Natural Language Inference for Question Answering

Gabor Angeli

Stanford University

April 14, 2015

My View on Question Answering

Two parts to question answering:

- Understanding the question
- Understanding the corpus (IE)

My View on Question Answering

Two parts to question answering:

- Understanding the question
  - Most Semantic Parsing work seems to be here
  - Can be arbitrarily hard; we only worry about easy factoids.
- Understanding the corpus (IE)

My View on Question Answering

Two parts to question answering:

- Understanding the question
  - Most Semantic Parsing work seems to be here
  - Can be arbitrarily hard; we only worry about easy factoids.
- Understanding the corpus (IE)

My View on Question Answering

Two parts to question answering:

- Understanding the question
  - Most Semantic Parsing work seems to be here
  - Can be arbitrarily hard; we only worry about easy factoids.
- Understanding the corpus (IE)
  - What facts are in the corpus?
    [Angeli et al., 2014a, Angeli et al., 2014b]

My View on Question Answering

Two parts to question answering:

- Understanding the question
  - Most Semantic Parsing work seems to be here
  - Can be arbitrarily hard; we only worry about easy factoids.
- Understanding the corpus (IE)
  - What facts are in the corpus?
    [Angeli et al., 2014a, Angeli et al., 2014b]
  - What facts are entailed by the corpus?
    [ACL OpenIE Submission]

My View on Question Answering

Two parts to question answering:

- Understanding the question
  - Most Semantic Parsing work seems to be here
  - Can be arbitrarily hard; we only worry about easy factoids.
- Understanding the corpus (IE)
  - What facts are in the corpus?
    [Angeli et al., 2014a, Angeli et al., 2014b]
  - What facts are entailed by the corpus?
    [ACL OpenIE Submission]
  - How do we find if a fact is entailed by the corpus?
    [Angeli and Manning, 2013, Angeli and Manning, 2014]

Challenges in IE

Questions don't always fit into a known schema

Challenges in IE

Questions don't always fit into a known schema

✓ *Where was Barack Obama born?*

Challenges in IE

Questions don't always fit into a known schema

- ✓ *Where was Barack Obama born?*
- ? *What city borders the largest river in California?*
- ? *Who is Chris Manning's academic advisor?*

Challenges in IE

Questions don't always fit into a known schema

- ✓ *Where was Barack Obama born?*
- ? *What city borders the largest river in California?*
- ? *Who is Chris Manning's academic advisor?*
- ✗ *Do cats have tails?*

Challenges in IE

Questions don't always fit into a known schema

- ✓ *Where was Barack Obama born?*
- ? *What city borders the largest river in California?*
- ? *Who is Chris Manning's academic advisor?*
- ✗ *Do cats have tails?*

Questions don't always talk about named entities

- ✓ *Where was Barack Obama born?*

Challenges in IE

Questions don't always fit into a known schema

- ✓ *Where was Barack Obama born?*
- ? *What city borders the largest river in California?*
- ? *Who is Chris Manning's academic advisor?*
- ✗ *Do cats have tails?*

Questions don't always talk about named entities

- ✓ *Where was Barack Obama born?*
- ? *What treatments are there for diabetes?*
- ✗ *Do cats have tails?*

Challenges in IE

Questions don't always fit into a known schema

- ✓ *Where was Barack Obama born?*
- ? *What city borders the largest river in California?*
- ? *Who is Chris Manning's academic advisor?*
- ✗ *Do cats have tails?*

Questions don't always talk about named entities

- ✓ *Where was Barack Obama born?*
- ? *What treatments are there for diabetes?*
- ✗ *Do cats have tails?*

Answers aren't always literally justified in text

- ✓ *Born in Hawaii, Barack Obama ...*
- ? *Doctors prescribe Insulin for Diabetes.*
- ✗ *All animals have tails.*

No entailment: What facts are in the corpus?

We can do QA from Freebase ... so let's make Freebase bigger

Relation Extraction + TAC-KBP

No entailment: What facts are in the corpus?

We can do QA from Freebase ... so let's make Freebase bigger

Relation Extraction + TAC-KBP

- Done: KBP 2013 – 2014
- Done: KBP + active learning
- Pending: KBP + deep learning / matrix factorization

No entailment: What facts are in the corpus?

We can do QA from Freebase ... so let's make Freebase bigger

Relation Extraction + TAC-KBP

- Done: KBP 2013 – 2014
- Done: KBP + active learning
- Pending: KBP + deep learning / matrix factorization

Does it address the challenges?

- Fixed relation schema
- Very named-entity-centric
- Zero entailments from text

Mini-Entailment: List facts entailed by corpus

Extract maximally informative, correct things

Natural Logic + OpenIE (ACL submission)

Mini-Entailment: List facts entailed by corpus

Extract maximally informative, correct things

Natural Logic + OpenIE (ACL submission)

1. Split sentence up into short, entailed clauses
2. Strip away excess information in clauses
3. Extract open IE triples

Mini-Entailment: List facts entailed by corpus

Extract maximally informative, correct things

Natural Logic + OpenIE (ACL submission)

1. Split sentence up into short, entailed clauses
2. Strip away excess information in clauses
3. Extract open IE triples

Does it address the challenges?

- Open relation schema
- Named entity agnostic
- Only limited entailment from text

Lots of Entailment: Is a fact entailed by corpus?

Given a fact, can we support it in the corpus?

Natural Logic

Lots of Entailment: Is a fact entailed by corpus?

Given a fact, can we support it in the corpus?

Natural Logic

- Done: Natural Logic for Common Sense facts
- Pending: Natural Logic for Factoid facts (EMNLP)
- Future: … stay tuned

Lots of Entailment: Is a fact entailed by corpus?

Given a fact, can we support it in the corpus?

Natural Logic

- Done: Natural Logic for Common Sense facts
- Pending: Natural Logic for Factoid facts (EMNLP)
- Future: … stay tuned

Does it address the challenges?

