

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]
- [Bowman, 2014]



Find logics that are better for ML

Natural Logic!

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



The Persians are Invading Greece

All heroes are Persian!



Clearly you are wrong. You see, all Gods live on Olympus. Some heroes are Gods. And no one who lives on Olympus is Persian.



How Did You Solve This?

All heroes are Persian!



Clearly you are wrong. You see, all Gods live on Olympus. Some heroes are Gods. And no one who lives on Olympus is Persian.



Show of Hands: First Order Logic?

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



Syllogisms: The First Natural Logic

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



Syllogisms: The First Natural Logic

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

...But syllogisms are crippledly unexpressive



Modern Natural Logic

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



Modern Natural Logic

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

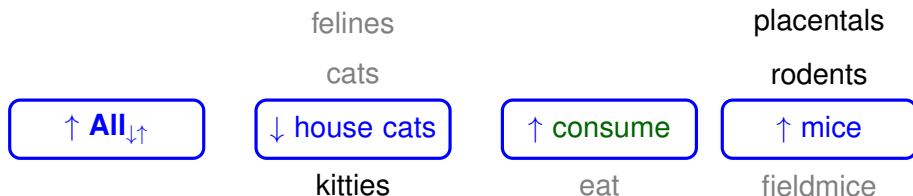
Mutations must respect *polarity*.



Modern Natural Logic

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

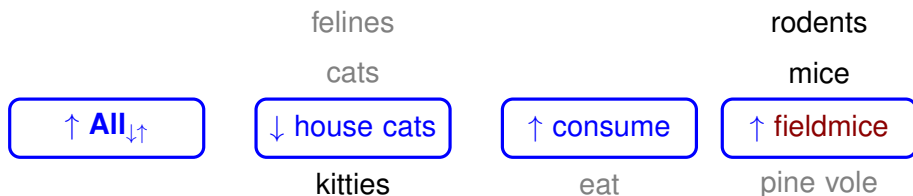
Mutations must respect *polarity*.



Modern Natural Logic

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

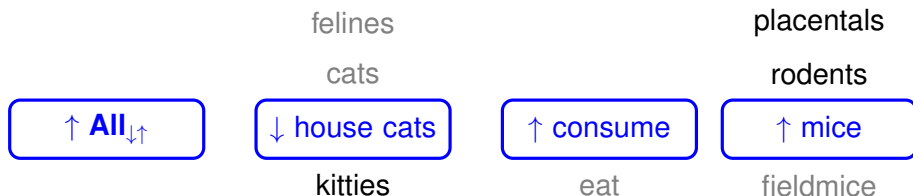
Mutations must respect *polarity*.



Modern Natural Logic

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

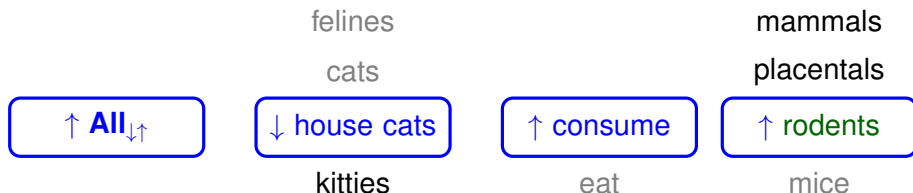
Mutations must respect *polarity*.



Modern Natural Logic

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

Mutations must respect *polarity*.

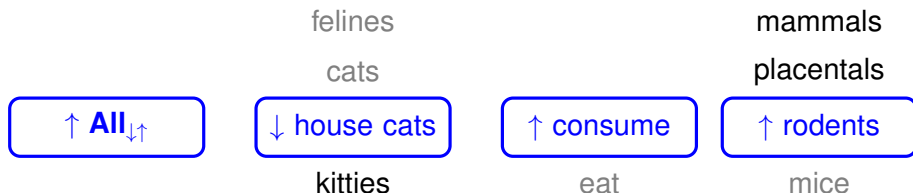


Modern Natural Logic

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

Mutations must respect *polarity*.

