Natural Language Inference with Monotonicity

Hai Hu

Indiana University

huhai@indiana.edu

2020 @ NALOMA @ WeSSLLI

Overview

Introduction

Natural Language Inference

Our goal is to solve inference problems in natural language such as the following:

entail, *contradict* or *neutral*? SICK dataset (2014)

P: A flute is being played by a girl
H: *There is no woman playing a flute*

Natural Language Inference

Our goal is to solve inference problems in natural language such as the following:

entail, *contradict* or *neutral*? SICK dataset (2014)

P: A flute is being played by a girl
H: *There is no woman playing a flute*

entail, *contradict* or *neutral*? FraCaS dataset (1996)

P1: Most Europeans are resident in Europe
P2: All Europeans are people
P3: All people who are resident in Europe can travel freely within Europe
H: *Most Europeans can travel freely within Europe*

Natural Language Inference

Our goal is to solve inference problems in natural language such as the following:

entail, *contradict* or *neutral*? SICK dataset (2014)

P: A flute is being played by a girl

H: *There is no woman playing a flute*

entail, *contradict* or *neutral*? FraCaS dataset (1996)

P1: Most Europeans are resident in Europe

P2: All Europeans are people

P3: All people who are resident in Europe can travel freely within Europe

H: *Most Europeans can travel freely within Europe*

Often referred to as **Natural Language Inference** (NLI) or **Recognizing Textual Entailment** (RTE).

2 Approaches

- **(Natural-)Logic-based**: tableau / translation into logical representations + a theorem prover / pure Natural Logic (MacCartney and Manning, 2008; Mineshima et al., 2015; Martínez-Gómez et al., 2017; Abzianidze, 2017; Yanaka et al., 2018; Kalouli et al., 2019; Hu et al., 2018, 2019)
- **Machine-learning-based**: many, e.g., RNN, ESIM, BERT family (Bowman et al., 2015; Chen et al., 2017; Devlin et al., 2019)

Our system: MonaLog

MonaLog = MOnotonicity and NAtural LOGic:

- Light-weight, no translation to logical representations
- Natural logic based, explainable, and easy to interpret
- Able to generate inferences for other purposes

Our system: MonaLog

MonaLog = MOnotonicity and NAtural LOGic:

- Light-weight, no translation to logical representations
- Natural logic based, explainable, and easy to interpret
- Able to generate inferences for other purposes

It can:

1. Solve natural language inference problems (e.g., SICK, FraCaS)
2. Generate natural language inferences (for data augmentation)
3. Create challenging monotonicity datasets

Monotonicity

Monotonicity

$$f(x)^{\uparrow} = 5 + x^{\uparrow}$$

$$f(x)^{\uparrow} = 5 - x^{\downarrow}$$

$$\text{every(man, walks)}^{\uparrow} = \text{every man}^{\downarrow} \text{ walks}^{\uparrow}$$

Monotonicity

$$f(x)^{\uparrow} = 5 + x^{\uparrow}$$

$$f(x)^{\uparrow} = 5 - x^{\downarrow}$$

$$\text{every(man, walks)}^{\uparrow} = \text{every man}^{\downarrow} \text{ walks}^{\uparrow}$$

Key intuition

Truth value holds if we:

replace *man* with a word/phrase denoting a **subset**,

or

replace *walk* with a word/phrase denoting a **superset**,

MonaLog

MonaLog pipeline

Task: Predict the semantic relation between an ordered sentence pair (Entailment, Neutral or Contradiction?)

- premise: *every dog dances*.
- hypothesis: *some cute poodle moves*.

MonaLog pipeline

Task: Predict the semantic relation between an ordered sentence pair (Entailment, Neutral or Contradiction?)

- premise: *every dog dances*.
- hypothesis: *some cute poodle moves*.

1) **Polarization**	$\rightarrow$ 2) **Generation**	$\rightarrow$ 3) **Search**
	E = {some cute animal moves, ... }	hypothesis $\in$ E?
every$^{\uparrow}$ *dog*$^{\downarrow}$ *dances*$^{\uparrow}$	N = {every animal moves, ... }	hypothesis $\in$ N?
	C = {no labrador dances, ...}	hypothesis $\in$ C?

