Appendix: Categorizing Sexism and Misogyny through Neural Approaches

PULKIT PARIKH and HARIKA ABBURI, IIIT-Hyderabad, India

NIYATI CHHAYA, Adobe Research, India and IIIT-Hyderabad, India

MANISH GUPTA* and VASUDEVA VARMA, IIIT-Hyderabad, India

ACM Reference Format:

Pulkit Parikh, Harika Abburi, Niyati Chhaya, Manish Gupta, and Vasudeva Varma. 2020. Appendix: Categorizing Sexism and Misogyny through Neural Approaches. 1, 1 (December 2020), 6 pages.

1 HYPER-PARAMETER VALUES

Using experiments on a validation set, which was merged into the training set during the test runs, for each method, we choose the values of (up to) four hyper-parameters: the LSTM dimension, the attention dimension, the number CNN filters for kernel sizes 2, 3, and 4 each, and the k value in k-max-over-time pooling. The values used for all proposed methods and deep learning baselines for which we report results in our paper are provided in Tables 1, 2, 3, 4, 5, 6, and 7. The same LSTM and attention dimensions are used in all parts of a model. Table 3 also contains the layer sizes for the stacked autoencoder models.

The hyper-parameter values for the traditional machine learning baselines for sexism classification are as follows. For SVM, we set soft margin (C) to 1.0. For RF (Random Forest) and GBT (Gradient Boosting Trees), we use 100 estimators. For extracting character and word n-grams, the maximum number of features used, word n-gram range, and character n-gram range are set to 10000, (1,2), and (1,5) respectively. The hyper-parameter values for misogyny classification are the same as above except that the number of estimators for RF and GBT is 50.

For misogyny detection, the hyper-parameter values for the traditional ML methods are as follows. For SVM, 1 is used as the value of soft margin (C). For RF and GBT, we set the number of estimators to 100. For extracting character and word n-grams, the maximum number of features used, word n-gram range, and character n-gram range are 5000, (1,3), and (1,5) respectively.

For the two methods that achieved the best results for Subtask A and Subtask B of the Evalita 2018 shared task on Automatic Misogyny Identification (AMI) that we use as baselines for misogyny detection and misogyny classification respectively, we use the hyper-parameter values mentioned in the corresponding papers.

Authors' addresses: Pulkit Parikh, pulkit.parikh@research.iiit.ac.in; Harika Abburi, harika.a@research.iiit.ac.in, IIIT-Hyderabad, India; Niyati Chhaya, nchhaya@adobe.com, Adobe Research, India, IIIT-Hyderabad, India; Manish Gupta, manish.gupta@iiit.ac.in; Vasudeva Varma, vv@iiit.ac.in, IIIT-Hyderabad, India.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Manuscript submitted to ACM

Manuscript submitted to ACM

^{*}The author is also an applied researcher at Microsoft.

^{© 2020} Association for Computing Machinery.

Table 1. Hyper-parameter values corresponding to multi-label sexism classification results with the EBCE loss for deep learning methods (the sub-columns for the proposed methods specify different configurations as to the ways in which the input is embedded and/or processed; different colors/styles represent different sentence-level concatenations in a method; colors/styles have no connections across rows)

			Approac	ch		dim.	dim.	#CNN filters of each ker- nel size	over-time pooling
Baselines (with ELMo)	L	biLSTM-Attention			200	300	N.A.	N.A.	
[N	- 1	Hierarchical-biLSTM-A				300	400	N.A.	N.A.
1 E		CNN-biLSTM-Attention	n			300 N.A.	400	100	N.A.
ıt.	- 1	C-biLSTM					N.A.	N.A.	N.A.
3		CNN-Kim					N.A.	150	N.A.
les		biLSTM					N.A.	N.A.	N.A.
1		BERT-biLSTM-Attention					500	N.A.	N.A.
ase	- 1	USE-biLSTM-Attention					600	N.A.	N.A.
В		InferSent-biLSTM-Atte	ntion			100	200	N.A.	N.A.
		biL-att applied on	c-k-max applied on	c-biL-att applied on	Text Encoder				
	- 1	ELMo			tBERT	200	400	N.A.	N.A.
	İ	ELMo, GloVe			tBERT	200	400	N.A.	N.A.
	Ì	concat(ELMo, GloVe)			tBERT	200	500	N.A.	N.A.
		Ling, ELMo, GloVe			tBERT	200	300	N.A.	N.A.
		ELMo, GloVe			tBERT, USE	200	400	N.A.	N.A.
S	Fla	Ling, fastText, ELMo,			tBERT, USE,	200	500	N.A.	N.A.
lod		GloVe			InferSent				
Proposed methods	Ì		ELMo		tBERT	N.A.	N.A.	100	1
H			ELMo, GloVe		tBERT, USE	N.A.	N.A.	200	1
sed				ELMo, GloVe	tBERT	200	400	150	1
bo			ELMo		tBERT	200	500	100	1
ro		ELMo, GloVe	ELMo, GloVe		tBERT, USE	100	300	100	1
					tBERT	300	600	N.A.	N.A.
	- 1	ELMo			tBERT	300	600	N.A.	N.A.
		ELMo, GloVe			tBERT	100	200	N.A.	N.A.
		ELMo, GloVe			tBERT	100	200	N.A.	N.A.
		ELMo, GloVe			tBERT	200	400	N.A.	N.A.
	g	concat(ELMo, GloVe)			tBERT	100	100	N.A.	N.A.
	逗	concat(ELMo, GloVe)			tBERT	300	600	N.A.	N.A.
	erarc		ELMo, GloVe		tBERT	300	500	100	1
	ier			ELMo, GloVe	tBERT	100	100	100	1
	Ŧ	Ling, ELMo, GloVe			tBERT	300	600	N.A.	N.A.
		ELMo, GloVe			tBERT, USE	300	600	N.A.	N.A.
		ELMo, GloVe			tBERT, USE	100	100	N.A.	N.A.
	L	ELMo, GloVe			tBERT, USE	100	200	N.A.	N.A.
			ELMo, GloVe		tBERT	300	500	100	1
		Ling, fastText, ELMo,			tBERT, USE,	200	400	N.A.	N.A.
		GloVe			InferSent				

