Combining Natural Logic and Shallow Reasoning for Question Answering.

Gabor Angeli, Neha Nayak, Chris Manning

Stanford University

August 8, 2016

Old Problem: Logic + ML often at odds

Old Problem: Logic + ML often at odds

ML gives us practical, generalizable systems:

P: *Ovaries are the female part of the flower, which produces eggs that are needed for making seeds.*

H: *A flower produces the seeds.*

...But struggles with logical subtleties

P: *Eating candy for dinner is an example of a poor health habit.*

H: *Eating candy is an example of a good health habit.*

Make ML more first-order-logic-like

Markov Logic Networks

- [Richardson and Domingos, 2006]
- [Niu et al., 2011]

Probabilistic Soft Logic

- [Kimmig et al., 2012]
- [Beltagy et al., 2014]

Deep Learning + Logic

- [Rocktäschel et al., 2014]

Find logics that are better for ML

Natural Logic!

[Sánchez Valencia, 1991, MacCartney and Manning, 2008, Icard III and Moss, 2014]

Find logics that are better for ML

Natural Logic!

Logic over natural language

- *Instantaneous* and *perfect* semantic parsing!
- Plays nice with lexical methods

[Sánchez Valencia, 1991, MacCartney and Manning, 2008, Icard III and Moss, 2014]

Find logics that are better for ML

Natural Logic!

Logic over natural language

- *Instantaneous* and *perfect* semantic parsing!
- Plays nice with lexical methods

Tractable

- Polynomial time entailment checking [MacCartney and Manning, 2008].

[Sánchez Valencia, 1991, MacCartney and Manning, 2008, Icard III and Moss, 2014]

Find logics that are better for ML

Natural Logic!

Logic over natural language

- *Instantaneous* and *perfect* semantic parsing!
- Plays nice with lexical methods

Tractable

- Polynomial time entailment checking [MacCartney and Manning, 2008].

Expressive (for common inferences)

- Second-order phenomena; *most*; quantifier scoping

[Sánchez Valencia, 1991, MacCartney and Manning, 2008, Icard III and Moss, 2014]

Find logics that are better for ML

Natural Logic!

Logic over natural language

- *Instantaneous* and *perfect* semantic parsing!
- Plays nice with lexical methods

Tractable

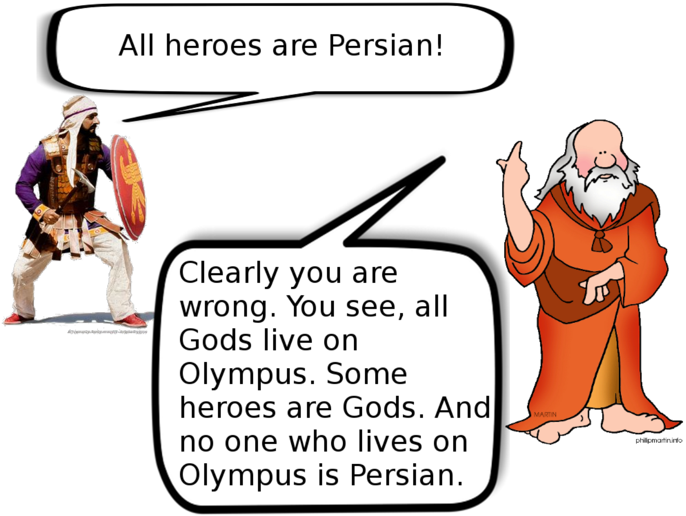
- Polynomial time entailment checking [MacCartney and Manning, 2008].

