Multi-Platform Authorship Verification

Abdulaziz Altamimi, Nathan Clarke*
Steven Furnell†
{Abdulaziz.altamimi,Nathan.clarke,Steven.furnell}@plymouth.ac.uk
Centre for Security, Communications and Network Research
University of Plymouth
Plymouth, UK

Fudong Li
Fudong.Li@port.ac.uk
School of Computing
University of Portsmouth
Portsmouth, UK

ABSTRACT

At the present time, there has been a rapid increase in the vari-ety and popularity of messaging systems such as social network messaging, text messages, email and Twitter, with users frequently exchanging messages across various platforms. Unfortunately, in amongst the legitimate messages, there is a host of illegitimate and inappropriate content - with cyber stalking, trolling and computer-assisted crime all taking place. Therefore, there is a need to identify individuals using messaging systems. Stylometry is the study of linguistic features in a text which consists of verifying an author based on his writing style that consists of checking whether a target text was written or not by a specific individual author. Whilst much research has taken place within authorship verification, studies have focused upon singular platforms, often had limited datasets and restricted methodologies that have meant it is difficult to appreciate the real-world value of the approach. This paper seeks to overcome these limitations through providing an analysis of authorship verification across four common messaging systems. This approach enables a direct comparison of recognition performance and provides a basis for analyzing the feature vectors across platforms to better understand what aspects each capitalize upon in order to achieve good classification. The experiments also include an investigation into the feature vector creation, utilizing population and user-based techniques to compare and contrast performance. The experiment involved 50 participants across four common platforms with a total 13,617; 106,359; 4,539; and 6,540 samples for Twitter, SMS, Facebook, and Email achieving an Equal Error Rate (EER) of 20.16%, 7.97%, 25% and 13.11% respectively.

CCS CONCEPTS

• Security and privacy → Authentication.

*Nathan Clarke is additionally affiliated to the Security Research Institute, Edith Cowan University, Perth, Western Australia.

†Steven Furnell is also affiliated to the Security Research Institute, Edith Cowan University, Perth, Western Australia and the Centre for Research in Information and Cyber Security, Nelson Mandela University, Port Elizabeth, South Africa

KEYWORDS

author attributing, SMS messaging, Twitter messaging, Facebook messaging, Email messaging, stylometry biometric, Static analysis, Dynamic analysis

ACM Reference Format:
Abdulaziz Altamimi, Nathan Clarke, Steven Furnell, and Fudong Li. 2019. Multi-Platform Authorship Verification. In CECC 2019, 14 - 15 November, 2019, Munich, Germany. ACM, New York, NY, USA, 7 pages. https://doi.org/ 10.1145/3306307.3328180

1 INTRODUCTION

Around 500 million tweets are sent per day, 4.3 billion Facebook messages are posted, more than 200 million emails are sent ev-ery day, approximately 2 million new blog posts are created daily and around 15 billion texts are sent every minute across the globe [24, 25]. Research has shown that it is popular for individuals to have multiple messaging systems [4]. For instance, a study by [9], reported that 64% of Facebook users also had accounts in Myspace, and LinkedIn. The number of monthly active users on several known social networks and messaging systems has grown considerably. For example, the number of active users of "Facebook" expanded from 1 billion in 2012 to 2.7 billion users in 2018 [21, 26]. These messaging systems often provide environment for users to connect with their friends and family, users get together in these messaging systems for text information sharing or making a new relationships.

However, those messaging systems are often utilised and targeted for criminal activities due to their popularity and anonymity [1, 2], ease of use and low cost [19]. This has led to a variety of direct and indirect criminal activities such as, sending spam texts to gain personal information [22, 27], grooming children, kidnap-ping, murder, terrorism and fomenting violence [19]. For example, in 2014, an analysis of the London riots shows that Twitter was used to provide key command and control functionality/services for criminals [12, 28]. This is the first documented example that a messaging system was used to facilitate widespread unlawful activities in the UK.

To verify the identity of the authors of these messages sent across these systems is an approach called authorship verification [14]. Unfortunately, merely relying on the details of the account to verify the author of that account could be misleading because messaging platforms usually do not enforce identity checking, thus, enabling the creation of fake accounts or accounts which are not easily traced back to an individual [18].