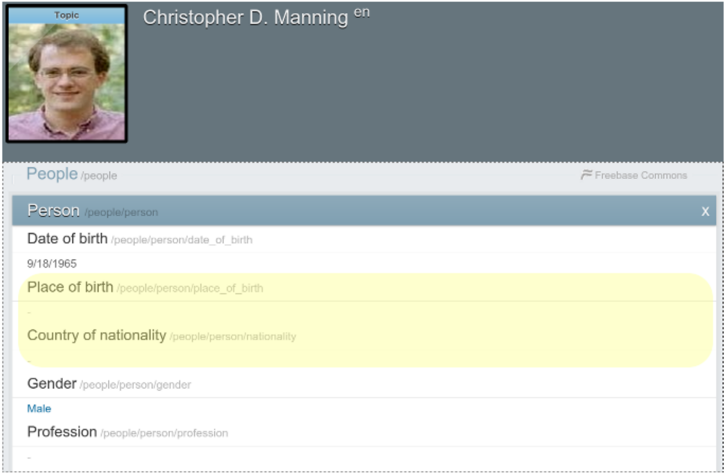
- Open relation schema
- Named entity agnostic
- Large class of entailments from corpus

Main Idea of Thesis

Natural language inference allows us to leverage latent information in text to improve recall of open-domain facts, without sacrificing [much] precision.

KBP: Directly find facts in text

KBP: Directly find facts in text



KBP: Directly find facts in text

Christopher Manning

Professor of Linguistics and Computer Science

Natural Language Processing Group, Stanford University

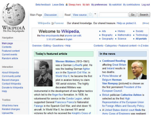
Brief Bio

- I'm Australian ("I come from a land of wide open spaces ...")
- BA (Hons) Australian National University 1989 (majors in mathematics, computer science and linguistics)
- PhD Stanford Linguistics 1995
- Asst Professor Carnegie Mellon University Computational Linguistics Program 1994-96
- Lecturer University of Sydney Dept of Linguistics 1996-99
- Asst Professor Stanford University Depts of Computer Science and Linguistics 1999-2006
- Assoc Professor Stanford University Depts of Linguistics and Computer Science 2006-2012
- Professor Stanford University Depts of Linguistics and Computer Science 2012-

KBP: Directly find facts in text

We’re filling a fixed relation schema

Unstructured Text



$\Longrightarrow$

Structured Knowledge Base

Barack Obama

44th President of the United States

Personal details	
Born	Barack Hussein Obama II August 4, 1961 (age 52) Honolulu, Hawaii, U.S.
Political party	Democratic
Spouse(s)	Michelle LaVaughn Robinson (m. 1992–present)
Children	Malia Ann Obama (b. 1998) Natasha Obama (b. 2001)

Active Learning for Relation Extraction

Input: Sentences containing (entity, slot value).
Output: Relation between entity and slot value.

Active Learning for Relation Extraction

Input: Sentences containing (entity, slot value).
Output: Relation between entity and slot value.

Consider two approaches:

- **Supervised:** Trivial as a supervised classifier.
  Training data: {(sentence, relation)}.
  But…

Active Learning for Relation Extraction

Input: Sentences containing (entity, slot value).
Output: Relation between entity and slot value.

Consider two approaches:

- **Supervised:** Trivial as a supervised classifier.
  Training data: {(sentence, relation)}.
  But… this training data is expensive to produce.

Active Learning for Relation Extraction

Input: Sentences containing (entity, slot value).
Output: Relation between entity and slot value.

Consider two approaches:

- **Supervised:** Trivial as a supervised classifier.
  Training data: {(sentence, relation)}.
  But… this training data is expensive to produce.

- **Distantly Supervised:** Artificially produce “supervised” data.
  Training data: {(entity, relation, slot value)}.
  But…

Active Learning for Relation Extraction

Input: Sentences containing (entity, slot value).
Output: Relation between entity and slot value.

Consider two approaches:

- **Supervised:** Trivial as a supervised classifier.
  Training data: {(sentence, relation)}.
  But... this training data is expensive to produce.

- **Distantly Supervised:** Artificially produce "supervised" data.
  Training data: {(entity, relation, slot value)}.
  But... this training data is much more noisy.

Active Learning: Combine Benefits of Both

Adding carefully selected supervision improves distantly supervised relation extraction.

Active Learning: Combine Benefits of Both

Adding carefully selected supervision improves distantly supervised relation extraction.

Old problem: Supervision is expensive, but very useful.

Old solution: Active learning!

Active Learning: Combine Benefits of Both

Adding carefully selected supervision improves distantly supervised relation extraction.

Old problem: Supervision is expensive, but very useful.

Old solution: Active learning!

- Select a subset of latent z to annotate.
- Fix these labels during training.

Active Learning: Combine Benefits of Both

Adding carefully selected supervision improves distantly supervised relation extraction.

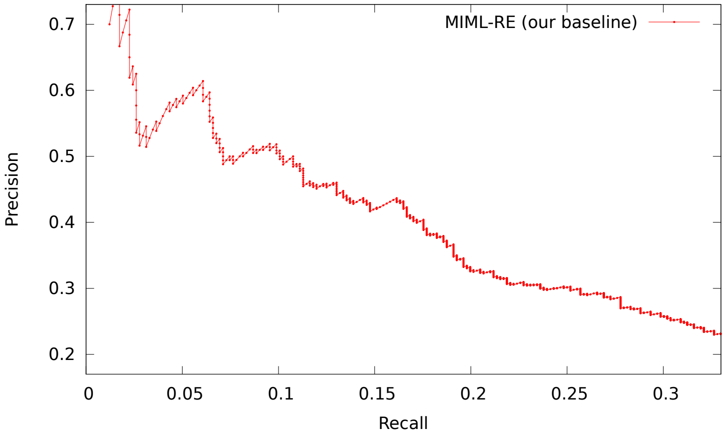
Old problem: Supervision is expensive, but very useful.

Old solution: Active learning!

- Select a subset of latent z to annotate.
- Fix these labels during training.
- Bonus: this creates a supervised training set.
    - We initialize from a supervised classifier on this training set.

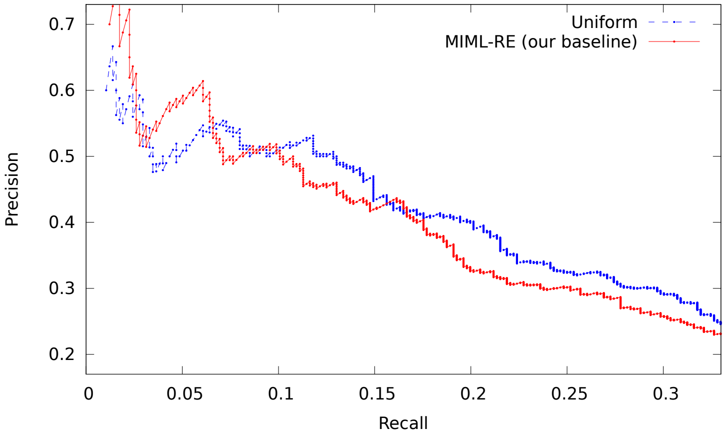
State-of-the-art results

Slot filling evaluation of Surdeanu et al. (2012).