Expressive (for common inferences)

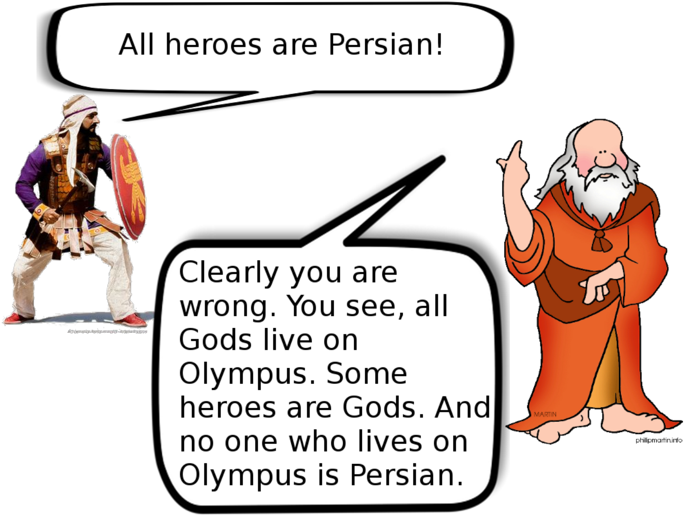
- Second-order phenomena; *most*; quantifier scoping
- No free lunch: shallow quantification; single-premise only

[Sánchez Valencia, 1991, MacCartney and Manning, 2008, Icard III and Moss, 2014]

The Persians are Invading Greece



How Did You Solve This?



Show of Hands: First Order Logic?

1	$\forall x\ \text{God}(x) \supset \text{LivesOnOlympus}(x)$	
2	$\exists x\ \text{Hero}(x) \land \text{God}(x)$	
3	$\neg\exists x\ \text{LivesOnOlympus}(x) \land \text{Persian}(x)$	
4	$\quad \forall x\ \text{Hero(x)} \supset \text{Persian}(x)$	
5	$\quad\quad a \mid \text{Hero}(a) \land \text{God}(a)$	$\exists$E, 2
6	$\quad\quad \text{Hero}(a)$	$\land$E, 5
7	$\quad\quad \text{Hero(a)} \supset \text{Persian}(a)$	$\forall$E, 4
8	$\quad\quad \text{Persian}(a)$	$\Rightarrow$E, 6, 7
9	$\quad\quad \text{God}(a)$	$\land$E, 5
10	$\quad\quad \text{God}(a) \supset \text{LivesOnOlympus}(a)$	$\forall$E, 1
11	$\quad\quad \text{LivesOnOlympus}(a)$	$\Rightarrow$E, 9, 10
12	$\quad\quad \text{LivesOnOlympus}(a) \land \text{Persian}(a)$	$\land$I, 8, 11
13	$\quad\quad \exists x\ \text{LivesOnOlympus}(x) \land \text{Persian}(x)$	$\exists$I, 12
14	$\quad \exists x\ \text{LivesOnOlympus}(x) \land \text{Persian}(x)$	R, 12
15	$\quad \bot$	$\bot$ I, 3, 14
16	$\neg\ \forall x\ \text{Hero(x)} \supset \text{Persian}(x)$	$\neg$ I, 4—15

Syllogisms: The First Natural Logic

1	*All Gods live on Mount Olympus*	
2	*Some heroes are Gods*	
3	*Nobody who lives on Mount Olympus is Persian*	
4	*Some heroes live on Mount Olympus*	AII (Darii), 1, 2
5	*Some heroes are not Persian*	EIO (Ferio), 4, 3
6	$\neg$ *All heroes are Persian*	SaP $\perp$ SoP, 5

Syllogisms: The First Natural Logic

1	*All Gods live on Mount Olympus*	
2	*Some heroes are Gods*	
3	*Nobody who lives on Mount Olympus is Persian*	
4	*Some heroes live on Mount Olympus*	AII (Darii), 1, 2
5	*Some heroes are not Persian*	EIO (Ferio), 4, 3
6	$\neg$ *All heroes are Persian*	SaP $\perp$ SoP, 5

...But syllogisms are cripplingly unexpressive

Modern Natural Logic

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

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

Modern Natural Logic

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

Mutations must respect *polarity*.

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

Modern Natural Logic

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

Mutations must respect *polarity*.

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

Modern Natural Logic

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

Modern Natural Logic

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

Mutations must respect *polarity*.

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

Modern Natural Logic

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

Modern Natural Logic

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

Mutations must respect *polarity*.

