Profiling Twitter users

using autogenerated features invariant to data distribution

Tiziano Fagni, IIT-CNR, Pisa - *tiziano.fagni@iit.cnr.it*
Maurizio Tesconi , IIT-CNR, Pisa - *maurizio.tesconi@iit.cnr.it*

Consiglio Nazionale delle Ricerche

"BOTS AND GENDER PROFILING" TASK

Given a Twitter user timeline (100 tweets), classify the user as "bot | human" and "male | female" if the user is identified as human

Multilingual problem: English and Spanish

AVAILABLE DATASETS

English

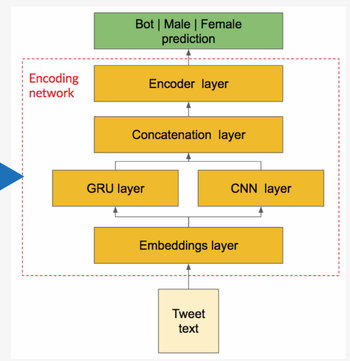
Split type	Bot	Female	Male	Total
train	1,440	720	720	2,880
validation	620	310	310	1,240

Spanish

Split type	Bot	Female	Male	Total
train	1040	520	520	2,080
validation	460	230	230	920

3 DIFFERENT TYPES OF TEXTUAL FEATURES

SYN encoder



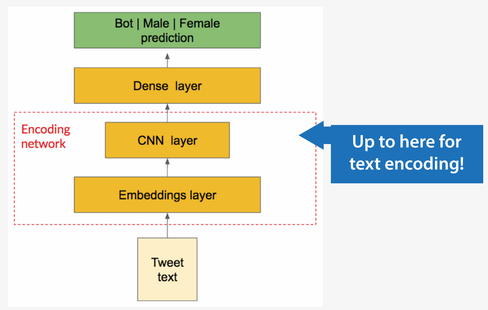
Encoding size: 300

SYN
Based on deep learning and exploiting only text's syntax.

W2V
Based on word2vec method.

CHARS
Based on deep learning and representing text at character level.

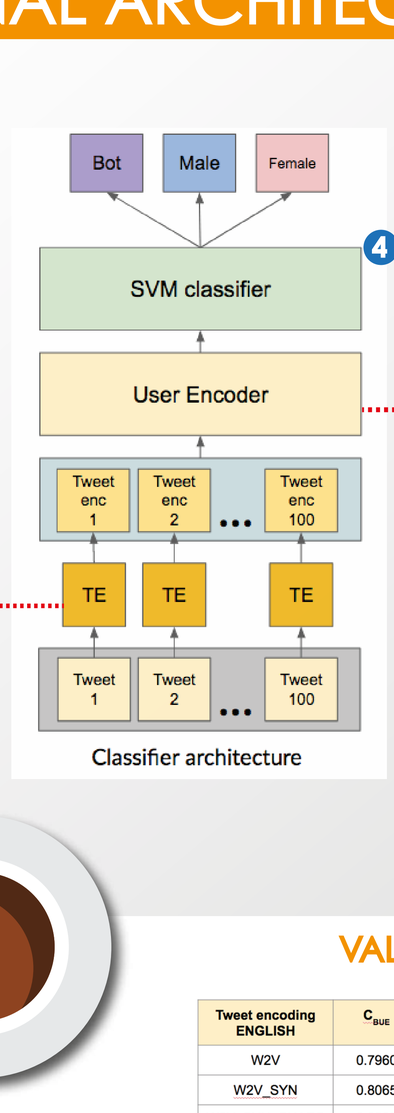
CHARS encoder

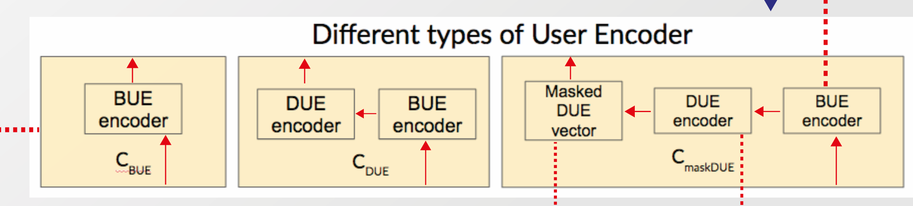


FINAL ARCHITECTURE

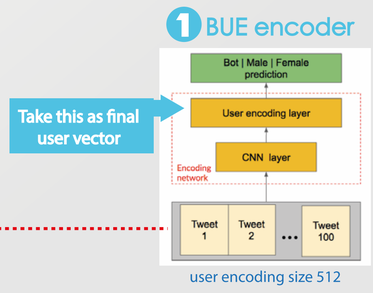
How we handle data mismatch

1. Learn a Twitter user encoder by processing its tweets and evaluate it on the final challenge task.
2. Highlight features which allow to distinguish users coming from training and test distribution.
3. Mask most important features able to distinguish users coming from training and test distribution.
4. Build final classifier skipping masked features.



