Sarcasm: Understanding & Generation

Smaranda Muresan (smara@columbia.edu)

Natural Language Processing
Columbia University

Understanding People's Attitudes is Important

News Item

SFGATE NEWS SPORTS BUSINESS ENTERTAINMENT FOOD LIVING TRAVEL REAL ESTATE C

Man shot to death near Occupy Oakland camp

By Matthai Kuruvila and Demian Bulwa Published 4:00 am, Friday, November 11, 2011

ADVERTISEMENT

Discussion Forum

User1: A shooting in Oakland? That NEVER happens.

User2: ***Shootings happen in Oakland all the time*** and it had nothing to do with the Occupy movement. […]

User3: This shooting does have something to do with the Occupy movement because many of the witnesses are the Occupiers and it happened only a few yards away from the encampment.

Very sad news, regardless of if this young man was w/ #OO or not URL… via @sfgate

Oh yay. Another shooting #sarcasm

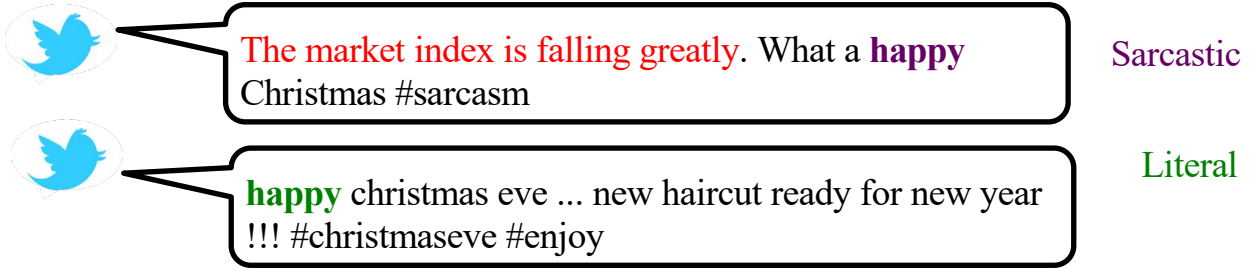
Unrecognized Sarcasm will lead to erroneous belief and sentiment detection

Verbal Irony/Sarcasm

Type of Figurative Language

Verbal Irony: the use of words to express something other than and especially the opposite of the literal meaning (Mirriam-Webster dictionary)

Sarcasm: the use of words that mean the opposite of what you really want to say especially in order to insult someone, to show irritation, or to be funny". (Mirriam-Webster dictionary)



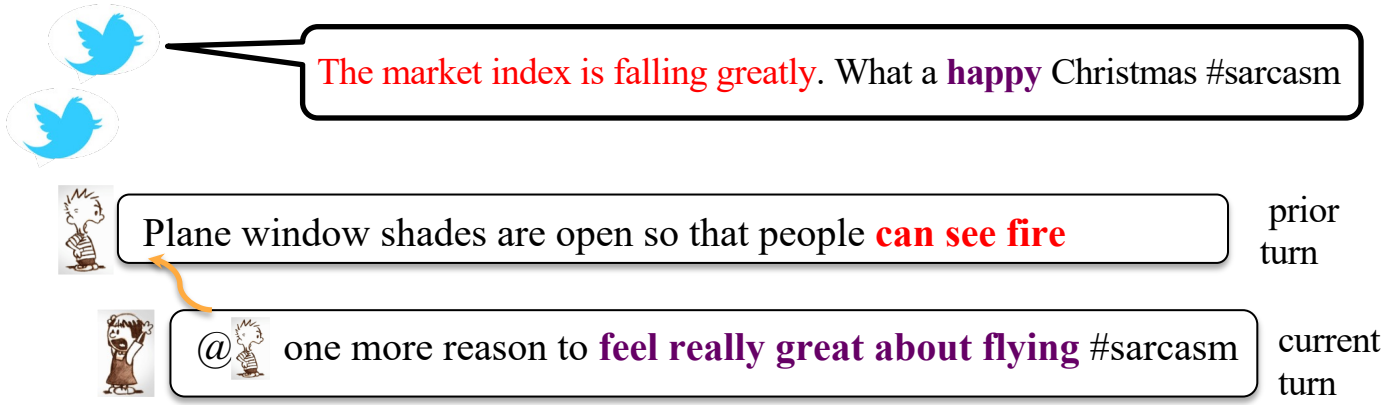
Verbal Irony/Sarcasm

Characteristics:

Irony Markers*: *"A shooting in Oakland? That **NEVER** happens"

- typographics, punctuation, hyperbole, interjection

***Irony Factors*: Evaluative, Reversal of Valence,** Semantic Incongruity with context



Overview

- Sarcasm Understanding: Textual Entailment + Explanation
- Sarcasm Generation
- Sarcasm Detection: Modeling the conversation context

FLUTE: Figurative Language Understanding through Textual Explanations (EMNLP 2022)

Collaborators:

Tuhin Chakrabarty

Arkadiy Saakyan

Debanjan Ghosh

Figurative Language Understanding as Textual Entailment

Understanding Figurative Language (e.g., sarcasm, metaphor, simile, idioms) can be framed as Recognizing Textual Entailment (a.k.a, Natural Language Inference) task (Chakrabarty et al., 2021, Stowe et al., 2022; Srivastava et al, 2022)

Hypothesis: *The place looked like a fortress.*
Premise: *The place looked impenetrable and inescapable.*

Entailment

Hypothesis: *I love going to the dentist.*
Premise: *I hate going to the dentist.*

Contradiction

Figurative Language Understanding as Textual Entailment

Issues in current datasets

- Similarly to RTE/NLI datasets they suffer from spurious correlations
- Sometimes the label is associated with the overall sentence and not the meaning of the figurative language expression
- Not all have entailment *and* contradictions pairs for same figurative language type

FLUTE: Figurative Language Understanding through Textual Explanations

- Desiderata:
  - Frame the task as natural language inference (NLI) *with explanation generation* for label prediction (similar to e-SNLI)
  - 4 figurative language types: sarcasm, simile, metaphors and idioms
  - Entailment and contradiction labels refer to the *meaning of the figurative language expression* rather than other aspects of the sentence
  - More varied rewriting of hypothesis (figurative) and premise (literal) pairs to minimize trivial predictions

Premise

I have not had any sleep for the past three days only to come back home and my neighbor is having a loud party

A neighbor having a loud party can be disruptive and keep someone from getting much-needed sleep, so being thankful about it is unrealistic

I am really thankful that my neighbor is having a loud party when I have not had any sleep for the past three days

Explanation

Loud Party	→	Disruptive
Disruption	→	Lack of Sleep
Lack of Sleep	→	Angry / Upset

Hypothesis

Implicit meaning that requires multiple reasoning steps to interpret

Premise

It's completely unacceptable and inappropriate that you believe you can blackmail me like this.

To be beyond the pale means to be completely unacceptable or inappropriate, and in this context the speaker is saying that it is unacceptable for the other person to try and blackmail them

Explanation

It's beyond the pale that you believe you can blackmail me like this.

Hypothesis

Non Compositional

beyond + the + pale

≠

completely unacceptable and inappropriate

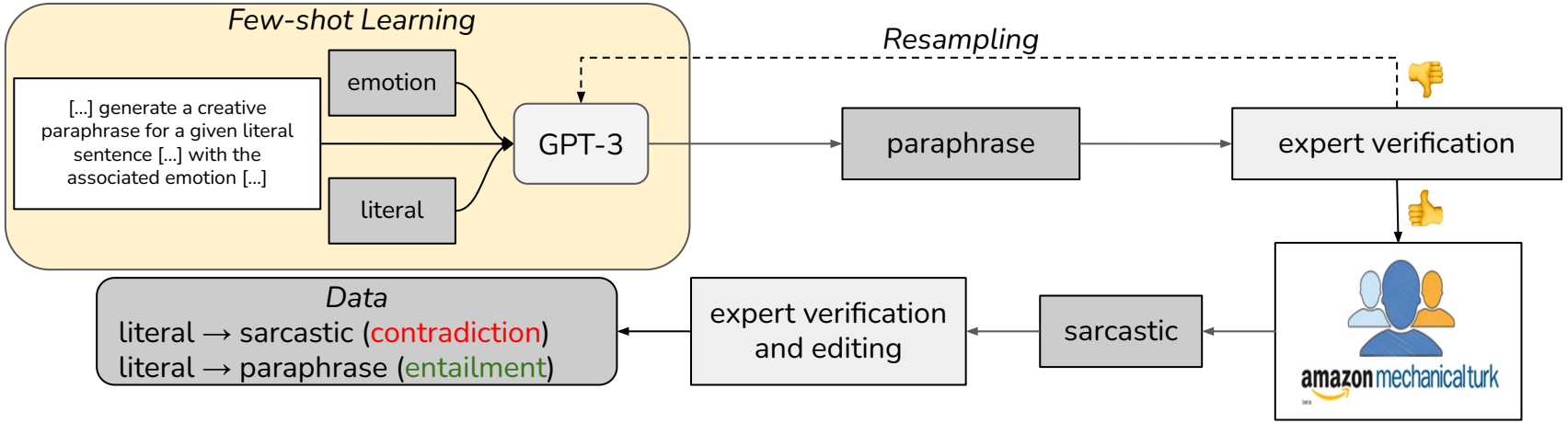
FLUTE: Figurative Language Understanding through Textual Explanations