Not pictured: also handles negation

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

Natural Logic Inference as Search

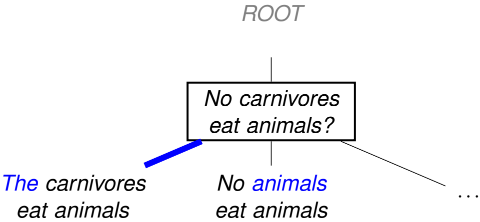
Shorthand for a node:

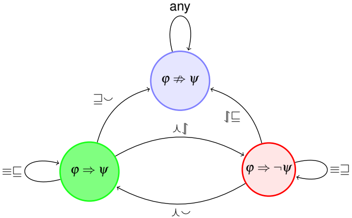
	organism		living thing
	animal	consume	organism
$\uparrow$ **No**$_{\downarrow\downarrow}$	$\downarrow$ carnivores	$\downarrow$ eat	$\downarrow$ animals
	felines	slurp	chordate

$\Downarrow$

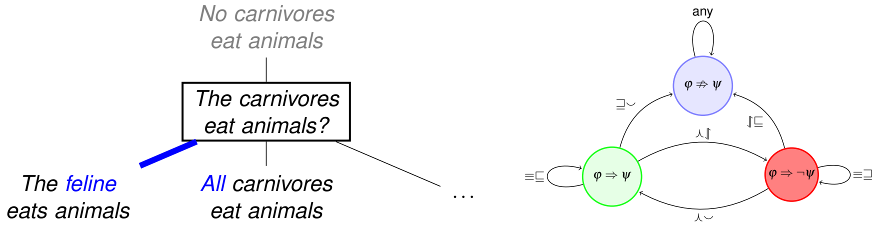
No carnivores eat animals?

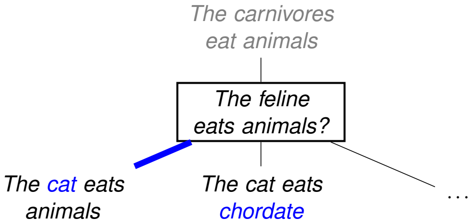
Natural Logic Inference as Search

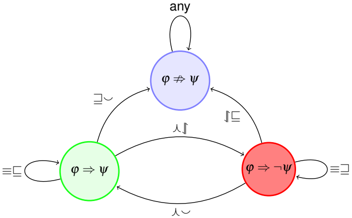




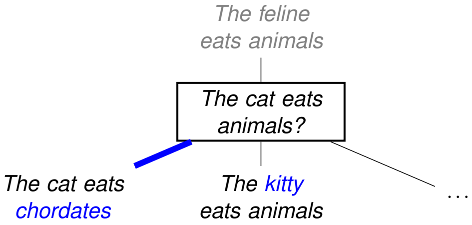
Natural Logic Inference as Search

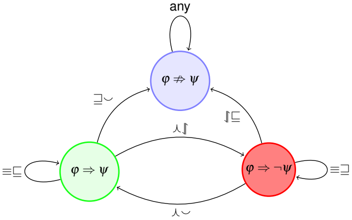


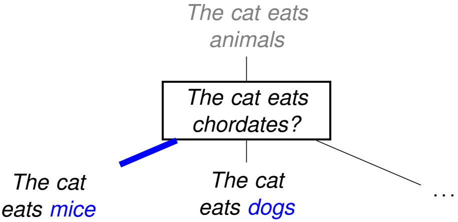

The carnivores eat animals
The feline eats animals?
The cat eats animals
The cat eats chordate
...


any
$\varphi \not\Rightarrow \psi$
$\varphi \Rightarrow \psi$
$\varphi \Rightarrow \neg\psi$
$\equiv\sqsubseteq$
$\equiv\sqsupseteq$

