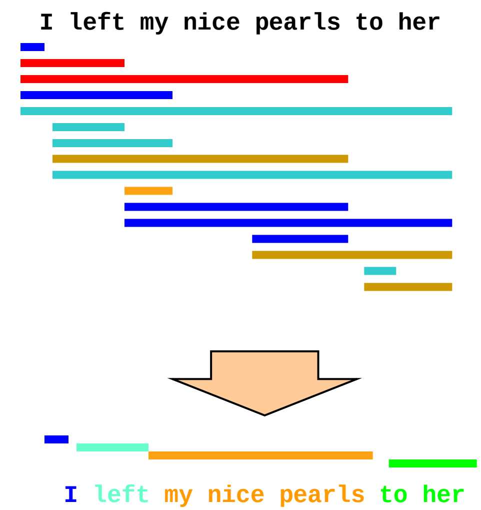
- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - **Binary classification**
- Classify argument candidates
  - Argument Classifier
    - **Multi-class classification**
- Inference

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

argmax $\sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$

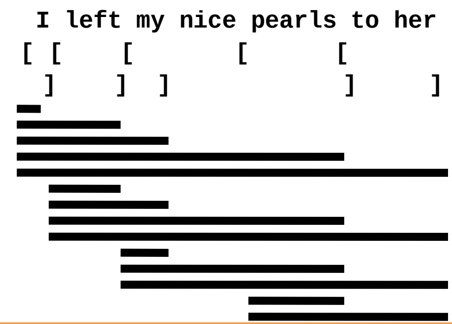
Subject to:

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



Algorithmic Approach

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



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

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

Subject to:

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

Abstract representation of expectations/knowledge

Algorithmic Approach

- Identify argument candid
  - Pruning [Xue&Palmer, EM
  - Argument Identifier
    - **Binary classification**
- Classify argument candid
  - Argument Classifier
    - **Multi-class classification**
- Inference

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

Unique labels

$$\forall y \in \mathcal{Y}, \sum_{i=0}^{n-1} 1_{\{y_i = y\}} \leq 1$$

$$\forall y \in \mathcal{Y}_R, \sum_{i=0}^{n-1} 1_{\{y_i = y = \text{“R-Ax”}\}} \leq \sum_{i=0}^{n-1} 1_{\{y_i = \text{“Ax”}\}}$$

$$\forall j, y \in \mathcal{Y}_C, 1_{\{y_j = y = \text{“C-Ax”}\}} \leq \sum_{i=0}^{j} 1_{\{y_i = \text{“Ax”}\}}$$

argmax $\sum_{a,t} y^{a,t} c^{a,t} = \sum_{a,t} 1_{a=t} c_{a=t}$

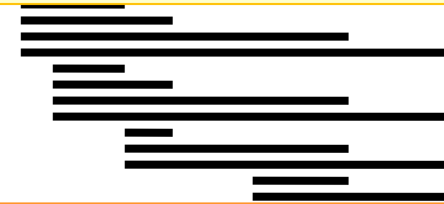
Subject to:

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

Abstract representation of expectations/knowledge

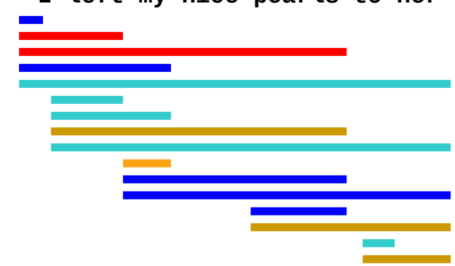
Algorithmic Approach

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



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

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



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

Subject to:

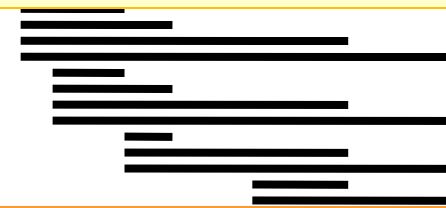
- One label per argument: $\sum_t y^{a,t} = 1$
- No overlapping or embedding
- Relations between verbs and arguments,….

Abstract representation of expectations/knowledge

Algorithmic Approach

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

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



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

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

Subject to:

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

Abstract representation of expectations/knowledge

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

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.

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.

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.

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.

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.