Studies have sought to determine the reliability of performing authorship verification but to date have tended to focus upon singular platforms [7, 8, 11, 20, 23]. Given individuals increasing tendency to use multiple platforms, understanding how these platforms per-form (in terms of classification) and analysing how the composition of the resultant feature vector might be similar or different across platforms are valid areas of exploration.

The rest of the paper is structured as follows: the related work is provided in Section 2, with the research methodology being described in Section 3. In Section 4, the experimental methodology is presented, followed by discussion being presented in Section 5. Section 6 concludes the paper and highlights future work.

2 RELATED WORK

According to [18], the long history of stylometry commenced in the 18th century when English logician Augustus de Morgan pro-posed that authorship of a person can be traced by analysing the length of two different texts. In the present time, the emergence of computer networks has led to a vast use of machine learning techniques, which facilitates stylometry analysis [5, 13]. Most re-searchers carried out studies primarily on stylometric features for the sake of author recognition. While a few studies focused on two or three stylometric features. The most detailed study on this topic was undertaken by Mosteller and Wallace [17] who studied the ambiguity of the authorship of the Federalist Papers. They also suggest that it is necessary to employ multiple stylometry features. Some researchers suggested that there are two fundamental classi-fications of stylometry techniques i.e. supervised and unsupervised [3]. In recent times, various courts of law in countries including the United Kingdom, the United States and Australia use stylometric as evidence [6]. According to [7], authorship analysis can be viewed from two different perspectives. Firstly, author identification seeks to identify the most likely author of a target document or post in question, given the samples of the writing of a number of authors. The prime goal is to determine the possibility of which author wrote the document or post in question; the author would be one of those whose samples were given [30]. Some researchers maintain that this kind of authorship identification may be not realistically suit-able and invalid. For example, if the number of suspected authors is large, conceivably the suspected authors are likely to be num-bered into the thousands. Second, there is no guarantee that the true-suspected of an anonymous text is among the known suspects. Also, the amount of the collected samples for each suspected maybe be limited and the anonymous document itself may be short and limited [14].

The second perspective, authorship verification entails checking if a target document was written or not by a specific person by investigating other pieces of writings from that person, it gives a binary answer of "Yes" or "No" to the question [7, 15]. This is ap-propriate for investigation because the suspect’s message would be available in the dataset and therefore the user’s stylometric features for a known verified account can be compared to the unknown account.

In line with previous studies and in terms of verification in most social messaging platforms, Table 1 shows the most recent stud-ies conducted for different messaging platforms. Broadly speaking, little research has been found on stylometry across many of the plat-forms. The seminal work in this field was conducted by [8]. Their research study achieved an Equal Error Rate (EER) of 16.73% for 10 users and 100 samples per user. Lexical, syntactic, and application-specific features were utilized in the features set. Their technique relied upon a n-gram technique to measure the degree of similarity between a block of characters and the profile of a user. On the SMS platform the seminal work was conducted by [23]. Their research study achieved an EER of 24%. Their findings were based upon 30 participants, with a minimum 15 samples per user, maximum sam-ples was not mentioned and a Radial Basis Function (RBF) neural network classifier was used. The EER was 24%, and several users experienced an EER of 0%.

The most prominent previous study in Facebook platform was conducted by [15]. They used posts for Facebook platform in or-der to determine whether user is authenticated among 30 users. Furthermore, they used Support Vector Machine (SVM) Light as the classifier with 233 features, a total of 9259 posts were applied and 12 tests were conducted. They achieved for 10 users with 233 features an accuracy rate of 81.6%. When the author number was increased to 20 and 30, the success rate slightly dropped to 79.8% and 79.6% respectively, with an EER of approximately of 20%.

