Combined CNN+RNN Bot and Gender Profiling

Notebook for PAN at CLEF 2019

Rafael Felipe Sandroni Dias and Ivandré Paraboni

School of Arts, Sciences and Humanities, University of São Paulo
Av. Arlindo Bettio, 1000. São Paulo, Brazil
rafaelsandroni@usp.br,ivandre@usp.br

Abstract This paper describes an approach to bot and gender author profiling that makes use of a weighted ensemble of convolution (CNN) and recurrent (RNN) neural networks based on char and word n-gram models alike. The proposed ensemble model is shown to outperform the use of its individual classifiers alone, and it was submitted for participation at the PAN-2019 author profiling shared task.

1 Introduction

Author profiling is generally understood as the computational task of inferring an individual's demographics from the text that they have written. The present work addresses two particular instances of author profiling - gender and bot versus human recognition - as developed in the context of the PAN-2019 bot and gender profiling task [8].

The tasks proposed at PAN-2019 consist of determining (in supervised fashion) whether the author of a given piece of text is a bot or a human and, in case of human authorship, to determine their gender. To this end, two datasets were provided, containing 412.000 tweets in English and 300.000 tweets in Spanish, respectively. Both datasets are labelled with author type (human or bot) and, in case of human authors, with gender information (male or female).

Central to our current work, we notice that deep learning methods, although generally successful in many fields, have been applied to author profiling tasks with varying degrees of success. In particular, we notice that the two best-performing systems at PAN-2017 author profiling task [2,5] did not resort to methods of this kind.

Based on these observations, the present work proposes a new take on deep learning methods for author profiling by presenting a combined approach that makes use of a weighted ensemble of convolution (CNN) and recurrent (RNN) neural networks based on char and word n-gram models alike. In doing so, our goal is to assess whether the combination of these resources may outperform the use of their individual classifier components alone, and whether the ensemble method may produce competitive results in bot and gender profiling as proposed at the PAN-2019 shared task.

2 Related Work

In this section we briefly review a number of recent studies that address popular author profiling tasks such as age and gender recognition by making use of deep learning methods.

The work in [1] addresses the issue of age profiling using convolutional networks. Unlike many similar tasks addressed at PAN-CLEF and elsewhere, however, age classification is modelled as a binary classification task (young / adult). The proposed model was evaluated using a Portuguese text corpus, and it was shown to outperform a number of standard baseline alternatives.

The use of deep learning methods is also the focus of the study in [3], which presents a character-based Convolutional Bidirectional Long Short-Term Memory (LSTM) and a word-based Bidirectional LSTM using Global Vectors (GloVe) for gender profiling on Twitter. A stacked architecture combining both models outperforms the use of the individual models alone, and also outperforms a number of bag of words and n-grams baseline models.

The work in [4] addresses gender, binary age bracket and user type ('individual', 'organisation' and 'other') profiling on Twitter data. All tasks were modelled as a graph vertex classification task based on two strategies: Naive Recursive Neural Unit (NRNU) and Long Short-Term Memory Unit (LSTMU). In both cases, the proposed models were found to outperform a number of baseline systems base on lexical information, logistic regression, label propagation, and others.

Finally, the PAN-CLEF 2018 shared task in [9] introduced a gender profiling task based on a combination of text and image data. Among the participant systems, the work in [10] was the overall winner of the competition by presenting a neural network model named 'Text Image Fusion Neural Network' (TIFNN) designed to leverage both text and image data sources.

3 Current Work

Our current work is motivated by the assumption that different learning strategies and text representations may provide multiple contributions to the tasks at hand, namely, bot and gender author profiling. To this end, the proposed model - hereby called *CNN+RNN-char-word* - combines four neural models into an ensemble setting: char- and word-level convolutional networks (CNN), and char- and word-level recurrent networks (RNN). To this end, both CNNs follow a multichannel architecture for variable-length n-gram representations, and both RNNs follow a long short term (LSTM) architecture with self-attention. The ensemble architecture is illustrated in Figure 1.