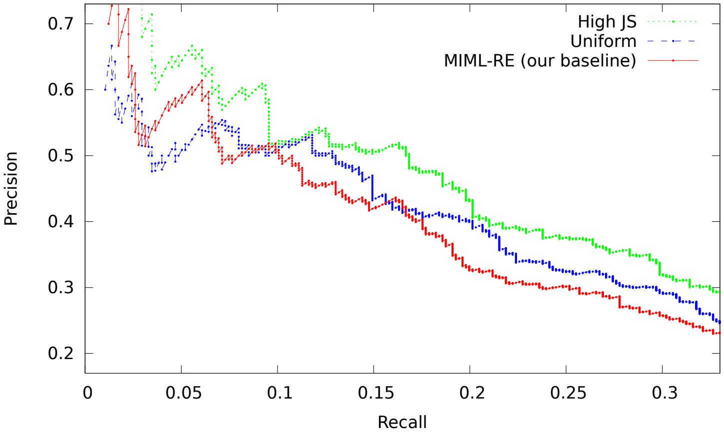
State-of-the-art results

Slot filling evaluation of Surdeanu et al. (2012).

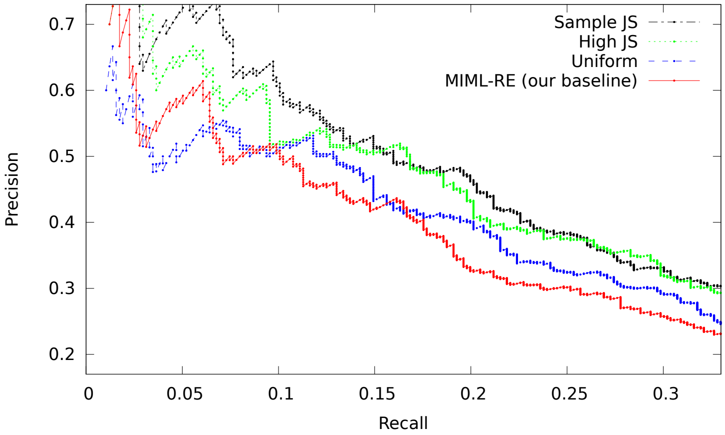


State-of-the-art results

Slot filling evaluation of Surdeanu et al. (2012).

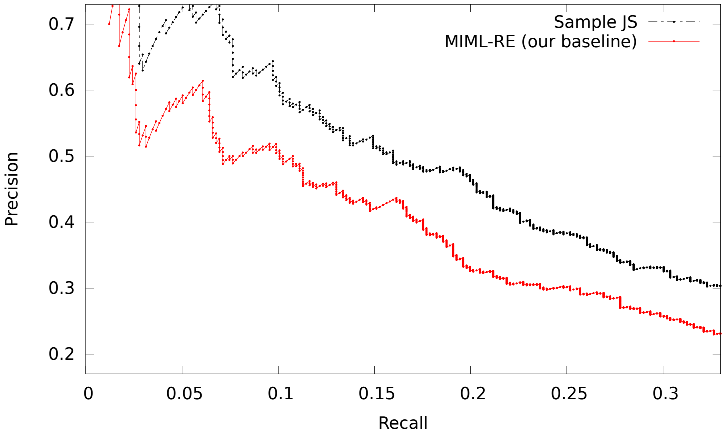


Slot filling evaluation of Surdeanu et al. (2012).

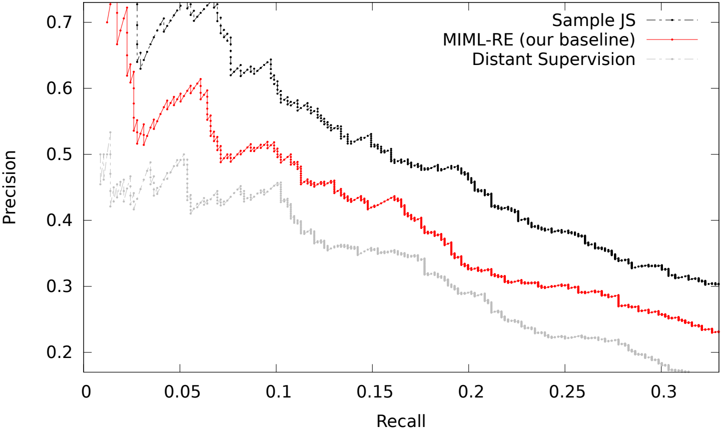


State-of-the-art results

Slot filling evaluation of Surdeanu et al. (2012).

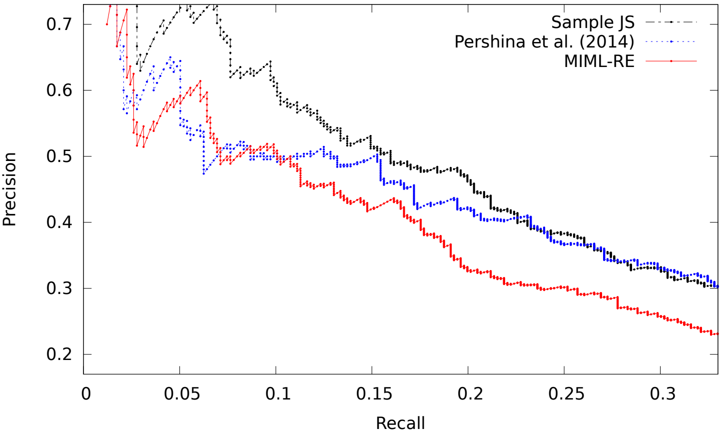


Slot filling evaluation of Surdeanu et al. (2012).



State-of-the-art results

Slot filling evaluation of Surdeanu et al. (2012).



