

Aggressive Language Detection using VGCN-BERT for Spanish Texts

Errol Mamani Condori
errol.mamani@ucsp.edu.pe

José Ochoa-Luna
jeochoa@ucsp.edu.pe

Outline

- overview
 - Motivation
 - task description and Aggressiveness in social Medias,
- Related work
 - transfer learning with BERT, GCN
- Methodology and Data
 - Our contribution, MEX-A3T, VGCN-BERT
- Experiments
- Conclusion

Motivation

**Donald J. Trump** @realDonaldTrump

Follow

Sadly, the overwhelming amount of violent crime in our major cities is committed by blacks and hispanics-a tough subject-must be discussed.

RETWEETS 3,217 LIKES 3,214

1:05 AM - 5 Jun 2013

**Donald J. Trump** @realDonaldTrump

Following

Mexico's court system corrupt.I want nothing to do with Mexico other than to build an impenetrable WALL and stop them from ripping off U.S.

RETWEETS 574 LIKES 690

574 690

**Miklos Lukacs** @mlukacs · Dec 31, 2017

El criminal proceder de la corrupta organizacion antifujimorista @IDL_R no acabara con lamentos sino con acciones legislativas concretas e inmediatas. El sicariato politico-mediatco que el secuestrado Gorriti y sus secuaces han institucionalizado debe ser duramente sancionado



37 257 416

37 257 416

**Donald J. Trump** @realDonaldTrump

Follow

Sorry losers and haters, but my I.Q. is one of the highest -and you all know it! Please don't feel so stupid or insecure,it's not your fault

1,105 RETWEETS 563 FAVORITES

1,105 563

**Jayda Fransen** @Jayda8F

Follow

VIDEO: Muslim migrant beats up Dutch boy on crutches!



**Serpa Olivos** † @ElenaOlivos

Seguir

@contab36 No tengo NADA contra San Marcos.Hablo de serranos como tú,que tienen twitter.En mi universidad no entra gente como tú,color puerta

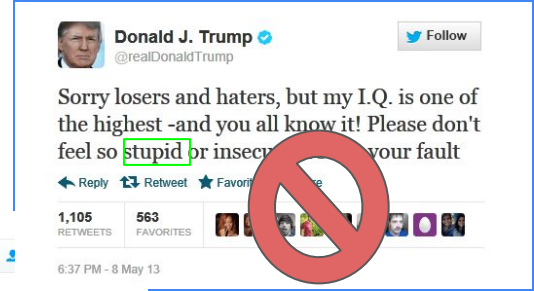
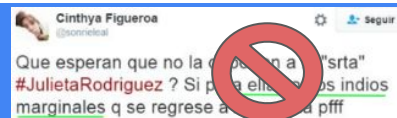
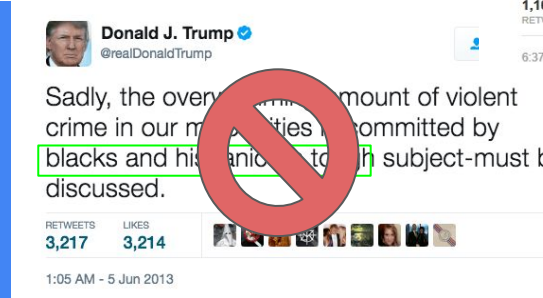
16:53 · 11 de dic. de 2014

**Cynthia Figueroa** @cynthiareal

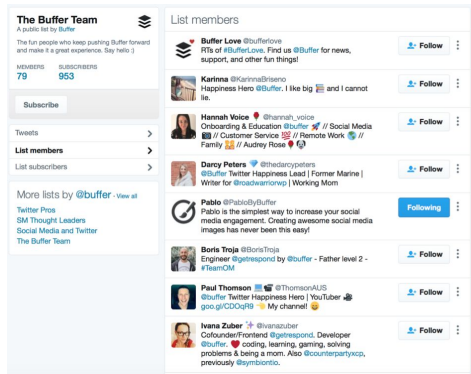
Seguir

Que esperan que no la deportan a la "srta" #JulietaRodriguez ? Si para ella somos indios marginales q se regrese a Argentina pfff

