

Bots and Gender Prediction Using Language Independent Stylometry-based Approach

Notebook for PAN at CLEF 2019

Shaina Ashraf, Omer Javed, Muhammad Adeel, Haider Ali
Rao Muhammad Adeel Nawab

Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.

shainaashraf@cuilahore.edu.pk, {[omerjaved11](mailto:omerjaved11@gmail.com),
[mirzaadeel6233.haideriqbalm11](mailto:mirzaadeel6233.haideriqbalm11@gmail.com)}@gmail.com,
adeelnawab@cuilahore.edu.pk

Abstract This paper describes our participation for the Bots and Gender Profiling task at *PAN 2019*¹. The aim of this task is to first classify a profile either as bot or human. If the profile is written by a human, it should be further classified as male or female. Our proposed approach is based on language independent stylometry features. A total of 27 language independent stylometry features (18 are character-based features and remaining 9 are emotion-based features) are used to build the system for Bots and Gender Profiling task. On training dataset, for English language, Accuracy scores of 0.97 and 0.80 are obtained for bot and human classification task and male / female classification task respectively. For Spanish language, Accuracy of 0.93 and 0.75 is obtained for bot and human classification task and male / female classification task respectively. On test dataset 1, for English language, Accuracy scores of 0.92 and 0.76 are obtained for bot and human classification task and male / female classification task. For Spanish language, Accuracy of 0.86 and 0.75 is obtained for bot and human classification task and male / female classification task respectively. On test dataset 2, for English language, bot and human classification task and male/female classification task obtained Accuracy scores of 0.92 and 0.76 respectively, whereas for Spanish language, bot and human classification task and male/female classification task obtained Accuracy scores of 0.88 and 0.72 respectively.

Keywords: Bot and Gender Profiling, Author Profiling, Stylometry-based Features, Emotion-based Features, Emojis

¹ Copyright (c) 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2019, 9-12 September 2019, Lugano, Switzerland.

1 Introduction

As the usage of social networking platforms such as Facebook, Twitter, Instagram, blogs and community forums is arising, the communication methods are changing. People feel free to talk, discuss and post their reviews, comments on such channels more frequently. Many people rely on social forums i.e. Reddit, Yelp, Quora and Amazon message boards, etc., to get information, feedback and recommendations for different products and services. However, a large number of users on social networking sites are taking miss-advantage of such forums by making fake profiles, spams and bots. In recent years, bots are being used to pose as humans on social networking platforms to influence other social media users with ideological, political or commercial purposes. Bots can exaggerate the popularity of products by writing positive reviews and rating them. They can also sabotage the reputation of competitive products through negative reviews and ratings. Furthermore, bots are also being widely used for fake news spreading. Therefore, it is important to develop author profiling systems which can discriminate bot profiles from human ones.

The study presents a stylometry-based approach to address the problem of Bots and Gender Profiling. A total of 27 language independent features are used, which can be broadly categorized into: (1) character-based features and (2) emotions-based features. A range of classifiers have been applied including Logistic Regression, Random Forest, Linear SVC, BernoulliNB, MultinomialNB and SVC (Support Vector Classifier) to train and test our proposed system. The developed system is deployed on TIRA [9] for final evaluation on test datasets. A detailed comparison of all the systems presented in the *PAN 2019 Bot and Gender Profiling task* can be found in [10].

The rest of this paper is organized as follows: Section 2 describes related work on author profiling, Section 3 presents our proposed approach, Section 4 describes the experimental setup, Section 5 presents results and their analysis. Finally, Section 6 concludes the paper with future work directions.

2 Related Work

In previous studies, many researchers have explored different methods i.e. stylometry-based, content-based, topic-based, emotion-based and deep learning for finding different demographics of an author on social media. In [1], the authors have applied stylometry-based approach for cross-genre author profiling. Their set of stylometry-based features included 6 vocabulary richness features, 26 character-based features, 16 syntactic features and 7 lexical features. Promising results were obtained using their proposed set of stylometry-based features (Accuracy of 0.576 for gender classification, 0.371 for age classification and 0.256 for combined classification of age and gender).