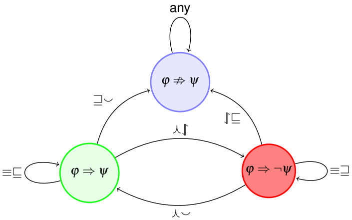
Natural Logic Inference as Search





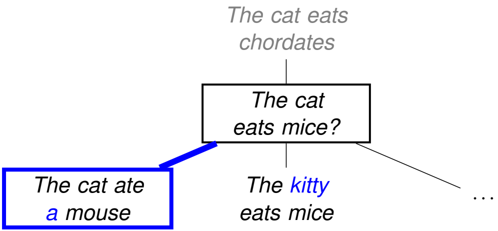


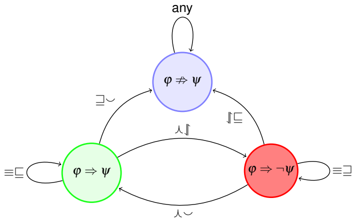
The cat eats animals
The cat eats chordates?
The cat eats mice
The cat eats dogs
...



any
φ ⇏ ψ
φ ⇒ ψ
φ ⇒ ¬ψ

Natural Logic Inference as Search





Three Contributions for Generalizable Inference

1. **Partial order over meronymy + relations**

Earth	
EU	have
$\uparrow$ Germany	$\uparrow$ own
Berlin	sell

Three Contributions for Generalizable Inference

1. **Partial order over meronymy + relations**

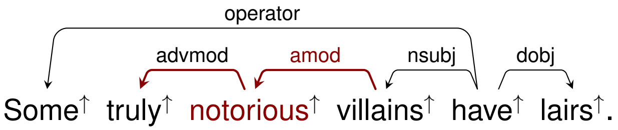
Earth

EU | have

↑ Germany | ↑ own

Berlin | sell

2. **Natural Logic over dependency trees**



Three Contributions for Generalizable Inference

1. **Partial order over meronymy + relations**

Earth

EU	have
$\boxed{\uparrow \text{Germany}}$	$\boxed{\uparrow \text{own}}$
Berlin	sell

2. **Natural Logic over dependency trees**

operator

advmod amod nsubj dobj

Some$^{\uparrow}$ truly$^{\uparrow}$ notorious$^{\uparrow}$ villains$^{\uparrow}$ have$^{\uparrow}$ lairs$^{\uparrow}$.

3. **Hybrid statistical / logical solver**

An Example Search

ROOT

Rain and snow are types of precipitation?

Rain and snow are forms of precipitation

Heavy rain and snow are types of precipitation

…

An Example Search

ROOT

Rain and snow are types of precipitation?

Rain and snow are forms of precipitation

Heavy rain and snow are types of precipitation

…

An Example Search

Rain and snow are types of precipitation

Rain and snow are forms of precipitation?

Rain and snow are forms of weather

Heavy rain and snow are forms of precipitation

. . .

An Example Search

Rain and snow are types of precipitation

Rain and snow are forms of precipitation?

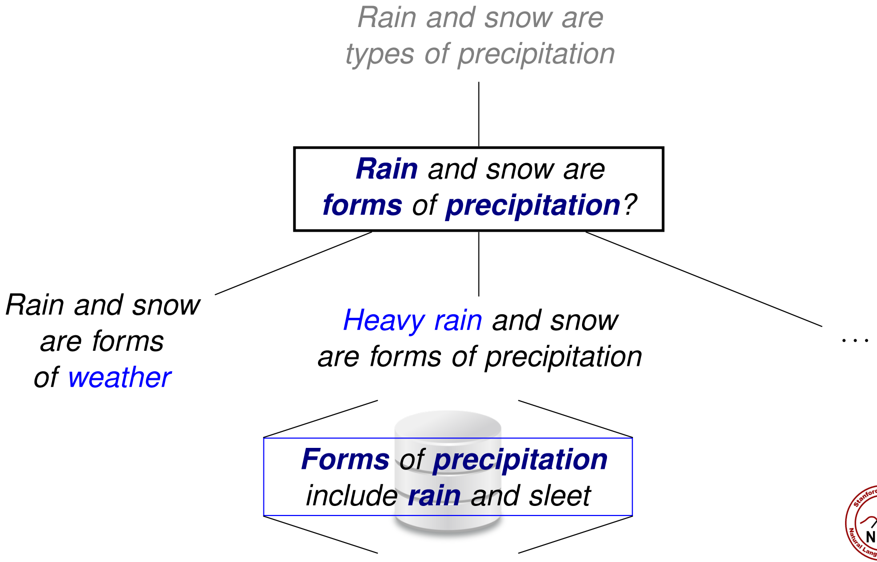
Rain and snow are forms of weather

Heavy rain and snow are forms of precipitation

. . .