For email, the most prominent previous study was [11] which yielded EERs ranging from 17.1% to 22.4%. The approach taken in their study was to cluster the anonymous e-mail using stylometric features and extracting the "writeprint" to verify the author. They extracted 292 different stylometry features from 158 users and then analysing these features using EM, k-means, and bisecting k-means classifiers and achieved an EER of 17.1%. However, their technique was based upon clustering and mining the writing styles from a col-lection of e-mails written by multiple anonymous authors and tried to group e-mails written by the same author. The Enron dataset was utilised which has been utilized extensively for authorship analysis research under a variety of different methodological methods, in-cluding text categorisation [20]. Another prominent previous study for email verification [7] yielded an EER of 14.35% using an N gram technique and the Enron corpus involving 87 authors. They used two steps, the first step; the user profile was derived by extracting n-grams from sample documents. The second step, a user specific threshold was computed and used later in the verification phase.

In general, the existing studies has focussed upon single plat-forms, and with limited datasets. There is also a lack of analysis of on the underling feature vectors that are appropriate for users within and across platforms.

3 RESEARCH METHODOLOGY

The goal of this research concerns understanding what the relative performance is across platforms utilising a common population (i.e. is it possible to verify individuals using email as well as it is on Twitter). Within this objective, several aspects of the domain were also examined, including, feature vector composition, classification approaches and the volume of data required to obtain reliable results. These were encapsulated within two core experiments focused upon the feature vector composition: population-based and user-based verification.

Table 1: Summary of literature review for messaging systems in verification studies

Study	Document Type	Authors/#Texts	Method/Approach	Feature type	Classifier	Performance (%)
[8]	Twitter	10/100 sample per user	N-gram	Lexical syntactic structural	Gaussian-Bernoulli	16.73% (EER)
[23]	SMS	30/min 15 sample per user	Calculate user word profiling & linguistic features	Lexical Syntactic Structure emotional keywords	Neural network (RBF Classification)	24% (EER)
[15]	Facebook	30 authors / 308.6 sample per user	Compare classifiers SVM and C4.5	Lexical syntactic structural short messages features	SVM	79.6% (Acc.) (EER)≈ 20%
[11]	Email	150/ Enron email 200,399 e-mails	Clustering	Lexical syntactic Structural amd content-specific	EM, k-means, and bisecting k-means	17.1-22.4 (EER)
[7]	Email	87/Enron email 200,399 emails	n-grams	Lexical syntactic structural and content specific	Ad hoc similarity. Distance(Percentage of shared n-grams)	14.35% (EER)

The proposed research methodology draws upon the prior art [2, 7, 15, 30]. As illustrated in Figure 1, each platform will go through a process to extract features, prioritise the features in terms of discriminative information prior to being applied to a standard supervised training methodology. The need to prioritise the feature vector was as a result of experimenting with the feature vector length upon performance. Whilst several techniques exist to do this, Random Forest algorithms (RF) was selected to perform the features ranking because it has the ability to assign feature importance [16, 29]. Only the top n ranked features are fitted into classifier. One based upon ranking features across the whole population and a second based upon ranking of individual users.

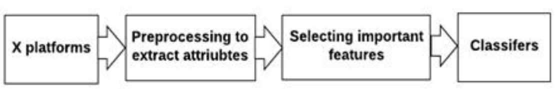

Figure 1: Research methodology

A portion of the 227 stylometric features have been selected from a subset of features from Zheng's research [30] and Li's re-search [15], to include character-based and word-based features, syntactic and structure features because these stylometric features have achieved good performances with online messaging systems and social media texts. Moreover, 48 additional features popularly used in social media such as emotional icons have been included (and it would be the first time that these emoticons features will be used and tested against different platforms). A total of 275 features that included 227 stylometric, and 48 social network specific fea-tures with emoticons features were extracted. Table 2 presents an overview of the feature groups within this.

Each of the messaging platforms required a process to be developed to parse the relevant messaging data, ensuring only relevant data was parsed. For example, in email, it was important to ensure the user's emails are parsed and not the replies to emails that are often appended to email replies. Given the nature of the sensitivity of the data, it was critical that this parsing did not simply extract the messages but automatically performed feature extraction on the fly and run on the participants computer. In this manner, the research team never had to directly store the original messages. Ethical approval was sought and awarded by the host Institution. An automated feature extraction program was developed using NetBeans. It reads the individual input files that have been extracted from Twitter, SMS, Facebook and email text messages. The only thing that appears when collecting data is the interface of an automated feature extraction software as shown in Figure 2.