The weights assigned to each of the individual classifiers are computed by searching for an optimal confidence estimate for each model so that the combined output error is minimised. Optimization proper is performed by making use of Simplex [6] using training data according to equation 1.

$$\tilde{w} = \min \sum_i f_i(x) * w_i \tag{1}$$

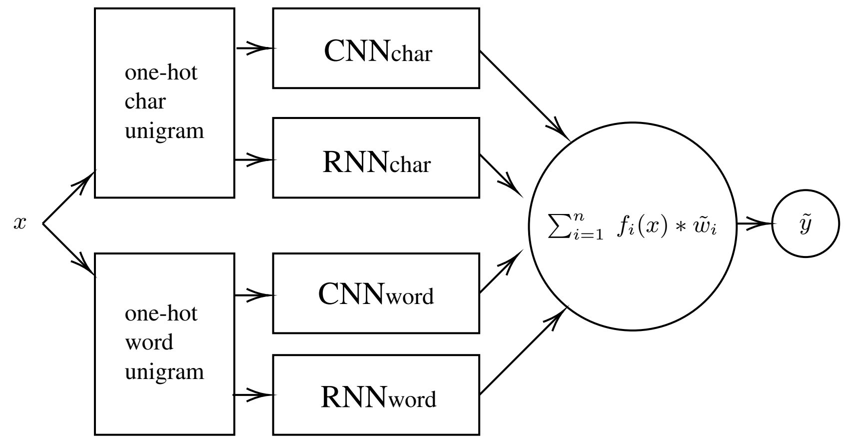

Figure 1: CNN+RNN char+word ensemble architecture.

In this equation, for each model i, we vary the voting weights $w \in \mathbb{R}$ between 0 and 1 so as to minimise the error rate of the combined models f given by the factor $\tilde{w}$, which represents the relative weight of the model in the overall results.

Input texts are individually transformed into one-hot character- and word-level vectors as required by each model. Vectors are taken as initial weights for the input layers in both CNN and RNN modules, and subsequently optimised through back-propagation.

The CNN modules make use of 1D convolutional layers with max pooling, using filters of size 3 and 4, both of which with mappings of size 64, and using ReLU as an activation function, a 0.003 L2 regularization max pooling of size 2. This is followed by a fully-connected layer containing 1024 neurons using ReLU and a 0.35 drop-out regularization and a softmax output layer. Training was performed in mini batches of size 32 with RMSProp optimization and using entropy loss as a cost function. Validation was performed over a 20% portion of the data until convergence with Early Stopping.

The RNN models make use of a self-attention mechanism and memory of size 64 with a 0.12 drop-out regularization. This is followed by a hidden layer conveying 1024 neurons using ReLU as an activation function, and a softmax output layer. Training is performed with AdaDelta optimization and using entropy loss as a cost function. Validation is performed over a 20% portion of the data until convergence with Early Stopping.

Results from the four models are combined using a weighted voting strategy based on a confidence estimate for each individual model according to their results during the training stage.

4 Evaluation

The combined *CNN+RNN-char-word* ensemble model described in the previous section was tested against its individual components, namely, *CNN-char*, *CNN-word*, *RNN-char* and *RNN-word*. In addition to that, a simple logistic regression baseline - hereby called

LogReg - using a bag-of-words model with LibLinear optimization, L2 regularization and α =100 was also implemented.

The ensemble voting mechanism is based on the weights assigned to each individual classifier during training by performing Simplex [6]. The actual weights obtained from the PAN-2019 corpus are summarised in Table 1 for the current tasks (bot and gender profiling) and target languages (English and Spanish.)

Table 1: Optimal weights for each individual classifier and task.

Task	CNN-char	CNN-word	RNN-char	RNN-word
Bot-English	0.24	0.11	0.64	0.01
Bot-Spanish	0.18	0.09	0.51	0.22
Gender-English	0.09	0.38	0.10	0.43
Gender-Spanish	-0.02	-0.05	0.27	0.71

From these optimal weights we notice that the bot recognition problem is more generally captured by the RNN-char model, whereas gender profiling seems to be best captured by the RNN-word model.

5 Results

In this section we report results of the proposed Ensemble approach and baseline systems applied to the bot and gender profiling tasks at PAN-2019, using the evaluation tool provided [7].

5.1 Bot Recognition

Table 2 summarises accuracy results for the bot recognition task based on the PAN-2019 development dataset.

Table 2: Bot recognition accuracy results using the PAN-2019 development dataset. Best results for each language are highlighted.

Model	English	Spanish
LogReg	0.87	0.83
CNN-char	0.86	0.82
CNN-word	0.67	0.67
RNN-char	0.90	0.86
RNN-word	0.89	0.85
CNN-RNN-char-word	**0.94**	**0.91**

From these results we notice that bot recognition was a relatively simple task, with accuracy scores above 90% for both languages. We also notice that the Ensemble *CNN-RNN-char-word* model generally outperforms all alternatives, and that the *RNN-char* model was the second best model. As anticipated in the previous section, this outcome provides further evidence that the results of the Ensemble approach are more influenced by the use of RNN models than by the use of CNNs.

