

Automatic classification of the Big-Five personality traits from texts using embeddings and Recurrent Neural Networks

Cristian David Ríos Urrego
Felipe Orlando López Pabón

cdavid.rios@udea.edu.co
forlando.lopez@udea.edu.co

University of Antioquia



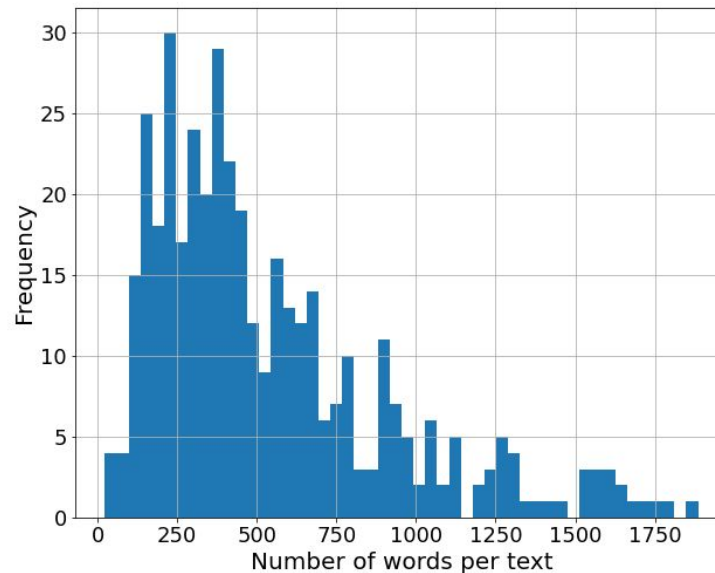
- Dataset
- Pre-processing
- Architectures
- Results GloVe
- Results ELMo
- Summary vs Baseline
- Conclusions

The transcripts of the audio from vlogs of a set of 404 YouTube vloggers that explicitly show themselves in front of the a webcam talking about a variety of topics including personal issues, politics, movies, books, etc are considered [1].

| Trait | Texts with presence of the trait | Texts with absence of the trait |
|-------------------------------|----------------------------------|---------------------------------|
| Openness to experience | 201 | 203 |
| Conscientiousness | 195 | 209 |
| Extraversion | 195 | 209 |
| Agreeableness | 186 | 218 |
| Emotional stability | 201 | 203 |

1. Biel, J. I., Tsiminaki, V., Dines, J., & Gatica-Perez, D. (2013, December). Hi YouTube! Personality impressions and verbal content in social video. In *Proceedings of the 15th ACM on International conference on multimodal interaction* (pp. 119-126).

The transcriptions contains a total of approximately 10K unique words and 240K word tokens. Regarding the gender of the YouTube vloggers, a balance is presented, where 52% of them are female (210 vloggers) and 48% are male (194 vloggers).



Before all, the data must be cleaned and standardized to avoid noise and getting them ready for analysis.

- Convert all the texts to lower case
- Crutches (i.e. words like “um”, “am”) are removed.
- Remove punctuation
- Remove numbers
- Remove Stopwords
- Lemmatization: to transform the words into their root form

Original text:

¡She is reading a Math book! um um She and her boyfriend are going to seE a movie , ¿are they going to dinner together? uh uh

Preprocessed text:

she be read math book she and her boyfriend be go to see movie be they go to dinner together

ELMo Architecture

| Layer (type) | Output Shape | Param # |
|-----------------------------|-------------------|---------|
| input_2 (InputLayer) | (None, 300) | 0 |
| lambda_2 (Lambda) | (None, 300, 1024) | 0 |
| BiLSTM (Bidirectional) | (None, 1024) | 6295552 |
| dense_2 (Dense) | (None, 1) | 1025 |
| Total params: 6,296,577 | | |
| Trainable params: 6,296,577 | | |
| Non-trainable params: 0 | | |

