AMPERSAND: Argument Mining for PERSuAsive oNline Discussions

Tuhin Chakrabarty, Christopher Hidey, Smaranda Muresan
Kathleen Mckeown and Alyssa Hwang

Natural Language Processing
Columbia University

ARGUMENTATIVE DISCUSSION FORUM

CMV: I think the Paris shooting makes a good case for culture of responsible gun ownership.

….

I also think this would be much better then increase in police numbers and rights as policemen are a very homogeneous group trained to stick together and the danger of even deepening the "not one of us" (pack) mentality and escalation of "police state".

….

….

Those time periods/places did not enjoy lower rates of violent death. […] but according [this study] the murder rate was "extraordinarily high by today's standard."

…

Looking at the rest of the thread, I think I see what might be a flaw in our collective logic. The incidents that stand out most in our heads are events like Paris or Aurora or Columbine- premeditated shootings with the intent of inflicting a lot of casualties.

ARGUMENTATIVE DISCUSSION FORUM

CMV: I think the Paris shooting makes a good case for culture of responsible gun ownership.

….

I also think this would be much better then increase in police numbers and rights as policemen are a very homogeneous group trained to stick together and the danger of even deepening the "not one of us" (pack) mentality and escalation of "police state".

….

….

Those time periods/places did not enjoy lower rates of violent death. […] but according [this study] **the murder rate was "extraordinarily high by today's standard."**

…

Looking at the rest of the thread, **I think I see what might be a flaw in our collective logic.** **The incidents that stand out most in our heads are events like Paris or Aurora or Columbine- premeditated shootings with the intent of inflicting a lot of casualties.**

Claim
Premise

3

ARGUMENTATIVE DISCUSSION FORUM

CMV: I think the Paris shooting makes a good case for culture of responsible gun ownership.

….

I also think this would be much better then increase in police numbers and rights **as policemen are a very homogeneous group trained to stick together and the danger of even deepening the "not one of us" (pack) mentality and escalation of "police state".**

….

Those time periods/places did not enjoy lower rates of violent death. […] but according [this study] **the murder rate was "extraordinarily high by today's standard."**

…

Looking at the rest of the thread, **I think I see what might be a flaw in our collective logic.** **The incidents that stand out most in our heads are events like Paris or Aurora or Columbine- premeditated shootings with the intent of inflicting a lot of casualties.**

Intra-relation
(micro-level)

Inter-relations
(macro-level)

4

TASK OF INTEREST

- Identify argument components in a full discussion thread
  - **Claim**
  - **Premise**

- Identify two kinds of argument relations:
  - **Inter-turn relations** (argumentative relations to support or attack the other person's argument)
  - **Intra-turn relations** (to support one's claim or premise).

CONTRIBUTIONS

- Introduce two novel distantly-labeled data-sets

- Fine-tune BERT on the distant- labeled data and demonstrate improvement

- Use discourse relations from RST for argument relations.

LABELLED DATASET STATISTICS

TOTAL NUMBER OF THREADS	112
NUMBER OF DIALOGUE TURNS	380
#CLAIM , #PREMISE #NONARGUMENT	1205 , 1536 , 799
TOTAL NO. OF INTRA-TURN RELATIONS	4.6(%) OF 27254 POSSIBLE PAIRS
TOTAL NO. OF INTER-TURN RELATIONS	3.2(%) OF 26695 POSSIBLE PAIRS

DISTANT LABELED DATA

- Challenge: small size labeled dataset

- Leverage distant-labeled data from Reddit and use transfer learning techniques

  - micro-level for intra-turn relations

  - macro-level (dialogue) for inter-turn relations.

MICRO-LEVEL CONTEXT DATA

IMHO, Calorie-counting is a crock what you have to look at is how whole- some are the foods you are eating. *Refined sugar is worse than just empty calories - I believe your body uses a lot of nutrients up just processing and digesting it.*

CLAIM **&** *PREMISE*

IMHO + Context Dataset (4.6 Million comments) inspired by Chakrabarty et al (2019)

MACRO-LEVEL CONTEXT DATA

CMV: A rise in female representation in elected government isn't a good or bad thing.