In [3], the authors have classified the humans and bots by learning tweets patterns and then further categorized bots in to classes i.e. spam bots, consumption and broadcast bots. They proposed a new profiling framework that consists of entropy-based features such as timings of tweets, hashtags, URL's and followers count etc. The author

worked on nearly 159 thousand bots and human data on Twitter. The experiments results show efficient results on malicious and benign bots to find the interesting behavior traits. In [14], authors have investigated content-based features (word and character n-grams) and 64 stylometry-based features (11 lexical word-based, 47 lexical character-based and 6 vocabulary measures) for the identification of gender and age traits on multilingual corpora.

In [18], the authors have focused on instance-based, prototype based and distance-based classification strategy. They have extracted different features i.e. frequency of negative and positive emoticons, mark of retweets, no of hashtags and part of speech tags for the identification gender and language task.

In [6], the authors have detected bots from Wikidata by extracting comment-based features of user. The comments-based features help to examine the editing behavior of registered and non-registered users. The author used the random forest classifier and a gradient boosting classifier and applied optimization by hyper parameter for both models. The performance of model is efficient against the registered user information.

In [19], the authors have used image and text-based combined features for gender identification. They have represented text using bag of terms (BoT) model and for CNN model for image representation.

3 Proposed Language Independent Stylometry-based Approach

Writing style of an author helps to identify various attributes of an author, for example, age, gender, personality type, occupation and political interest etc. It is expected that the writing style of a human is significantly different from a bot. Therefore, stylometry features [13] are likely to be very helpful in discriminating bot profiles from human ones. Another major difference between a human profile and a bot profile is the usage of emotions. The profile generated by a bot is likely to be plain text, whereas on the other hand, a human profile is likely to be a mixture of both text and emotions. Considering the above two factors, our proposed approach uses a combination of character-based stylometry features and emotions-based features to distinguish human from bot. Note that our proposed approach uses language independent stylometry features i.e. they can be applied on any language for bot and human profiling.

In our proposed system, a total of 27 stylometry-based features are used (18 features are character-based and 9 are emotion-based). The set of character-based features includes: (1) url_count, (2) space_count, (3) capital_count, (4) text_length, (5) curly_brackets_count, (6) round_brackets_count, (7) underscore_count, (8) question_mark_count, (9) exclamation_mark_count, (10) dollar_mark_count, (11) ampersand_mark_count, (12) hash_count, (13) tag_count, (14) slashes_count, (15) operator_count, (16) punc_count, (17) line_count, (18) word_count. The set of emotion-based features includes: (1) emoji_count, (2) face_smiling, (3) face_affection, (4) face_tongue, (5) face_hand, (6) face_neutral_skeptical, (7) face_concerned, (8) monkey_face, (9) emotions (for details see Table 3.1).

No	Feature	Description
15	round_brackets_count	Count ()
16	underscore_count	Count _
17	question_mark_count	Count ?
18	exclamation_mark_count	Count !
19	dollar_mark_count	Count \$
20	ampersand_mark_count	Count &
21	hash_count	Count #
22	tag_count	Count @
23	slashes_count	Count Slashes // \
24	operator_count	Count Operators +-*/%<>^
25	punc_count	Count Punctuations ",,:;`
26	line_count	Count next lines \n
27	word_count	Count Words A-Za-z

Table 4.1 Distribution of data in the *PAN19-author-profiling-training* corpus for *Bot and Gender Profiling task*

	Total Profiles	Bot	Male	Female
English	4120	2060	1030	1030
Spanish	3000	1500	750	750

4 Experimental Setup

This section describes the main statistics of the training corpus, evaluation methodology and evaluation measures.

4.1 Training Corpus

We used *PAN19-author-profiling-training* dataset to train our proposed system. We have performed author profiling task for both languages i.e. English and Spanish. The English training corpus contains 4,120 author profiles and each profile contains 100 tweets in English, whereas Spanish training corpus contains 3,000 author profiles and each profile consists of 100 tweets in Spanish (see Table 4.1 for detailed statistics of both corpora). Note that, in our proposed approach, no pre-processing or cleaning operations were performed on both training and test datasets because URL's and hashtags were used as features in the classification task.

4.2 Evaluation Methodology

The tasks of predicting an author's type as bot or human and determining gender from his/her text are treated as supervised document classification tasks. We performed binary classification tasks for distinguishing bot from human and then identification of its gender. A range of classifiers were explored including Logistic Regression, Random Forest classifier, LinearSVC, BernoulliNB, MultinomialNB and SVC to train and test our proposed system. The numeric values generated by the 27 stylometry features (see Section 3) were used as input to these classifiers.