GloVe Architecture

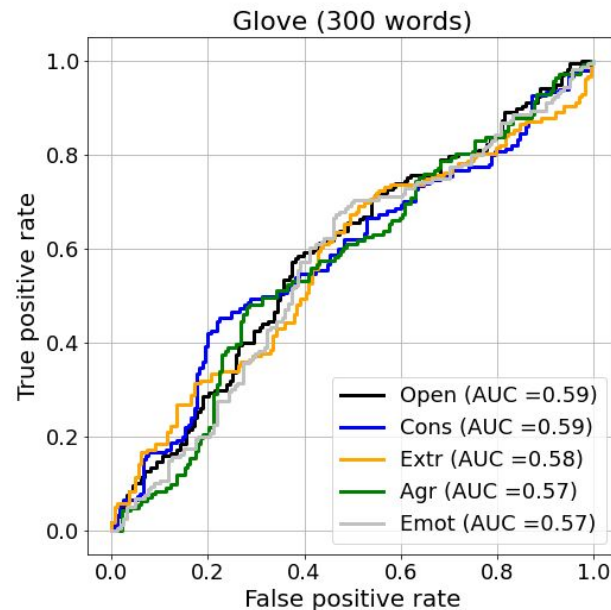
| Layer (type) | Output Shape | Param # |
|---------------------------------|------------------|---------|
| input_1 (InputLayer) | (None, 300) | 0 |
| embedding_1 (Embedding) | (None, 300, 300) | 1500000 |
| BiLSTM (Bidirectional) | (None, 1024) | 3330048 |
| dense_1 (Dense) | (None, 1) | 1025 |
| Total params: 4,831,073 | | |
| Trainable params: 3,331,073 | | |
| Non-trainable params: 1,500,000 | | |

Results Glove (300 words)



UNIVERSIDAD DE ANTIOQUIA

| | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|-------------------------------|--------------|-----------------|-----------------|
| Openness to experience | 58,1 | 63,1 | 53,4 |
| Conscientiousness | 62,2 | 45,3 | 77,1 |
| Extraversion | 58,4 | 65,4 | 51,8 |
| Agreeableness | 61,3 | 48,2 | 71,5 |
| Emotional stability | 59,1 | 61,0 | 57,1 |

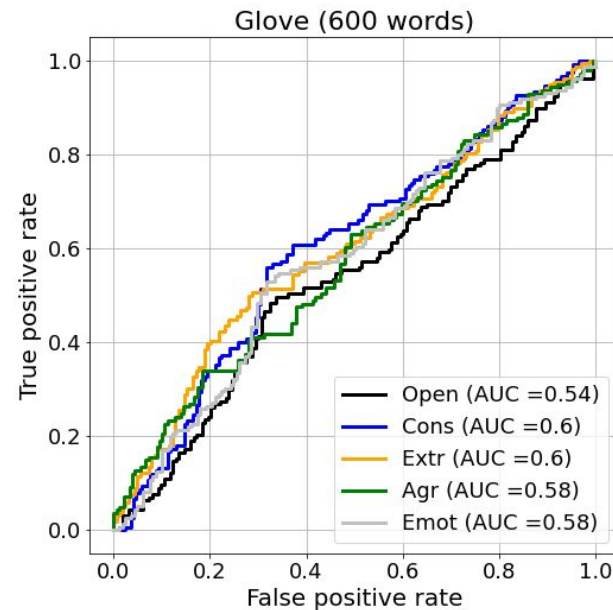


Results Glove (600 words)



UNIVERSIDAD DE ANTIOQUIA

| | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|-------------------------------|--------------|-----------------|-----------------|
| Openness to experience | 57,8 | 46,5 | 68,7 |
| Conscientiousness | 61,3 | 56,7 | 65,3 |
| Extraversion | 60,3 | 41,7 | 78,1 |
| Agreeableness | 60,6 | 34,0 | 81,6 |
| Emotional stability | 60,3 | 54,1 | 66,5 |

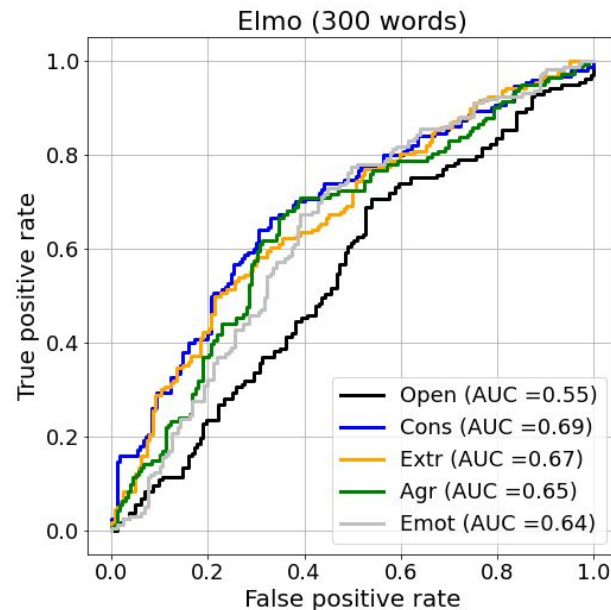


Results Elmo (300 words)