- Create a dataset, FLUTE, using a model-in-loop approach (GPT-3) containing ~9000 pairs across 4 figurative language types
- Implement a baseline model for figurative language understanding as e-NLI
- Introduce a new evaluation metric to combine label accuracy with explanation quality
- Automatic and Human Evaluation

Generating NLI pairs for Sarcasm

- Where to start?
  - Getting sarcastic messages from Twitter seems restrictive by the #
    - => what about starting from literal sentences?
    - Empathetic Dialogues (Literal sentences annotated with negative emotions)
  - Crowdworkers on AMturk are
    - Not good at doing large edits to create diverse entailment pairs
    - Good at doing minimum edits and making sure sentence is entail or contradicted
  - GPT-3 Is not not good at generating sarcasm (yet) but is great at paraphrasing

Generating NLI pairs for Sarcasm



Empathetic Dialogues dataset

My next door neighbors are always arguing in our shared hallway.
+
Angry

GPT-3

It's so *annoying* to have to hear my next door neighbors argue all the time in our shared hallway.

MTurk

It's so *pleasant* to have to hear my next door neighbors argue all the time in our shared hallway.

Prompt Used for GPT-3 to generate Paraphrases

You will be presented with examples of some literal input sentences and their creative paraphrases. For each example, we also provide the associated emotion. Your task is then to generate a creative paraphrase for a given literal sentence where the creative paraphrase should reflect the associated emotion without changing its meaning. Make sure to use some sort of humor and commonsense about everyday events and concepts

1) **Literal**: A lot of people have got engaged recently.

Emotion: surprised

Creative Paraphrase: The way all the couples are pairing off lately and naming the big day, I think Cupid's really busy.

2) **Literal**: We have enough candles mom

Emotion annoyed

Creative Paraphrase: I think the Catholic church is going to have to canonize a whole new generation of saints to justify our candle use mom

. . .

Generating Explanations + Data Stats

- We generate explanations for each (hypothesis, premise) pair (either entailment and contradiction) using GPT-3 (~13 examples)
- Expert validation and correction: 21% for sarcasm, 20% for simile, 40% metaphor and 10% for idioms

	Entails	Contradicts	Total
Paraphrase	1339	-	1339
+ Sarcasm	-	2678	2678
Simile	750	750	1500
Metaphor	750	750	1500
Idiom	1000	1000	2000

Prompts Used To Generate Explanations for Sarcasm

Explanations for Entailment

You will be presented with examples of two sentences typically a premise along with an entailing paraphrase of the premise called the hypothesis. Your task is to generate natural language explanations to justify the Entailment between the premise and the hypothesis.

1) **Premise**: Awful seeing a naked man run through my neighborhood.

Hypothesis: The sight of a man running through my neighborhood sans clothes was pretty disgusting.

Explanation: It is socially unacceptable to not wear clothes and step out of one's house so seeing a man who is running naked in the neighborhood is pretty shameful and disgusting.

Explanations for Contradiction

You will be presented with examples of some literal and sarcastic sentences. Your task is then to write explanations to justify why it is sarcastic w.r.t the literal

1) **Literal**: When I moved into my apartment it was full of bugs

Sarcasm: I absolutely loved when I moved into my apartment and found it crawling with bugs.

Explanation: Bugs are usually disgusting and most people are terrified of them therefore it is unlikely to love seeing someone's apartment infested by them.

Type	Premise (literal)	Hypothesis (figurative*)	Label	Explanation
Paraphrase + Sarcasm	My next door neighbors are *always arguing* in our shared hallway.	It's *so annoying* to have to hear my next door neighbors *argue all the time* in our shared hallway.	E	The sound of arguing neighbors can often be very disruptive and if it happens all the time in a common space like a shared hallway it is natural to find it annoying.
		It's *so pleasant* to have to hear my next door neighbors *argue all the time* in our shared hallway.	C	The sound of arguing neighbors can often be very disruptive and so someone considering it to be pleasant is not really accurate.
Simile	The assembly hall was now *hot and moist*, more so than usual.	In fact, the assembly hall was now *like a steam sauna*.	E	A sauna is a hot and moist environment, so the simile is saying that the hall is even hotter and more moist than usual.
	The assembly hall was now *cold and dry*, more so than usual.		C	A steam sauna is a small room or hut where people go to sweat in steam, so it would be hot and humid, not cold and dry.
Metaphor	He *mentally assimilated* the knowledge or beliefs of his tribe.	He *absorbed the knowledge* or beliefs of his tribe.	E	To absorb something is to take it in and make it part of yourself.
	He *utterly decimated* his tribe's most deeply held beliefs.		C	Absorbed typically means to take in or take up something, while "utterly decimated" means to destroy completely.
Idiom	Lady Southridge was wringing her hands, *trying hard and desperately to salvage* the bleak and miserable situation so that it somehow looks positive.	Lady southridge was wringing her hands, trying *to grasp at straws.*	E	To grasp at straws means to make a desperate attempt to salvage a bad situation, which is exactly what Lady Southridge is trying to do.
	Lady Southridge was wringing her hands, *doing absolutely nothing to overturn* the bleak and miserable situation so that it somehow looks positive.		C	To grasp at straws means to make a desperate attempt to salvage a bad situation, but the sentence describes not doing anything to change the situation

Experimental Setup

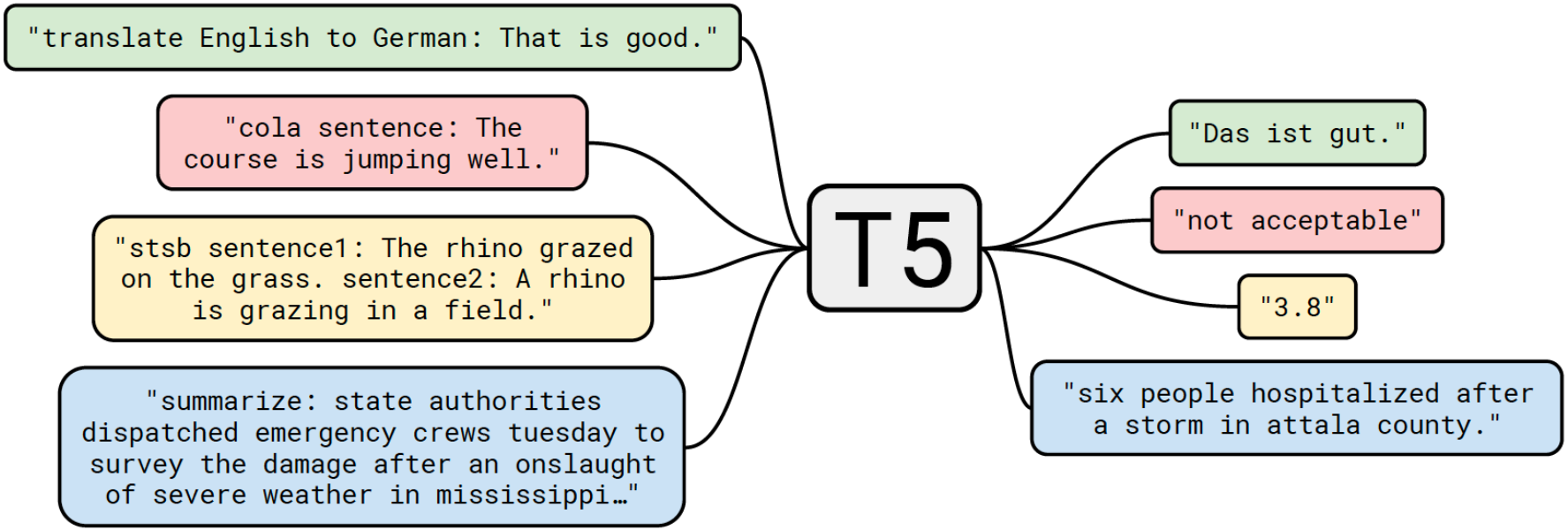
- Train a joint self-rationalizing model (I->OR) using a T5 model (Text-to-Text Transfer Transformer) (Raffel et al, 2020) using the following instruction

  Does the sentence "P" entail or contradict the sentence "H"? Please answer between "Entails" or "Contradicts" and explain your decision in a sentence.