…

According to this new story, a record number of women are seeking office in this year's US midterm elections. While some observers hail this phenomenon as a step in the right direction, **I don't think it's good thing one way or the other: a politician's sex has zero bearing on their ability to govern or craft effective legislation**. As such...
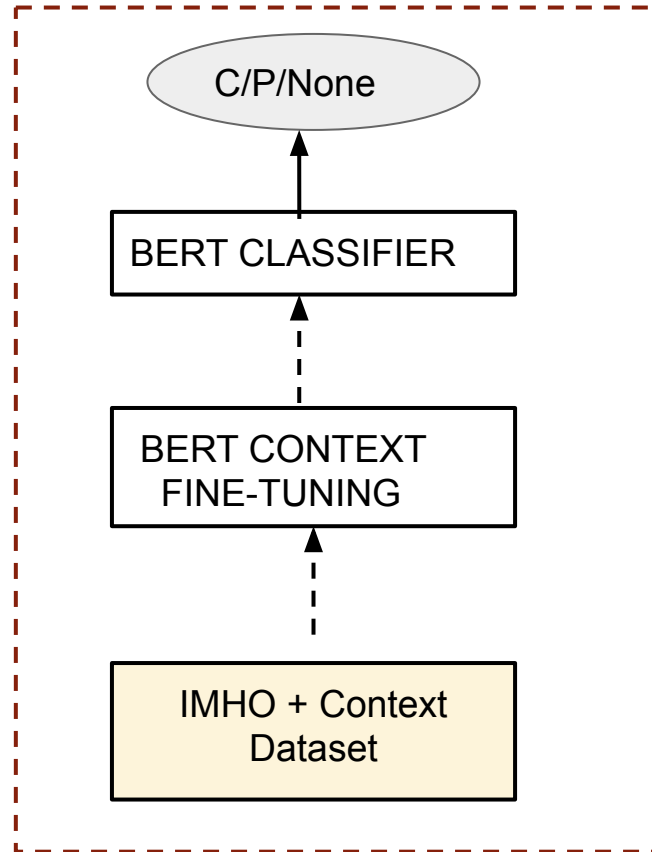
"**I don't think it's good thing one way or the other: a politician's sex has zero bearing on their ability to govern or craft effective legislation**"