UNIVERSIDAD DE ANTIOQUIA

| | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|-------------------------------|--------------|-----------------|-----------------|
| Openness to experience | 57,8 | 70,7 | 45,4 |
| Conscientiousness | 66,9 | 64,0 | 69,4 |
| Extraversion | 64,1 | 53,5 | 74,4 |
| Agreeableness | 65,9 | 66,7 | 65,4 |
| Emotional stability | 64,1 | 67,3 | 60,9 |

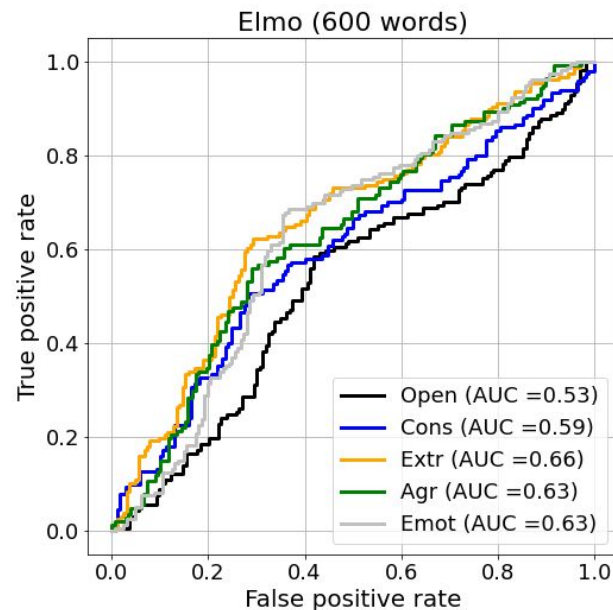


Results Elmo (600 words)



UNIVERSIDAD DE ANTIOQUIA

| | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|-------------------------------|--------------|-----------------|-----------------|
| Openness to experience | 57,8 | 59,9 | 55,8 |
| Conscientiousness | 61,9 | 50,7 | 71,8 |
| Extraversion | 65,3 | 62,8 | 67,7 |
| Agreeableness | 64,4 | 56,0 | 76,9 |
| Emotional stability | 64,7 | 68,6 | 60,9 |



| Best results vs Baseline (SVM with Gaussian Kernel) | | | | |
|---|-----------------------|--------------|-----------------|-----------|
| | Baseline Accuracy (%) | Accuracy (%) | Number of words | Embedding |
| Openness to experience | 56.50 | 58,13 | 300 | GloVe |
| Conscientiousness | 63.40 | 66,86 | 300 | ELMo |
| Extraversion | 63.80 | 65,31 | 600 | ELMo |
| Agreeableness | 60.90 | 65,94 | 300 | ELMo |
| Emotional stability | 56.70 | 64,69 | 600 | ELMo |

- As shown in the results, using the proposed architecture it was possible to improve up to 8% of the accuracy of the models with respect to the baseline, which proves the capacity of architectures like ELMo for the classification of personality traits.
- Preprocessing is an important phase in text analysis because depending on the content of our data when introduced into a neural network, the performance of the model will change, since words with similar semantic and syntactic meaning according to the context should be represented by nearby word embeddings.
- One of the great advantages of bidirectional recurrent neural networks is that these allow embeddings as input (GloVe), and also capture context forward and backward. This considerably improves the performance of current models for the classification and prediction of personality traits.

Automatic classification of the Big-Five personality traits from texts using embeddings and Recurrent Neural Networks

Cristian David Ríos Urrego
Felipe Orlando López Pabón

cdavid.rios@udea.edu.co
forlando.lopez@udea.edu.co

University of Antioquia