- 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 [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]
- Hard Co-reference Problems [NAACL'15]

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

All together, the outcome is the best end-to-end coreference results on CoNLL data and on ACE [CoNLL'15]

- 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

All together, the outcome is the best end-to-end coreference results on CoNLL data and on ACE [CoNLL'15]

- 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

Knowledge representation called "**predicate schemas**"

- 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"**

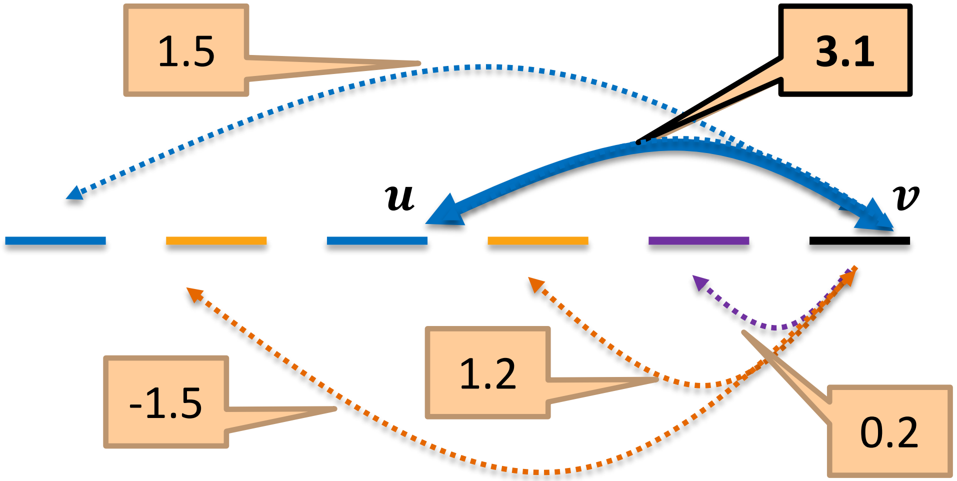
Hard Co-reference Problems

Knowledge representation called "**predicate schemas**"

- 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

- $y = \arg\max_y \sum_{uv} \mathrm{w_{uv}} \cdot y_{uv}$
  s.t $\sum_{u<v} yuv \;<= 1,\ \forall \mathrm{v}$
  $y_{uv}\ \varepsilon\ \{0,1\}$