- T5:e-SNLI: fine-tune the above model on e-SNLI (removing the Neutral class)
- T5:FLUTE: fine-tune the above model on on our FLUTE data

Brief Intro to T5 model idea

- Text-to-Text Models: T5 (Text-To-Text Transfer Transformer) (Raffel et al., 2020)



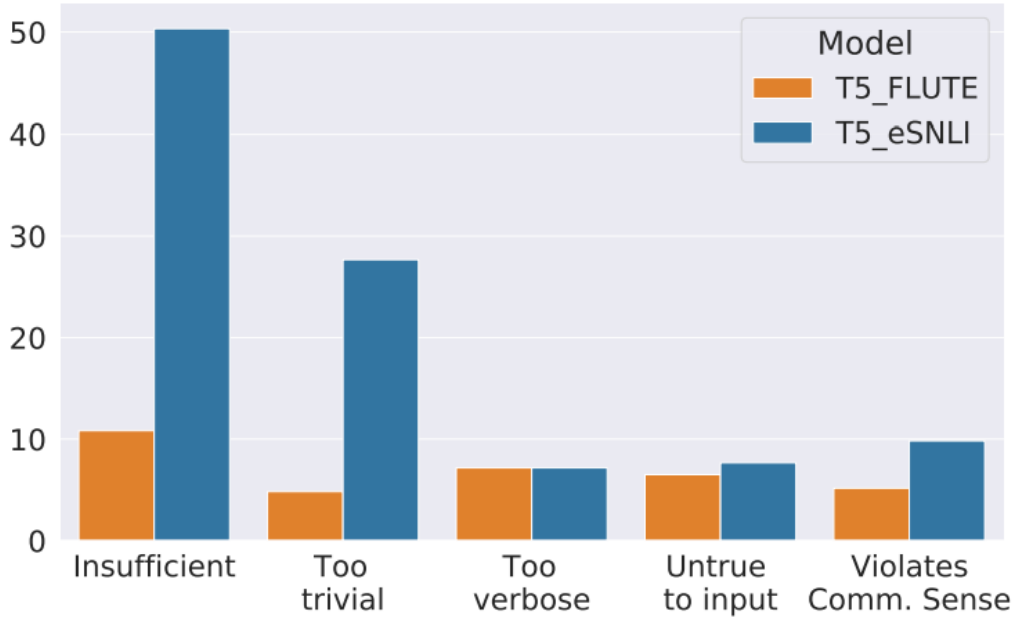
Evaluation

- Automatic Evaluation
  - Blind test set of 1500 instances (750 sarcasm, 250 each for simile/metaphor/idioms)
  - Accuracy@ExplantionScoreThreshold
  - Explanation score: average between BERTScore and BLEURT (0-100)
  - Accuracy@0, Accuracy@50, Accuracy@60
- Human Evaluation
  - Given the two sentences, does the explanation justify the answer above?
    - Yes (1), WeakYes (2/3), WeakNo (1/3), No(0)
    - Hscore = sample average for each explanation across 3 Turkers
    - IAA among Turkers 0.45 Krippendorf alpha (moderate agreement)

Results

	$T5_{e\text{-}SNLI}$						$T5_{FLUTE}$					
	Acc @0	Acc @50	Acc @60	H_{score}	Yes%	No%	Acc @0	Acc @50	Acc @60	H_{score}	Yes%	No%
Sarcasm	60.6	15.7	2.4	34.2	14.7	52.0	**91.6**	**86.2**	**56.2**	**85.3**	**75.3**	**8.7**
Simile	61.2	22.8	3.6	43.6	22.0	40.7	**62.8**	**57.2**	**30.4**	**84.9**	**74.7**	**8.0**
Metaphor	**81.8**	31.8	11.6	55.3	36.0	28.0	73.3	**55.6**	**23.7**	**80.2**	**64.0**	**6.0**
Idiom	**84.8**	46.4	7.6	60.9	37.3	24.7	79.2	**77.2**	**66.8**	**83.1**	**69.3**	**8.7**

Results


Model
T5_FLUTE
T5_eSNLI
50
40
30
20
10
0
Insufficient
Too trivial
Too verbose
Untrue to input
Violates Comm. Sense

Premise	No one came to my wedding.
Hypothesis	No big deal that there were no guests at my wedding reception considering it's the only one I'll ever have.
Label	Contradiction.
Gold_Exp	A wedding reception is supposed to be a joyous event with many guests, so not having any guests at it would make it a very uneventful and pointless one.
$T5_{e\text{-}SNLI}$	No one is same as No big deal.
$T5_{FLUTE}$	A wedding reception is a celebration of the couple's union and is usually attended by family and friends so not having guests there would be seen as snub.

SHARED task at FigLang 2022 workshop @EMNLP

https://figlang2022sharedtask.github.io/task.html

- We ran a shared task at FigLang 2022 workshop at EMNLP
- 5 teams participated (several beat our baselines ☺)
- Please check the Figurative Language Workshop papers and our report

- One interesting highlight from one of the teams

Setting	Acc@0	Acc@50	Acc@60
Regular	92.16	87.92	66.14
Hyp-Only	65.47	60.96	45.95
Prem-Only	56.31	47.81	33.74

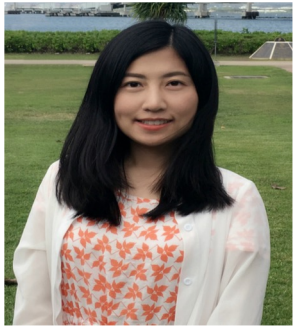
Table 4: T5-large performance on the FigLang dataset with either the hypothesis or premise removed.

Overview

- Sarcasm Interpretation: Textual Entailment + Explanation
- Sarcasm Generation
- Sarcasm Detection: Modeling the conversation context

R^3 : Reverse, Retrieve, and Rank for Sarcasm Generation with Commonsense Knowledge (ACL 2020)

Tuhin Chakrabarty Debanjan Ghosh Nanyun (Violet) Peng



Insight: Sarcasm Factors

1. Be evaluative
2. Be based on a *reversal of valence* between the *literal* and *intended* meaning
3. Be based on a *semantic incongruity with the context*, which can include shared *common sense or world knowledge* between the speaker and addressee
4. Be aimed at some target
5. Be relevant to the communicative situation in some way