KBP is already "entailment"

Chris, a tenured professor at Stanford, is friends with Fei-Fei

$\Rightarrow$ *Chris is a tenured professor at Stanford*

$\Rightarrow$ *Chris is a professor at Stanford*

$\Rightarrow$ *Chris is employed by Stanford*

$\Rightarrow$ (*Chris*; *employee_of*; *Stanford*)

KBP is already "entailment"

Chris, a tenured professor at Stanford, is friends with Fei-Fei

$\Rightarrow$ *Chris is a tenured professor at Stanford*

$\Rightarrow$ *Chris is a professor at Stanford*

$\Rightarrow$ *Chris is employed by Stanford*

$\Rightarrow$ (*Chris*; *employee_of*; *Stanford*)

Born in a small town, she took the midnight train going anywhere

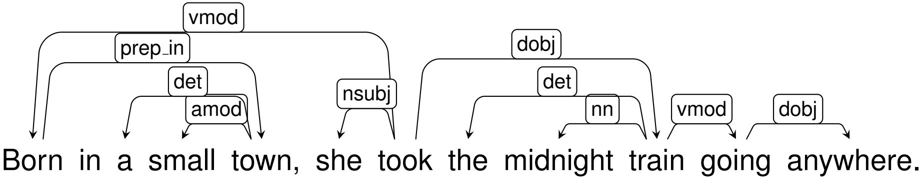
$\Rightarrow$ *She took the midnight train going anywhere*

$\Rightarrow$ *She took the midnight train*

$\Rightarrow$ *She took midnight train*

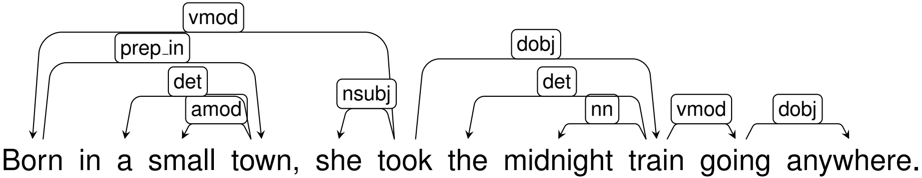
$\Rightarrow$ (*She*; *took*; *midnight train*)

Treat As Entailment



(input)

Treat As Entailment

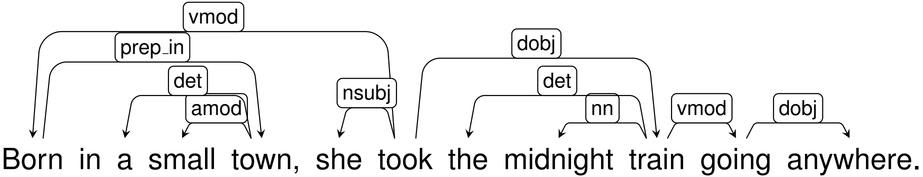


(input)

↓

she took the midnight train going anywhere	*she took the midnight train*
Born in a small town, she took the midnight train	***she took midnight train***
Born in a town, she took the midnight train	. . .

Treat As Entailment



(input)

↓

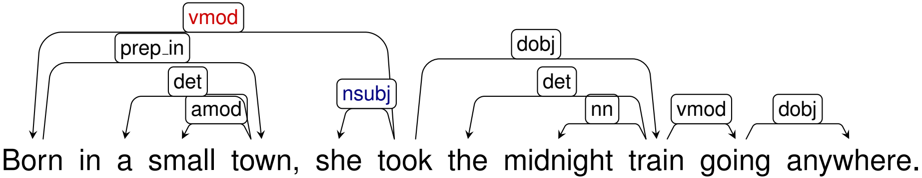
she took the midnight train going anywhere
Born in a small town, she took the midnight train
Born in a town, she took the midnight train

she took the midnight train
she took midnight train
...

↓

(she; took; midnight train)

Treat As Entailment



(input)

↓

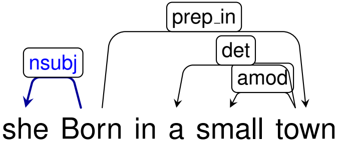
she took the midnight train going anywhere
Born in a small town, she took the midnight train
Born in a town, she took the midnight train

she took the midnight train
she took midnight train
. . .

↓

(she; took; midnight train)

Treat As Entailment



(extracted clause)

↓

she Born in small town

she Born in a town

she Born in town

↓

(she; born in; small town)

(she; born in; town)

OpenIE Summary

Chris, a tenured professor at Stanford, is friends with Fei-Fei
$\Rightarrow$ (*Chris*; *is friends with*; *Fei-Fei*)

1. Split a long sentence into short entailed sentences.
   (e.g., *Chris is a tenured professor at Stanford*)

OpenIE Summary

Chris, a tenured professor at Stanford, is friends with Fei-Fei
⇒ (*Chris*; *is friends with*; *Fei-Fei*)

1. Split a long sentence into short entailed sentences.
   (e.g., *Chris is a tenured professor at Stanford*)

2. Shorten each clause maximally
   ✓ *Chris is friends with Fei-Fei*
   ✓ *Chris is professor*
   ✗ *Chris is friends*

OpenIE Summary

Chris, a tenured professor at Stanford, is friends with Fei-Fei
$\Rightarrow$ (*Chris*; *is friends with*; *Fei-Fei*)

1. Split a long sentence into short entailed sentences.
   (e.g., *Chris is a tenured professor at Stanford*)

2. Shorten each clause maximally
   ✓ *Chris is friends with Fei-Fei*
   ✓ *Chris is professor*
   ✗ *Chris is friends*

3. Optionally segment into triples

State-of-the-art OpenIE

KBP 2013 end-to-end evaluation:

System	P	R	F_1
UW Official	**69.8**	11.4	19.6
Ollie	57.4	4.8	8.9
NaturalLI – Nominals	66.7	7.7	13.8
Ollie + Nominal Rels	56.8	12.1	19.9
NaturalLI	60.3	**14.3**	**23.1**

State-of-the-art OpenIE

KBP 2013 end-to-end evaluation:

System	**P**	**R**	$\mathbf{F_1}$
UW Official	**69.8**	11.4	19.6
Ollie	57.4	4.8	8.9
NaturalLI − Nominals	66.7	7.7	13.8
Ollie + Nominal Rels	56.8	12.1	19.9
NaturalLI	60.3	**14.3**	**23.1**

- If we add alternate names and websites: 28.6 F_1
- Context: NYU *trained on KBP*: 27.7 F_1

State-of-the-art OpenIE

KBP 2013 end-to-end evaluation:

System	**P**	**R**	**F_1**
UW Official	**69.8**	11.4	19.6
Ollie	57.4	4.8	8.9
NaturalLI – Nominals	66.7	7.7	13.8
Ollie + Nominal Rels	56.8	12.1	19.9
NaturalLI	60.3	**14.3**	**23.1**

- If we add alternate names and websites: 28.6 F_1
- Context: NYU *trained on KBP*: 27.7 F_1
- MIML-RE: 36.2 F_1; Top system: 40.2 F_1

Natural Logic subsumes Syllogisms

Some cat ate a mouse

(all mice are rodents)

