

Appendices

A Features Analysis

The list of the investigated features is summarized in Table 1.

B Ablation study

Tables 2 and 3 report on the contribution of each individual feature on the performances obtained by the classifier.

C SemEval’20 T11 dev set on the most frequent 5 labels

Table 4 shows an improvement of the obtained results for all labels but “Repetition”. For additional investigations, we select the most frequent propaganda techniques in the training set, and we run the proposed methods on such subset. The workshop SemEval’20 T11 has not published gold labels of the test set. To perform an error analysis, we proceed this section based on label(s) of the most frequent propaganda techniques in the development set. Results on this subtask show an improvement of the obtained results for all labels but “Repetition” technique that does not outperform over such reduced-classes setting. Due to the template-span evaluation, we can only perform our error analysis on false negative predicted instances.

Factor	Feature	Resource	Vector	Technique
Persuasion				
Influencing factors	Speech style	General Inquirer Dictionary	182	Sum of occurrence of each style found in a word of a sentence
	Lexical complexity	BERT sentence embedding	768	Extract a vector of 768 dimensions per token in a sentence, then compute the average of all token to produce a single set of vector.
Tone	Concreteness	Concreteness lexicon	1	Sum of the standardized word scores
	Subjectivity	Subjectivity lexicon	2	#Count frequency of word labeled as “weaksuj” and “strongsubj”
Sentiment analysis				
Semantic features	Sentiment labels	SentiWordnet	3	Sum of positive, negative, and neutral of word scores individually
	VAD labels	Warriner lexicon	3	Sum of the standardized scores of valence, arousal, dominance
	Emotion labels	DepecheMood++ lexicon	8	Sum of 8 emotions
	Connotation	Connotation lexicon	3	#Count words labeled as positive, negative, and neutral
	Politeness	Politeness lexicon	2	#Count words labeled as positive and negative
Message simplicity				
Exaggeration	Imageability	Imageability resource	2	Sum of abstract and concreteness scores
Style	Length	-	4	#Count actual char-length, word length, punctuation frequency, capital-case frequency
		-	27	Length encoding (char-level)
	Pronouns	Our pronouns lexicon	123	#Count all types of pronouns

Table 1: List of semantic features and techniques used.

Semantic Features	NLP4IF’19 Test Set		
	F1	Precision	Recall
Persuasion techniques.			
Speech style	0.63	0.66	0.61
Lexical complexity	0.67	0.67	0.66
Concreteness	0.65	0.69	0.62
Subjectivity	0.66	0.70	0.62
Sentiment techniques			
SentiWordnet	0.68	0.71	0.65
Warriner VAD	0.64	0.67	0.61
DepecheMood++	0.64	0.66	0.61
Connotation	0.65	0.69	0.62
Politeness	0.65	0.68	0.62
Message simplicity techniques			
Imageability	0.66	0.69	0.63
Lexical Length	0.66	0.71	0.62
Lexical Encoding	0.66	0.70	0.62
Pronouns	0.60	0.63	0.58

Table 2: Results on semantic features on BERT + Featured Logistic Regression.

Argumentation Features	NLP4IF’19 Test Set		
	F1	Precision	Recall
Argument Detection			
(1)	0.70	0.71	0.70
(2)	0.69	0.70	0.68
(1)+(2)	0.71	0.72	0.70
Argumentation Components			
(3)	0.70	0.71	0.68
(4)	0.70	0.70	0.69
(3)+(4)	0.69	0.70	0.68
All features			
(1)+(3)	0.70	0.71	0.69
(1)+(4)	0.70	0.70	0.69
(2)+(3)	0.70	0.69	0.67
(2)+(4)	0.68	0.70	0.67
(1)+(2)+(3)+(4)	0.69	0.70	0.68

Table 3: Results on argumentation features on proposed BERT + Featured Logistic Regression. (1) Argumentative, (2) Non-Argum., (3) Claim, (4) Premise

	SemEval’20 T11					
	Average	Appeal to fear-prejudice	Doubt	Exaggeration, Minimisation	Name-Calling, Labeling	Repetition
Proposed Architecture						
Fine-tuned RoBERTa (3 epochs)	0.60	0.41	0.55	0.59	0.75	0.41
Fine-tuned RoBERTa (5 epochs)	0.62	0.39	0.57	0.59	0.78	0.45
Fine-tuned RoBERTa (10 epochs)	0.60	0.45	0.48	0.53	0.77	0.43
Fine-tuned RoBERTa (15 epochs)	0.61	0.52	0.28	0.55	0.77	0.54
Fine-tuned RoBERTa (20 epochs)	0.63	0.55	0.32	0.57	0.78	*0.56
Proposed Architecture + Semantic Features						
Fine-tuned RoBERTa (3 epochs)	0.66	0.54	0.56	0.63	0.79	0.52
Fine-tuned RoBERTa (5 epochs)	0.62	0.49	0.57	0.54	0.79	0.41
Fine-tuned RoBERTa (10 epochs)	0.61	0.42	0.48	0.52	0.76	0.45
Fine-tuned RoBERTa (15 epochs)	0.57	0.18	0.45	0.48	0.77	0.41
Fine-tuned RoBERTa (20 epochs)	0.62	0.47	0.47	0.59	0.76	0.51
Proposed Architecture + Argumentation Features						
Fine-tuned RoBERTa (3 epochs)	0.60	0.41	0.51	0.60	0.76	0.40
Fine-tuned RoBERTa (5 epochs)	0.63	0.43	0.59	0.59	0.77	0.46
Fine-tuned RoBERTa (10 epochs)	0.63	0.47	0.47	0.59	0.78	0.49
Fine-tuned RoBERTa (15 epochs)	0.58	0.44	0.44	0.47	0.73	0.51
Fine-tuned RoBERTa (20 epochs)	0.58	0.53	0.40	0.54	0.74	0.46
Proposed Architecture + All Features						
Fine-tuned RoBERTa (3 epochs)	0.61	0.43	0.54	0.53	0.75	0.50
Fine-tuned RoBERTa (5 epochs)	0.65	0.48	0.54	0.59	0.81	0.50
Fine-tuned RoBERTa (10 epochs)	0.63	0.40	0.51	0.58	0.79	0.48
Fine-tuned RoBERTa (15 epochs)	0.61	0.35	0.47	0.53	0.77	0.53
Fine-tuned RoBERTa (20 epochs)	0.63	0.51	0.50	0.60	0.79	0.47

Table 4: Experiments on SemEval’20 T11 dev set on the most frequent 5 labels.