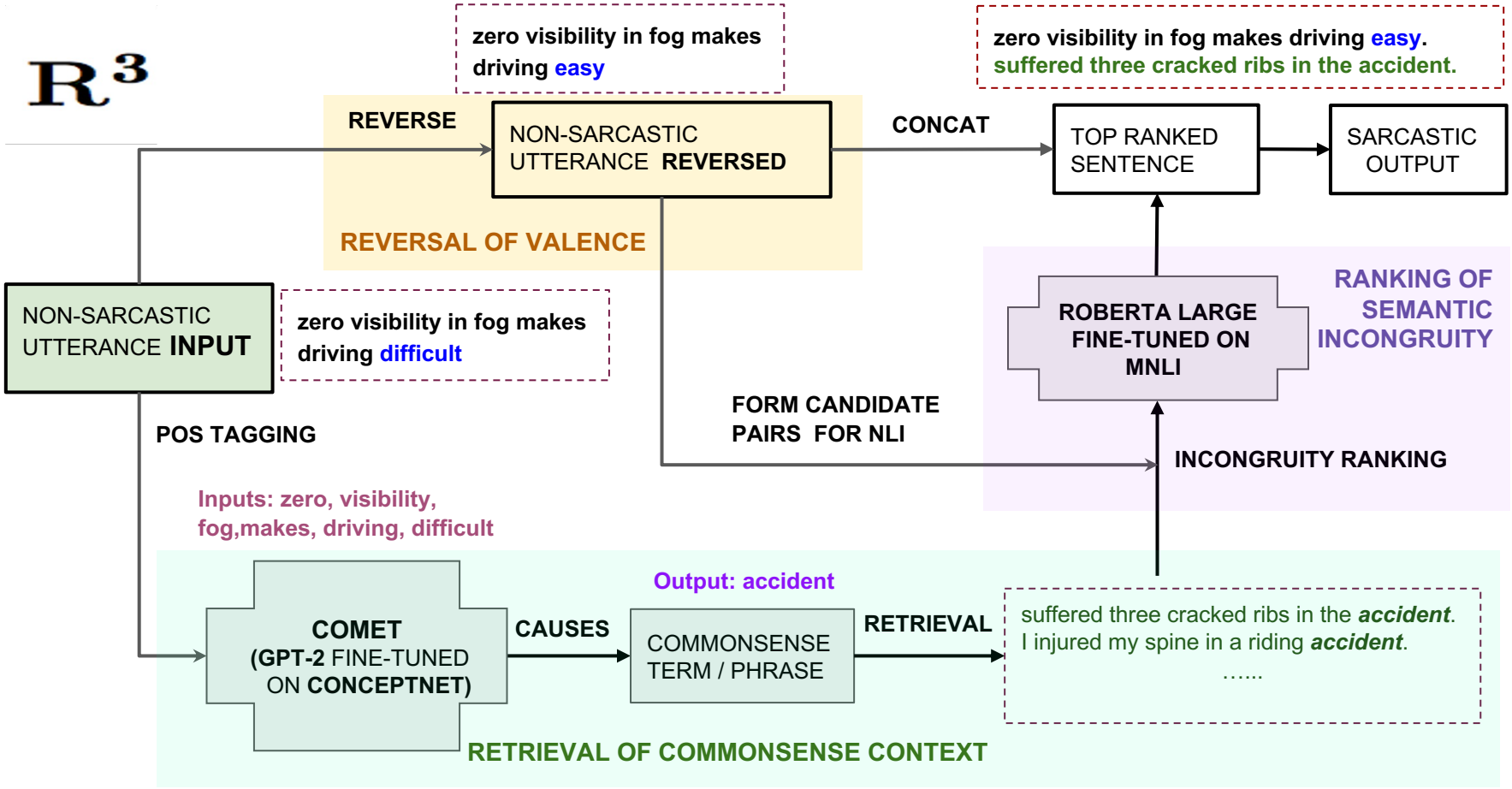
Task and Approach

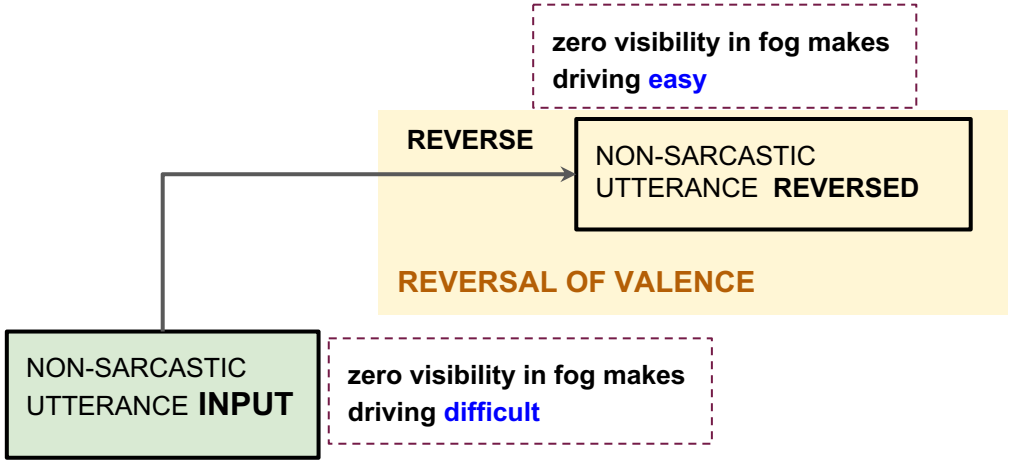
Task: given an evaluative non-sarcastic utterance, generate a sarcastic utterance that keeps the speaker's intended meaning.

Given the lack of training data for sarcasm generation, we propose a novel *unsupervised* approach that has three main components inspired by the *sarcasm factors*

$\mathbf{R^3}$

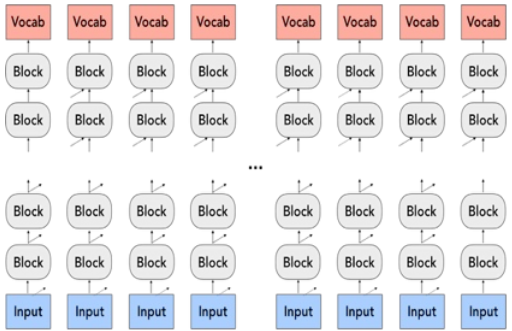
- **R**eversal of Valence
- **R**etrieval of Common sense Context
- **R**anking of Semantic Incongruity


R3
zero visibility in fog makes driving easy
zero visibility in fog makes driving easy. suffered three cracked ribs in the accident.
REVERSE
NON-SARCASTIC UTTERANCE REVERSED
CONCAT
TOP RANKED SENTENCE
SARCASTIC OUTPUT
REVERSAL OF VALENCE
NON-SARCASTIC UTTERANCE INPUT
zero visibility in fog makes driving difficult
ROBERTA LARGE FINE-TUNED ON MNLI
RANKING OF SEMANTIC INCONGRUITY
FORM CANDIDATE PAIRS FOR NLI
POS TAGGING
INCONGRUITY RANKING
Inputs: zero, visibility, fog,makes, driving, difficult
Output: accident
COMET (GPT-2 FINE-TUNED ON CONCEPTNET)
CAUSES
COMMONSENSE TERM / PHRASE
RETRIEVAL
suffered three cracked ribs in the accident.
I injured my spine in a riding accident.
…...
RETRIEVAL OF COMMONSENSE CONTEXT


zero visibility in fog makes driving easy
REVERSE
NON-SARCASTIC UTTERANCE REVERSED
REVERSAL OF VALENCE
NON-SARCASTIC UTTERANCE INPUT
zero visibility in fog makes driving difficult

Quick Idea behind Adapted Knowledge Models

- Language models **implicitly represent some level of knowledge**
- Re-train them on knowledge graphs to **learn structure of knowledge**
- Resulting knowledge model **generalizes structure** to other concepts