Motivation



Task description



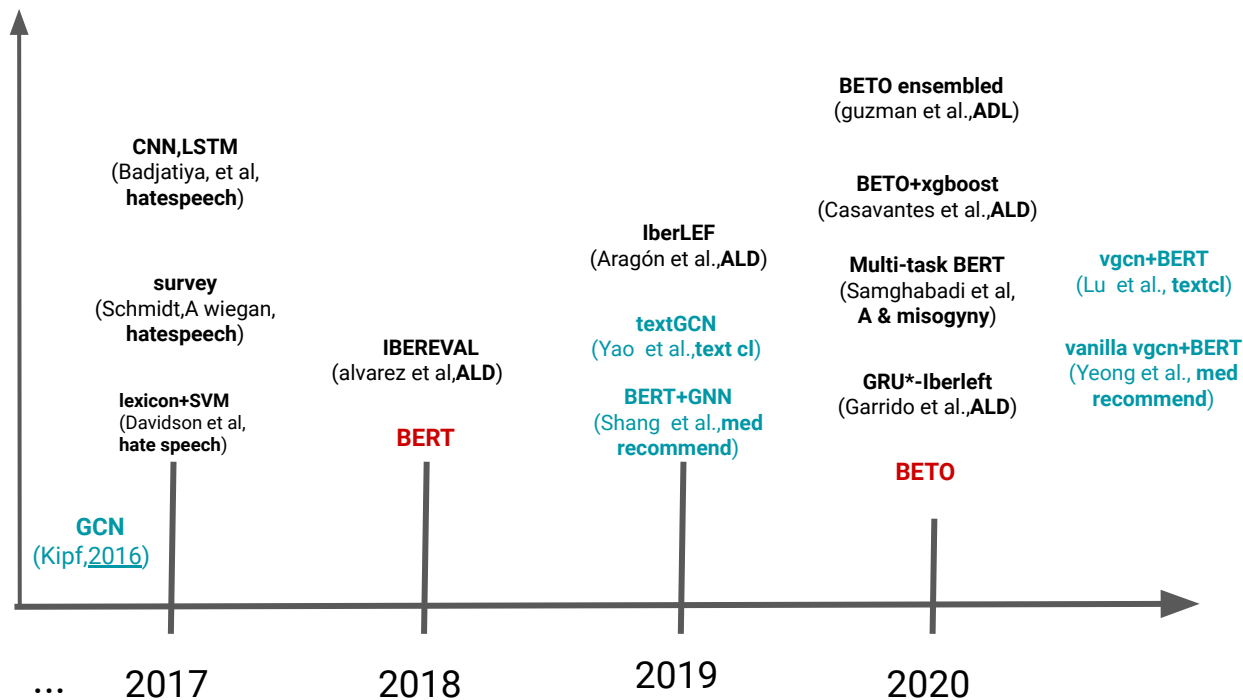
Hate
speech

Aggressive

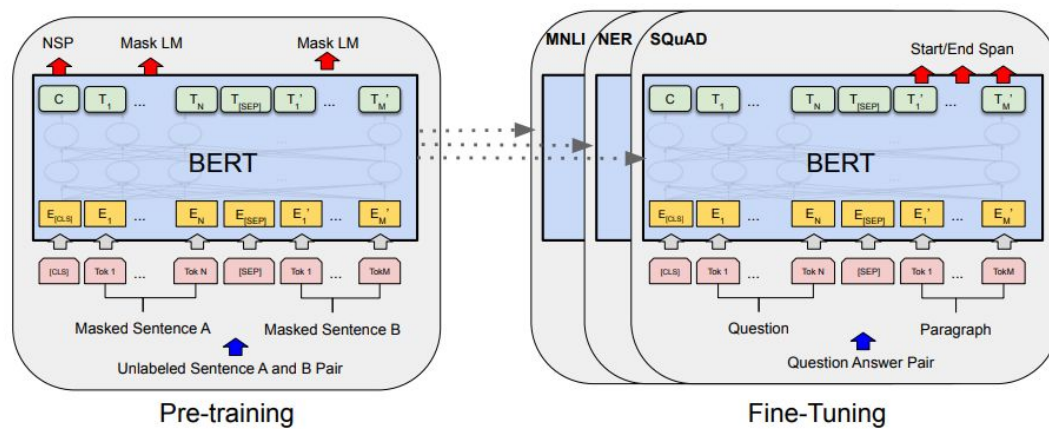


“AGGRESSIVE”

Literature



Transfer Learning with BERT



BERT
(Devlin et al., [2018](#))