Forms of precipitation include rain and sleet

An Example Search



Lexical Alignment Classifier

Forms *of* ***precipitation*** ***include*** ***rain*** *and* ***sleet***

Rain *and* ***snow*** *are* ***forms*** *of* ***precipitation***

Lexical Alignment Classifier

Forms *of* ***precipitation*** ***include*** ***rain*** *and* ***sleet***

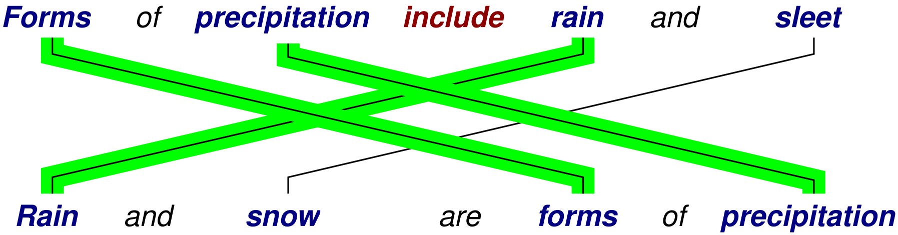
Rain *and* ***snow*** *are* ***forms*** *of* ***precipitation***

Lexical Alignment Classifier

Forms *of* ***precipitation*** ***include*** ***rain*** *and* ***sleet***

Rain *and* ***snow*** *are* ***forms*** *of* ***precipitation***

Lexical Alignment Classifier

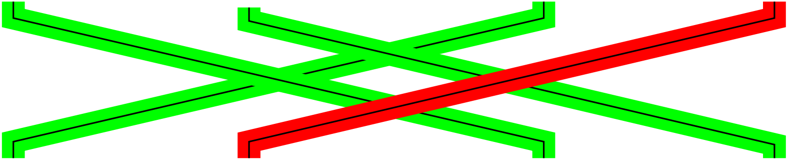


Features

1. Matching words

Lexical Alignment Classifier

Forms *of* ***precipitation*** ***include*** ***rain*** *and* ***sleet***

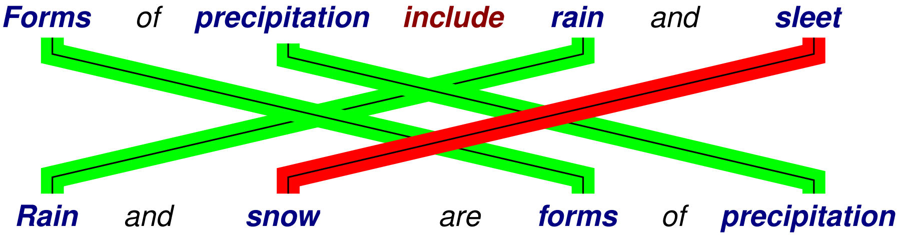


Rain *and* ***snow*** *are* ***forms*** *of* ***precipitation***

Features

1. Matching words
2. Mismatched words

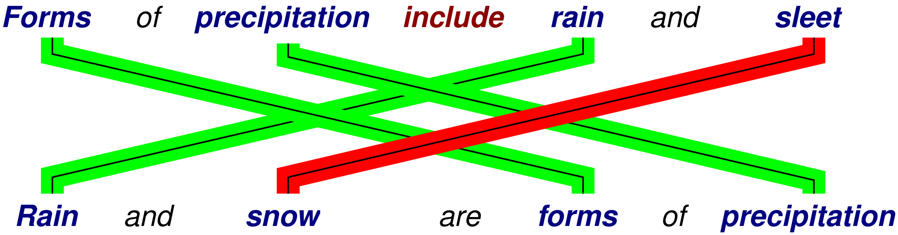
Lexical Alignment Classifier



Features

1. Matching words
2. Mismatched words
3. Unmatched words in premise/consequent

Lexical Alignment Classifier



Features

1. Matching words
2. Mismatched words
3. Unmatched words in premise/consequent

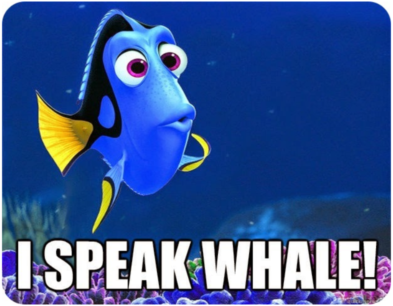
Competitive with Stanford RTE system (63% on RTE3)