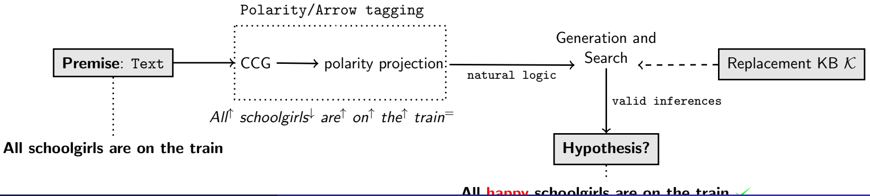
MonaLog pipeline

Task: Predict the semantic relation between an ordered sentence pair (Entailment, Neutral or Contradiction?)

- premise: *every dog dances*.
- hypothesis: *some cute poodle moves*.

1) **Polarization**	$\rightarrow$ 2) **Generation**	$\rightarrow$ 3) **Search**
every$^{\uparrow}$ *dog*$^{\downarrow}$ *dances*$^{\uparrow}$	E = {some cute animal moves, ... } N = {every animal moves, ... } C = {no labrador dances, ...}	hypothesis $\in$ E? hypothesis $\in$ N? hypothesis $\in$ C?

In detail:



Polarization algorithm

Input: raw sentences: *every man walks*

Step 1: get CCG parse tree using CandC or EasyCCG parser (Clark and Curran, 2007; Lewis and Steedman, 2014)

Step 2: *mark* and *polarize* (van Benthem, 1986; Sánchez-Valencia, 1991; Hu and Moss, 2018)

Output: every$^{\uparrow}$ man$^{\downarrow}$ walks$^{\uparrow}$

Provably correct compared to MacCartney and Manning (2008).

Hu, H. and Moss, L. S. (2018). Polarity computations in flexible categorial grammar.
In *Proceedings of *SEM*, pages 124–129

Examples of polarized sentences

$No^{\uparrow}$ $man^{\downarrow}$ $walks^{\downarrow}$

$Every^{\uparrow}$ $man^{\downarrow}$ $and^{\uparrow}$ $no^{\uparrow}$ $woman^{\downarrow}$ $sleeps^{=}$

$If^{\uparrow}$ $some^{\downarrow}$ $man^{\downarrow}$ $walks^{\downarrow}$, $then^{\uparrow}$ $no^{\uparrow}$ $woman^{\downarrow}$ $runs^{\downarrow}$

$Every^{\uparrow}$ $man^{\downarrow}$ $does^{\downarrow}$ $n't^{\uparrow}$ $hit^{\downarrow}$ $every^{\downarrow}$ $dog^{\uparrow}$

$Every^{\uparrow}$ $young^{\downarrow}$ $man^{\downarrow}$ $that^{\uparrow}$ $no^{\uparrow}$ $young^{\downarrow}$ $woman^{\downarrow}$ $hits^{\uparrow}$ $cried^{\uparrow}$

$At^{\uparrow}$ $least^{\uparrow}$ $seven^{\downarrow}$ $fish^{\uparrow}$ $died^{\uparrow}$ $yesterday^{\uparrow}$ $in^{\uparrow}$ $Morocco^{\uparrow}$

$A^{\uparrow}$ $dog^{\uparrow}$ $who^{\uparrow}$ $ate^{\uparrow}$ $two^{\downarrow}$ $rotten^{\downarrow}$ $biscuits^{\downarrow}$ $was^{\uparrow}$ $sick^{\uparrow}$ $for^{\uparrow}$ $three^{\downarrow}$ $days^{\downarrow}$

Generation

MonaLog:

Input: sentence pair: *every*$^{\uparrow}$ *man*$^{\downarrow}$ *walks*$^{\uparrow}$?? *some young man moves*

Generation

MonaLog:

Input: sentence pair: *every*$^{\uparrow}$ *man*$^{\downarrow}$ *walks*$^{\uparrow}$?? *some young man moves*