Table 2. Hyper-parameter values corresponding to multi-label sexism classification results for the proposed ensemble approach (the hierarchical architecture with one sentence-level group is used; the number of learners equals <code>max_label_subsets</code> where our automatic label subset selection method is used)

biL-att applied on	Text Encoder	Ensemble settings	LSTM	Attention	#CNN filters	k in k-max-
			dim.	dim.	of each ker-	over-time
					nel size	pooling
ELMo, GloVe	tBERT	max_label_subsets = 2, label_membership = 1	200	400	N.A.	N.A.
ELMo, GloVe, fastText,	tBERT, USE, InferSent	max_label_subsets = 2, label_membership = 1	100	100	N.A.	N.A.
Ling						
ELMo, GloVe	tBERT	max_label_subsets = 3, label_membership = 1	200	400	N.A.	N.A.
ELMo, GloVe	tBERT	max_label_subsets = 3, label_membership = 2	200	400	N.A.	N.A.
ELMo, GloVe	tBERT	8 omnipresent labels, max_label_subsets = 2,	200	400	N.A.	N.A.
		label_membership = 1				
ELMo, GloVe, fastText,	tBERT, USE, InferSent	8 omnipresent labels, max_label_subsets = 2,	200	400	N.A.	N.A.
Ling		label_membership = 1				
ELMo, GloVe	tBERT	8 omnipresent labels, max_label_subsets = 3,	200	500	N.A.	N.A.
		label_membership = 1				
ELMo, GloVe	tBERT	8 omnipresent labels, max_label_subsets = 3,	100	200	N.A.	N.A.
		$label_membership = 2$				
ELMo, GloVe	tBERT	all <i>omnipresent labels</i> , 2 learners	300	500	N.A.	N.A.
ELMo, GloVe, fastText,	tBERT, USE, InferSent	all omnipresent labels, 2 learners	200	300	N.A.	N.A.
Ling						
ELMo, GloVe	tBERT	all omnipresent labels, 3 learners	300	600	N.A.	N.A.

Table 3. Hyper-parameter values corresponding to multi-label sexism classification results with the autoencoder-based method for using unlabeled data (using the hierarchical approach with *biL-att* on ELMo and GloVe and BERT for embedding sentences)

Encoder Settings	Layer Sizes	LSTM	Attention	#CNN filters	k in k-max-
		dim.	dim.	of each ker-	over-time
				nel size	pooling
5 stacked layers, quickly decreasing layer size	922, 819, 717, 614, 512	300	500	N.A.	N.A.
3 stacked layers, quickly decreasing layer size	922, 819, 717	200	400	N.A.	N.A.
5 stacked layers, slowly decreasing layer size	973, 922, 870, 819, 768	300	400	N.A.	N.A.
3 stacked layers, slowly decreasing layer size	973, 922, 870	100	200	N.A.	N.A.

Table 4. Hyper-parameter values corresponding to multi-label sexism classification results under various settings

Multi-Label Setting	Approach	LSTM	Attention	#CNN filters	k in k-max-
_		dim.	dim.	of each ker-	over-time
				nel size	pooling
	biLSTM-Attention	300	600	N.A.	N.A.
Label Powerset with class imbalance correction	Hierarchical-biLSTM-Attention	200	400	N.A.	N.A.
	Best proposed method	100	100	N.A.	N.A.
	biLSTM-Attention	200	400	N.A.	N.A.
Label Powerset without class imbalance correction	Hierarchical-biLSTM-Attention	300	600	N.A.	N.A.
	Best proposed method	200	300	N.A.	N.A.
Binary Relevance	biLSTM-Attention	100	300	N.A.	N.A.
Billary Relevance	Hierarchical-biLSTM-Attention	100	300	N.A.	N.A.
	Best proposed method	100	300	N.A.	N.A.
NCE loss	biLSTM-Attention	200	500	N.A.	N.A.
INCE 1088	Hierarchical-biLSTM-Attention	200	400	N.A.	N.A.
	Best proposed method	100	200	N.A.	N.A.