Old Problem: Logic + Lexical Classifiers

FOL and lexical classifiers don't speak the same language

Old Problem: Logic + Lexical Classifiers

FOL and lexical classifiers don't speak the same language
...but natural logic does!



Big Picture

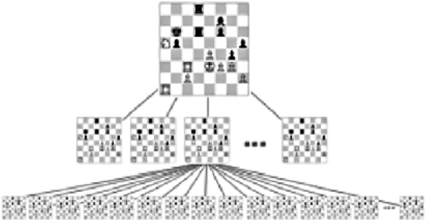
Run our usual search

1. If we find a premise, great!

Big Picture

Run our usual search

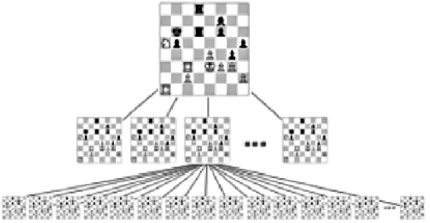
1. If we find a premise, great!
2. If not, use lexical classifier as an *evaluation function*



Big Picture

Run our usual search

1. If we find a premise, great!
2. If not, use lexical classifier as an *evaluation function*



Visit 1M nodes / second: We have to be fast!

Dissecting Our Classifier

Anatomy of a Classifier

- Features f (matching / mismatched / unmatched words)
- Weights w
- Entailment pair x

$$p(\text{entail} \mid x) = \frac{1}{1+\exp(-w^{\mathrm{T}} f(x))}$$

Dissecting Our Classifier

Anatomy of a Classifier

- Features f (matching / mismatched / unmatched words)
- Weights w
- Entailment pair x

$$p(\text{entail} \mid x) = \frac{1}{1+\exp(-w^{\mathrm{T}} f(x))}$$

$$p(\text{entail} \mid x) \text{ monotone w.r.t. } \left(w^{\mathrm{T}} f(x)\right)$$

Dissecting Our Classifier

Anatomy of a Classifier

- Features f (matching / mismatched / unmatched words)
- Weights w
- Entailment pair x

$$p(\text{entail} \mid x) = \frac{1}{1+\exp(-w^{\mathrm{T}} f(x))}$$

$$p(\text{entail} \mid x) \text{ monotone w.r.t. } \left(w^{\mathrm{T}} f(x)\right)$$

- Only need $w^{\mathrm{T}} f(x)$ during search to compute $\max p(\text{entail} \mid x)$
- $w^{\mathrm{T}} f(x)$ is our evaluation function

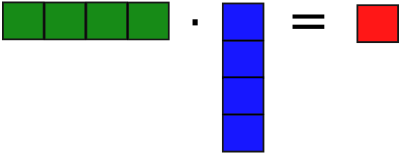
Incorporating our Evaluation Function

Anatomy of a Search Step

1. Mutate a word, or
2. Delete a word, or
3. Insert a word.

Each step updates a small number of features

$$w^{\mathrm{T}} f(x) = v$$



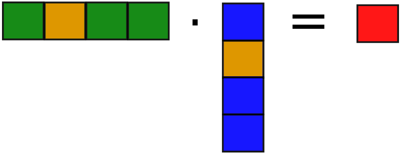
Incorporating our Evaluation Function

Anatomy of a Search Step

1. Mutate a word, or
2. Delete a word, or
3. Insert a word.

Each step updates a small number of features

$$w^{\mathrm{T}} f(x) = v$$



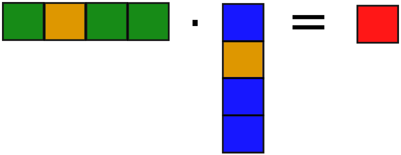
Incorporating our Evaluation Function

Anatomy of a Search Step

1. Mutate a word, or
2. Delete a word, or
3. Insert a word.

Each step updates a small number of features

$$v' = v - w_i \cdot f_i + w_i \cdot f_i$$



Why is this Important?