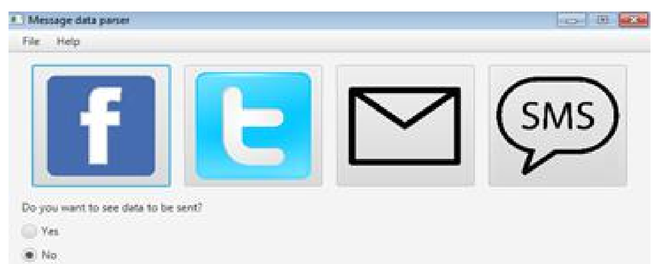

Figure 2: A screenshot of an interface of an automated feature extraction software

Each platform required a bespoke solution:

- E-mail: Outlook e-mails from the folders "sent" and "sent items" within each user's folder were selected; all duplicate e-mails and the email signature were removed. Each e-mail was parsed to extract the body of the message and remove received texts when existing. All e-mails that contain, title, tables and web were removed.
- SMS: Software called Jihosoft Phone Transfer was used to export the data from participants. The export was used to

Table 2: Summary of stylometric features

Feature type	Features	Description
Lexical features	Char based (F1-50)	Character-based features (features 1-50), which count the frequency of specific characters such as number of Alphabetic, characters, Special characters and uppercase that will be tested
	Word based (F210-227)	Word based features (features 210-227), such as counting the frequency of long words or short words will be tested.
Syntactic features	Punctuations (F51-58)	A set of punctuations listed from (features 51-58) will be tested.
	Function words (F59-208)	A set of function words listed from (features 59-208) will be tested
Structural features	No of sentences (F209)	(Feature 209), which shows the number of sentences will be tested. F213 and F214 can calculate sentences feature and can be categorized for structural feature or word based lexical feature.
Social network and emotional specific features	Social network specific features (F228-275)	Such as smiley faces, emotional icons and missing proper punctuation listed from features.

parse each SMS text to extract the body of the message and remove received texts if they existed. All SMS that contain, numbers, title, tables and addresses were removed.

- Twitter: A data crawler from the Twitter API was used to return a list of all tweets of given user. All duplicated tweets, Re-Tweet (RT) tweets, hashtag such as"#word" were replaced by a meta tag "#hash" and removed. In addition, all @user reference was replaced by a meta tag"@cite" and removed.
- Facebook: A graph Facebook API was used to return a list of all posts of given user. All posts that contain, numbers, title, tables and web addresses were removed.

Table 3: Summary of stylometric features

Description	Platforms			
	SMS	Twitter	Facebook	Email
Number of participants	26	41	46	47
Number of text messages	106,359	13,617	4,539	6,540
Average number of text messages per user (mean)	4091	332	99	139
The maximum length of text messages	30.6 words	35 words	1147 words	3712 words
Average length of messages per user	10 words	13 words	15 words	74 words

A total of 50 participants were recruited with the conditions that they had to have a least 2 of the 4 identified platforms and were of consenting age (18 years +). Table 3 presents an overview of the resulting dataset that was acquired.

Three different classification algorithms were applied to investigate an optimum algorithm for classification. The selection of the techniques was determined based upon the anlaysis of the prior art [11, 15, 23]. This included Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting (GB). Each classifier was tested with a different set of features and the corpus of the dataset has been divided for train/test into the ratios 70/30, 60/40, 40/60, 50/50, 20/80 and 10/90 respectively in order to evaluate how much data was required for training to still achieve a reliable result.

4 EXPERIMENTAL RESULTS

4.1 Experiment 1. Population-based verification approach

The results of using population-based approach is shown in Table 4. The results illustrated the best features found in the Train/Test ratios for all platforms in the population was 70/30. The purpose of these tests was to find the classifier with the lowest EER subject to the number of features applied. Therefore, it can be observed that the best performance is achieved by the SMS and email platforms with EERs of 7.97% and 13.11% respectively; with the performance significantly increasing for the other two platforms EERs of (20.16% and 25% for Twitter and Facebook respectively). An analysis of the dataset shows in terms of size and composition, SMS and email repositories represent either end of the spectrum, suggesting volume is less likely to be a determining factor over the composition of the message itself. The results certainly do show a significant difference in performance depending upon the platform utilized.