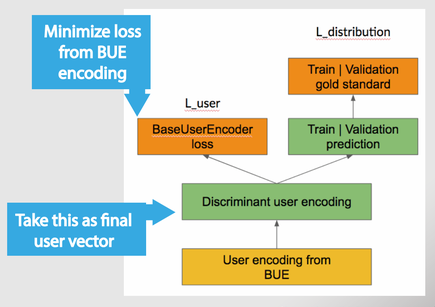


1 BUE encoder



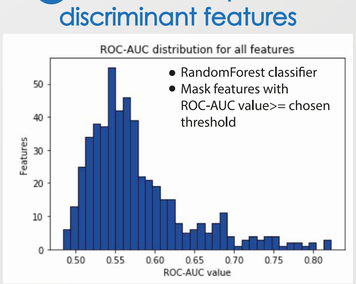
user encoding size 512

2 DUE encoder



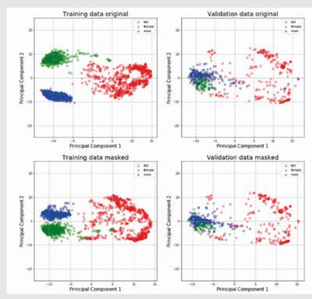
user encoding size 512

3 Mask most important discriminant features



Features obtained by DUE encoder

3 Data clustering with masking



VALIDATION SET

Tweet encoding ENGLISH	C_{BUE}	C_{DUE}	Mask cut threshold	$C_{maskDUE}$
W2V	0.7960	**0.8032**	0.65	0.7998
W2V_SYN	0.8065	0.8225	0.65	**0.8282**
W2V_SYN_CHARS	0.8203	0.8169	0.7	**0.8266**

Tweet encoding SPANISH	C_{BUE}	C_{DUE}	Mask cut threshold	$C_{maskDUE}$
W2V	0.6920	**0.6967**	0.65	0.6945
W2V_SYN	0.7670	0.7695	0.65	**0.7741**
W2V_SYN_CHARS	0.7789	0.7771	0.7	**0.7793**

User encoding DL net: CNN
SVM classifier Params 'C': 0.1, 'gamma': 0.01, 'kernel': 'rbf'

EARLY TEST DATASET

English

Tweet encoding	User encoding	Val type accuracy	Val gender accuracy	Best val accuracy	Test type accuracy	Test gender accuracy	Test accuracy
W2V	C_{DUE}	0.8693	0.7370	0.8032	0.8902	0.7576	0.8239
W2V_SYN	$C_{maskDUE}$	**0.9080**	0.7483	**0.8282**	**0.9091**	**0.7955**	**0.8523**
W2V_SYN_CHARS	$C_{maskDUE}$	0.8967	**0.7564**	0.8266	0.8939	0.7424	0.8181

Spanish

Tweet encoding	User encoding	Val type accuracy	Val gender accuracy	Best val accuracy	Test type accuracy	Test gender accuracy	Test accuracy
W2V	C_{DUE}	0.8630	0.5304	0.6967	0.8778	0.6889	0.7833
W2V_SYN	$C_{maskDUE}$	**0.8863**	0.6619	0.7741	**0.8944**	**0.7556**	**0.8250**
W2V_SYN_CHARS	$C_{maskDUE}$	0.8826	**0.6760**	**0.7793**	0.8778	0.6722	0.7750

FINAL TEST DATASET

	"Type" task		*"Gender" task*		
Method	**English**	**Spanish**	**English**	**Spanish**	**Global**
MAJORITY	0.5000	0.5000	0.5000	0.5000	0.5000
RANDOM	0.4905	0.4861	0.3716	0.3700	0.4295
CHAR N-GRAMS	**0.9360**	0.8972	0.7920	0.7289	0.8385
WORD N-GRAMS	0.9356	0.8833	**0.7989**	0.7244	0.8355
WORD EMBEDDINGS	0.9030	0.8444	0.7879	0.7156	0.8127
LDSE	0.9054	0.8372	0.7800	0.6900	0.8031
W2V_SYN - $C_{maskDUE}$	0.9148	**0.9144**	0.7670	**0.7589**	**0.8387**