Faster Search ⇒ Deeper Reasoning

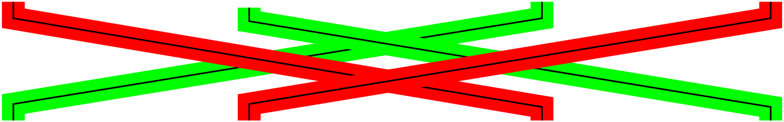
- **Speed:** Around 1M search states visited per second
- **Memory:** 32 byte search states

Speed: Don't re-featurize at every timestep.

Memory: Never store intermediate fact as String.

Another Example Search

Forms *of* ***precipitation*** ***include*** ***rain*** *and* ***sleet***



Rain *and* ***snow*** *are* ***types*** *of* ***precipitation***

Score $w^{\mathrm{T}} f(x)$: -0.5

Feature	w	$f(x)$
Matching words	2.0	2
Mismatched words	-1.0	2
Unmatched premise	-0.5	1
Unmatched consequent	-0.75	0
Bias	-2.0	1

Another Example Search

ROOT

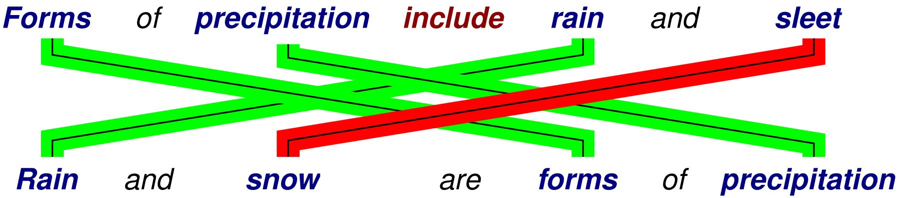
Rain and snow are types of precipitation?

*Rain and snow are **forms** of precipitation*

*Rain and snow are types of **weather***

. . .

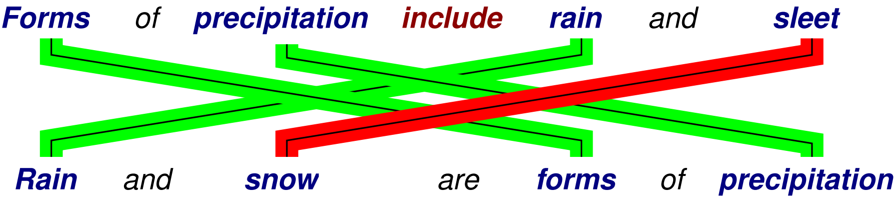
Another Example Search



Score $w^{\mathrm{T}} f(x)$**:** -0.5 + 2 – -1

Feature	w	$f(x)$
Matching words	2.0	**3**
Mismatched words	-1.0	**1**
Unmatched premise	-0.5	1
Unmatched consequent	-0.75	0
Bias	-2.0	1

Another Example Search



Score $w^{\mathrm{T}} f(x)$**:** 2.5

Feature	w	$f(x)$
Matching words	2.0	**3**
Mismatched words	-1.0	**1**
Unmatched premise	-0.5	1
Unmatched consequent	-0.75	0
Bias	-2.0	1

Solving 4^{th} Grade Science

Multiple choice questions from real 4^{th} grade science exams

Solving 4^{th} Grade Science

Multiple choice questions from real 4^{th} grade science exams

Which activity is an example of a good health habit?

(A) Watching television
(B) Smoking cigarettes
(C) Eating candy
(D) Exercising every day

Solving 4^{th} Grade Science

Multiple choice questions from real 4^{th} grade science exams

Which activity is an example of a good health habit?