Nobody is saying that women are better politicians than men, and thus, more female representation is inherently better for our political system. Rather, the argument is that...
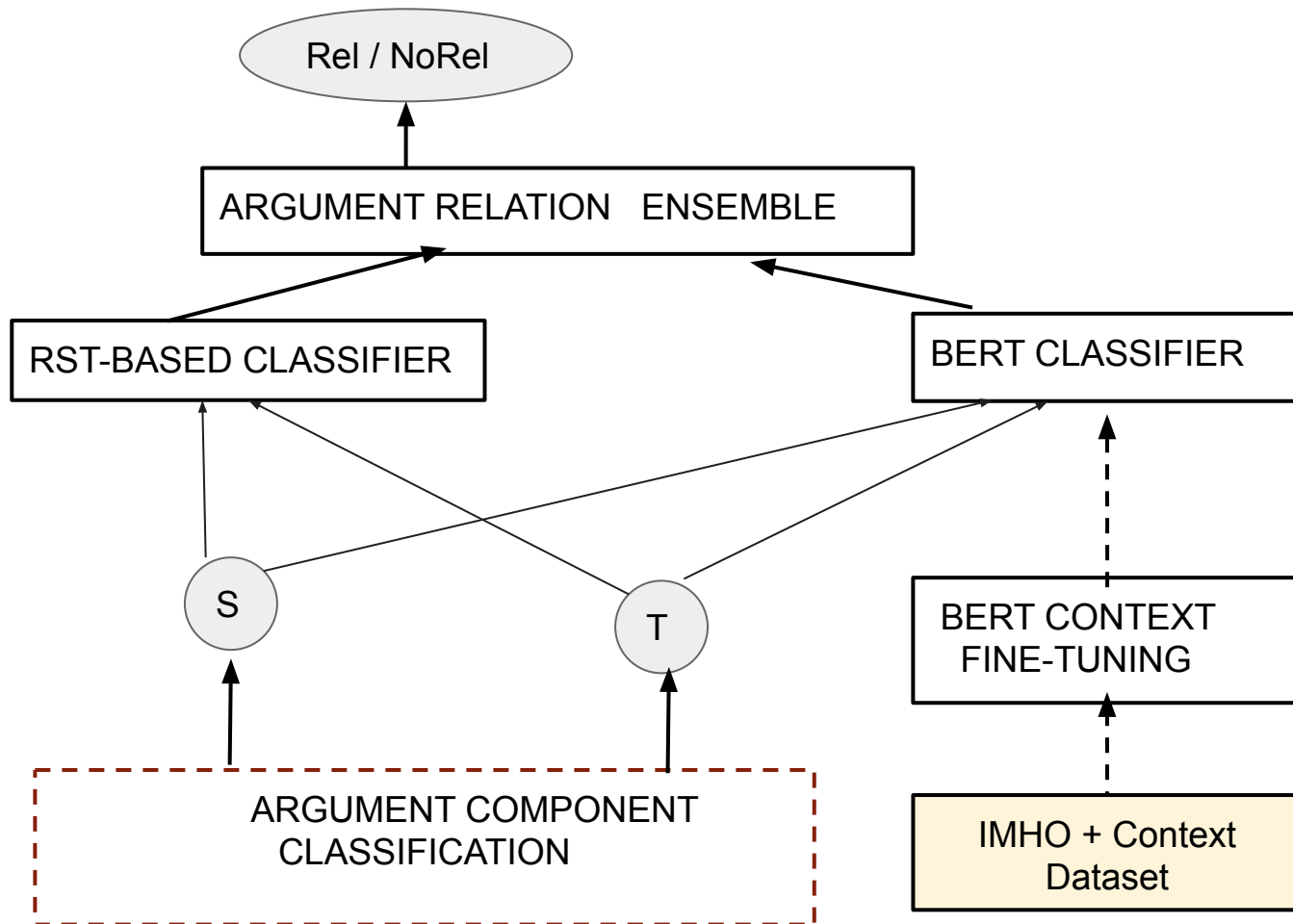
MACRO-LEVEL CONTEXT DATA

- Feature of Reddit: Users can easily *quote* another user's response used to highlight exactly what part of someone's argument a particular user is targeting.

- 97,636 pairs from 19,413 threads

- Language model fine- tuning ⟶ sentence level,
  - (quoted text ,following sentence) ⟶ distant-labeled inter-post pairs