Not pictured: also handles negation



Natural Logic Q/A as Search

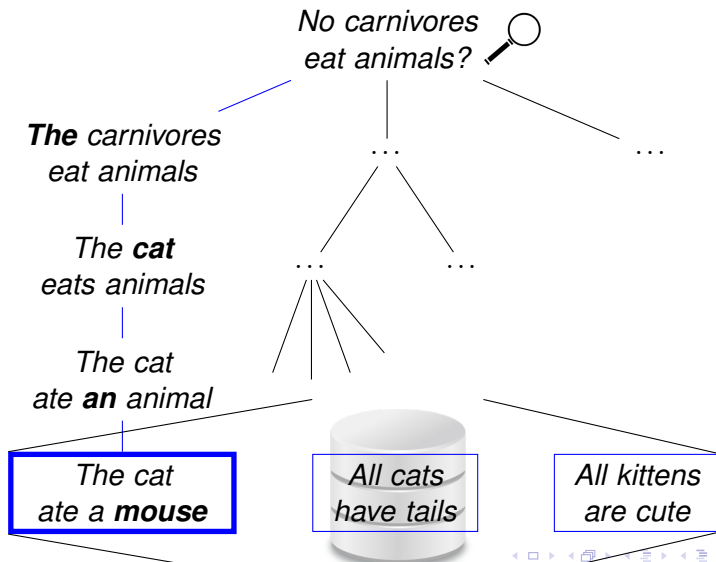
Shorthand for a node:



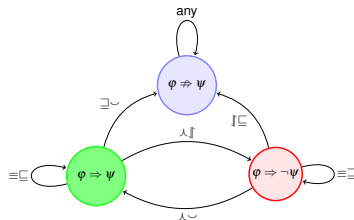
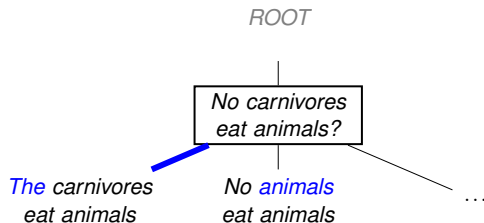
*No carnivores
eat animals?*



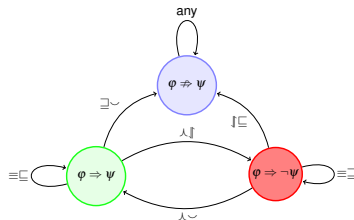
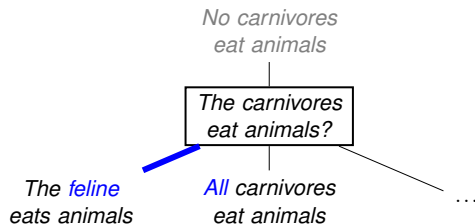
Natural Logic Q/A as Search



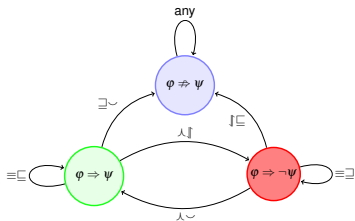
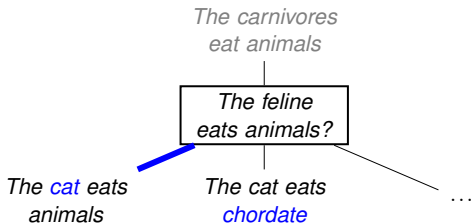
Natural Logic Q/A as Search



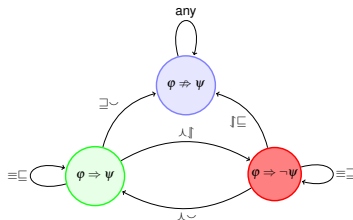
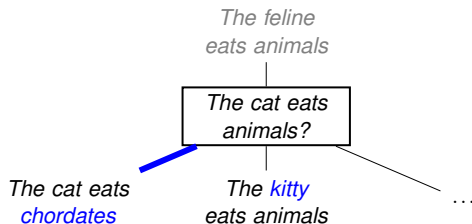
Natural Logic Q/A as Search



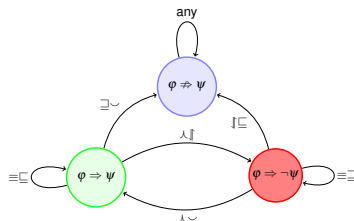
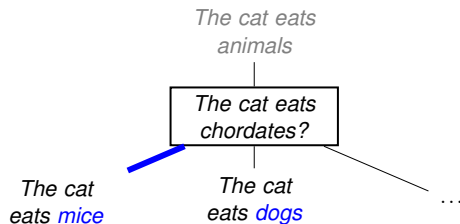
Natural Logic Q/A as Search



