Supplementary Material for AMPERSAND: Argument Mining for PERSuAsive oNline Discussions

A Hyper-Parameters and Other Experimental Settings

A.1 Language Model Fine-Tuning

IMHO+context : We train the language model fine-tuning using the default parameters metioned in https://github.com/huggingface/pytorch-pretrained-BERT. However, we train the fine-tuning step for 2 epochs. Training 1 epoch took approximately 3 days on 2 K-80 GPU's

QR : Trained using default language model fine-tuning parameters for 3 epochs. Training 1 epoch took approximately 1 hour.

A.2 Argumentative Component Classification

Baseline Model For the baseline model we use XGBoost classifier with our handcrafted features. We found the default setting worked best: https://xgboost.readthedocs.io/en/latest/parameter.html

Pretrained BERT We trained our model for 5 epochs with a learning rate of $2e-5$, a batch size of 64 and a max sequence length of 128. We lower case all tokens and use "bert-base-uncased." We observed statistically significant performance drop on using "bert-large-uncased" models.

IMHO Fine-tuned BERT The IMHO Fine-tuned BERT is trained for 10 epochs with the same parameters as pretrained BERT.

A.3 Intra-turn Relation Prediction

Baseline Model For the baseline model we use again use XGBoost classifier with our handcrafted features and found the default setting worked best for us.

Pretrained BERT We trained our model for 5 epochs using the same parameters as for Argumentative Component Classification.

IMHO Finetuned BERT We trained our model for 8 epochs with the same parameters.

A.4 Inter-turn Relation Prediction

Baseline Model For the baseline model we again use XGBoost classifier with our handcrafted features using the default setting.

Pretrained BERT We trained our model for 3 epochs with the same parameters as Argumentative Component Classification.

QR Fine-tuned BERT For QR Fine-tuned BERT we trained our model for 7 epochs with the same parameters.

A.5 RST classifiers

We use the pre-trained models and code given in https://github.com/jiyfeng/DPLP from obtaining RST trees from the argument pairs. For obtaining the root relations we use the https://www.nltk.org/_modules/nltk/tree.html libraries. We again train an XGBoost classifier with the default settings.

A.6 Extractive Summarization

We use the BERT-Sum model from the publicly available implementation at https://github.com/nlpyang/BertSum. Because we had 19.4k document-summary pairs we used a batch size of 300. We trained the model for 50000 steps with a warmup of 10000 steps. We saved our model checkpoints at every 1000 steps for final evaluation. The learning rate for training was $2e-3$.

B Additional Qualitative Analysis

B.1 Argument Relations vs Sentence Pair Tasks

BERT (Devlin et al., 2019) has been shown to work well on sentence pair tasks like Natural Language Inference (Bowman et al., 2015) or Semantic Textual Similarity (STS). However, it hasn't been used for predicting argumentative relations. Table 1 shows that a model trained on entailment and textual similarity says the sentence pair is semantically related although from an argumentative point of view there is no relation between them. Our experimental results show that detecting argumentative relations is a hard task and while they perform better than baseline models there is plenty of room for improvement.

NLI	STS	Pair
Entails	3.8	[My main reason against the marines was that they are very homogeneous.][From all the marines that I've met they all think and act the same way to the point where it freaks me out.]

Table 1: Predictions on a Pair with No Relation

B.2 RST parse tree visualizations

We provide visualizations of RST parse trees on argumentative relations from the CMV data.

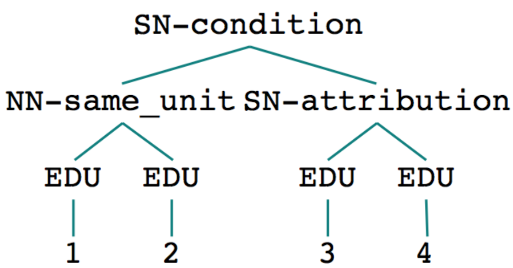

Figure 1: RST parse tree obtained from concatenation of arguments in 2

ARG1	ARG2
(If existence from your perspective)$_1$(lies solely on your consciousness)$_2$	(after you die)$_3$(it doesn't matter what you left)$_4$

Table 2: Example of an argumentative relation from CMV. The parentheses delineate EDUs.

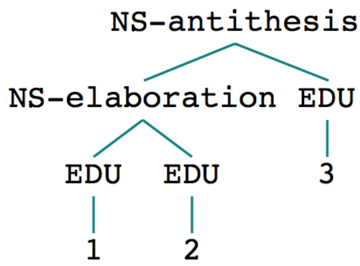

Figure 2: RST parse tree obtained from concatenation of arguments in 3

ARG1	ARG2
(Joseph was just a regular Jew)$_1$ (without the same kind of holiness as the other two)$_2$	(Aren't Mary and Joseph, two holy people especially perfect virgin Mary, both Jews? Wasn't Jesus a Jew?)$_3$

Table 3: Example of an argumentative relation from our data-set.The parenthesis is given to represent EDU's.

References

Samuel Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Jacob Devlin, Ming-Wei Chang, Lee Kenton, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 17th Annual Meeting of the North American Association for Computational Linguistics*.