  - **QR**, for quote-response pairs

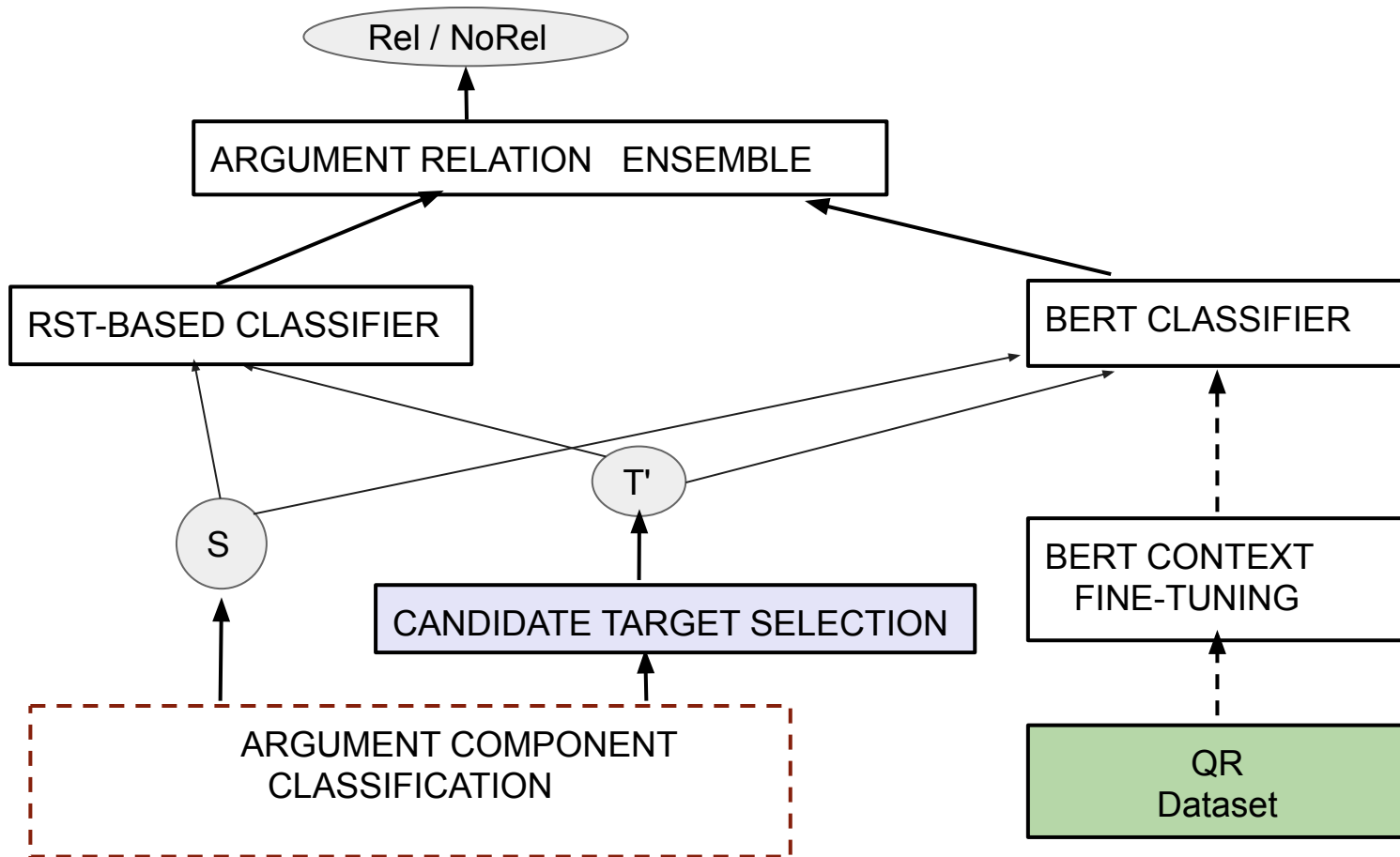# ARGUMENT COMPONENT CLASSIFICATION

INTRA RELATION CLASSIFICATION

Rel / NoRel

ARGUMENT RELATION   ENSEMBLE

RST-BASED CLASSIFIER

BERT CLASSIFIER

S

T

BERT CONTEXT FINE-TUNING

ARGUMENT COMPONENT CLASSIFICATION

IMHO + Context Dataset

INTER-RELATION CLASSIFICATION

Rel / NoRel

ARGUMENT RELATION   ENSEMBLE

RST-BASED CLASSIFIER

BERT CLASSIFIER

S

T'