(A) Watching television
(B) Smoking cigarettes
(C) Eating candy
(D) Exercising every day

In our corpus:

- *Plasma TV's can display up to 16 million colors … great for watching TV … also make a good screen.*
- *Not smoking or drinking alcohol is good for health, regardless of whether clothing is worn or not.*
- *Eating candy for diner is an example of a poor health habit.*
- *Healthy is exercising*

Solving 4^{th} Grade Science

Multiple choice questions from real 4^{th} grade science exams

System	Train	Test
KNOWBOT	45	
KNOWBOT (ORACLE)	57	

[Hixon et al., 2015]

Solving 4^{th} Grade Science

Multiple choice questions from real 4^{th} grade science exams

System	Train	Test
KNOWBOT	45	
KNOWBOT (ORACLE)	57	
IR Baseline	49	
This Work	52	

[Hixon et al., 2015]

Solving 4^{th} Grade Science

Multiple choice questions from real 4^{th} grade science exams

System	Train	Test
KNOWBOT	45	
KNOWBOT (ORACLE)	57	
IR Baseline	49	
This Work	52	
More Data + IR Baseline	62	
More Data + This Work	**65**	

[Hixon et al., 2015]

Solving 4^{th} Grade Science

Multiple choice questions from real 4^{th} grade science exams

System	Train	Test
KNOWBOT	45	–
KNOWBOT (ORACLE)	57	–
IR Baseline	49	42
This Work	52	51
More Data + IR Baseline	62	58
More Data + This Work	**65**	**61**

[Hixon et al., 2015]

Solving 4^{th} Grade Science

Multiple choice questions from real 4^{th} grade science exams

System	Train	Test
KNOWBOT	45	–
KNOWBOT (ORACLE)	57	–
IR Baseline	49	42
This Work	52	51
More Data + IR Baseline	62	58
More Data + This Work	65	61
This Work +  + 	**74**	**67**

[Hixon et al., 2015]

Solving 4^{th} Grade Science

Multiple choice questions from real 4^{th} grade science exams

System	Train	Test
KNOWBOT	45	–
KNOWBOT (ORACLE)	57	–
IR Baseline	49	42
This Work	52	51
More Data + IR Baseline	62	58
More Data + This Work	65	61
This Work +  + 	**74**	**67**

We're able to pass 4^{th} grade science!

[Hixon et al., 2015]

Conclusions

Natural Logic

- A logic over the syntax of natural language
- Expressive but efficient

Natural Logic "plays nice" with statistical (/deep?) methods

- Both operate directly over text
- Use lexical classifier as evaluation function

NaturalLI + Evaluation Function

- Also detects *likely* entailment / contradictions
- 3% improvement on science exam questions

References I

Beltagy, I., Erk, K., and Mooney, R. (2014).
Semantic parsing using distributional semantics and probabilistic logic.
In *Association for Computational Linguistics (ACL)*, page 7.

Hixon, B., Clark, P., and Hajishirzi, H. (2015).
Learning knowledge graphs for question answering through conversational dialog.
NAACL.

Icard III, T. and Moss, L. (2014).
Recent progress on monotonicity.
Linguistic Issues in Language Technology.

References II

Kimmig, A., Bach, S., Broecheler, M., Huang, B., and Getoor, L. (2012).
A short introduction to probabilistic soft logic.
In *Proceedings of the NIPS Workshop on Probabilistic Programming: Foundations and Applications*, pages 1–4.

MacCartney, B. and Manning, C. D. (2008).
Modeling semantic containment and exclusion in natural language inference.
In *Coling*.

Niu, F., Ré, C., Doan, A., and Shavlik, J. (2011).
Tuffy: Scaling up statistical inference in markov logic networks using an rdbms.
VLDB.

References III

Richardson, M. and Domingos, P. (2006).
Markov logic networks.
Machine learning, 62(1-2):107–136.

Rocktäschel, T., Bošnjak, M., Singh, S., and Riedel, S. (2014).
Low-dimensional embeddings of logic.
In *Proceedings of the ACL 2014 Workshop on Semantic Parsing*.

Sánchez Valencia, V. M. (1991).
Studies on natural logic and categorial grammar.
PhD thesis, University of Amsterdam.