Transfer Learning with BERT

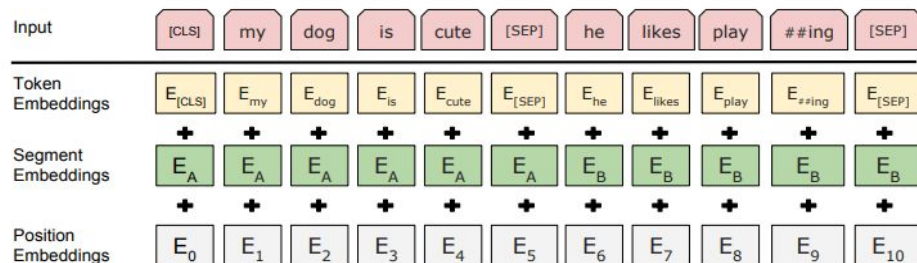
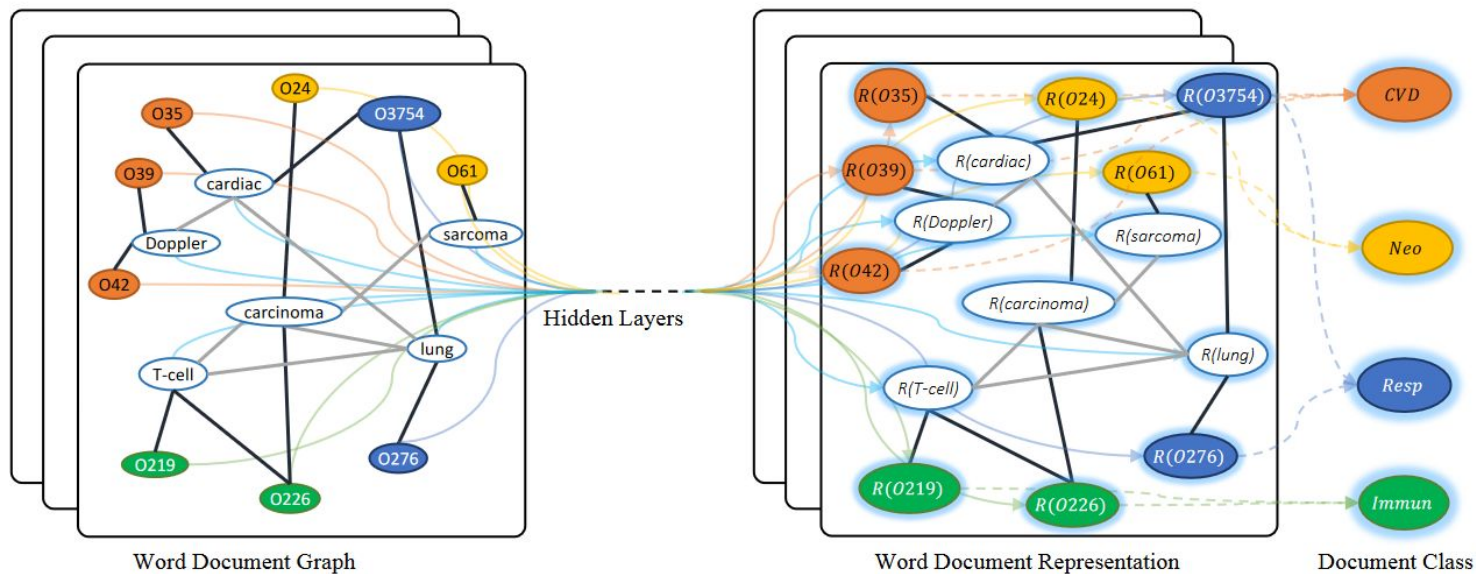


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

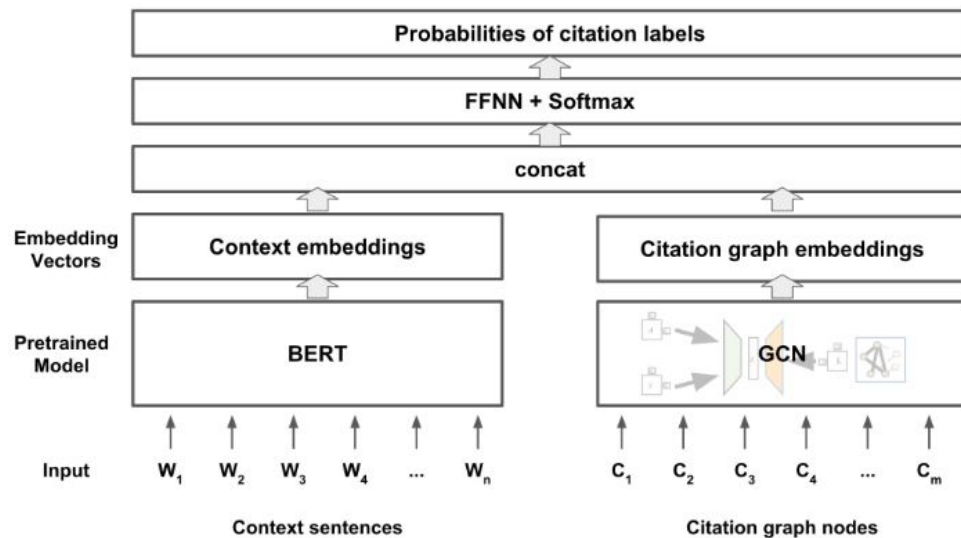
BERT
(Devlin et al., [2018](#))