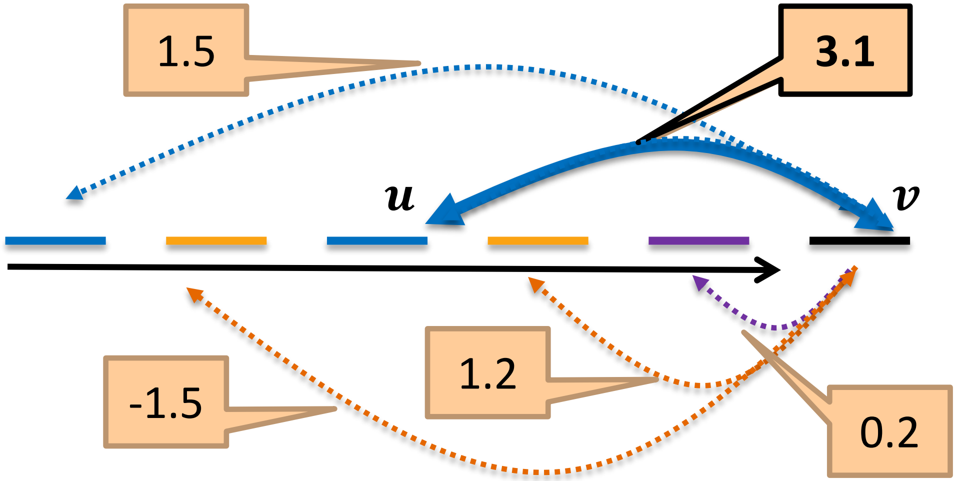


ILP Formulation of Coreference Resolution

- $y = \arg\max_y \sum_{uv} \mathrm{w}_{\mathrm{uv}} \cdot y_{uv}$

  s.t $\sum_{u < v} yuv <= 1, \ \forall \mathrm{v}$

  $y_{uv} \ \varepsilon \ \{0,1\}$

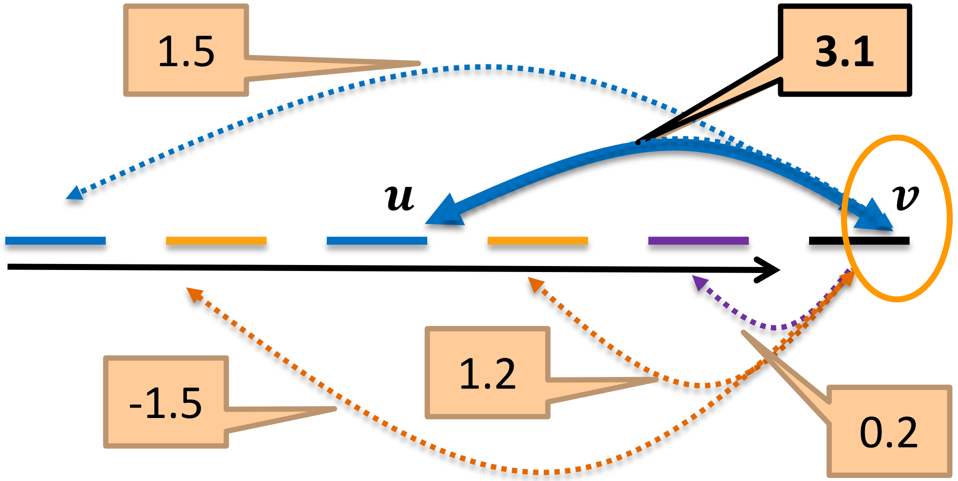


ILP Formulation of Coreference Resolution

- $y = \arg\max_y \sum_{uv} \mathrm{w}_{\mathrm{uv}} \cdot y_{uv}$

  s.t $\sum_{u < v} yuv \;$ <= 1, $\forall$v

  $y_{uv} \;\varepsilon\;$ {0,1}



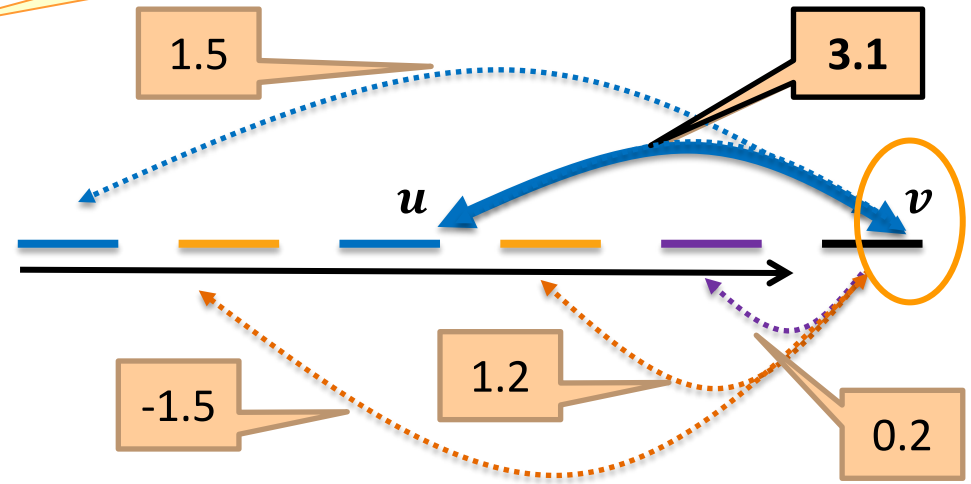
ILP Formulation of Coreference Resolution

Variable $\boldsymbol{y_{uv}}$ indicates a coreference link u→v

- $y = \arg\max_y \sum_{uv} \mathrm{w}_{\mathrm{uv}} \cdot y_{uv}$

  s.t $\sum_{u<v} yuv$ <= 1, $\forall$v

  y_{uv} ε {0,1}



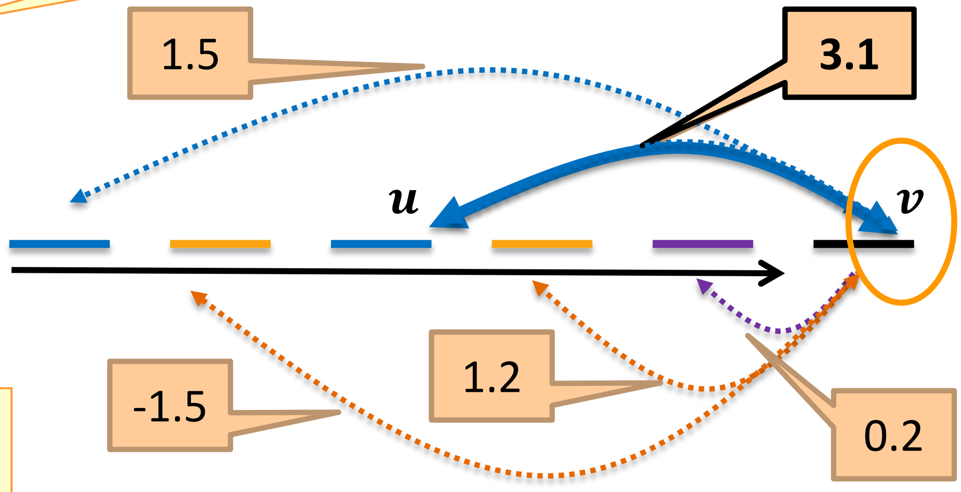
ILP Formulation of Coreference Resolution