Step 1: Build **knowledge base**: extract all adjs, nouns, adverbs, verbs, RC, and add

Generation

MonaLog:

Input: sentence pair: *every*$^{\uparrow}$ *man*$^{\downarrow}$ *walks*$^{\uparrow}$?? *some young man moves*

Step 1: Build **knowledge base**: extract all adjs, nouns, adverbs, verbs, RC, and add

1. $a\ n \leq n$, $n\ p \leq n$, and $n\ r \leq n$. (*small dog* $\leq$ *dog*, *dog from France* $\leq$ *dog*, *dog that barks* $\leq$ *dog*)
2. $v\ a \leq v$. (*walk fast* $\leq$ *walk*)
3. WordNet information: *poodle* $\leq$ *dog*, *dog* $|$ *cat*, *big* $\perp$ *small*

Generation

MonaLog:

Input: sentence pair: *every*$^{\uparrow}$ *man*$^{\downarrow}$ *walks*$^{\uparrow}$?? *some young man moves*

Step 1: Build **knowledge base**: extract all adjs, nouns, adverbs, verbs, RC, and add

1. $a\ n \leq n$, $n\ p \leq n$, and $n\ r \leq n$. (*small dog* $\leq$ *dog*, *dog from France* $\leq$ *dog*, *dog that barks* $\leq$ *dog*)
2. $v\ a \leq v$. (*walk fast* $\leq$ *walk*)
3. WordNet information: *poodle* $\leq$ *dog*, *dog* $|$ *cat*, *big* $\perp$ *small*

Step 2: use "substitution" to generate *entailments* and *neutrals*, and other simple rules for *contradictions*

Generation

MonaLog:

Input: sentence pair: $every^{\uparrow}$ $man^{\downarrow}$ $walks^{\uparrow}$?? *some young man moves*

Step 1: Build **knowledge base**: extract all adjs, nouns, adverbs, verbs, RC, and add

1. $a\ n \leq n$, $n\ p \leq n$, and $n\ r \leq n$. (*small dog* $\leq$ *dog*, *dog from France* $\leq$ *dog*, *dog that barks* $\leq$ *dog*)
2. $v\ a \leq v$. (*walk fast* $\leq$ *walk*)
3. WordNet information: *poodle* $\leq$ *dog*, *dog* $|$ *cat*, *big* $\perp$ *small*

Step 2: use "substitution" to generate *entailments* and *neutrals*, and other simple rules for *contradictions*

Step 3: check if the hypothesis is in one of the generated sets

Generation

MonaLog:

Input: sentence pair: $every^{\uparrow}$ $man^{\downarrow}$ $walks^{\uparrow}$?? *some young man moves*

Step 1: Build **knowledge base**: extract all adjs, nouns, adverbs, verbs, RC, and add

1. $a\ n \leq n$, $n\ p \leq n$, and $n\ r \leq n$. (*small dog* $\leq$ *dog*, *dog from France* $\leq$ *dog*, *dog that barks* $\leq$ *dog*)
2. $v\ a \leq v$. (*walk fast* $\leq$ *walk*)
3. WordNet information: *poodle* $\leq$ *dog*, *dog* $\mid$ *cat*, *big* $\perp$ *small*

Step 2: use "substitution" to generate *entailments* and *neutrals*, and other simple rules for *contradictions*

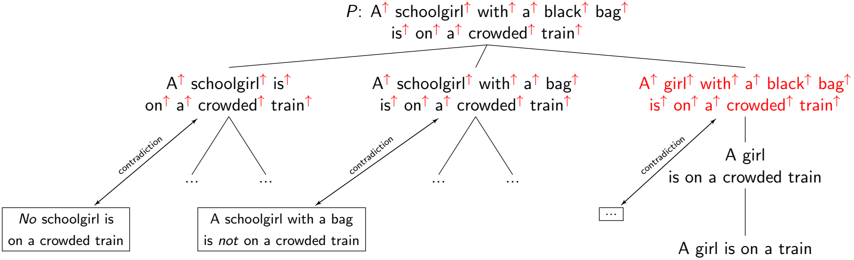
Step 3: check if the hypothesis is in one of the generated sets