A comparison of these results with previous studies' performance, as shown in Table 1 for study [23] for SMS and studies [6, 26] of email platforms, shows that the performance in this research for both platforms has outperformed the prior art. The feature vector composition and feature ranking procedure is a likely consequence of this. It is also worth noting, these performances where achieved using higher proportions of participants and samples.

In terms of the number of feature tested, as shown in Table 4, it can be observed that apparently some features have more discriminative value and it is not necessary to include all features e.g

Table 4: Summary of stylometric features

Test ID	Feature tested	Performance EER (%)											
		Twitter			SMS			FB			Email		
		SVM	GB	RF	SVM	GB	RF	SVM	GB	RF	SVM	GB	RF
Test 1	Top 10	24.88	23.24	24.06	14.78	9.81	10.53	27.68	28.31	25.88	22.43	16.84	18.18
Test 2	Top 20	24	21.03	25.51	13.67	8.56	10.78	27.97	27.25	31.97	22.43	13.31	15.61
Test 3	Top 30	24.07	20.16	23.8	15.58	8.35	11.36	26.56	26.5	31.11	24.37	14.44	16.77
Test 4	Top 50	25.78	20.77	27.3	15.9	8.19	11.42	27.89	26.69	28.42	24.8	13.65	17.09
Test 5	Top 100	26.34	20.38	27.3	17.65	7.97	12.58	29.37	25.18	32.28	27.55	13.11	19.81
Test 6	All	31.41	20.47	29.37	21.11	8.1	13.82	38.44	25	33.39	32.78	13.5	22.71

Twitter, SMS and Email. Therefore, subsets of stylometric features would be more reliable in determining authorship across these platforms. In order to investigate some subsets of stylometric features of the relative performances are across platforms utilising a com-mon population, the first top 10 most discriminative features in each platform has been explored as shown in Table 5.

Table 5: Results of the top 10 population feature for each platform

Tw	SMS	FB	Email
219	27	52	29
27	232	55	38
55	231	54	50
3	52	1	55
1	219	2	39
2	215	27	102
52	233	213	51
54	274	212	52
213	1	3	27
39	3	214	42

It shows that all platforms Twitter, SMS, Facebook and Email platform are shared these features: F52 number of punctuation (syntactic) and F27 number of alphabet a-z (lexical). While, F55 number of punctuation (syntactic feature) is shared between Twitter, Facebook and Email. F3 number of uppercase character (lexical) is shared between Twitter, SMS and Facebook. F1 number of charac-ters (lexical feature) is shared between Twitter, SMS and Facebook. F54 number of punctuation (syntactic Feature) is shared between Twitter and Facebook. F2 number of alphabets (lexical feature) is shared between Twitter and Facebook. F39 Number of special char-acter (lexical) is shared between Twitter and Email, lastly, F213 average sentence length in terms of character (structure) is shared between Twitter and Facebook.

It appears that populationaly lexical feature seems to play a large role than other features across platforms as shown in Table 6. Lexical features appeared six times, syntactic features appeared three times, and structure features appeared only once time. Lexical features are the most common feature for population in multi-platforms, even if the number of features rises to above the top 10 features, top 20 features and to top 30 features, because they are involved in more than one platform.

Table 6: Results of the top 10 population feature for each platform

#features	Features	Platforms			
		Twitter	SMS	FB	Email
F52	#punctuation (syntactic)	✓	✓	✓	✓
F27	#alphabets (lexi-cal)	✓	✓	✓	✓
F55	#punctuation (syntactic)	✓		✓	✓
F3	# uppercase char-acters (lexical)	✓	✓	✓	
F1	#characters (lexi-cal)	✓	✓	✓	
F54	#punctuation (syntactic)	✓		✓	
F2	#alphabets (lexi-cal)	✓		✓	
F39	#special charac-ter (lexical)	✓			✓
F213	Average sentence length in terms of char (structure)	✓		✓	
F219	#words with 5 chars (lexical)	✓	✓		

4.2 Experiment 2. User-Based verification approach

For the user-based feature composition, the best results were also found using Train/Test ratios of "70/30" (as illustrated in Table 7). This included selecting a varying number of features ranging from 10 to 275 as an input vector for a classification algorithm.