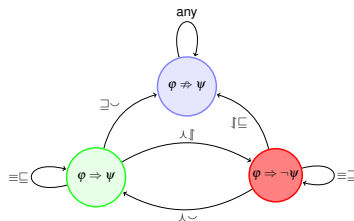
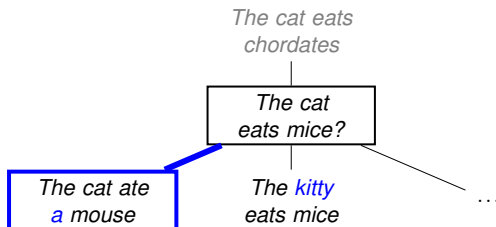
Natural Logic Q/A as Search



Natural Logic Q/A as Search

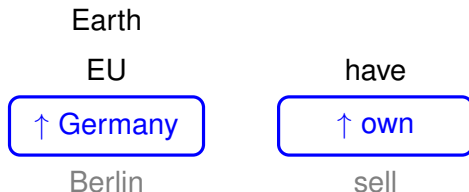


Natural Logic Q/A as Search



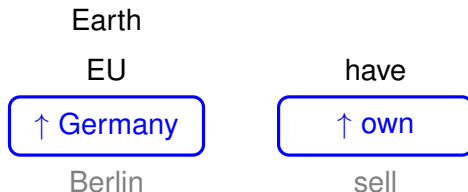
Three Contributions for Generalizable Inference