CANDIDATE TARGET SELECTION

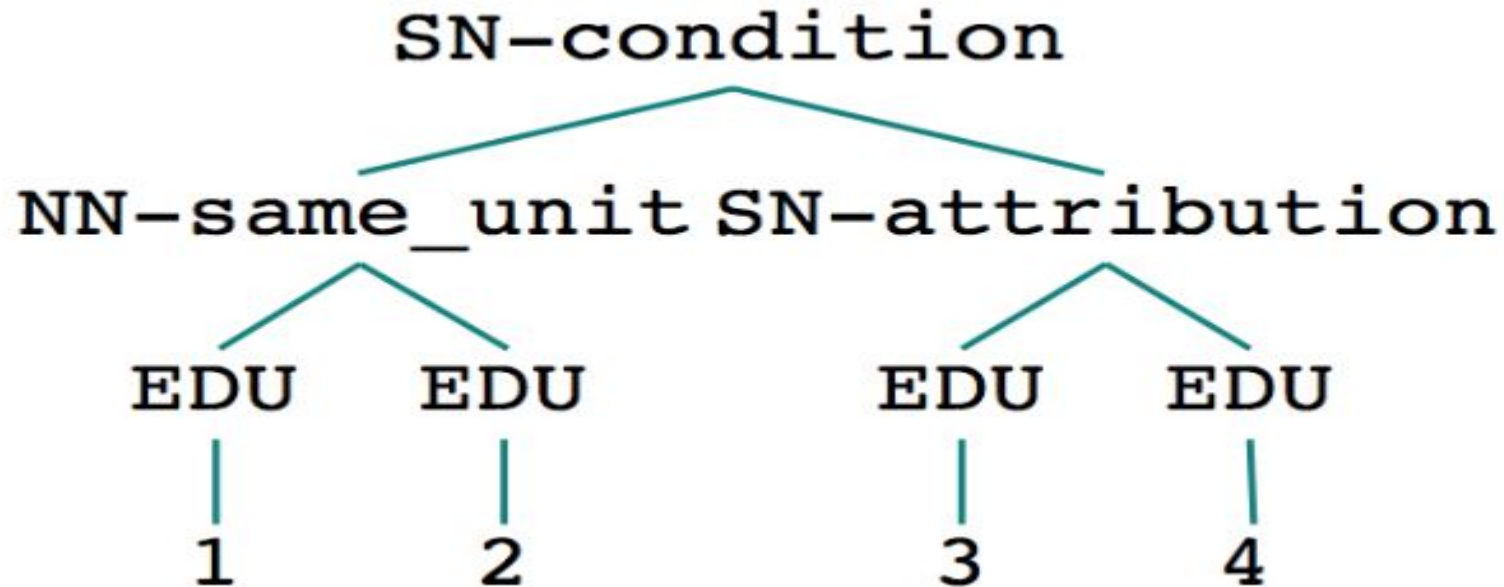BERT CONTEXT FINE-TUNING

ARGUMENT COMPONENT CLASSIFICATION

QR Dataset

14

DISCOURSE RELATIONS

Argument Pair where each argument is broken into EDU's (Elementary Discourse Units)

ARGUMENT 1	ARGUMENT 2
(If existence from your perspective)$_1$(lies solely on your consciousness,)$_2$	(after you die)$_3$(it doesn't matter what you left)$_4$

DISCOURSE RELATIONS

DISCOURSE RELATIONS

To extract features from a pair of argumentative components

- Concatenate 2 components ⟶ single text input

- Use state-of-the-art RST discourse parser (Ji and Eisenstein, 2014)
  - Create parse trees
  - Use predicted discourse relation @ root as a categorical feature

- Use a one-hot encoding of these relations as features and train an XGBoost Classifier and predict Relation (1) vs No-Relation(0)
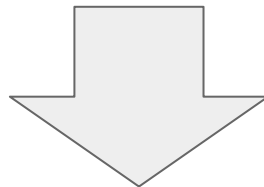
CANDIDATE TARGET SELECTION

CMV: I think the Paris shooting makes a good case for culture of responsible gun ownership.

….

I also think this would be much better then increase in police numbers and rights as policemen are a very homogeneous group trained to stick together and the danger of even deepening the "not one of us" (pack) mentality and escalation of "police state".

….

EXTRACTIVE SUMMARIZER

CMV: I think the Paris shooting makes a good case for culture of responsible gun ownership.

….

I also think this would be much better then increase in police numbers and rights

CANDIDATE TARGET SELECTION

- Use the **QR** data to train & identify candidate targets using techniques from extractive summarization (BERTSUM *),

- Treat the quoted sentences as gold labels, resulting in 19,413 pairs of document (post) and summary (quoted sentences).