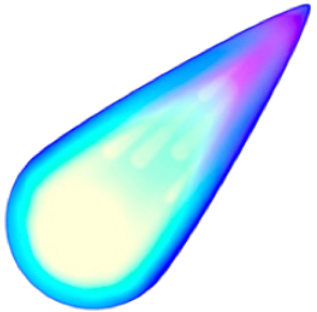


Pre-trained Language Model

+

Seed Knowledge Graph Training

=



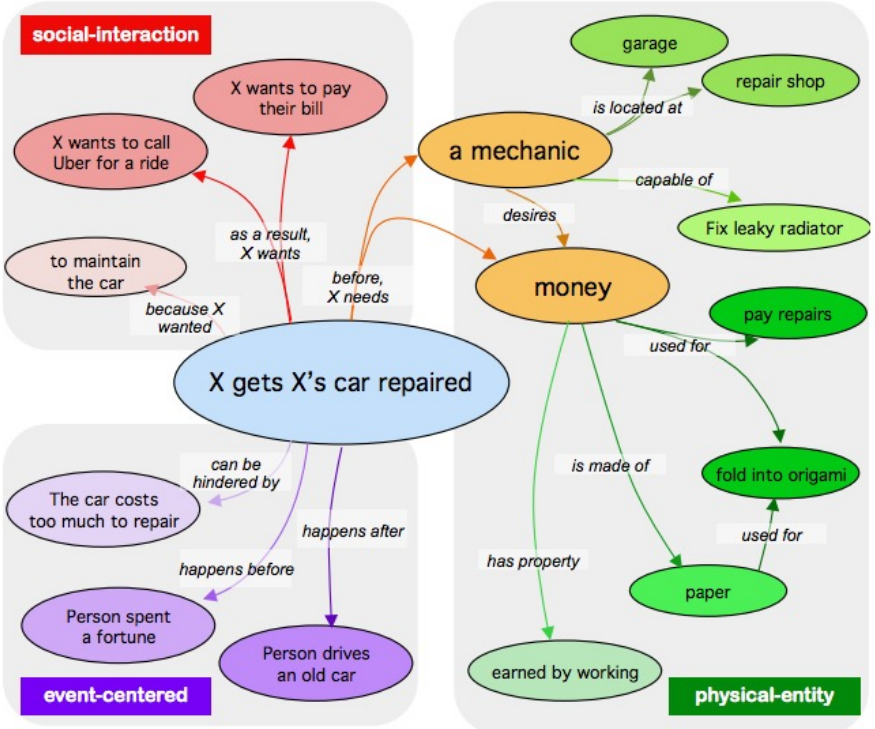
COMET

mango, used for -> salsa
person sails across oceans, requires -> buy a boat

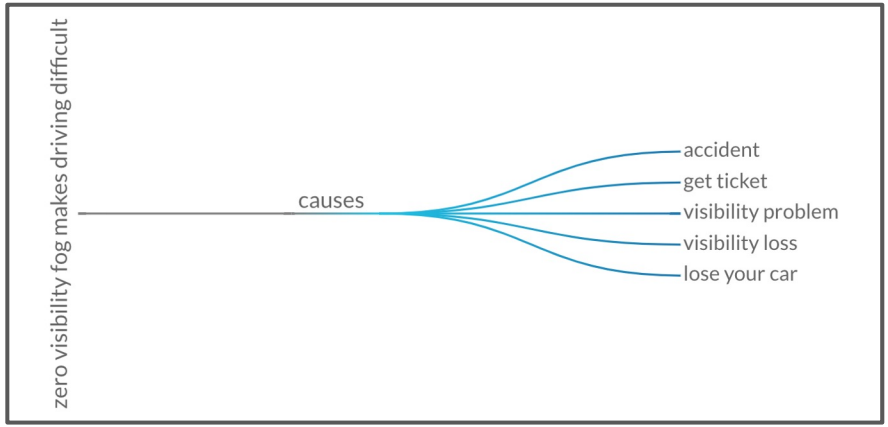
(Bosselut et al., 2019)

Quick Idea behind Adapted Knowledge Models

- COMET (COMmonsEnse Transformers) – ConceptNet, ATOMIC (Sap et al 2019)



(Hwang et al, 2020)



COMET (Bosselut et al 2019)