1. Partial order over meronymy + relations

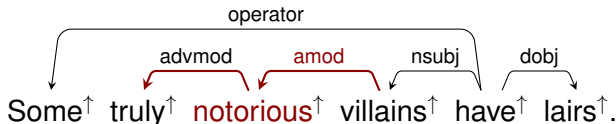


Three Contributions for Generalizable Inference

1. Partial order over meronymy + relations

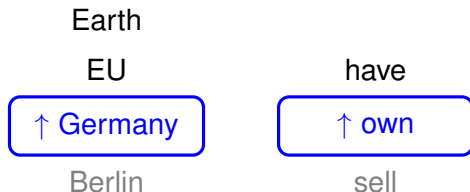


2. Natural Logic over dependency trees

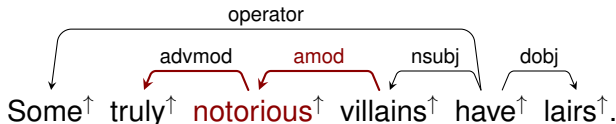


Three Contributions for Generalizable Inference

1. Partial order over meronymy + relations



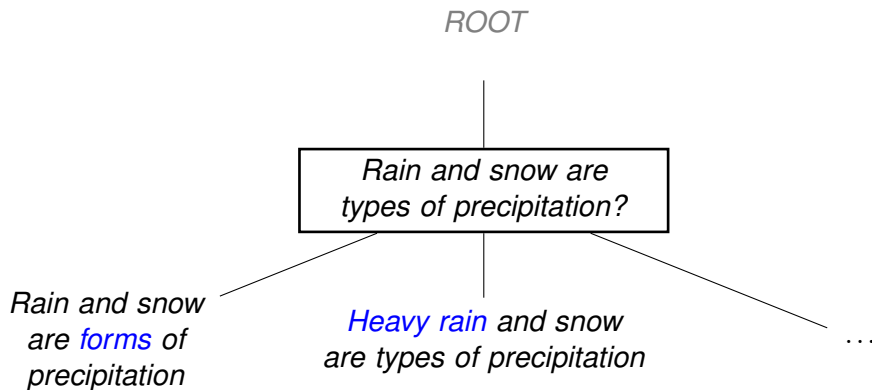
2. Natural Logic over dependency trees



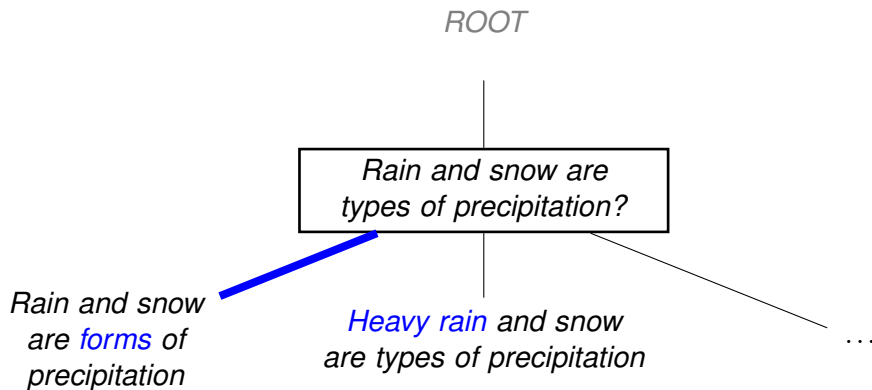
3. Hybrid statistical / logical solver



An Example Search



An Example Search



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

*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



Lexical Alignment Classifier

Forms of precipitation include rain and sleet

Rain and snow are forms of precipitation



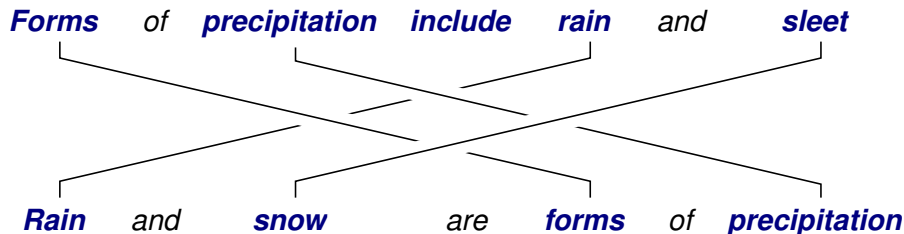
Lexical Alignment Classifier

Forms of precipitation include rain and sleet

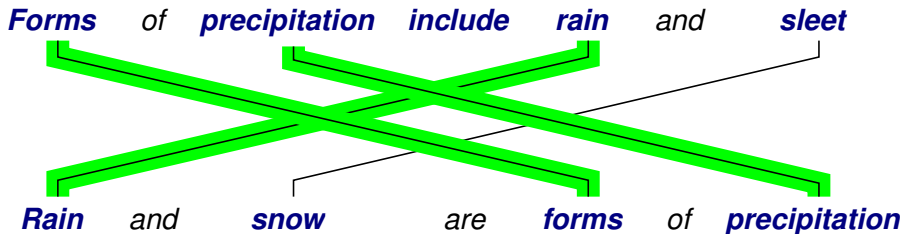
Rain and snow are forms of precipitation



Lexical Alignment Classifier



Lexical Alignment Classifier

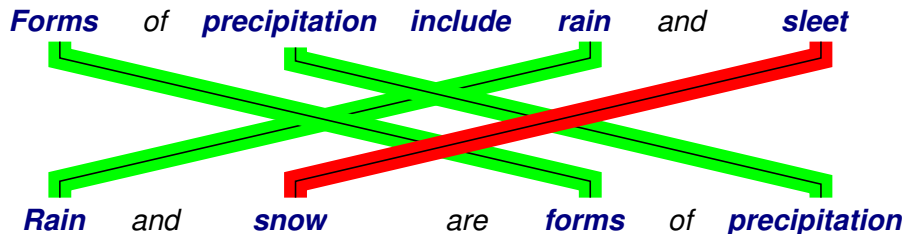


Features

1. Matching words



Lexical Alignment Classifier

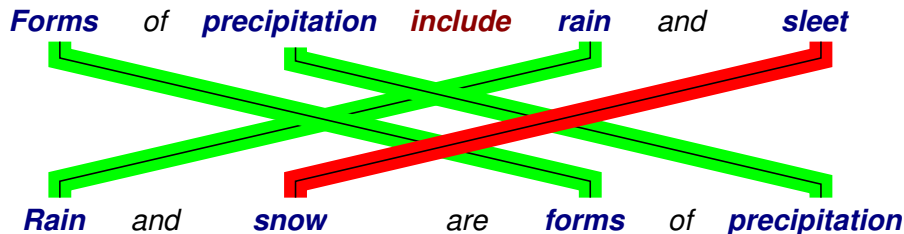


Features

1. Matching words
2. Mismatched words



Lexical Alignment Classifier

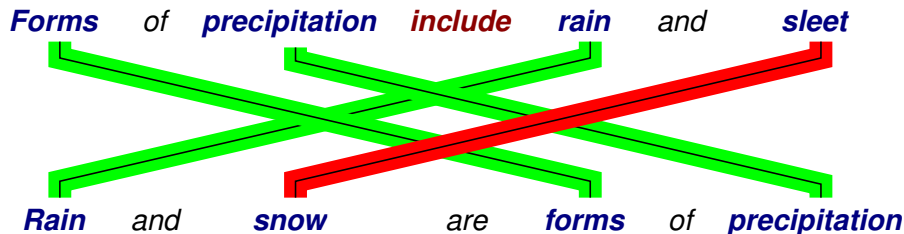


Features

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)