It has been noticed that SMS and email have achieved good performances of 7.97% and 12.03% respectively. While, Twitter and Facebook messages have achieved poor performance an EER of 20.28% and 23.78% respectively as shown in Table 7.

An analysis of the data set in terms of size and composition of individual users shows that the individual user is likely use the same spectrum that share writing styles in the email and SMS, and there is a clear indication that the writing style used between these two platforms is likely to be similar. For example, it is characterized by common features, one of which is for example to be private and

Table 7: User-based verification experimental (one vs. all authorship verification)

Test ID	Feature tested	Performance EER (%)											
		Twitter			SMS			FB			Email		
		SVM	GB	RF	SVM	GB	RF	SVM	GB	RF	SVM	GB	RF
Test 1	Top 10	23.78	22.02	24.22	14.38	9.04	9.69	24.59	25.1	26.31	18.04	12.95	14.41
Test 2	Top 20	23.4	21.16	24.01	14.23	8.17	10.37	24.95	23.78	28.33	18.75	12.03	14.46
Test 3	Top 30	23.12	20.53	24.95	15.5	8.18	10.46	26.81	24.08	27.69	19.12	12.16	15.06
Test 4	Top 50	24.39	20.42	25.74	15.72	7.99	10.54	26.14	24.64	31.45	24.15	12.33	17.66
Test 5	Top 100	26.95	20.45	26.48	17.27	7.97	13.23	33.27	25.13	32.13	26.41	13.05	21.03
Test 6	All	31.41	20.28	29.66	21.07	8.07	16	37.92	25.09	33.58	32.85	13.87	20.36

personal platforms for the user and texts may be directed to specific people. Suggesting that size is less likely to be a determining factor in the composition of the message itself. The results certainly do show a significant difference in performance depending upon the platform utilized.

In contrast, and in terms of feature tested, subsets of stylometric features would be more reliable in determining authorship using a few features as they have lower performance with a few fea-tures such as EERs of 7.97%, 12.03%, and 23.78% for SMS, email and Facebook platforms, respectively. Facebook and email are more verifiable and can be verified with only a few features (the top 20 features) since the user often writes a longer document on these platforms, and linguistic tendencies are often determined and vice versa with Twitter and SMS, which require more features to be verified for short texts.

In the previous experiment, Twitter for example, in the feature test, had the top 30 features, while user-based involved 275 features. This is might be because the population often has equally similar linguistic feature traits, such as the number of special characters. While for user-based features, the user is often diverse concerning own use of language and thus this has ramifications.

In addition, the GB classifier has outperformed other classifiers in both cases population and user-based. Maybe that's because the GB classifier allows optimization of an arbitrary differentiable loss function [10].

performance was achieved, determining of the best features for each platform was attempted. Table 8 demonstrates the subset features for user 1 on different platforms, including features that are shared by more than one platform. It seems that individually lexical feature seems to play a large role than other features as shown in Table 9. Lexical features appeared seven times, structure features appeared only once time.

5 DISCUSSION

In this paper, subsets of stylometric features in each platforms have been verified in order to find out what the most stylometric features in these platforms. This includes a comprehensive survey on their interrelationships linguistically with other platforms for which of these subsets of stylometric features would be more reliable in determining authorship. It has been found out that the first top 10 feature in population based experiment showed the lexical feature type appeared with a greater frequency than other types, followed by syntactic. It can be said that syntactic and lexical features were in

Table 8: User-based stylometric feature for user 1 in different platforms

Twitter	SMS	Facebook	Email
F1	F1	F1	F1
F32	F28	F220	F40
F2	F3	F224	F20
F33	F214	F275	F213
F55	F213	F19	F3
F3	F220	F33	F2
F24	F2	F24	F228
F4	F4	F3	F4
F19	F24	F2	F59
F224	F211	F219	F214