NON-SARCASTIC UTTERANCE **INPUT**

zero visibility in fog makes driving difficult

POS TAGGING

Inputs: zero, visibility, fog,makes, driving, difficult

COMET (GPT-2 FINE-TUNED ON **CONCEPTNET)**

CAUSES

Output: accident

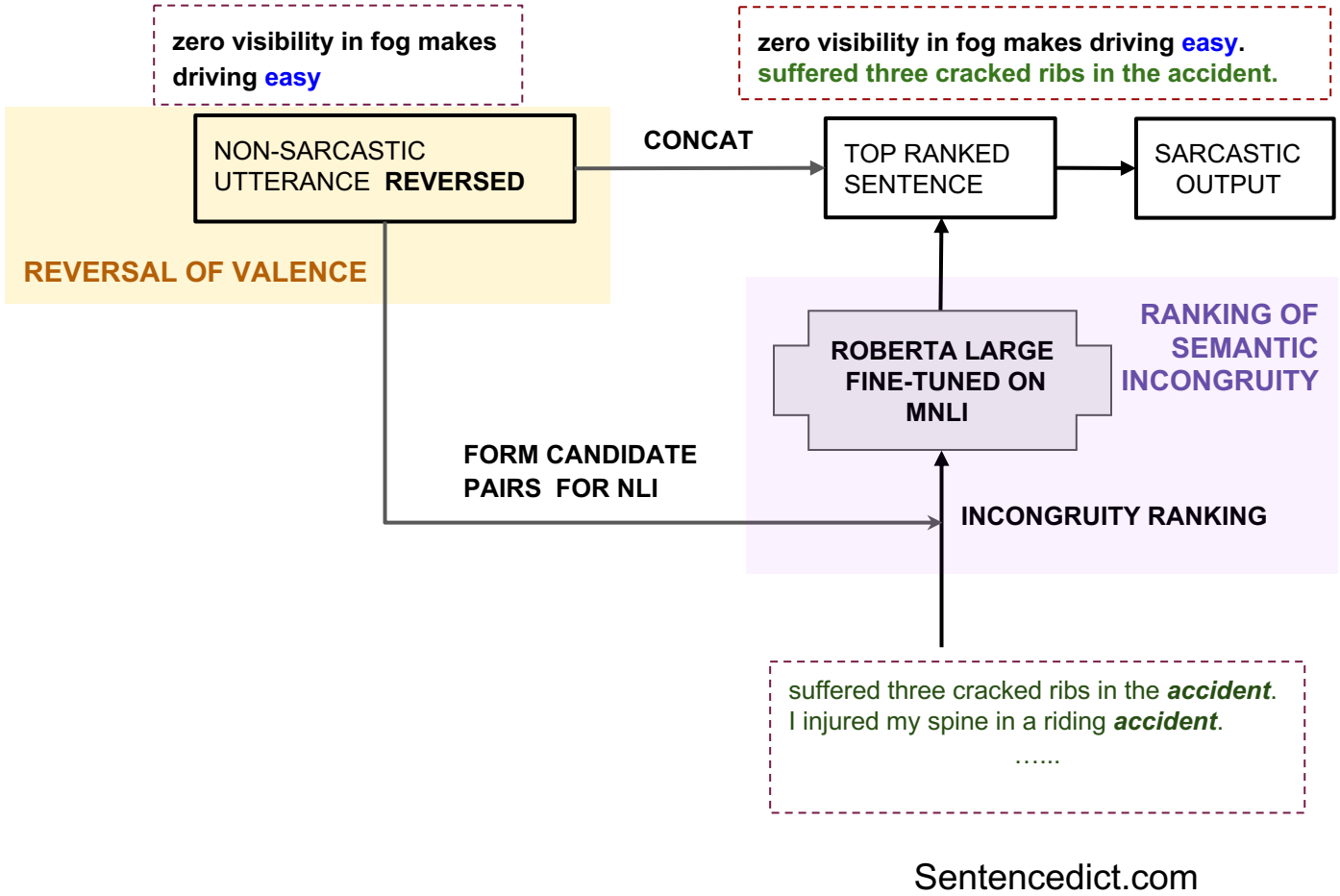
COMMONSENSE TERM / PHRASE

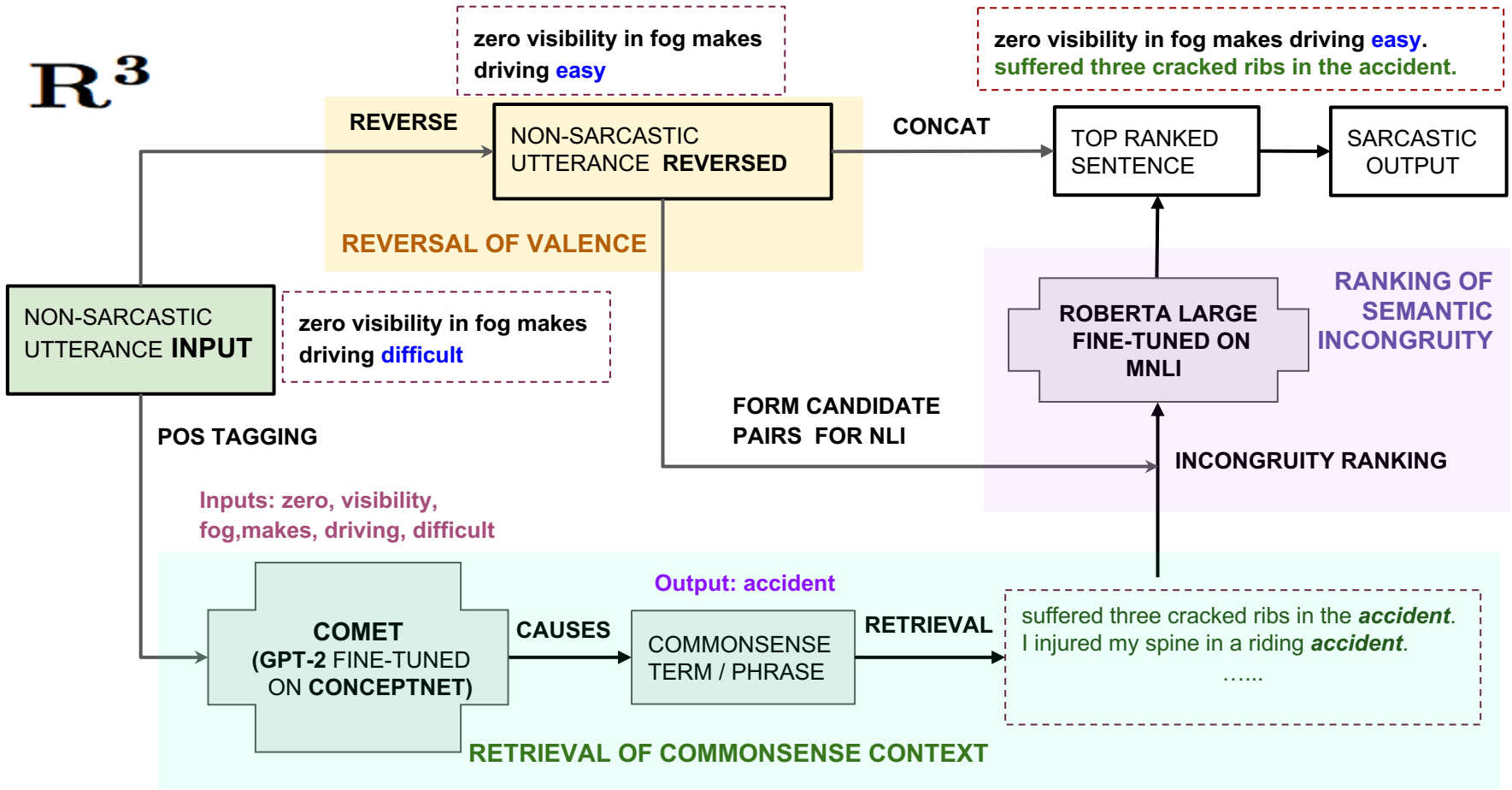
RETRIEVAL

suffered three cracked ribs in the ***accident***.
I injured my spine in a riding ***accident***.
…...

RETRIEVAL OF COMMONSENSE CONTEXT

Sentencedict.com


zero visibility in fog makes driving easy
zero visibility in fog makes driving easy.
suffered three cracked ribs in the accident.
NON-SARCASTIC UTTERANCE REVERSED
CONCAT
TOP RANKED SENTENCE
SARCASTIC OUTPUT
REVERSAL OF VALENCE
ROBERTA LARGE FINE-TUNED ON MNLI
RANKING OF SEMANTIC INCONGRUITY
FORM CANDIDATE PAIRS FOR NLI
INCONGRUITY RANKING
suffered three cracked ribs in the accident.
I injured my spine in a riding accident.
…...
Sentencedict.com


R3
zero visibility in fog makes driving easy
zero visibility in fog makes driving easy. suffered three cracked ribs in the accident.
REVERSE
NON-SARCASTIC UTTERANCE REVERSED
CONCAT
TOP RANKED SENTENCE
SARCASTIC OUTPUT
REVERSAL OF VALENCE
NON-SARCASTIC UTTERANCE INPUT
zero visibility in fog makes driving difficult
ROBERTA LARGE FINE-TUNED ON MNLI
RANKING OF SEMANTIC INCONGRUITY
POS TAGGING
FORM CANDIDATE PAIRS FOR NLI
INCONGRUITY RANKING
Inputs: zero, visibility, fog,makes, driving, difficult
Output: accident
COMET (GPT-2 FINE-TUNED ON CONCEPTNET)
CAUSES
COMMONSENSE TERM / PHRASE
RETRIEVAL
suffered three cracked ribs in the accident.
I injured my spine in a riding accident.
…...
RETRIEVAL OF COMMONSENSE CONTEXT

Evaluation Setup

Test Set

Test on 150 randomly selected non sarcastic utterance

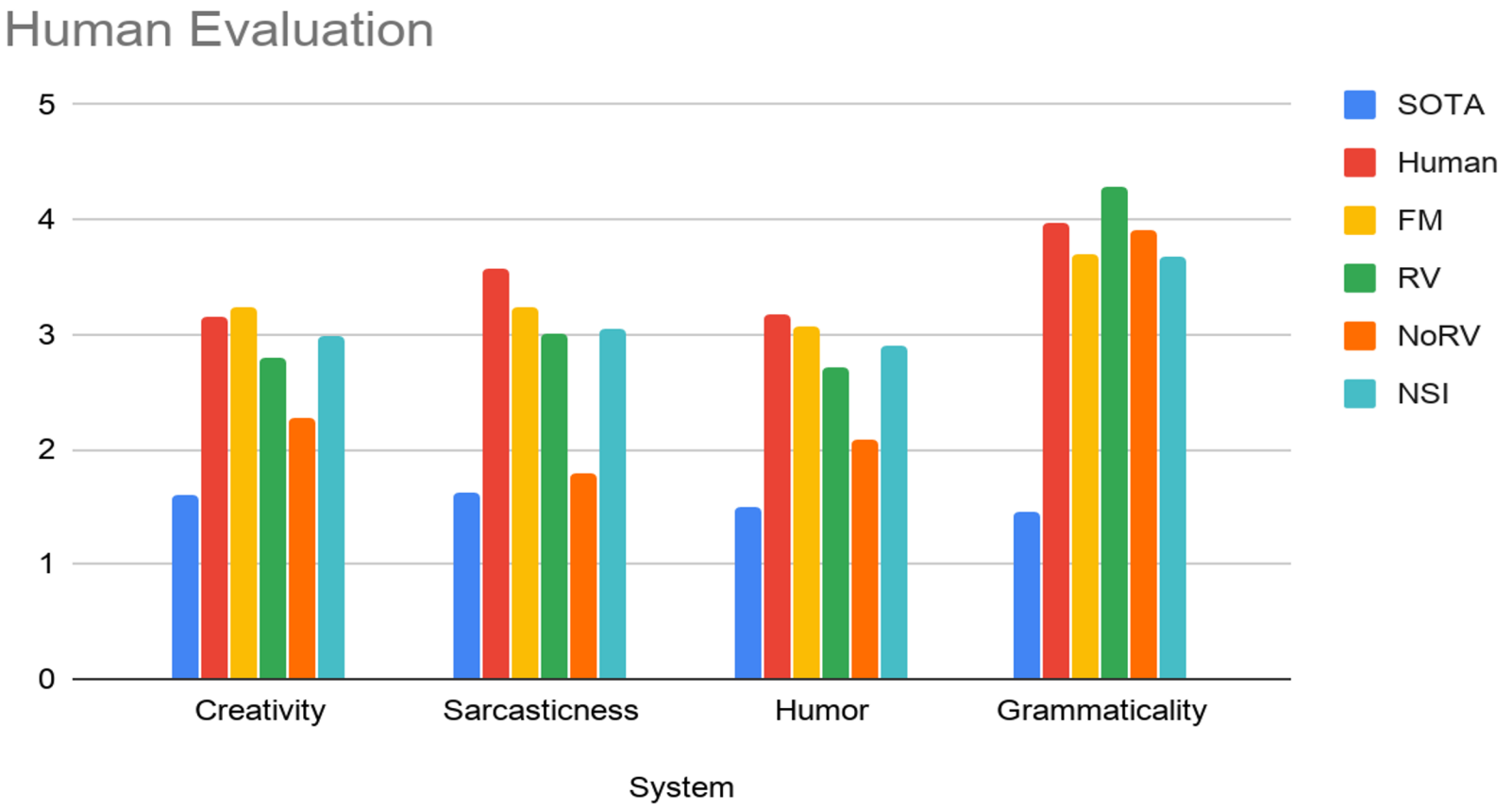
Evaluation Criterion

- Creativity
- Sarcasticness
- Humor
- Grammaticality

Ablations

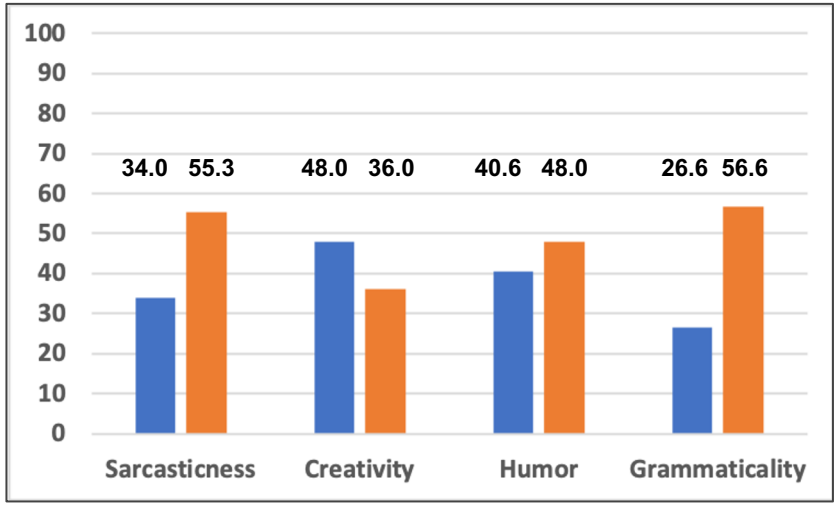
- Reversal of Valence (RV)
- No Reversal of Valence (NoRV)
- No Semantic Incongruity (NSI)
- SOTA (Hybrid reinforced seq2seq model *Mishra et al 2019*)
- Our Full Model (FM)
- Human (Gold) Sarcasm