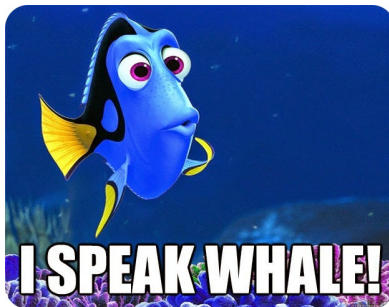
Old Problem: Logic + Lexical Classifiers

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

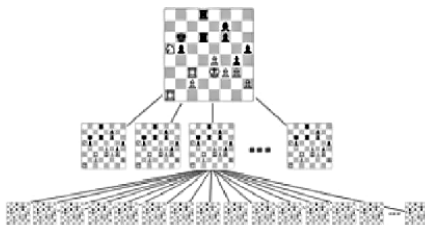
1. If we find a premise, great!



Big Picture

Run our usual search

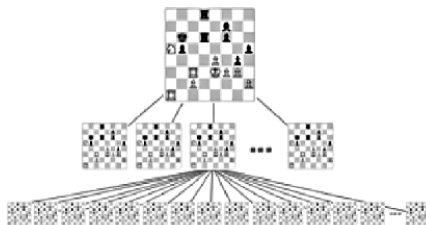
1. If we find a premise, great!
2. If not, use lexical classifier as an *evaluation function*



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

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



Dissecting Our Classifier

Anatomy of a Classifier

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

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

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



Dissecting Our Classifier

Anatomy of a Classifier

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

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

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

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

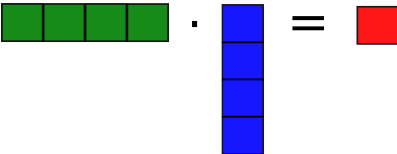


Incorporating our Evaluation Function

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

$$w^T f(x) = v$$


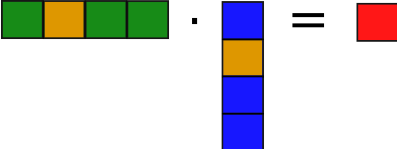


Incorporating our Evaluation Function

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

$$w^T f(x) = v$$


The diagram illustrates the dot product $w^T f(x) = v$. It shows a horizontal row of four colored squares (green, orange, green, green) representing the weight vector w^T . This is followed by a dot operator. To the right of the dot is a vertical column of four colored squares (blue, orange, blue, blue) representing the feature vector $f(x)$. This is followed by an equals sign. To the right of the equals sign is a single red square representing the result v .

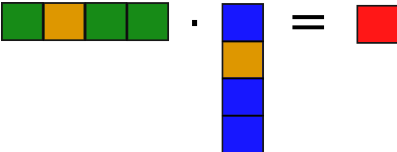


Incorporating our Evaluation Function

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

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


The diagram shows a horizontal vector of four colored squares (green, orange, green, green) multiplied by a vertical vector of four colored squares (blue, orange, blue, blue). The result is a single red square, representing the dot product of the two vectors.



Why is this Important?



Faster Search $\Rightarrow$ Deeper Reasoning

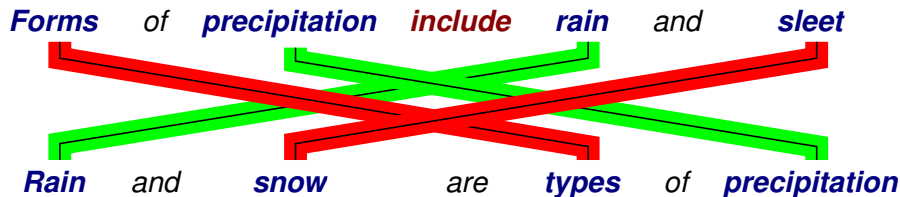
- **Speed:** Around 1M search states visited per second
- **Memory:** 32 byte search states

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

Memory: Never store intermediate fact as String.