Graph Convolutional Network



Text GCN
(Yao et al., [2019](#))

Vanilla VGCN-BERT



Text Med citation
(Jeong et al., [2019](#))

Our contribution

- Combining the local and global information using VGCN-BERT without any external word embeddings or knowledge which is suitable and novel to detect the aggressiveness words with a global notion for the Spanish language.
- The combination of VGCN with a BERT model (a Spanish pre-trained version called BETO [9]) allows to obtain comparable results regarding the ensemble BERT models in Spanish Aggressiveness detection.

DataSet Spanish

(MEXT-A3T)

Aggresssive Samples:

“Sólo a las Pu#..aS MOCOS#S GORDAS Y feas les gusta ese al%man xd”

“Profe hijo de las mil pu#..as 6 de calificación como es posible. ”

Non Aggressive Samples:

“Put#s Madres ahora comprendo todo, tu tan lava y yo tan frio ”

“ Segunda vez que me pasa. Estoy hasta la madre”

DataSet Spanish

(MEXT-A3T)

Las fans de odiseo se ven bien bonitas en sus fotos de twitter y estan bien feas en persona.

*(Odiseo fans look really pretty in their twitter photos and **they** are pretty ugly in person.)*



Odisseo fans look really pretty...

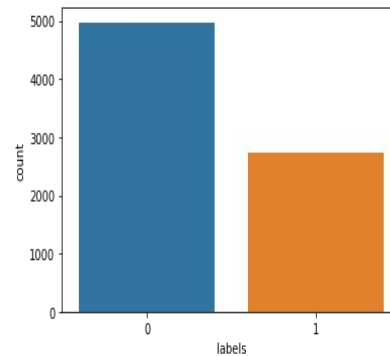
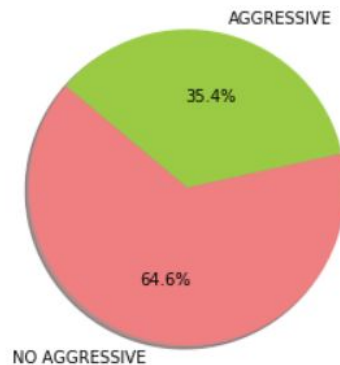


... They are pretty ugly in person.

dataset

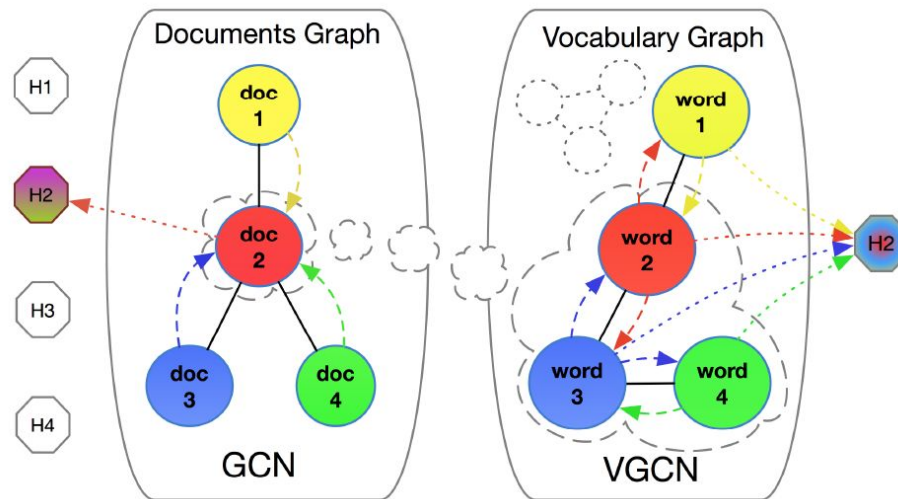
Table 1. MEX-A3T tweets with Aggressive data set distribution of classes.