4.3 Evaluation Measure

Evaluation is carried out using Accuracy measure. Accuracy is defined as ratio of correctly predicted profiles to total number of profiles.

$$\text{Accuracy} = \frac{\text{Number of correctly classified profiles}}{\text{Total number of profiles}}$$

5 Results and Analysis

5.1 Results on Training Dataset

Table 5.1 presents the Accuracy results of our proposed approach on *PAN19-author-profiling-training* dataset using 6 different machine learning algorithms. The best results are obtained using Random Forest classifier for both English (0.970 Accuracy for bot/human & 0.802 for gender prediction) and Spanish (0.935 Accuracy for bot/human & 0.755 for gender prediction) languages. As can be noted that these results are very promising, highlighting the fact that language independent character-based, and emotion-based features used in our proposed approach are useful in discriminating a bot from human as well as distinguishing a male profile from a female one.

Table 5.1 Results obtained on *PAN19-author-profiling-training* corpus using our proposed approach for *PAN 2019 Bot and Gender Profiling task*

Classifier	English		Spanish	
	Bot/Human	Male/Female	Bot/Human	Male/Female
Logistic Regression	0.906	0.7303	0.871	0.678
Random Forest	0.970	0.802	0.935	0.755
LinearSVC	0.869	0.5209	0.749	0.577
BernoulliNB	0.904	0.629	0.822	0.603
MultinomialNB	0.813	0.577	0.796	0.657
SVC	0.479	0.490	0.505	0.469

5.2 Results on Test Datasets

In *PAN 2019 Bot and Gender Profiling task*, final evaluation is carried out on two test corpora: (1) *PAN19-author-profiling-test-dataset1* corpus and (2) *PAN19-author-profiling-test-dataset2* corpus. Table 5.2 shows results obtained using our proposed language independent stylometry-based approach on both test corpora. On *PAN19-author-profiling-test-dataset1* corpus, for English language, Accuracy scores of 0.9280 and 0.7652 are obtained for bot/human and male/female classification tasks respectively, whereas for Spanish language, 0.8611 and 0.7556 Accuracy scores are obtained for human/bot and male/female classification tasks respectively. Similarly, on *PAN19-author-profiling-test-dataset2* corpus, for English language, Accuracy scores of 0.9227 and 0.7583 are obtained for bot/human and male/female classification tasks respectively, whereas for Spanish language, 0.8839 and 0.7261 Accuracy scores are obtained for human/bot and male/female classification tasks respectively.

It can be noted that Accuracy results for English tweets are higher compared to Spanish, even though same language independent features are extracted for both languages. The possible reason for this is that Spanish profiles in the train and test

Table 5.2 Results obtained on *PAN19-author-profiling-test-dataset1* and *PAN19-author-profiling-test-dataset2* corpora using our proposed approach for *PAN 2019 Bot and Gender Profiling task*

Corpus	English		Spanish	
	Type: Bot/Human	Gender: Male/Female	Type: Bot/Human	Gender: Male/Female
<i>PAN19-author-profiling-test-dataset1</i>	0.9280	0.7652	0.8611	0.7556
<i>PAN19-author-profiling-test-dataset2</i>	0.9227	0.7583	0.8839	0.7261

datasets of the *PAN 2019 Bot and Gender Profiling task* may contain text in more than one language since the datasets provided by the *PAN* organizers contain raw tweets and re-tweets i.e. no pre-processing and / or cleaning is performed. Consequently, performance drops for the Spanish language. These results also show that the Accuracy for the identification of type i.e. human/bot is very high compared to gender prediction which shows that our proposed stylistic features are more suitable for discriminating bot from human than gender discrimination. This is likely to happen because bots are likely to generate profiles without emotions and humans generate profiles with a combination of emotions and texts. Consequently, it makes it easier for the classifiers to distinguish human from bot.

6 Conclusion

This paper presents a language independent stylometry-based approach for the *PAN 2019 Bot and Gender Profiling task*. A total of 27 stylistic features were used to build the proposed system (18 are character-based and 9 emotion-based). A range of classifiers were also applied including Logistic Regression, Random Forest, LinearSVC, BernoulliNB, MultinomialNB and SVC. Promising results were obtained on both test datasets in the final evaluation.