Table 9: User-based stylometric feature for user 1 in different platforms

#features	Features	Platforms			
		Twitter	SMS	FB	Email
F1	#characters (lexical)	✓	✓	✓	✓
F2	#alphabets (lexical)	✓	✓	✓	✓
F3	#uppercase characters (lexical)	✓	✓	✓	✓
F24	#alphabet a-z (lexical)	✓	✓	✓	
F4	#alphabet a-z (lexical)	✓	✓		✓
F213	Average sentence length in terms of char (structure)		✓		✓
F19	#alphabet a-z (lexical)	✓		✓	
F33	#special character (lexical)	✓		✓	

the top of features between platforms for population when first top 10 of feature were captured and are the most distinctive stylometric features, thus lexical and syntactic features would be more reliable in authorship in these platforms. It can be suggested that SMS and Email these platforms behavioural often similar and thus common in terms of number of (alphabet a-z- lexical feature; punctuation-syntactic feature) and may contain the same words and alphabet and thus similar in the same style of writing context, including the use of the punctuation features as shown in Table 6.

In the user based-feature vector approach, in terms of classifica-tion, it has also been observed that Facebook and Email messages can be more verifiable in terms of user feature base, because they need only top the 20 features to be verified, and since the author often writes a long document and variety of words on in these platforms, which can be easy to verify using a low number of fea-tures other than platforms that have limited texts to verify such as Twitter and SMS.

In terms of similarities and differences in classification between population-based and user-based approaches, it can be noticed the user and population based have a little difference and the user based outperformed population based.

6 CONCLUSIONS AND FUTURE WORK

This research is the first study to explore and investigate multi-platform linguistic user profiling across four major messaging sys-tems (SMS, Twitter, Facebook and Email). Many stylometry features have been recommended in previous work for authorship verifi-cation and it is not clear which one of stylometric features would be robust and trusted. In this research, population-based feature verification method played a pivotal role to establish that each plat-form has its own linguistic behaviour features and can be similar to other platforms for the same behaviours and indeed this facili-tates the process of the verification. Lexical feature show robust in various different messaging system platforms and are is applicable in population-based verification. In the user-based feature compo-sition, each individual author has its own linguistic features, and the performance was seen to outperform the population based. In the future work, research will explore the extent to which a user's profile created from one messaging platform can be reliably applied to data from other platforms. This would enable investigators to take a known set of messages from a suspect (e.g. text messages from his personal phone) and apply the profile to data from other anonymous platforms to determine the probability of the suspect being the individual who created them. This will build towards the development of a unified authorship verification approach.

REFERENCES

[1] Ahmed Abbasi and Hsinchun Chen. 2005. Applying authorship analysis to extremist-group web forum messages. IEEE Intelligent Systems 20, 5 (2005), 67–75.

[2] Ahmed Abbasi and Hsinchun Chen. 2008. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. ACM Transactions on Information Systems (TOIS) 26, 2 (2008), 7.

[3] Ahmed Abbasi and Hsinchun Chen. 2008. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. ACM Transactions on Information Systems (TOIS) 26, 2 (2008), 7.

[4] Mishari Almishari, Ekin Oguz, and Gene Tsudik. 2016. Trilateral large-scale osn account linkability study. In 2016 AAAI Fall Symposium Series.

[5] Shlomo Argamon, Marin Šarić, and Sterling S Stein. 2003. Style mining of elec-tronic messages for multiple authorship discrimination: first results. In Proceed-ings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 475–480.

[6] Michael Brennan, Sadia Afroz, and Rachel Greenstadt. 2012. Adversarial stylom-etry: Circumventing authorship recognition to preserve privacy and anonymity. ACM Transactions on Information and System Security (TISSEC) 15, 3 (2012), 12.

[7] Marcelo Luiz Brocardo, Issa Traore, Sherif Saad, and Isaac Woungang. 2014. Verifying online user identity using stylometric analysis for short messages. Journal of networks 9, 12 (2014), 3347.

[8] Marcelo Luiz Brocardo, Issa Traore, Isaac Woungang, and Mohammad S Obaidat. 2017. Authorship verification using deep belief network systems. International Journal of Communication Systems 30, 12 (2017), e3259.

[9] Cheng Ta Chung, Chia Jui Lin, Chih Hung Lin, and Pu Jen Cheng. 2014. Person identification between different online social networks. In Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 01. IEEE Computer Society, 94–101.