Output: Entailment b/c hypothesis $\in$ *entailments*

Hu, H., Chen, Q., Richardson, K., Mukherjee, A., Moss, L. S., and Kuebler, S. (2020). MonaLog: a lightweight system for natural language inference based on monotonicity.
In *Proceedings of the SCiL*, pages 319–329

Generation: a search tree

Use “substitution” to generate *entailments* and *contradictions*



Experiments using MonaLog

Experiments on SICK

- SICK (Sentences Involving Compositional Knowledge)
- 10,000 English sentence pairs, generated from image, video descriptions, annotated by crowd workers (Marelli et al., 2014).

Experiments on SICK

- SICK (Sentences Involving Compositional Knowledge)
- 10,000 English sentence pairs, generated from image, video descriptions, annotated by crowd workers (Marelli et al., 2014).

premise	hypothesis	orig. label	corr. label
There is no girl in white dancing	A girl in white is dancing	C	C
Two girls are lying on the ground	**Two girls are sitting on the ground**	**N**	**C**
A couple who have just got married are walking down the isle	The bride and the groom are leaving after the wedding	E	N
A girl is on a jumping car	One girl is jumping on the car	E	N (?)

Table: Examples from SICK and corrected SICK (Kalouli et al., 2018).

Exp 1: set-up

Solve SICK, using MonaLog (+ BERT)

- MonaLog:

1. Syntactic transformations:

a) pass2act; b) there be no N doing sth. $\rightarrow$ No N is doing sth.

2. Generate entailments and contradictions from *premise*.
3. If *hypothesis* in E/C, then return E/C, else return Neutral.

- MonaLog + BERT:

If MonaLog returns E/C, then use MonaLog, else use BERT.

Exp 1: Results

system	P	R	acc.
On **uncorrected** SICK			
majority baseline	–	–	56.36
MonaLog (this work)			
MonaLog + all transformations	83.75	70.66	77.19
Hybrid: MonaLog + BERT	83.09	85.46	85.38
ML/DL-based systems			
BERT (base, uncased)	86.81	85.37	86.74
Yin and Schütze (2017)	–	–	**87.1**
Logic-based systems			
Abzianidze (2015)	97.95	58.11	81.35
Yanaka et al. (2018)	84.2	77.3	84.3
On **corrected** SICK			
MonaLog + all transformations	89.91	74.23	81.66
Hybrid: MonaLog + BERT	85.65	87.33	**85.95**
BERT (base, uncased)	84.62	84.27	85.00

Decent performance on uncorrected SICK. Need to fully correct SICK.

Exp 2: Data augmentation

Experiment 2

1. Pair the generated entailments/contradictions with the input premise.
2. Add newly generated pairs to SICK.train. Fine-tune BERT.

Exp 2: generated NLI pairs

Sentence pairs generated by MonaLog, lemmatized:

label	premise	hypothesis	comm.
E	A woman be not cooking something	A person be not cooking something	correct
E	A man be talk to a woman who be seat beside he and be drive a car	A man be talk	correct
E	A south African plane be not fly in a blue sky	A south African plane be not fly in a very blue sky in a blue sky	unnat.
C	No panda be climb	Some panda be climb	correct
C	A man on stage be sing into a microphone	A man be not sing into a microphone	correct
C	No man rapidly be chop some mushroom with a knife	Some man rapidly be chop some mushroom with a knife with a knife	unnat.
E	Few$^{\uparrow}$ people$^{\downarrow}$ be$^{\downarrow}$ eat$^{\downarrow}$ at$^{\downarrow}$ red$^{\downarrow}$ table$^{\downarrow}$ in$^{\downarrow}$ a$^{\downarrow}$ restaurant$^{\downarrow}$ without$^{\downarrow}$ light$^{\uparrow}$	Few$^{\uparrow}$ large$^{\downarrow}$ people$^{\downarrow}$ be$^{\downarrow}$ eat$^{\downarrow}$ at$^{\downarrow}$ red$^{\downarrow}$ table$^{\downarrow}$ in$^{\downarrow}$ a$^{\downarrow}$ Asian$^{\downarrow}$ restaurant$^{\downarrow}$ without$^{\downarrow}$ light$^{\uparrow}$	correct

No incorrect labels, but $\sim$10% unnatural.