$\therefore$ *Some cat ate a* ***rodent***

Natural Logic

Natural Logic subsumes Syllogisms

Some cat ate a mouse
(all mice are rodents)
∴ *Some cat ate a **rodent***

Cognitively easy inferences are easy:

- Most cats eat mice
  ∴ *Most cats eat **rodents***

Natural Logic

Natural Logic subsumes Syllogisms

Some cat ate a mouse
(all mice are rodents)
∴ *Some cat ate a **rodent***

Cognitively easy inferences are easy:

- Most cats eat mice
  ∴ *Most cats eat **rodents***
- "*All students who know a foreign language learned it at university.*"

Natural Logic

Natural Logic subsumes Syllogisms

Some cat ate a mouse
(all mice are rodents)
∴ *Some cat ate a **rodent***

Cognitively easy inferences are easy:

- Most cats eat mice
  ∴ *Most cats eat **rodents***
- "*All students who know a foreign language learned it at university.*"
  ∴ *"They learned it at school."*

Natural Logic

Natural Logic subsumes Syllogisms

Some cat ate a mouse
(all mice are rodents)
$\therefore$ *Some cat ate a **rodent***

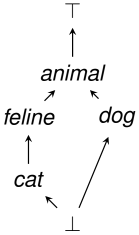
Cognitively easy inferences are easy:

- Most cats eat mice
  $\therefore$ *Most cats eat **rodents***
- "*All students who know a foreign language learned it at university.*"
  $\therefore$ *"They learned it at school."*

Facts are text; inference is lexical mutation

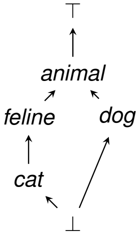
Natural Logic and Polarity

Treat hypernymy as a *partial order*.



Natural Logic and Polarity

Treat hypernymy as a *partial order*.



***Polarity* is the direction a lexical item can move in the ordering.**

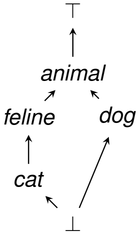
animal

feline

cat

house cat

Treat hypernymy as a *partial order*.



***Polarity* is the direction a lexical item can move in the ordering.**

animal

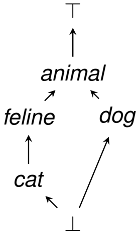
feline

$\boxed{\uparrow \text{cat}}$

house cat

Natural Logic and Polarity

Treat hypernymy as a *partial order*.



***Polarity* is the direction a lexical item can move in the ordering.**

living thing

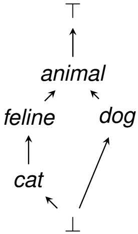
animal

$\boxed{\uparrow \text{feline}}$

cat

Natural Logic and Polarity

Treat hypernymy as a *partial order*.



***Polarity* is the direction a lexical item can move in the ordering.**

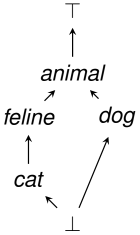
thing

living thing

$\boxed{\uparrow \text{animal}}$

feline

Treat hypernymy as a *partial order*.



***Polarity* is the direction a lexical item can move in the ordering.**

thing

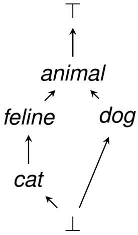
living thing

$\boxed{\downarrow \text{animal}}$

feline

Natural Logic and Polarity

Treat hypernymy as a *partial order*.



***Polarity* is the direction a lexical item can move in the ordering.**

living thing

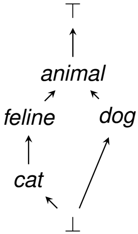
animal

$\boxed{\downarrow \text{feline}}$

cat

Natural Logic and Polarity

Treat hypernymy as a *partial order*.



***Polarity* is the direction a lexical item can move in the ordering.**

animal

feline

$\boxed{\downarrow \text{cat}}$

house cat

An Example Inference

Quantifiers determines the *polarity* (↑ or ↓) of words.

	carnivores		placentals
	felines	consume	rodents
↑ **All**$_{\downarrow\uparrow}$	↓ cats	↑ eat	↑ mice
	house cats	slurp	fieldmice

An Example Inference

Quantifiers determines the *polarity* ($\uparrow$ or $\downarrow$) of words.

Mutations must respect *polarity*.

	felines		placentals
	cats	consume	rodents
$\uparrow$ **All**$_{\downarrow\uparrow}$	$\downarrow$ house cats	$\uparrow$ eat	$\uparrow$ mice
	kitties	slurp	fieldmice

An Example Inference

Quantifiers determines the *polarity* (↑ or ↓) of words.

Mutations must respect *polarity*.

	felines		placentals
	cats		rodents
↑ $\textbf{All}_{\downarrow\uparrow}$	↓ house cats	↑ consume	↑ mice
	kitties	eat	fieldmice

An Example Inference

Quantifiers determines the *polarity* (↑ or ↓) of words.

Mutations must respect *polarity*.

	felines		rodents
	cats		mice
↑ **All**$_{\downarrow\uparrow}$	↓ house cats	↑ consume	↑ fieldmice
	kitties	eat	pine vole

An Example Inference

Quantifiers determines the *polarity* (↑ or ↓) of words.

Mutations must respect *polarity*.

	felines		placentals
	cats		rodents
$\uparrow$ **All**$_{\downarrow\uparrow}$	$\downarrow$ house cats	$\uparrow$ consume	$\uparrow$ mice
	kitties	eat	fieldmice

An Example Inference

Quantifiers determines the *polarity* (↑ or ↓) of words.

Mutations must respect *polarity*.

	felines		mammals
	cats		placentals
↑ **All**$_{\downarrow\uparrow}$	↓ house cats	↑ consume	↑ rodents
	kitties	eat	mice

An Example Inference

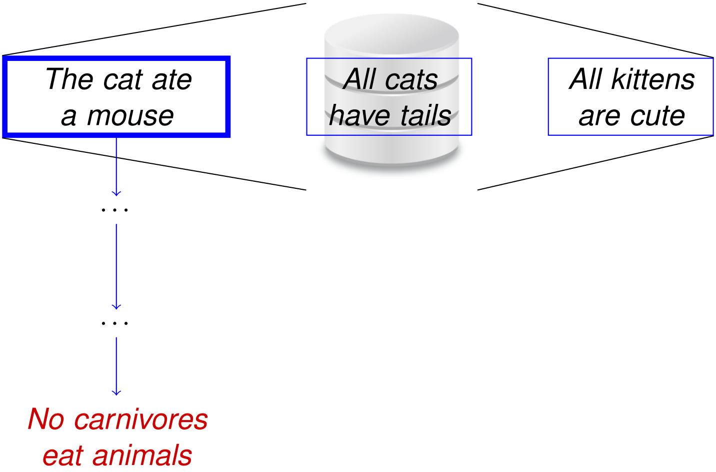
Quantifiers determines the *polarity* (↑ or ↓) of words.