Variable y_{uv} indicates a coreference link u→v

- $y = \arg\max_y \sum_{uv} \mathrm{w_{uv}} \cdot y_{uv}$

  s.t $\sum_{u<v} yuv \;<= 1, \;\forall$v

  $y_{uv} \;\varepsilon \;\{0,1\}$



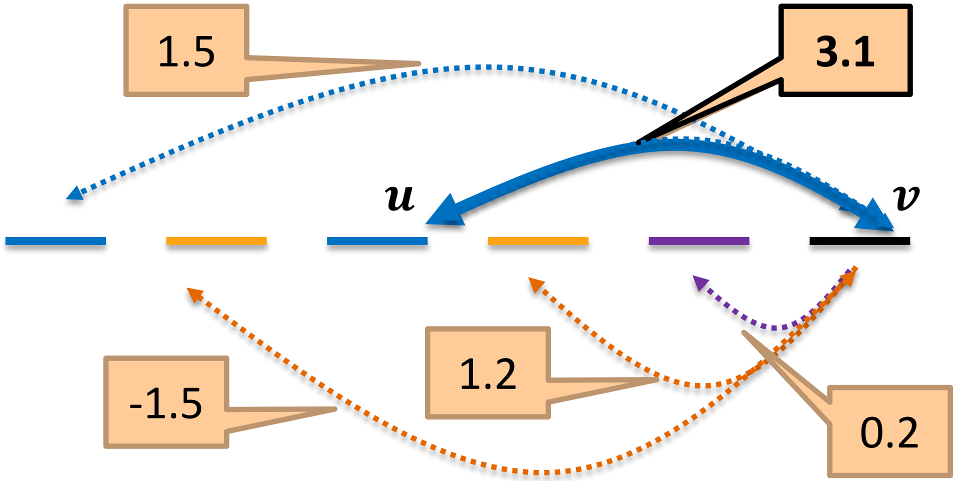
Best Link Approach: only one of the antecedents **u** is linked to **v**