Exp 2: Results

Results of BERT trained on MonaLog-augmented data:

training data	# E	# N	# C	acc.
SICK.train: baseline	1.2k	2.5k	0.7k	85.00
all gen. + SICK.train	30k	2.5k	14k	86.51
E, C prob. threshold = 0.95	30k	2.5k	14k	86.71
Hybrid baseline	1.2k	2.5k	0.7k	85.95
Hybrid + all gen.	30k	2.5k	14k	87.16
Hybrid + all gen. + threshold	30k	2.5k	14k	**87.49**

Shows the usefulness and high-quality of MonaLog generated data.
(Observation: BERT is insensitive to skewed dataset.)

Exp 3: Creating NLI Datasets

Motivation

- How difficult is monotonicity inference for machine learning models?

Exp 3: Creating NLI Datasets

Motivation

- How difficult is monotonicity inference for machine learning models?
- MonaLog $\rightarrow$ free monotonicity inferences.

Exp 3: Creating NLI Datasets

Motivation

- How difficult is monotonicity inference for machine learning models?
- MonaLog $\rightarrow$ free monotonicity inferences.

Richardson, K., Hu, H., Moss, L. S., and Sabharwal, A. (2020). Probing Natural Language Inference Models through Semantic Fragments.

In *Proceedings of AAAI*

Research questions

Can models (neural and transformers) learn monotonicity?

Research questions

Can models (neural and transformers) learn monotonicity?

- Can monotonicity be learned from scratch?
- Can models trained on general NLI datasets do monotonicity (zero-shot)?
- Can these models be re-trained to master monotonicity?

Also other semantic phenomena: *negation*, *quantifier*, *counting*, ...
$\rightarrow$ semantic fragments

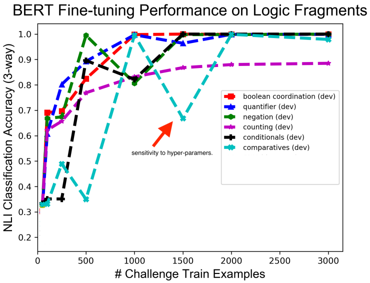
Creating monotonicity datasets

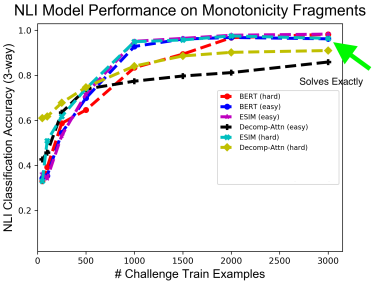
1. Write grammar rules and determine the vocabulary:
   All black mammals saw exactly 5 stallions who danced
2. Use MonaLog to generate Entailments, Neutrals and Contradictions:
   A brown or black poodle saw exactly 5 stallions who danced

Question 1: Can We Learn Fragments from Scratch?

- Training *task-specific* models without special NLI pre-training

- Training *task-specific* models without special NLI pre-training

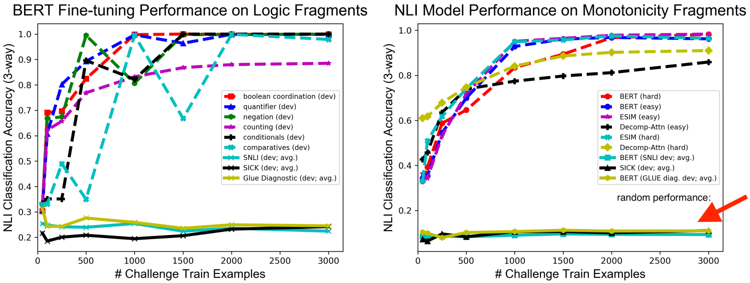




- BERT (+ ESIM, Decomposable-Attention) can easily learn most fragments. Difficult for other LSTM-based models/baselines.