5.2 Gender Profiling

Table 3 summarises accuracy results for the gender profiling task based on the PAN-2019 development dataset.

Table 3: Gender profiling accuracy results using the PAN-2019 development dataset. Best results for each language are highlighted.

Model	English	Spanish
Baseline	**0.74**	**0.72**
CNN-char	0.59	0.58
CNN-word	0.55	0.56
RNN-char	0.64	0.63
RNN-word	0.72	0.71
CNN-RNN-char-word	**0.74**	**0.72**

Gender profiling turned out to be much more challenging than bot recognition, as evidenced by the overall lower accuracy scores if compared to those in the previous Table 2. From these results, we notice also that the Ensemble *CNN-RNN-char-word* once again outperforms all CNN and RNN alternatives in both languages, but results in this case are the same as those obtained by the *LogReg* baseline model (at a much lower cost.)

Finally, Table 4 shows results of the Ensemble *CNN-RNN-char-word* approach for both tasks based on the PAN-CLEF 2019 validation dataset, as obtained in the final submission to the shared task.

Table 4: Bot and gender profiling accuracy results using the PAN-2019 validation dataset.

Task	English	Spanish
Bot	0.84	0.82
Gender	0.58	0.65

Results based on validation data are considerably lower than those observed in the development data, and particularly so in the case of the gender profiling task, which suggest a certain degree of overfitting.

6 Final Remarks

This paper presented an approach to bot and gender profiling that makes use of a weighted ensemble of convolution (CNN) and recurrent neural networks (RNN) based on char and word n-gram models alike. The Ensemble approach outperformed a number of baseline alternatives making use of single network and language models, but a certain accuracy loss was observed in the final validation.

As future work, we intend to investigate how the individual components of the model interact, and extend the current architecture by making use of word and character embeddings.

Acknowledgements

The authors acknowledge support by FAPESP grant 2016/14223-0 and from the University of São Paulo.

References

1. aes, R.G.G., Rosa, R.L., de Gaetano, D., Rodríguez, D.Z., Bressan, G.: Age groups classification in social network using deep learning. IEEE Access 5, 10805–10816 (2017)
2. Basile, A., Dwyer, G., Medvedeva, M., Rawee, J., Haagsma, H., Nissim, M.: N-GrAM: New groningen author-profiling model. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum. Dublin (2017)
3. Gopinathan, M., Berg, P.C.: A deep learning ensemble approach to gender identification of tweet authors (2017)
4. Kim, S.M., Xu, Q., Qu, L., Wan, S., Paris, C.: Demographic inference on Twitter using recursive neural networks. In: Proceedings of ACL-2017. pp. 471–477. Vancouver, Canada (2017)
5. Martinc, M., Skrjanec, I., Zupan, K., Pollak, S.: PAN 2017: Author profiling - gender and language variety prediction. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum. Dublin (2017)
6. Nelder, J., Mead, R.: A simplex method for function minimization comput. The Computer Journal 7 (01 1965)
7. Potthast, M., Gollub, T., Wiegmann, M., Stein, B.: TIRA Integrated Research Architecture. In: Ferro, N., Peters, C. (eds.) Information Retrieval Evaluation in a Changing World - Lessons Learned from 20 Years of CLEF. Springer (2019)
8. Rangel, F., Rosso, P.: Overview of the 7th Author Profiling Task at PAN 2019: Bots and Gender Profiling. In: Cappellato, L., Ferro, N., Losada, D., Müller, H. (eds.) CLEF 2019 Labs and Workshops, Notebook Papers. CEUR-WS.org (Sep 2019)
9. Rangel, F., Rosso, P., Montes-y-Gómez, M., Potthast, M., Stein, B.: Overview of the 6th Author Profiling Task at PAN 2018: Multimodal Gender Identification in Twitter. In: Cappellato, L., Ferro, N., Nie, J.Y., Soulier, L. (eds.) Working Notes Papers of the CLEF 2018 Evaluation Labs. CEUR Workshop Proceedings, CLEF and CEUR-WS.org (2018)
10. Takahashi, T., Tahara, T., Nagatani, K., Miura, Y., Taniguchi, T., Ohkuma, T.: Text and image synergy with feature cross technique for gender identification. In: Working Notes Papers of the Conference and Labs of the Evaluation Forum (CLEF-2018) vol.2125. Avignon, France (2018)