Another Example Search

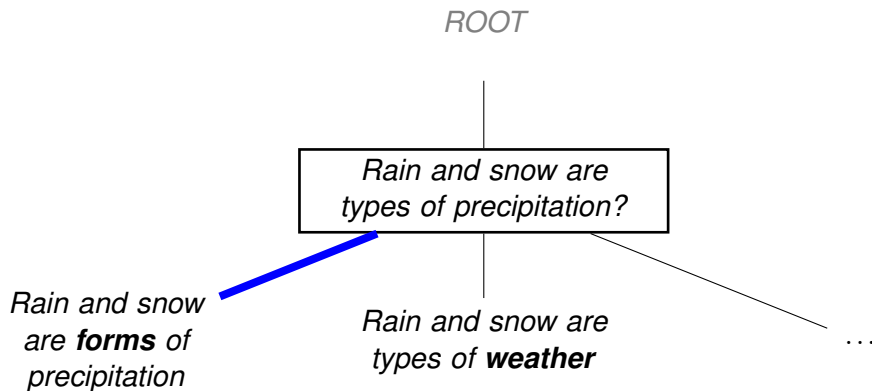


Score $w^T f(x)$: -0.5

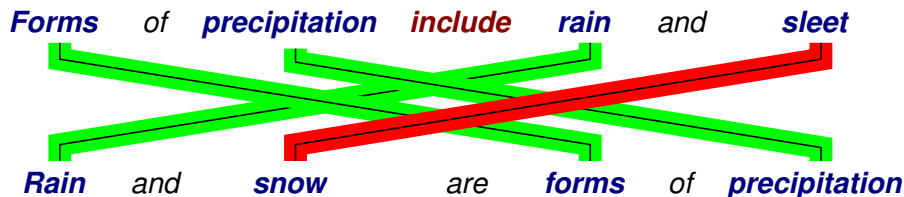
Feature	w	$f(x)$
Matching words	2.0	2
Mismatched words	-1.0	2
Unmatched premise	-0.5	1
Unmatched consequent	-0.75	0
Bias	-2.0	1



Another Example Search



Another Example Search

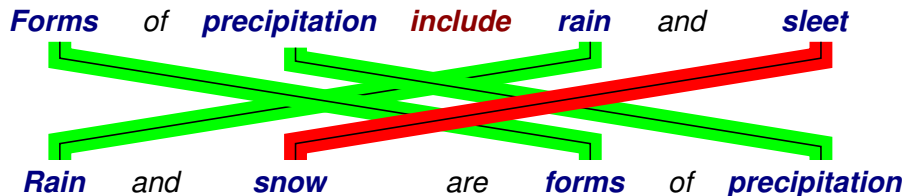


Score $w^T f(x)$: $-0.5 + 2 - 1$

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 $w^T f(x)$: 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 4th Grade Science

Multiple choice questions from real 4th grade science exams



Solving 4th Grade Science

Multiple choice questions from real 4th 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 4th Grade Science

Multiple choice questions from real 4th 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 4th Grade Science

Multiple choice questions from real 4th grade science exams

System	Train	Test
KNOWBOT	45	
KNOWBOT (ORACLE)	57	

[Hixon et al., 2015]



Solving 4th Grade Science

Multiple choice questions from real 4th grade science exams

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

[Hixon et al., 2015]



Solving 4th Grade Science

Multiple choice questions from real 4th 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 4th Grade Science

Multiple choice questions from real 4th 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 4th Grade Science

Multiple choice questions from real 4th 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 4th Grade Science

Multiple choice questions from real 4th 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 4th grade science!
(Though AI2's recent systems outperform this)

[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.



Bowman, S. (2014).

Can recursive neural tensor networks learn logical reasoning?

ICLR (arXiv:1312.6192).



Hixon, B., Clark, P., and Hajishirzi, H. (2015).

Learning knowledge graphs for question answering through conversational dialog.

NAACL.



References II



Icard III, T. and Moss, L. (2014).

Recent progress on monotonicity.

Linguistic Issues in Language Technology.



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.



References III

 Niu, F., Ré, C., Doan, A., and Shavlik, J. (2011).

Tuffy: Scaling up statistical inference in markov logic networks using an rdbms.

VLDB.

 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.