Class	Train Corpus	Test Corpus
Non Aggressive	5222	2238
Aggressive	2110	905
Total	7332	3143



Methodology

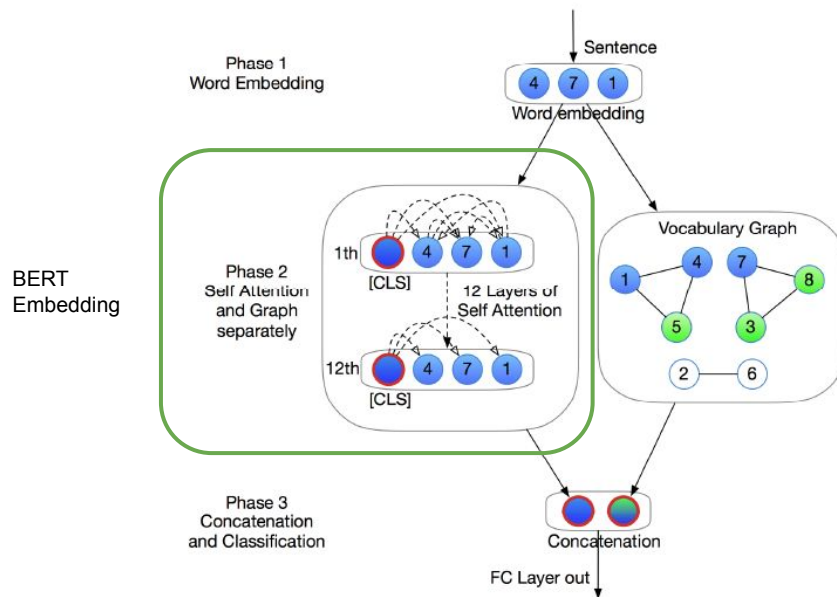
VGCN-BERT



VGCN-BERT
(Lu et al., [2019](#))

Methodology

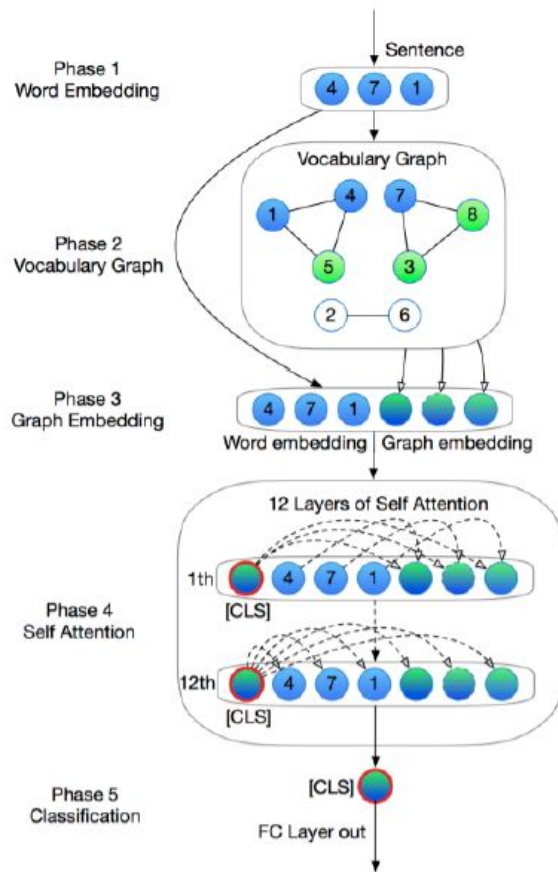
VGCN-BERT



VGCN-BERT based on
Jeon, 2019
(Lu et al., [2019](#))

Methodology

VGCN-BERT



VGCN-BERT
(Lu et al., [2019](#))

Experiments

Table 2. Results of the Aggressive Detection using F1-score in test set.

Model	F1 Aggressive	F1 non-agressive	F1 macro
BoW-SVM	0.6760	0.8780	0.7770
BI-GRU	0.7124	0.8841	0.7983
BETO+msg	0.7720	0.9042	0.8381
bert (multi*)	0.7809	0.9094	0.8452
bert (1 BETO)	0.7998	0.9195	0.8596
bert (20 BETOs)	<u>0.7994</u>	0.92.23	<u>0.8608</u>
VGCN-BERT *	0.8124	0.9169	0.8642

* meas our developed system model.

conclusion

- We have shown that adding a vocabulary graph, even in a small set of vocabulary for Spanish BERT, could improve the context local information.
- Our experimental results allowed us to state that using VGCN-BERT could be beneficial to detect Aggression content, specially for the Spanish language.
- a vocabulary and global information of a language is a path for improving the BERT model as a classier.
- we believe that a word within a sentence could contribute a lot if its relationship within a tweet contributes to detect aggressiveness

Aggressive Language Detection using VGCN-BERT for Spanish Texts

Errol Mamani Condori
errol.mamani@ucsp.edu.pe

José Ochoa-Luna
jeochoa@ucsp.edu.pe