Mutations must respect *polarity*.

Inference is reversible.

	felines		mammals
	cats		placentals
↑ **All**$_{\downarrow\uparrow}$	↓ house cats	↑ consume	↑ rodents
	kitties	eat	mice

Natural Logic Inference is Search



Natural Logic Inference is Search

No carnivores eat animals?

***The** carnivores eat animals*

*The **cat** eats animals*

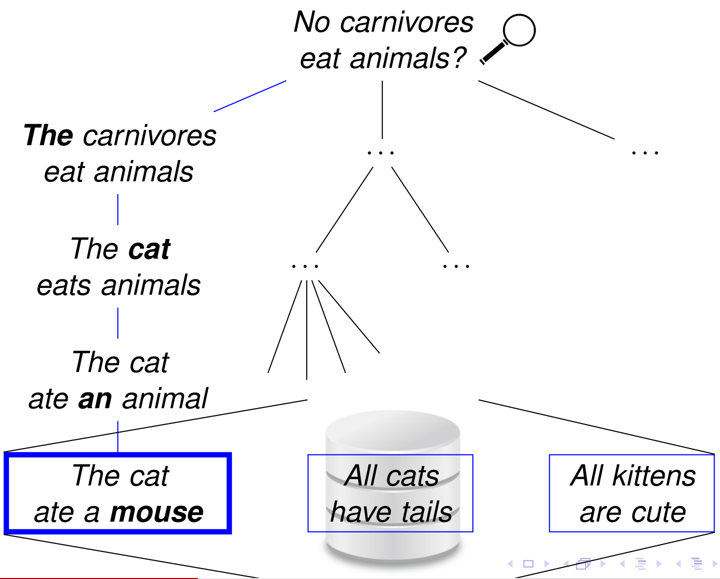
*The cat ate **an** animal*

*The cat ate a **mouse***

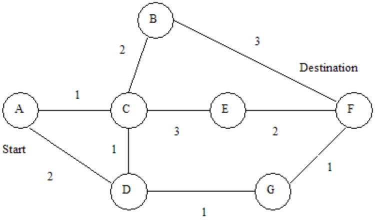
All cats have tails

All kittens are cute

Natural Logic Inference is Search

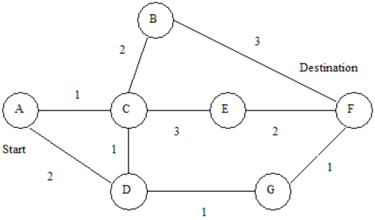


Natural Logic Inference is Search



Nodes (*fact*, truth maintained $\in \{\text{true}, \text{false}\}$)

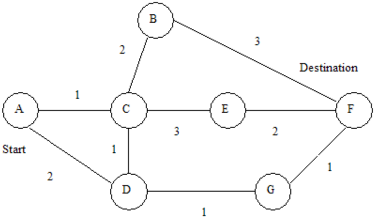
Natural Logic Inference is Search



Nodes (*fact*, truth maintained $\in \{\text{true}, \text{false}\}$)

Start Node (*query fact*, *true*)
End Nodes *any known fact*

Natural Logic Inference is Search



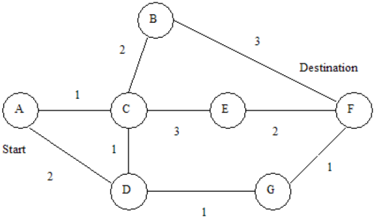
Nodes (*fact*, truth maintained $\in \{\text{true}, \text{false}\}$)

Start Node (*query fact*, *true*)

End Nodes *any known fact*

Edges Mutations of the current fact

Natural Logic Inference is Search



Nodes (*fact*, truth maintained $\in \{\mathrm{true}, \mathrm{false}\}$)

Start Node (*query fact*, *true*)
End Nodes *any known fact*

Edges Mutations of the current fact
Edge Costs How "wrong" an inference step is (learned)

An Example Search (as reverse inference)

Search mutates *opposite* to polarity

organism		organism
animal	→	animal
$\downarrow$ carnivores		$\downarrow$ carnivores
felines		felines

An Example Search (as reverse inference)

Truth maintained: *true*

Current Node:

	organism		living thing
	animal	consume	organism
↑ **No**$_{\downarrow\downarrow}$	↓ carnivores	↓ eat	↓ animals
	felines	slurp	chordate

An Example Search (as reverse inference)

Truth maintained: *false*

		organism		living thing
		animal	consume	organism
Current Node:	$\uparrow$ **The**$_{\uparrow\uparrow}$	$\uparrow$ carnivores	$\uparrow$ eat	$\uparrow$ animals
	All$_{\downarrow\uparrow}$	felines	slurp	chordate

An Example Search (as reverse inference)

Truth maintained: *false*

Current Node:

↑ **The**$_{\uparrow\uparrow}$	↑ **felines**	↑ eat	↑ animals
All$_{\downarrow\uparrow}$	cats	slurp	chordate

animals
carnivores
consume
living thing
organism

An Example Search (as reverse inference)

Truth maintained: *false*

Current Node:

	carnivores		living thing
	felines	consume	organism
$\uparrow$ **The**$_{\uparrow\uparrow}$	$\uparrow$ **cats**	$\uparrow$ eat	$\uparrow$ animals
All$_{\downarrow\uparrow}$	kitties	slurp	chordate

An Example Search (as reverse inference)

Truth maintained: *false*

		carnivores		organisms
		felines	consume	animals
Current Node:	$\uparrow$ **The**$_{\uparrow\uparrow}$	$\uparrow$ cats	$\uparrow$ eat	$\uparrow$ **chordates**
	All$_{\downarrow\uparrow}$	kitties	slurp	mice

An Example Search (as reverse inference)

Truth maintained: *false*

		carnivores		animals
		felines	consume	chordates
Current Node:	↑ **The**$_{\uparrow\uparrow}$	↑ cats	↑ eat	↑ **mice**
	All$_{\downarrow\uparrow}$	kitties	slurp	fieldmice

An Example Search (as reverse inference)