Question 1: Can We Learn Fragments from Scratch?

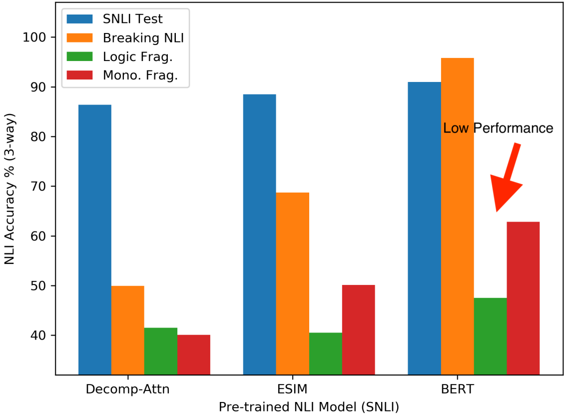
- Training *task-specific* models without special NLI pre-training



- **The Problem**: models are just **idiot savants**, cannot solve any other tasks (common probing strategy **but not always insightful**).

Question 2: Zero-shot Evaluation

- How do models trained on NLI benchmarks perform?



- Pre-trained NLI **models perform poorly**, provides a new task that break models; but does this tell us much?

The Biggest Challenge

Can we build models that are simultaneously good at our diagnostic tasks and their original benchmarks?

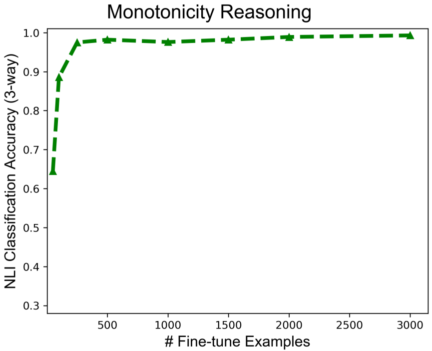
Assumption: A model's ability to quickly learn new tasks with limited *cost* (i.e., forgetting of original task) provides evidence of competence.

Question 3: Can Models be Fixed? (Most interesting)

- **Model Inoculation** (Liu et al. (2019)): Continue training models on small amounts of diagnostic data; aim to (quickly/cheaply) fix model.

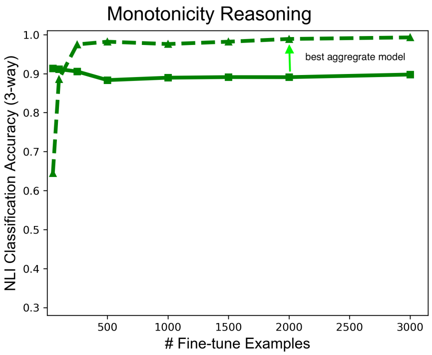
Question 3: Can Models be Fixed? (Most interesting)

- **Model Inoculation** (Liu et al. (2019)): Continue training models on small amounts of diagnostic data; aim to (quickly/cheaply) fix model.



Question 3: Can Models be Fixed? (Most interesting)

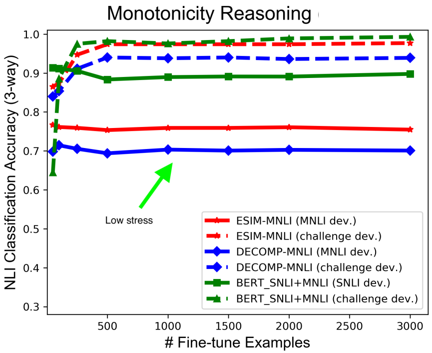
- **Model Inoculation** (Liu et al. (2019)): Continue training models on small amounts of diagnostic data; aim to (quickly/cheaply) fix model.



- **Loss-less Inoculation**: Models should be penalized for forgetting (a sign of stress), take best aggregate model.