- Thus, for a candidate argument pair A → B, where B is the quoted sentence, if B is not extracted by the summarization model we predict that there is no relation between A and B.

*Fine-tune BERT for Extractive Summarization , YangLiu 2019

PIPELINE SUMMARY

- **Argument Component Classification**

- **Intra-relations**
    - Fine tuning (MICRO)
    - Discourse relation

- **Inter-relations**
    - Fine tuning (MACRO)
    - Discourse relation
    - Candidate Selection

ARGUMENT COMPONENT CLASSIFICATION

METHOD	CLAIM	PREMISE	NON ARGUMENT
Stab and Gurevych (2017) + EWE	56.0	65.9	69.6
Morio and Fujita (2018)	54.2	68.5	73.2
Chakrabarty et al. (2019)	57.8	70.8	70.5
BERT Devlin et al.(2019)	62.0	72.2	71.3
IMHO + Context Fine-Tuned BERT	**67.1**	**72.5**	**75.7**

INTRA-TURN RELATION PREDICTION

METHOD	PRECISION		RECALL		F-SCORE	
	GOLD	PRED	GOLD	PRED	GOLD	PRED
All Relations	5.0	-	100.0	-	9.0	-
Morio and Fujita (2018)	10.0	-	48.8	-	16.6	-
BERT Devlin et al (2019)	12.0	11.0	67.0	60.0	20.3	18.5
IMHO Context Fine-Tuned BERT	14.3	13.2	69.0	65.0	23.7	21.8
+ RST Ensemble	**16.7**	**15.5**	**73.0**	**70.2**	**27.2**	**25.4**

*More baselines in paper

INTER-TURN RELATION PREDICTION

METHOD	PRECISION		RECALL		F-SCORE	
	GOLD	PRED	GOLD	PRED	GOLD	PRED
All Relations	5.0	-	100.0	-	9.0	-
Morio and Fujita (2018)	7.6	-	40.0	-	12.7	-
BERT Devlin et al (2019)	8.8	7.9	76.0	70.0	15.8	14.1
QR Context Fine-Tuned BERT	11.0	10.0	75.3	72.5	19.1	17.6
+ RST Ensemble	11.0	12.2	**79.0**	**75.5**	21.2	19.1
Candidate Target Selection	**18.9**	**17.5**	79.0	74.0	**30.5**	**28.3**

*More baselines in paper

IMPACT OF FINE-TUNING

TABLE	PAIR
IMHO	[IMHO, you should not **quantify** it as good or bad.][**Tons of people** have monogamous relationships without issue.]
CMV	[how would you even **quantify** that.][there are **many people** who want close relationships without romance.]

IMPACT OF FINE-TUNING

QR	[It might be that egalitarians, anti-feminists, MRAs & redpillers, groups that I associate with opposing feminism - might be in fact very **distinct** & **different** groups, but I don't know that] [**I do see** all four of these as **distinct** groups].
CMV	[I may have a different stance on seeing no **difference** between companion animals and farm animals.] [**I do see distinction** between a pet and livestock]

IMPACT OF DISCOURSE RELATIONS

Discourse Relation	Argument 1	Argument 2
Evaluation	The only way your life lacks meaning is if you give it none to begin with	Life is ultimately meaningless and pointless
Antithesis	Joseph was just a regular Jew without the same kind of holiness as the other two	Aren't Mary and Joseph, two holy people especially perfect virgin Mary, both Jews? Wasn't Jesus a Jew?

CONCLUSION

- Show how fine-tuning on data-sets similar to the task of interest is often beneficial

- Show how to use transfer learning by leveraging discourse and dialogue context

- Show structure of the fine-tuning corpus is essential for improved performance on pre-trained language models

- Predictions that take advantage of RST discourse cues are complementary to BERT predictions

- Demonstrated methods to reduce the search space and improve precision

QUESTIONS

tc2896@columbia.edu

chidey@cs.columbia.edu

https://github.com/tuhinjubcse/AMPERSAND-EMNLP2019