Truth maintained: *false*

Current Node:

	carnivore			animal
	feline	consumed		chordate
↑ **The**$_{\uparrow\uparrow}$	↑ cat	↑ ate	↑ **a**$_{\uparrow\uparrow}$	↑ mouse
All$_{\downarrow\uparrow}$	kitty	slurped	**All**$_{\downarrow\uparrow}$	fieldmouse

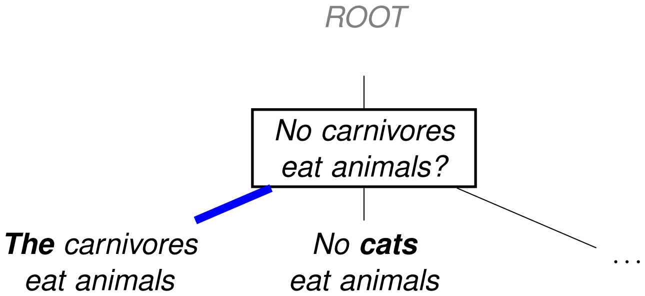
An Example Search (as graph search)

Shorthand for a node:

	organism		living thing
	animal	consume	organism
↑ **No**$_{\downarrow\downarrow}$	↑ carnivores	↓ eat	↑ animals
	felines	slurp	chordate

No carnivores eat animals?

An Example Search (as graph search)



An Example Search (as graph search)

No carnivores eat animals

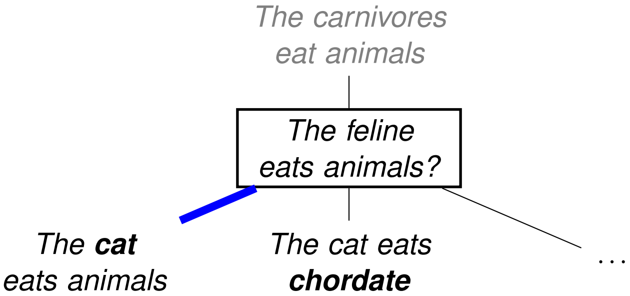
The carnivores eat animals?

*The **feline** eats animals*

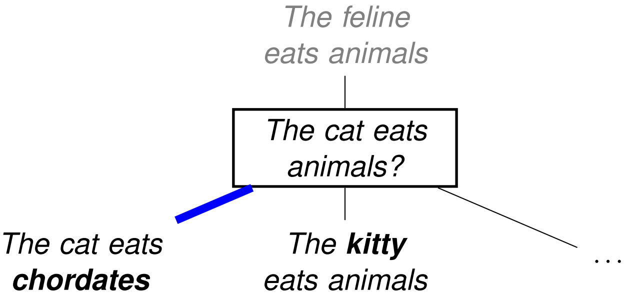
***All** carnivores eat animals*

...

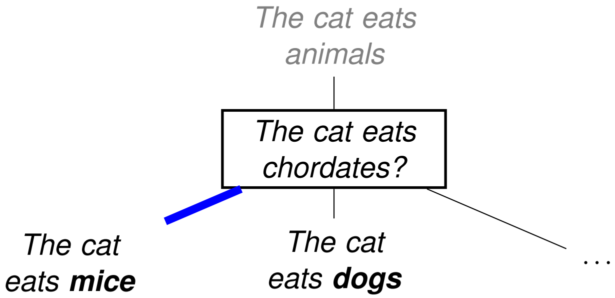
An Example Search (as graph search)



An Example Search (as graph search)



An Example Search (as graph search)



An Example Search (as graph search)

The cat eats chordates

The cat eats mice?

The cat ate ***a*** *mouse*

The ***kitty*** *eats mice*

. . .

“Soft” Natural Logic

Want to make likely (but not certain) inferences.

- Same motivation as Markov Logic, Probabilistic Soft Logic, etc.

“Soft” Natural Logic

Want to make likely (but not certain) inferences.

- Same motivation as Markov Logic, Probabilistic Soft Logic, etc.
- Each *edge template* has a cost $\theta \geq 0$.

"Soft" Natural Logic

Want to make likely (but not certain) inferences.

- Same motivation as Markov Logic, Probabilistic Soft Logic, etc.
- Each *edge template* has a cost $\theta \geq 0$.

Detail: Variation among *edge instances* of a template.

- WordNet: *cat* $\rightarrow$ *feline* **vs.** *cup* $\rightarrow$ *container*.
- Nearest neighbors distance.
- Each *edge instance* has a distance f.

"Soft" Natural Logic

Want to make likely (but not certain) inferences.

- Same motivation as Markov Logic, Probabilistic Soft Logic, etc.
- Each *edge template* has a cost $\theta \geq 0$.

Detail: Variation among *edge instances* of a template.

- WordNet: *cat* $\rightarrow$ *feline* **vs.** *cup* $\rightarrow$ *container*.
- Nearest neighbors distance.
- Each *edge instance* has a distance f.

Cost of an edge is $\theta_i \cdot f_i$.

"Soft" Natural Logic

Want to make likely (but not certain) inferences.

- Same motivation as Markov Logic, Probabilistic Soft Logic, etc.
- Each *edge template* has a cost $\theta \geq 0$.

Detail: Variation among *edge instances* of a template.

- WordNet: *cat* $\rightarrow$ *feline* **vs.** *cup* $\rightarrow$ *container*.
- Nearest neighbors distance.
- Each *edge instance* has a distance f.

Cost of an edge is $\theta_i \cdot f_i$.
Cost of a path is $\theta \cdot \mathbf{f}$.

“Soft” Natural Logic

Want to make likely (but not certain) inferences.

- Same motivation as Markov Logic, Probabilistic Soft Logic, etc.
- Each *edge template* has a cost $\theta \geq 0$.

Detail: Variation among *edge instances* of a template.

- WordNet: *cat* $\rightarrow$ *feline* **vs.** *cup* $\rightarrow$ *container*.
- Nearest neighbors distance.
- Each *edge instance* has a distance f.

Cost of an edge is $\theta_i \cdot f_i$.
Cost of a path is $\theta \cdot \mathbf{f}$.
Can learn parameters θ.

Summary: NaturalLI

High-Level Takeaways

- *Deep* inferences from a *large* knowledge base.
- Leverage arbitrarily large plain-text knowledge bases.
- "Soft" logic with probability of truth.

Summary: NaturalLI

High-Level Takeaways

- *Deep* inferences from a *large* knowledge base.
- Leverage arbitrarily large plain-text knowledge bases.
- “Soft” logic with probability of truth.

Results

- ConceptNet facts: 12% recall → 49% recall @ 91% precision.
- Checks logical entailment (not just fuzzy query).