Question 3: Can Models be Fixed? (Most interesting)

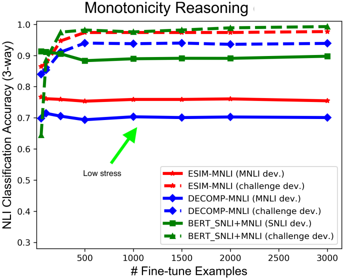
- **Model Inoculation** (Liu et al. (2019)): Continue training models on small amounts of diagnostic data; aim to (quickly/cheaply) fix model.

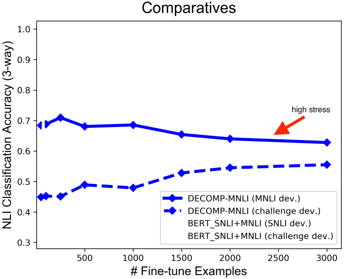


- Mastering diagnostic tasks with little loss gives evidence of competence and strong correspondence to training distribution.

Question 3: Can Models be Fixed? (Most interesting)

- **Model Inoculation** (Liu et al. (2019)): Continue training models on small amounts of diagnostic data; aim to (quickly/cheaply) fix model.

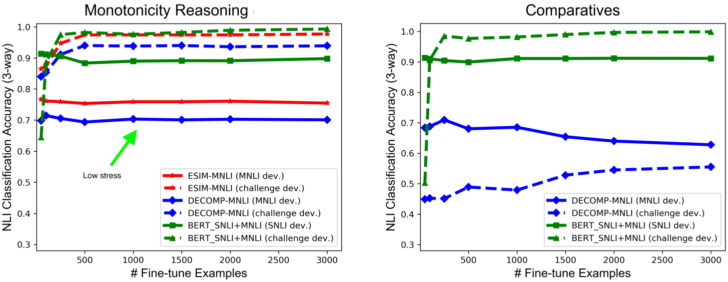




- **Not all fragments are the same**: some stress models (i.e., lead to forgetting) more than others; indicate lack of competence.

Question 3: Can Models be Fixed? (Most interesting)

- **Model Inoculation** (Liu et al. (2019)): Continue training models on small amounts of diagnostic data; aim to (quickly/cheaply) fix model.



- **General finding**: more robust models (e.g., BERT) learn fast and with less forgetting; indication of higher competence.

Summary

Summary

- We built a light-weight, interpretable Natural-Logic-based NLI system with decent performance on NLI datasets.
- Our system can generate high-quality NLI sentence pairs which are useful for data augmentation and dataset creation.
- Future work:
  - evaluation of the polarization accuracy;
  - extend to wider natural logic phenomena;
  - fully corrected SICK dataset;
- Questions & comments?

Links for our programs

- Github repository for polarization algorithm: ccg2mono
- Github repository for MonaLog: MonaLog
- Github repository for code and data for experiment 3: semantic_fragments

Acknowledgments

This series of work will not be possible without our wonderful collaborators:

Larry Moss	Kyle Richardson
Sandra Kübler	Qi Chen
Atreyee Mukherjee	Thomas Icard
Ashish Sabharwal	

and colleagues at Indiana University.

Hai Hu is partly supported by China Scholarship Council.

Examples from other fragments

Fragments	Example (premise,label,hypothesis)
Negation	*Laurie has only visited Nephi, Marion has only visited Calistoga.* CONTRADICTION *Laurie didn't visit Nephi*
Boolean	*Travis, Arthur, Henry and Dan have only visited Georgia* ENTAILMENT *Dan didn't visit Rwanda*
Quantifier	*Everyone has visited every place* NEUTRAL *Virgil didn't visit Barry*
Counting	*Nellie has visited Carrie, Billie, John, Mike, Thomas, Mark, .., and Arthur.* ENTAILMENT *Nellie has visited more than 10 people.*
Conditionals	*Francisco has visited Potsdam and if Francisco has visited Potsdam then Tyrone has visited Pampa* ENTAILMENT *Tyrone has visited Pampa.*
Comparatives	*John is taller than Gordon and Erik..., and Mitchell is as tall as John* NEUTRAL *Erik is taller than Gordon.*
Monotonicity	*All black mammals saw exactly 5 stallions who danced* ENTAILMENT *A brown or black poodle saw exactly 5 stallions who danced*

References I

Abzianidze, L. (2015). A tableau prover for natural logic and language. In *Proceedings of EMNLP*, pages 2492–2502.