Results

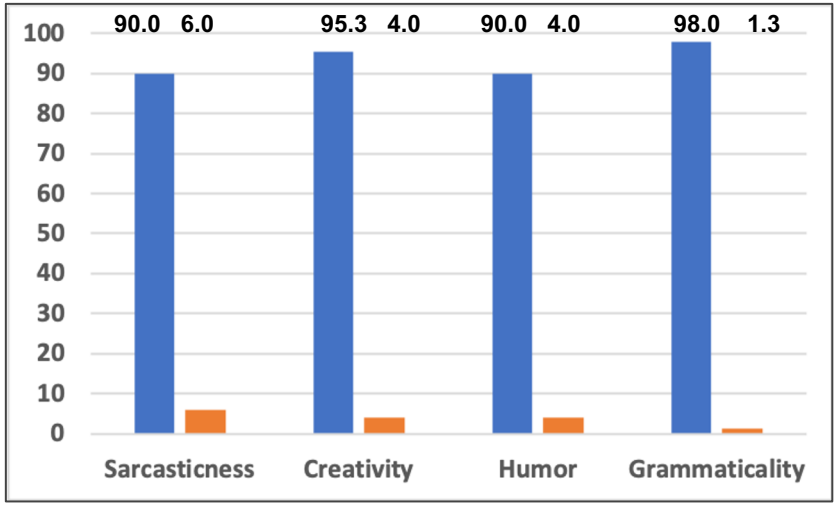

Human Evaluation
5
4
3
2
1
0
Creativity
Sarcasticness
Humor
Grammaticality
System
SOTA
Human
FM
RV
NoRV
NSI

Pairwise Game

FM Vs. Human



FM Vs. SOTA



Win

Loss

Takeaways

- *Use theoretically-inspired approaches*
- *Context is important: Dialogue context; Common sense knowledge; What else?*
- Neither generation nor detection is a solved problem
- Model-in-the-loop seems a promising approach for building datasets
- Sarcasm understanding/generation is hard and interesting! Come work on it!

THANK YOU!!

Tuhin Chakrabarty **Alexander Fabbri** **Debanjan Ghosh** **Arkadiy Saakyan** **Nanyu (Violet)Peng**





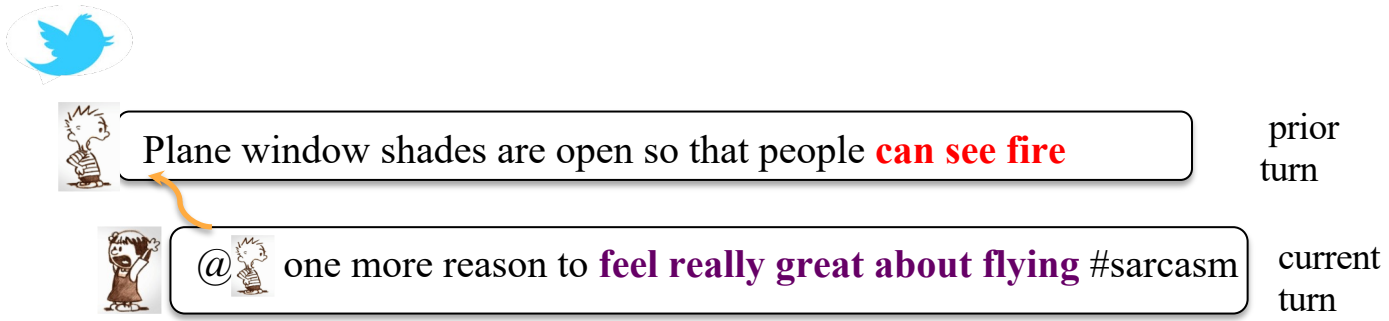
Data and Code: https://github.com/tuhinjubcse/

Overview

- Sarcasm Interpretation: Textual Entailment + Explanation
- Sarcasm Generation
- **Sarcasm Detection: Modeling the conversation context**

Sarcasm: Modeling Conversational Context

- Characteristics:
  - **Irony Factors:** Incongruency with context



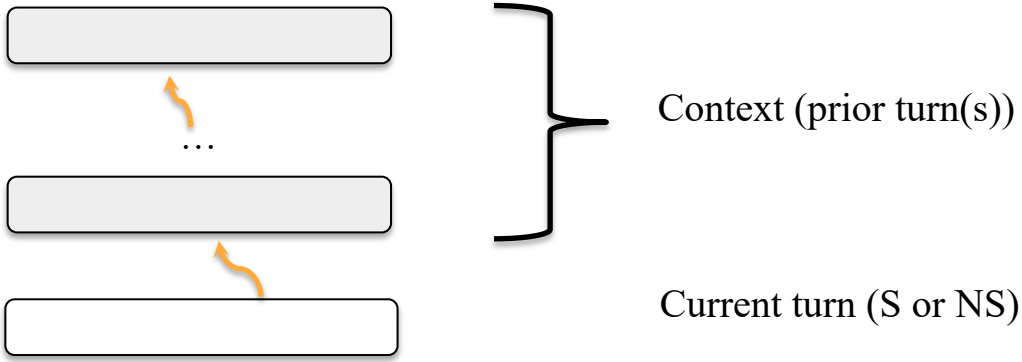
(Ghosh, Fabrri and Muresan, 2017;2018)
best paper award at SIGDIAL 2017

Sarcasm: Modeling Conversational Context

- **RQ1: Can *conversational context* help in sarcasm detection**
- RQ2: Can we identify ***what part*** of the context triggers the sarcastic reply?

Data Annotations

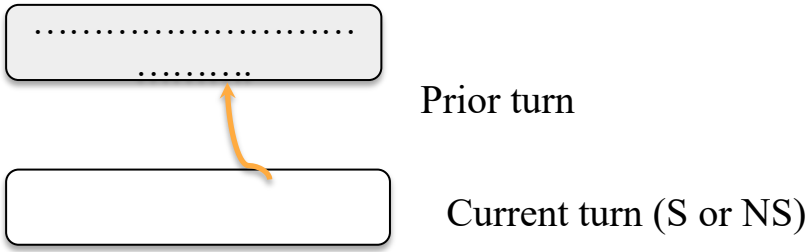
Twitter Corpus



- ``reply_to(@user)'' to detect reply; collect the full thread
- **Self-labeled corpus** (provided by the tweet authors using hashtags)
  - S: #sarcasm, #sarcastic
  - NS: #love, #hate, … [González-Ibáñez et al. 2011]
- 25K instances (12K S/13K NS; 30% with > 1 context utterance)

Data Annotations

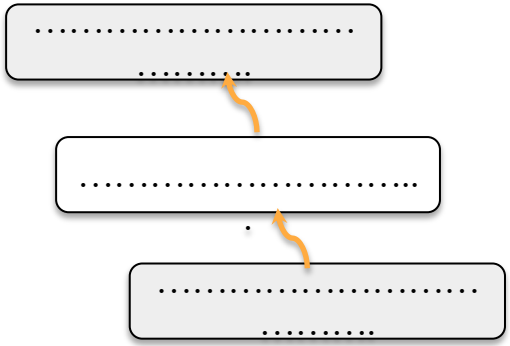
Internet Argument Corpus$_{v2}$ ([Oraby et al, 2016], Discussion forum



- **Annotated by crowdsourcing** at *comment level*: ***perceived sarcasm***! (4950 Instances)
  - Balanced between S/NS
  - Comments between 3-7 sentences long
  - 3 types of sarcasm: ***General, Rhetorical Question, Hyperbole***

Data Annotations

IAC_{v2} and $IAC+_{v2}$ (Discussion forum)



Prior turn

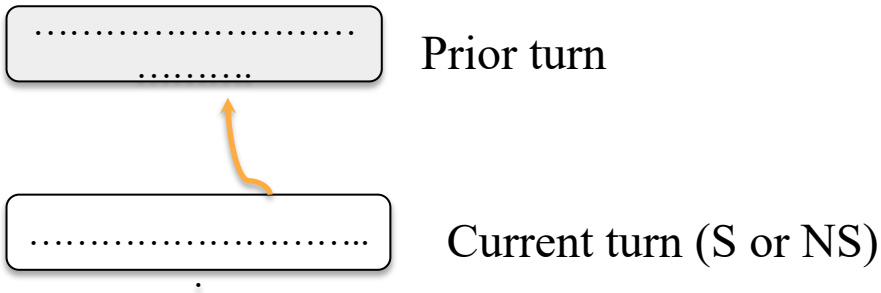
Current turn (S or NS)

Succeeding turn

- Built a subset that includes succeeding turns (2900 Instances)
  - Balanced between Sarcastic/Non-Sarcastic
  - Comments between 3-7 sentences long
  - 3 types of sarcasm: General, Rhetorical Question, Hyperbole

Data Annotations

Reddit Corpus ([Khodak et al. 2017], Discussion forum)



- **Self-labeled** corpus, annotated at comment level (``\s marker added by speaker)
- Collected subset of 50K instances
  - Balanced between Sarcastic/Non-Sarcastic
  - Comments between 3-7 sentences long

RQ1: Can Conversational Context Help in Sarcasm Detection?

