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 (April 2020), 5 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 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, and 6. 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.

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.

© 2020 Association for Computing Machinery.

Manuscript submitted to ACM

Manuscript submitted to ACM

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

Table 1. Multi-label sexism classification results with the EBCE loss for deep learning methods (different colors/styles represent different sentence-level concatenations)

	rent sentence-level concate	Approac			ICTM	Attention	#CNN filters	lz in lz mov
		Арргоас	:n		l	dim.	of each ker-	
<u></u>	biLSTM-Attention	biLSTM-Attention			200	300	N.A.	N.A.
Ž	Hierarchical-biLSTM-Attention			300	400	N.A.	N.A.	
E	CNN-biLSTM-Attention			300			N.A.	
Baselines (with ELMo)					N.A.	N.A.	N.A.	N.A.
3	CNN-Kim						150	N.A.
ıes	biLSTM							N.A.
li.	BERT-biLSTM-Attenti	BERT-biLSTM-Attention					N.A.	N.A.
ase	USE-biLSTM-Attention	n			300			N.A.
B	InferSent-biLSTM-Atte				100	200	N.A.	N.A.
	biL-att Applied on	c-k-max Applied on	c-biL-att Applied on					
	ELMo	1		tBERT	200			N.A.
	ELMo, GloVe			tBERT	200			N.A.
	concat(ELMo, GloVe)				200			N.A.
	Ling, ELMo, GloVe				200			N.A.
,	ELMo, GloVe	<u></u>		tBERT, USE	200			N.A.
st E	Ling, fastText, ELMo),		tBERT, USE,	200	500	N.A.	N.A.
100	GloVe			InferSent				
et		ELMo					100	1
Proposed methods		ELMo, GloVe			1		200	1
sec			· ·		200		150	1
bo	ELMo	ELMo			200		100	1
ro	ELMo, GloVe	ELMo, GloVe		tBERT, USE	100		100	1
_				tBERT	300			N.A.
	ELMo			redtBERT	300			N.A.
	ELMo, GloVe			tBERT	100			N.A.
	ELMo, GloVe			<u>tBERT</u>	100		N.A.	N.A.
	ELMo, GloVe				200			N.A.
. ?	concat(ELMo, GloVe) concat(ELMo, GloVe) Ling, ELMo, GloVe			tBERT	100			N.A.
1	concat(ELMo, GloVe)			tBERT	300			N.A.
		ELMo, GloVe		tBERT	300	500	100	1
. [3	lei		ELMo, GloVe	tBERT	100		100	1
2	Ling, ELMo, GloVe				300			N.A.
	ELMo, Glove				300			N.A.
	ELMo, GloVe			,	100			N.A.
	ELMo, GloVe				100		N.A.	N.A.
	ELMo, GloVe	ELMo, GloVe			300		100	1
	Ling, fastText, ELMo),		tBERT, USE,	200	400	N.A.	N.A.
	GloVe			InferSent				

Table 2. 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 omnipresent labels, 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. 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. 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.
L	Hierarchical-biLSTM-Attention	200	400	N.A.	N.A.
	Best proposed method	100	100	of each kernel size N.A. 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.
biliary 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 IOSS	Hierarchical-biLSTM-Attention	200	400	N.A.	N.A.
	Best proposed method	100	200	N.A.	N.A.

157 158 159 160 161 162 163 164 165 168 169 170 171 172 173 174 175 176 177 178 179 181 182 183 184 185 186

188

189 190 191

207 208

LSTM Attention #CNN filters k in k-max-Approach dim. dim. of each ker-over-time nel size pooling CNN-Kim N.A. N.A. 100 N.A. biLSTM 200 N.A. N.A. N.A. biLSTM-Attention N.A. N.A. 200 400 N.A. C-biLSTM N.A. 100 100 BERT N.A. N.A. N.A. N.A. USE N.A. N.A. N.A. N.A. InferSent N.A. N.A. N.A. N.A. c-k-max Applied on c-biL-att Applied on Text Encoder biL-att Applied on tBERT N.A. N.A. N.A. N.A. ELMo BERT 100 200 N.A. N.A. tBERT ELMo 100 300 N.A. N.A. ELMo USE 200 400 N.A. N.A. ELMo BERT N.A. N.A. 150 4 ELMo, GloVe BERT 100 300 N.A. N.A. ELMo, GloVe BERT, USE 300 600 N.A. N.A. concat(ELMo, GloVe) N.A. N.A. BERT, USE 100 300 N.A. N.A. ElMo, GloVe, Ling BERT 100 100 ElMo, GloVe, Ling BERT, USE N.A. 100 300 N.A. ElMo, GloVe, fastText, 100 200 100 N.A. Ling BERT, USE, In-100 ElMo, GloVe, fastText, N.A. 200 N.A. ferSent Ling ElMo, GloVe, fastText, BERT, USE, In-N.A. N.A. 100 1

ferSent

ferSent

ferSent

BERT, USE, In-200

BERT, USE, In-300

400

600

N.A.

100

N.A.

1

Table 5. Results for misogyny detection (baselines use ELMo embeddings; proposed methods use the flat architecture)

Ling

Ling

ElMo, GloVe, fastText, ElMo, GloVe, fastText,

concat(ELMo, GloVe,

fastText, Ling)

Ling

Table 6. Results for misogyny classification (baselines use ELMo embeddings; proposed methods use the flat architecture)

		Approac					#CNN filters	k in k-max-
		• •			dim.	dim.	of each ker-	over-time
							nel size	pooling
	CNN-Kim			N.A.	N.A.	100	N.A.	
s	biLSTM			300	N.A.	N.A.	N.A.	
ne	biLSTM-Attention			100	300	N.A.	N.A.	
eli	C-biLSTM			300	N.A.	100	N.A.	
Baselines	BERT			N.A.	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.	N.A.
ŧ	concat(ELMo, GloVe)			USE	100	200	N.A.	N.A.
me		ELMo, GloVe		tBERT	N.A.	N.A.	100	1
pe	ELMo, GloVe			USE	100	100	N.A.	N.A.
Proposed	ELMo, GloVe			tBERT	100	100	N.A.	N.A.
do.			ELMo, GloVe	tBERT	100	300	100	2
P	, , , ,			tBERT	100	200	N.A.	N.A.
	concat(ELMo, GloVe,			tBERT	100	200	N.A.	N.A.
	Ling)							
	ELMo, GloVe			tBERT, USE	200	400	N.A.	N.A.
	concat(ELMo, GloVe)			tBERT, USE	100	200	N.A.	N.A.
		ELMo		tBERT	200	400	100	1
	,	ELMo, GloVe		tBERT	200	500	200	1
	ELMo, GloVe, fastText,			tBERT, USE,	200	400	N.A.	N.A.
	Ling			InferSent				
	ELMo, GloVe, fastText,			tBERT, USE,	100	200	N.A.	N.A.
	Ling			InferSent				