In future, we plan to apply deep learning methods for the *PAN 2019 Bot and Gender Profiling task*.

References:

1. Ashraf, S., Iqbal, H. R., & Nawab, R. M. A. (2016, September). Cross-Genre Author Profile Prediction Using Stylometry-Based Approach. In *CLEF (Working Notes)* (pp. 992-999).
2. Ferrara, E., Varol, O., Menczer, F., & Flammini, A. (2016, March). Detection of promoted social media campaigns. In *tenth international AAAI conference on web and social media*.
3. Oentaryo, R. J., Murdopo, A., Prasetyo, P. K., & Lim, E. P. (2016, November). On profiling bots in social media. In *International Conference on Social Informatics* (pp. 92-109). Springer, Cham.
4. Shu, K., Wang, S., & Liu, H. (2018, April). Understanding user profiles on social media for fake news detection. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 430-435). IEEE.
5. Rangel, F., Rosso, P., Potthast, M., & Stein, B. (2017). Overview of the 5th author profiling task at pan 2017: Gender and language variety identification in twitter. *Working Notes Papers of the CLEF*.
6. Hall, A., Terveen, L., & Halfaker, A. (2018). Bot Detection in Wikidata Using Behavioral and Other Informal Cues. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 64.
7. Rangel, Francisco, Paolo Rosso, Manuel Montes-y-Gómez, Martin Potthast, and Benno Stein. "Overview of the 6th author profiling task at pan 2018: multimodal gender identification in Twitter." *Working Notes Papers of the CLEF* (2018).
8. Daelemans, W., Kestemont, M., Manjavancas, E., Potthast, M., Rangel, F., Rosso, P., Specht, G., Stamatatos, E., Stein, B., Tschuggnall, M., Wiegmann, M., Zangerle, E.: Overview of PAN 2019: Author Profiling, Celebrity Profiling, Cross-domain Authorship Attribution and Style Change Detection. In: Crestani, F., Braschler, M., Savoy, J., Rauber, A., Müller, H., Losada, D., Heinatz, G., Cappellato, L., Ferro, N. (eds.) Proceedings of the Tenth International Conference of the CLEF Association (CLEF 2019). Springer (Sep 2019)
9. 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)
10. 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)
11. Rangel, F., Rosso, P., Verhoeven, B., Daelemans, W., Potthast, M., Stein, B.: Evaluations Concerning Cross-genre Author Profiling. In: Working Notes Papers of the CLEF

2016 Evaluation Labs. CEUR Workshop Proceedings, CLEF and CEUR-WS.org (2016)

12. Soler, J., and Wanner, L. 2016. A semi-supervised approach for gender identification. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC-2016), Portoroz', Slovenia, European Language Resources Association (ELRA).
13. Flekova, L., Ungar, L., and Preotiu-Pietro, D. 2016. Exploring stylistic variation with age and income on Twitter. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016), Berlin, Germany.
14. Fatima, M., Hasan, K., Anwar, S., and Nawab, R. M. A. 2017. Multilingual author profiling on Facebook. *Information Processing & Management* 53(4): 886–904.
15. Przybyla, P., and Teisseyre, P. 2015. What do your look-alikes say about you? Exploiting strong and weak similarities for author profiling—Notebook for PAN at CLEF 2015. In *Evaluation Labs and Workshop – Working Notes Papers (CLEF-2015)*, Toulouse, France. CEUR-WS.org.
16. Rangel, F., Rosso, P., Franco, M. A Low Dimensionality Representation for Language Variety Identification. In: *Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'16)*, Springer-Verlag, LNCS(9624), pp. 156-169, 2018
17. Shrestha, P., Rey-Villamizar, N., Sadeque, F., Pedersen, T., Bethard, S., and Solorio, T. 2016. Age and gender prediction on health forum data. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC-2016)*. European Language Resources Association (ELRA).
18. Adame-Arcia, Y., Castro-Castro, D., Ortega-Bueno, R., Muñoz, R.,: Author Profiling, instance-based Similarity Classification. *Notebook for PAN at CLEF 2017 (2017)*
19. Taniguchi, T., Sakaki, S., Shigenaka, R., Tsuboshita, Y., Ohkuma, T.: A Weighted Combination of Text and Image Classifiers for User Gender Inference, pages 87–93. *Association for Computational Linguistics (2015)*