Summary: NaturalLI

High-Level Takeaways

- *Deep* inferences from a *large* knowledge base.
- Leverage arbitrarily large plain-text knowledge bases.
- "Soft" logic with probability of truth.

Results

- ConceptNet facts: 12% recall $\rightarrow$ 49% recall @ 91% precision.
- Checks logical entailment (not just fuzzy query).

Complexity doesn't grow with knowledge base size.

The “Proposal” Part

Possible Projects

What is NaturalLI bad at?

Possible Projects

What is NaturalLI bad at?

- **Real world Q/A (WebQuestions, TREC, AI2 Biology).** Most people don't really care about common-sense.

Possible Projects

What is NaturalLI bad at?

- **Real world Q/A (WebQuestions, TREC, AI2 Biology).**
  Most people don't really care about common-sense.
- **Reasoning with multiple premises.**
  E.g., DeMorgan's laws.

Possible Projects

What is NaturalLI bad at?

- **Real world Q/A (WebQuestions, TREC, AI2 Biology).**
  Most people don't really care about common-sense.
- **Reasoning with multiple premises.**
  E.g., DeMorgan's laws.
- **Fully leveraging training data.**
  What if the entailment rules aren't in WordNet?

Real World Q/A (EMNLP 2015)

Goal: Make NaturalLI useful, not just cute.

Real World Q/A (EMNLP 2015)

Goal: Make NaturalLI useful, not just cute.

Start with NaturalLI...

+ Dependency trees (for passivization, etc.)
+ Meronymy (Hawaii is in USA)
+ Relational entailment (*works* $\rightarrow$ *employed*)
+ Paraphrases

Real World Q/A (EMNLP 2015)

Goal: Make NaturalLI useful, not just cute.

Start with NaturalLI. . .

+ Dependency trees (for passivization, etc.)
+ Meronymy (Hawaii is in USA)
+ Relational entailment (*works* $\rightarrow$ *employed*)
+ Paraphrases

Evaluate on:

- AI2 Biology tests
- WebQuestions
- TREC QA

Multiple Premises (logic paper?)

Goal: Make Natural Logic look more appealing.

Multiple Premises (logic paper?)

Goal: Make Natural Logic look more appealing.

The Trivial Case: Parse premises that define edges in a partial order (e.g., *cats are animals*, *Hawaii is in the USA*).

Multiple Premises (logic paper?)

Goal: Make Natural Logic look more appealing.

The Trivial Case: Parse premises that define edges in a partial order (e.g., *cats are animals*, *Hawaii is in the USA*).

Propositional + Natural Logic

- E.g., $a \vee b; \neg a \quad \Rightarrow \quad b$

Multiple Premises (logic paper?)

Goal: Make Natural Logic look more appealing.

The Trivial Case: Parse premises that define edges in a partial order (e.g., *cats are animals*, *Hawaii is in the USA*).

Propositional $+$ Natural Logic

- E.g., $a \vee b; \neg a \quad \Rightarrow \quad b$
- Shallow syntactic parse of sentence for conjunction / disjunction

Multiple Premises (logic paper?)

Goal: Make Natural Logic look more appealing.

The Trivial Case: Parse premises that define edges in a partial order (e.g., *cats are animals*, *Hawaii is in the USA*).

Propositional $+$ Natural Logic

- E.g., $a \vee b; \neg a \;\; \Rightarrow \;\; b$
- Shallow syntactic parse of sentence for conjunction / disjunction
- Operators in hypothesis are trivial (run all atoms and combine)
- Conjunction in premise is trivial (NaturalLI OpenIE can do this)
- Disjunction and material implication in premise is a bit tricky

Multiple Premises (logic paper?)

Goal: Make Natural Logic look more appealing.

The Trivial Case: Parse premises that define edges in a partial order (e.g., *cats are animals*, *Hawaii is in the USA*).

Propositional $+$ Natural Logic

- E.g., $a \vee b; \neg a \Rightarrow b$
- Shallow syntactic parse of sentence for conjunction / disjunction
- Operators in hypothesis are trivial (run all atoms and combine)
- Conjunction in premise is trivial (NaturalLI OpenIE can do this)
- Disjunction and material implication in premise is a bit tricky

Probabilistic logic of "some:" *some cats have tails*; *some cats are male* $\Rightarrow$ *some males have tails* (with some probability)

Fully Leveraging Training Data (??? paper)

Goal: Make NaturalLI less hand-coded.

Fully Leveraging Training Data (??? paper)

Goal: Make NaturalLI less hand-coded.

Research by Analogy:

- MIML-RE is to NaturalLI as Neural QA is to X.
- High level idea: Natural Logic mutations are transforms in vector space.

Fully Leveraging Training Data (??? paper)

Goal: Make NaturalLI less hand-coded.

Research by Analogy:

- MIML-RE is to NaturalLI as Neural QA is to X.
- High level idea: Natural Logic mutations are transforms in vector space.

Implementation 1: NaturalLI in vector space

- Same search, but each rule (e.g., go up WordNet) is a matrix operation. “Hit” a premise if we get close enough.
- Can train from SNLI corpus $+$ MT alignments.

Fully Leveraging Training Data (??? paper)

Goal: Make NaturalLI less hand-coded.

Research by Analogy:

- MIML-RE is to NaturalLI as Neural QA is to *X*.
- High level idea: Natural Logic mutations are transforms in vector space.

Implementation 1: NaturalLI in vector space

- Same search, but each rule (e.g., go up WordNet) is a matrix operation. “Hit” a premise if we get close enough.
- Can train from SNLI corpus $+$ MT alignments.

Implementation 2: Learn hyperplane between true and false facts.

- Positives from the internet.
- Negatives from mutating *each* positive, $+$ unrelated facts.

Thanks!

References I

Angeli, G., Chaganty, A., Chang, A., Reschke, K., Tibshirani, J., Wu, J. Y., Bastani, O., Siilats, K., and Manning, C. D. (2014a).
Stanford's 2013 KBP system.
In *TAC-KBP*.

Angeli, G. and Manning, C. D. (2013).
Philosophers are mortal: Inferring the truth of unseen facts.
In *CoNLL*.

Angeli, G. and Manning, C. D. (2014).
Naturalli: Natural logic inference for common sense reasoning.
In *EMNLP*.

Angeli, G., Tibshirani, J., Wu, J. Y., and Manning, C. D. (2014b).
Combining distant and partial supervision for relation extraction.
In *EMNLP*.