- Baseline (SVM with discrete features)
  - ngrams
  - Sentiment and pragmatic features
    - Linguistic Inquiry and Word Count (LIWC) lexicon
    - MPQA Subjectivity lexicon
    - Change of sentiment [Joshi et al. 2015]
  - Sarcasm Markers [Burgers et al. 2012]
    - Morpho-syntactic
      - (interjections: ``yeah''; ``uh''), Tag questions (``isn't it?"), Exclamations
    - Typographic
      - Capitalization (``NEVER''), quotation marks, emoticons
    - Tropes: figurative or metaphorical uses
      - Intensifiers (``greatest'', ``best''…)

Computational Models

- Long Short-Term Memory (LSTM) Networks [Hochreiter & Schmidhuber 1997]
  - Type of RNN; able to learn long-distance dependencies
  - One LSTM reads the context and another LSTM reads the response
- Attention-based LSTM Networks
  - Word and sentence level attention (hierarchical model; Yang et al. 2016) or only sentence level (avg. word embeddings)

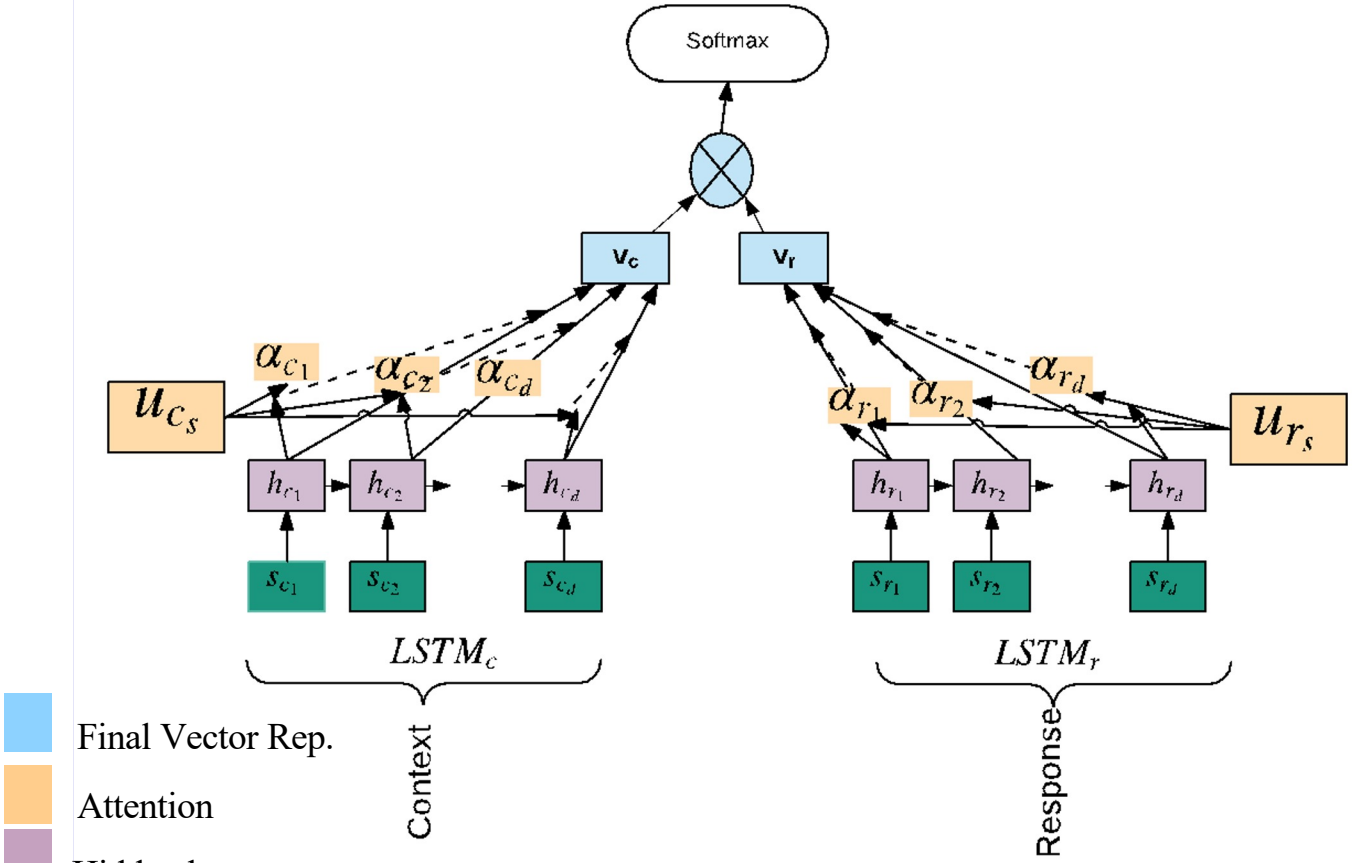


Fig. inspired by Yang et al. 2016

Results

Data Split: 80/10/10

Dev data is used for parameter tuning

Model	Twitter F1	IAC_{v2} F1	Reddit F1
SVM^{ct}	66.5	65.8	72.54
SVM^{ct+pt}	67.9	63.1	67.00
$LSTM^{ct}$	67.2	67.5	68.5
$LSTM^{ct+pt}$	68.0	69.9	74.22
$LSTM_a^{ct}$	73.4	69.8	74.68
$LSTM_a^{ct+pt}$	**75.1**	**71.3**	**75.42**

Twitter: using only immediate prior turn: $LSMT_a^{ct+pt\_last}$: 73.71

Results: Cross-corpora training

- Reddit data is self labeled and is larger
- Train on Reddit and Test on IAC_{v2}

Model	F1
$LSTM_a^{ct}$	65.11
$LSTM_a^{ct+pt}$	63.23

Possible issues:

- self-labeled vs. crowdsourced labeled
- topics

Train and Test on IAC^{+}_{v2}

	IAC^{+}_{v2}
SVM^{ct}	77.89
SVM^{ct+pt}	76.43
SVM^{ct+st}	68.83
$SVM^{ct+ps+st}$	72.77
$LSTM^{ct}$	79.25
$LSTM^{ct+pt}$	**83.32**
$LSTM^{ct+st}$	82.60
$LSTM^{ct+pt+st}$	**83.33**
$LSTM_a^{ct}$	80.05
$LSTM_a^{ct+pt}$	81.52
$LSTM_a^{ct+st}$	80.67
$LSTM_a^{ct+pt+st}$	81.08

Even if dataset is smaller the results are higher….

Possible answer: IAC^{+}_{v2} contains mostly Generic type of Sarcasm (95%) while IAC_{v2} contains an equal distribution of Generic, Rhetorical Questions and Hyperbole

Error Analysis

- Can capture cases of context incongruity

- Misses:
  - Use of contextual information outside the text (shared common ground)
  - Sarcastic turns with more than 5 sentences
  - User of profanity and slang
  - Use of rhetorical questions

Attention Weights

- Sarcasm markers
  - Explicit indicators of sarcasm
  - Attention put more weights on markers, such as emoticons (``:p'') and interjections (``yeah'', ``hmm'')

- However,
  - Interpretations based on attention weights have to be taken with care: classification should not rely solely on attention weights [Rocktäschel et al. 2015]

Conclusion

- Conversation context helps, particularly prior context

- Results might differ depending on corpora
  - Twitter vs. Discussion Forums
  - Self-labeled vs. Crowdsourcing labeled
  - Topics could have an influence
  - Size of data