ILP Formulation of Coreference Resolution

- $y = \arg\max_y \sum_{uv} \mathrm{w}_{\mathrm{uv}} \cdot y_{uv}$

  s.t $\sum_{u < v} yuv < 1, \ \forall \mathrm{v}$

  $y_{uv} \ \varepsilon \ \{0,1\}$

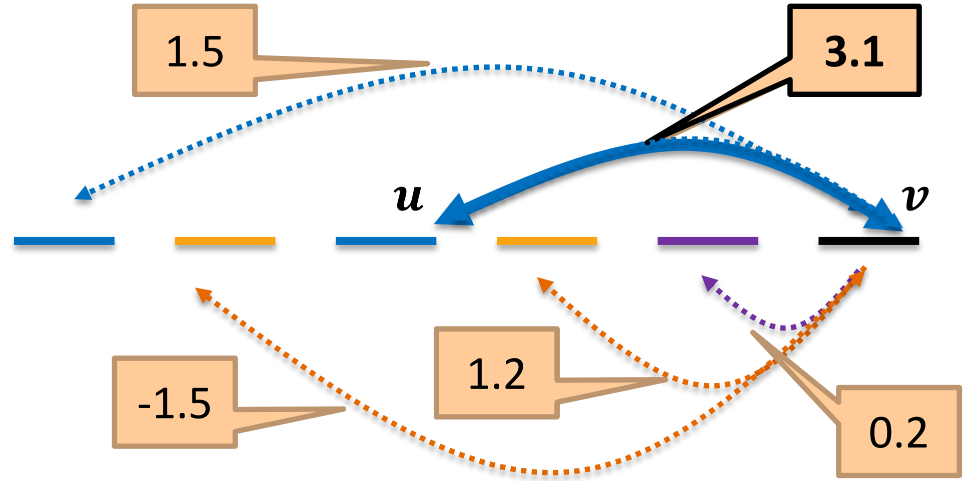


ILP Formulation of Coreference Resolution

- $y = \arg\max_y \sum_{uv} \mathrm{w}_{\mathrm{uv}} \cdot y_{uv}$

  s.t $\sum_{u<v} yuv < 1, \ \forall \mathrm{v}$

  $y_{uv} \ \varepsilon \ \{0,1\}$



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

ILP Formulation of Coreference Resolution

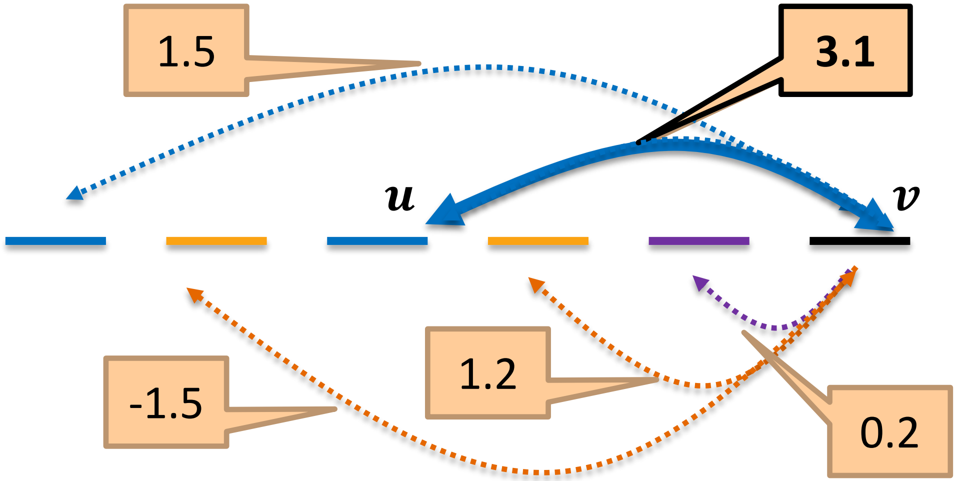
- $y = \arg\max_y \sum_{uv} \mathrm{w}_{\mathrm{uv}} \cdot y_{uv}$

  s.t $\sum_{u<v} yuv < 1, \ \forall v$

  $y_{uv} \ \varepsilon \ \{0,1\}$

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

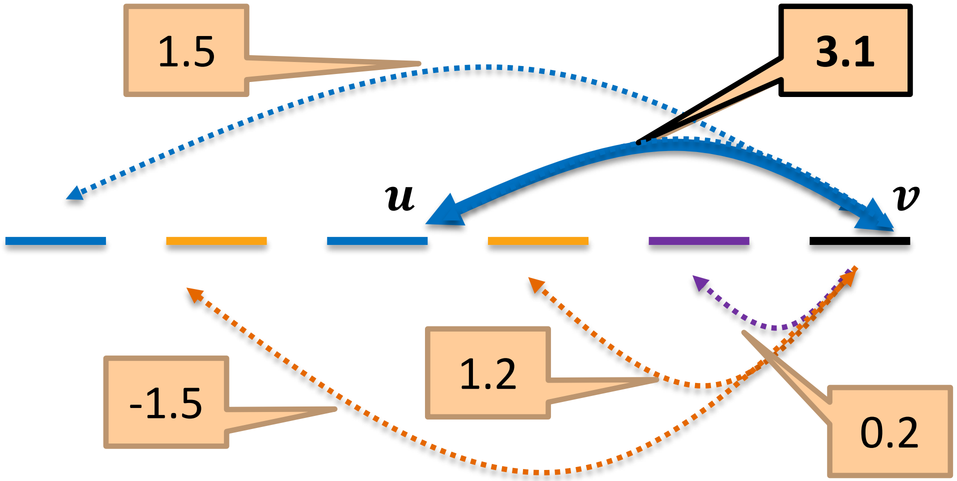
- $y = \arg\max_y \sum_{uv} \mathrm{w}_{\mathrm{uv}} \cdot y_{uv}$

  s.t $\sum_{u\ <\ v} yuv\ < 1,\ \forall \mathrm{v}$

  $y_{uv}\ \varepsilon\ \{0,1\}$

  predicate schemas

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



- 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.

COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

ILP Formulation of Coreference Resolution

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

- $y = \arg\max_y \sum_{uv} \mathrm{w}_{\mathrm{uv}} \cdot y_{uv}$

  s.t $\sum_{u < v} yuv < 1, \ \forall \mathrm{v}$

  $y_{uv} \ \varepsilon \ \{0,1\}$

  predicate schemas

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

- 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: $E =\sim\sim 3 \sum_i o_i$

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: $E =\sim\sim 3 \sum_i o_i$

  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: $E =\sim\sim 3 \sum_i o_i$

  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: $E =\sim\sim 3 \sum_i o_i$

  share it with

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

Mapping Natural Language Text to Expressions

- *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?*

Mapping Natural Language Text to Expressions

- *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*

Mapping Natural Language Text to Expressions

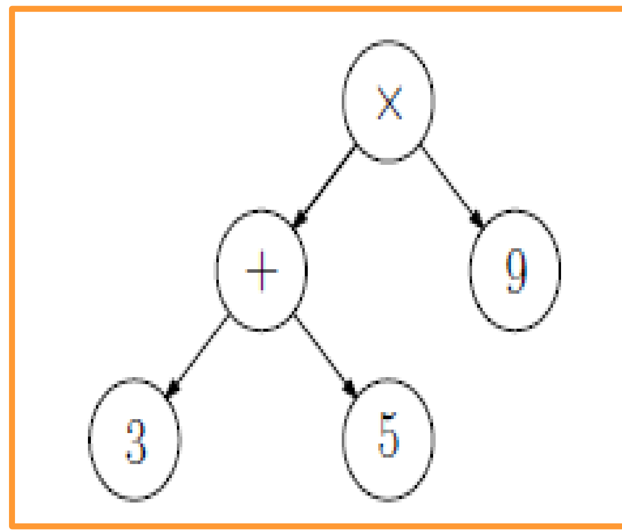
- *Gwen was organizing her book case making sure each of the shelves had exactly 9 books on it. She has ~~2~~ types of books – mystery books and picture books. If she had 3 shelves of mystery books and 5 shelves of picture books, how many books did she have total?*

- *[Roy & Roth'15] suggests a solution that involves "parsing" the problem into an expression tree*

Mapping Natural Language Text to Expressions

- *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 $\mathcal{T}$(E) implies that it is determined by the unique $\mathcal{T}$–operation between pairs of relevant quantities.

$$E^* = \text{argmax} \sum_q R(q)\, \mathbf{1}_{\mathbf{q}} + \square_{(q,\, q')}\, \text{Pair}(q, q', \odot(q, q'))\, \mathbf{1}_{\mathbf{q,q'}}$$

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.

Score of q being irrelevant to E

$$E^* = \operatorname{argmax} \sum_q R(q)\, \mathbf{1}_{\mathbf{q}} + \square_{(q,\, q')} \mathrm{Pair}(q, q', \odot(q, q'))\, \mathbf{1}_{\mathbf{q,q'}}$$

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.

Score of q being irrelevant to E

Score of $\odot$ being the unique operation between (q_i, q_j)

$$E^* = \text{argmax} \sum_q R(q)\ \mathbf{1}_{\mathbf{q}} + \square_{(q,\ q')}\ \text{Pair}(q,\ q',\ \odot(q,\ q'))\ \mathbf{1}_{\mathbf{q,q'}}$$

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.

Score of q being irrelevant to E

Score of $\odot$ being the unique operation between (q_i, q_j)

$$E^* = \text{argmax} \sum_q R(q)\, \mathbf{1}_{\mathbf{q}} + \square_{(q,\, q')} \text{Pair}(q, q', \odot(q, q'))\, \mathbf{1}_{\mathbf{q,q'}}$$

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

Inferring the Best Expression Tree

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

Score of q being irrelevant to E

Score of $\odot$ being the unique operation between (q_i, q_j)

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

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

Abstract Expectations developed given a text snippet

Inferring the Best Expression Tree

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

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

Score of q being irrelevant to E

Score of $\odot$ being the unique operation between (q_i, q_j)

$$E^* = \text{argmax} \sum_q R(q)\, \mathbf{1}_{\mathbf{q}} + \square_{(q,\, q')} \text{Pair}(q, q', \odot(q, q'))\, \mathbf{1}_{\mathbf{q,q'}}$$

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

Abstract Expectations developed given a text snippet

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, …

Thank You!

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, …