Table 5. Hyper-parameter values corresponding to results for misogyny detection (baselines use ELMo embeddings; proposed methods use the flat architecture)

		Approac	ch		LSTM	Attention	#CNN filters	k in k-max-
					dim.	dim.	of each ker-	over-time
							nel size	pooling
	CNN-Kim					N.A.	100	N.A.
Baselines	biLSTM	200	N.A.	N.A.	N.A.			
	biLSTM-Attention	200	400	N.A.	N.A.			
eli	C-biLSTM					N.A.	100	N.A.
3as	BERT					N.A.	N.A.	N.A.
	USE					N.A.	N.A.	N.A.
	InferSent				N.A.	N.A.	N.A.	N.A.
	biL-att applied on	c-k-max applied on	c-biL-att applied on	Text Encoder		•		
				tBERT	N.A.	N.A.	N.A.	N.A.
	ELMo			BERT	100	200	N.A.	N.A.
	ELMo			tBERT	100	300	N.A.	N.A.
methods	ELMo			USE	200	400	N.A.	N.A.
thc		ELMo		BERT	N.A.	N.A.	150	4
me	ELMo, GloVe			BERT	100	300	N.A.	N.A.
	ELMo, GloVe			BERT, USE	300	600	N.A.	N.A.
Proposed	concat(ELMo, GloVe)			BERT, USE	100	300	N.A.	N.A.
do.	ElMo, GloVe, Ling			BERT	100	100	N.A.	N.A.
Pr	ElMo, GloVe, Ling			BERT, USE	100	300	N.A.	N.A.
			ElMo, GloVe, fastText,		100	200	100	N.A.
			Ling					
	ElMo, GloVe, fastText,			BERT, USE, In-	100	200	N.A.	N.A.
	Ling			ferSent				
		ElMo, GloVe, fastText,		BERT, USE, In-	N.A.	N.A.	100	1
		Ling		ferSent				
	concat(ELMo, GloVe,			BERT, USE, In-	200	400	N.A.	N.A.
	fastText, Ling)			ferSent				
		ElMo, GloVe, fastText,		BERT, USE, In-	300	600	100	1
	Ling	Ling		ferSent				

288
289
290
291
292
293
294
295
296
297
298
300
301
302
303
304
305
306

Table 6. Hyper-parameter values corresponding to results for misogyny classification (baselines use ELMo embeddings; proposed methods use the flat architecture)

	Approach				LSTM	Attention	#CNN filters	k in k-max-
					dim.	dim.	of each ker-	over-time
							nel size	pooling
	CNN-Kim	CNN-Kim					100	N.A.
Baselines	biLSTM					N.A.	N.A.	N.A.
	biLSTM-Attention					300	N.A.	N.A.
eli	C-biLSTM					N.A.	100	N.A.
Bas	BERT					N.A.	N.A.	N.A.
_	USE				N.A.	N.A.	N.A.	N.A.
	InferSent				N.A.	N.A.	N.A.	N.A.
	biL-att applied on	c-k-max applied on	c-biL-att applied on	Text Encoder				
				tBERT	N.A.	N.A.	N.A.	N.A.
	ELMo			tBERT	200	500	N.A.	N.A.
		ELMo		tBERT	N.A.	N.A.	150	1
			ELMo	tBERT	100	300	100	1
	ELMo			USE	100	200	N.A.	N.A.
methods	concat(ELMo, GloVe)			tBERT	100	100		N.A.
È	concat(ELMo, GloVe)			USE	100	200		N.A.
lue l		ELMo, GloVe		tBERT	N.A.	N.A.	100	1
pa	ELMo, GloVe ELMo, GloVe ELMo, GloVe, Ling			USE	100	100		N.A.
OS	ELMo, GloVe			tBERT	100	100		N.A.
Cop			ELMo, GloVe	tBERT	100	300	100	2
P	ELMo, GloVe, Ling			tBERT	100	200		N.A.
	concat(ELMo, GloVe,			tBERT	100	200	N.A.	N.A.
	Ling)							
	ELMo, GloVe			tBERT, USE	200	400		N.A.
	concat(ELMo, GloVe)	77.) (tBERT, USE	100	200		N.A.
		ELMo		tBERT	200	400	100	1
		ELMo, GloVe		tBERT	200	500	200	1
	ELMo, GloVe, fastText,			tBERT, USE,	100	200	N.A.	N.A.
	Ling			InferSent				

Table 7. Hyper-parameter values corresponding to multi-label sexism classification results with the EBCE loss without weight-based class imbalance correction for deep learning methods (the sub-columns for the proposed methods specify different configurations as to the ways in which the input is embedded and/or processed; different colors/styles represent different sentence-level concatenations in a method; colors/styles have no connections across rows)

P	LSTM	Attention	#CNN filters	k in k-max-			
	dim.	dim.	of each ker-	over-time			
						nel size	pooling
biL-att applied on	c-k-max applied on	c-biL-att applied on	Text Encode	r			
Ling, fastText, ELMo,			tBERT, USE	, 100	100	N.A.	N.A.
GloVe			InferSent				