[10] Benoît Cadre Gérard Biau. 2019. OPTIMIZATION BY GRADIENT BOOSTING. https://arxiv.org/pdf/1707.05023.pdf,month=May,lastaccessed=

[11] Farkhund Iqbal, Liaquat A Khan, Benjamin Fung, and Mourad Debbabi. 2010. E-mail authorship verification for forensic investigation. In Proceedings of the 2010 ACM Symposium on Applied Computing. ACM, 1591–1598.

[12] Paul Lewis James Ball. 2011. Twitter and the riots: how the news spread. Retrieved Aug 06, 2011 from https://www.theguardian.com/uk/2011/dec/07/twitter-riots-how-news-spread

[13] Dmitri V Khmelev and Fiona J Tweedie. 2001. Using Markov chains for identifi-cation of writer. Literary and linguistic computing 16, 3 (2001), 299–307.

[14] Moshe Koppel, Jonathan Schler, and Shlomo Argamon. 2013. Authorship Attri-bution: What's Easy and What's Hard? Available at SSRN 2274891 (2013).

[15] Jenny S Li, John V Monaco, Li-Chiou Chen, and Charles C Tappert. 2014. Author-ship authentication using short messages from social networking sites. In 2014 IEEE 11th International Conference on e-Business Engineering. IEEE, 314–319.

[16] Promita Maitra, Souvick Ghosh, and Dipankar Das. 2016. Authorship Verification-An Approach based on Random Forest. arXiv preprint arXiv:1607.08885 (2016).

[17] Frederick Mosteller and David L Wallace. 2012. Applied Bayesian and classical inference: the case of the Federalist papers. Springer Science & Business Media.

[18] SM Nirkhi, Rajiv V Dharaskar, and VM Thakre. 2012. Analysis of online mes-sages for identity tracing in cybercrime investigation. In Proceedings Title: 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec). IEEE, 300–305.

[19] Manoj Parameswaran and Andrew B Whinston. 2007. Social computing: An overview. Communications of the Association for Information Systems 19, 1 (2007), 37.

[20] Fuchim Peng, Dale Schuurmans, Vlado Keselj, and Shaojun Wang. 2003. Au-tomated authorship attribution with character level language models. In 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2003). 267–274.

[21] Kleiner Perkins. [n. d.]. KPCB INTERNET TRENDS 2016. https://www.slideshare. net/kleinerperkins/2016-internet-trends-report/99.

[22] Rachel Roberts. 2017.Online trolling laws under consideration following abuse of MPs. Retrieved Aug 06, 2011 from https: //www.independent.co.uk/news/uk/politics/trolling-laws-online-abuse-mps-under-consideration-lord-bew-a7857891.html

[23] Hataichanok Saevanee, Nathan Clarke, and Steven Furnell. 2011. SMS linguistic profiling authentication on mobile device. In2011 5th International Conference on Network and System Security. IEEE, 224–228.

[24] Jeff Schultz. 2013. How Much Data is Created on the Internet Each Day. Retrieved Aug 06, 2011 from https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day

[25] Kit Smith. 2019. 126 Amazing Social Media Statistics and Facts. Retrieved Aug 06, 2011 from https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts/

[26] Statista. 2019. Number of social network users worldwide from 2010 to 2021 (in billions). Retrieved Aug 06, 2011 from https://www.statista.com/statistics/ 278414/number-of-worldwide-social-network-users/

[27] Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. 2010. Detecting spammers on social networks. In Proceedings of the 26th annual computer security applications conference. ACM, 1–9.

[28] Emma Tonkin, Heather D Pfeiffer, and Greg Tourte. 2012. Twitter, information sharing and the London riots? Bulletin of the American Society for Information Science and Technology 38, 2 (2012), 49–57.

[29] Luis Torgo. 2011. Data mining with R: learning with case studies. Chapman and Hall/CRC.

[30] Rong Zheng, Jiexun Li, Hsinchun Chen, and Zan Huang. 2006. A framework for authorship identification of online messages: Writing-style features and classification techniques. Journal of the American society for information science and technology 57, 3 (2006), 378–393.