Abzianidze, L. (2017). Langpro: Natural language theorem prover. *CoRR*, abs/1708.09417.

Bowman, S. R., Angeli, G., Potts, C., and Manning, C. D. (2015). A large annotated corpus for learning natural language inference. In *Proceedings of EMNLP*, pages 632–642.

Chen, Q., Zhu, X., Ling, Z.-H., Wei, S., Jiang, H., and Inkpen, D. (2017). Enhanced LSTM for natural language inference. In *Proceedings of ACL*, volume 1, pages 1657–1668.

Clark, S. and Curran, J. R. (2007). Wide-coverage efficient statistical parsing with CCG and log-linear models. *Computational Linguistics*, 33(4):493–552.

References II

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Hu, H., Chen, Q., and Moss, L. S. (2019). Natural language inference with monotonicity. In *Proceedings of IWCS*.

Hu, H., Chen, Q., Richardson, K., Mukherjee, A., Moss, L. S., and Kuebler, S. (2020). MonaLog: a lightweight system for natural language inference based on monotonicity. In *Proceedings of the SCiL*, pages 319–329.

Hu, H., Icard, T. F., and Moss, L. S. (2018). Automated reasoning from polarized parse trees. In *Proceedings of the Fifth Workshop on Natural Language and Computer Science*.

Hu, H. and Moss, L. S. (2018). Polarity computations in flexible categorial grammar. In *Proceedings of *SEM*, pages 124–129.

References III

Kalouli, A.-L., Crouch, R., and de Paiva, V. (2019). GKR: Bridging the gap between symbolic/structural and distributional meaning representations. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 44–55, Florence, Italy. Association for Computational Linguistics.

Kalouli, A.-L., Real, L., and de Paiva, V. (2018). Wordnet for "easy" textual inferences. In *Proceedings of LREC*, Miyazaki, Japan.

Lewis, M. and Steedman, M. (2014). A* CCG parsing with a supertag-factored model. In *Proceedings of EMNLP*, pages 990–1000.

Liu, N. F., Schwartz, R., and Smith, N. A. (2019). Inoculation by fine-tuning: A method for analyzing challenge datasets. *arXiv preprint arXiv:1904.02668*.

MacCartney, B. and Manning, C. D. (2008). Modeling semantic containment and exclusion in natural language inference. In *Proceedings of COLING*, pages 521–528. Association for Computational Linguistics.

References IV

Marelli, M., Menini, S., Baroni, M., Bentivogli, L., Bernardi, R., and Zamparelli, R. (2014). A SICK cure for the evaluation of compositional distributional semantic models. In *Proceedings of LREC*.

Martínez-Gómez, P., Mineshima, K., Miyao, Y., and Bekki, D. (2017). On-demand injection of lexical knowledge for recognising textual entailment. In *Proceedings of EACL*, pages 710–720.

Mineshima, K., Martínez-Gómez, P., Miyao, Y., and Bekki, D. (2015). Higher-order logical inference with compositional semantics. In *Proceedings of EMNLP*, pages 2055–2061.

Richardson, K., Hu, H., Moss, L. S., and Sabharwal, A. (2020). Probing Natural Language Inference Models through Semantic Fragments. In *Proceedings of AAAI*.

Sánchez-Valencia, V. (1991). *Studies on Natural Logic and Categorial Grammar*. PhD thesis, Universiteit van Amsterdam.

van Benthem, J. (1986). *Essays in Logical Semantics*. Reidel, Dordrecht.

References V

Yanaka, H., Mineshima, K., Martínez-Gómez, P., and Bekki, D. (2018). Acquisition of phrase correspondences using natural deduction proofs. In *Proceedings of NAACL*, pages 756–766, New Orleans, LA.

Yin, W. and Schütze, H. (2017). Task-specific attentive pooling of phrase alignments contributes to sentence matching. In *Proceedings of EACL*, pages 699–709.

The End