DAFFODIL INTERNATIONAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY, VOLUME 15, ISSUE 1, JANUARY, 2020
ISSN 1818-5878 (Print) 2408-8498 (Online)

EFFECT OF PSYCHOMETRIC FEATURES IN CLASSIFICATION OF INFORMAL TEXT

* Md. Waliur Rahman Miah, Mst. Sumaya Khatun, Amina Shaikh, Md. Jakirul Islam
Md. Nasim Akhtar, and Md. Jamal Uddin
Department of Computer Science and Engineering,
Dhaka University of Engineering and Technology, Gazipur
Email: walimiah@duet.ac.bd

Abstract: The anti-social behaviours in online social media follow some documented psychological trends. Informal texts used to perpetrate the anti-social act contain information of the psychological trends. That information can be useful in the task of identifying an offensive text in the social media. In this regard, we used psychometric information as a feature-set in conventional classifiers for the classification task of informal texts used in online social-media. In this paper, we investigated whether this has any positive effect on the performance of those classifiers. The results of our experiments show some promising outcomes. It appears that the psychometric information enriched the data set, which improved the performance of some classifiers in the classification of online informal text.

Keywords: *anti-social behaviour; online child exploitation; informal text; chat processing; classifiers; data mining.*

1. INTRODUCTION

Research on anti-social behaviour through online social media is comparatively new and evolving. Social media such as online chat massaging, facebook, twiter and others like these have become common communication tools nowadays. They are cheap and convenient. Usually informal short-texts are used as the medium of correspondence in those platforms. Due to the easy access of internet and anonymity of the users, anti-social behaviours through those media are increasing day by day. By the term "online anti-social behaviour" we include cybercrimes such as cyber bullying, child exploitation, trolling and harassment etc. All these online anti-social behaviours follow some documented psychological profiles [1-7].

To distinguish between an offensive text (of an anti-social behaviour) from a non-offensive text, a text classifier can be a potential tool. Using mere terms as the feature-set of the classifiers may not be able to capture the psychological traits. Instead, a feature-set with psychometric information would be more appropriate in this particular situation. Under these circumstances, we pose the following research questions:

1) Can we use psychometric features in classification task of informal chat text?
2) Do psychometric features make any improvement in the classification task of informal chat-text?

In this paper, we will progressively answer the above-mentioned questions through experiments and analyses. Our contributions in this paper include using psychometric information as features in classification of informal chat texts as well as analysing the effect of it. As a sample of anti-social behaviour, we used the informal texts of child exploitation chats. An elaborated explanation of the data set is provided in the section titled "data preparation and pre-processing". The result and performances of the classifiers are compared in the result section. Our experiments and results suggest that the answers of the above-mentioned research questions are interestingly affirmative.

The organization of this paper is as follows. The current section introduces the objective and motivation of this research. Section 2 reviews the related works. Section 3 compares the special characteristic of informal chat text with formal text. The psychometric features of informal chat text are explained in section 4. Section 5 describes the experiment. Result and analysis are provided in section 6. Finally, this paper is concluded in section 7 with putting some lights on the future directions.

2. RELATED WORKS

Text Classification (TC) techniques are popular tools for processing formal text documents to analyse the content. Recently, it is also being widely used in the informal online texts processing tasks.

Bengel et al. in [8] used text classifiers to identify the topics of a chat room or of an individual participant. The concepts of the chat text message are categorised with vector space classifiers. No psychological impact was discussed in that research.

In [9], Kucukyilmaz et al. worked on attribution and characterization of authorship in chat messages. Different information from the chat messages are extracted using a number of supervised classification techniques. The author of the chat message is recognized by using term- and writing-

style-based approach. The authors did not use or require any psychological information which is used in our approach. Also the chat messages in [9] were in Turkish and not in English.

Bifet and Frank in [6] analysed sentiment in twitter messages by using text classifiers. The twitter messages are small informal texts. The authors used Kappa statistics to evaluate the predictive accuracy of classifiers. No psychometric indicators are used.

In [10], Pendar performed an analysis on child-grooming chat-logs to distinguish predators' posts from victims' posts. The author used a set of 701 chat-conversations obtained from the pjfi.org website [11], which is the same website from where we collected part of our data. The author manually separated the set of victims' posts from the set of predators' posts. As feature sets the author used word unigrams, bigrams and trigrams. Those were fed into support vector machine (SVM) and k nearest neighbours (kNN) classifiers. A number of experiments were conducted by using different number of features varying from 5,000 to 10,000. The best result was with the f-measure of 0.943. The author concluded that 10,000 features provided a satisfactory performance. The most effective classification achieved with trigrams in a kNN algorithm with k equal to 30. The experiments appear to have some similarities with our experiments in regards of detection of online paedophilia. However the focus and experimental setup is different. The author in [10] did not use any psychological markers hence is different than our current focus. The task of classification is also different. Pender manually split the text of a chat-log into separate posts of predators and victims, and then applied classifiers. We used the whole text including both predators' and victims' texts with the hypothesis that the victims response is also necessary to recognize the grooming activity as it contains victims psychological and emotional traits. This aspect was absent in Pender's experiments.

An International Sexual Predator Identification Competition was organized at PAN in CLEF 2012. One of the problems of that competition titled as "identify the predators among all users in the different conversations" is close to our current experiment. One of the participants Villatoro-Tello et al. [12] achieved the highest performance in solving that problem. The authors' system in [12] was based on lexical features and a two-step classification. The first step was pre-filtering by removing those conversations containing: (a) only one user, (b) less than 6 posts per-user and (c) long sequences of unrecognised characters. This significantly reduced the number of chat conversations, users, and sexual predators. The reduction ratio is 90% approximately for conversations and users, and, only 8% for predators. The second step was a classification task. The authors used bag of words representation employing either a boolean or a TFiDF weighting scheme with Neural Network (NN) and SVM classifiers. The system achieved a recall (R) of 0.7874, precision (P) of 0.9804 and F1 measure of 0.8734. The experimental setup was different than ours. The authors used PAN data which was originally collected from the same source of pjfi.org webpage [11] and then customized according to the need of the competition. Whereas we used the raw chats from that website, the PAN authority split the raw chats into different chunks. The first step of filtering in [12] though worked in that particular task of PAN, however there is serious uncertainty about its performance in a real-world social-media. Also the authors did not use any psychometric features. Still if we want to compare our results with theirs, we can see from the result section of this paper that, with the psychometric features our experiment achieved better results in recall (R) and F1 measure. Our results achieved a recall (R) of 0.940 which is 15.26% higher than the recall in [12], and an F1 measure of 0.920 with an improvement of 4% than the F1 in [12]. The precision (P = 0.90) of our experiment was slightly lower than the precision (0.9804) in [12].

Drouin and Boyd in [13] used LIWC [14] to analyse chat transcripts. The authors examined that the kinds of words used by the offenders and the chid-victims follow different trends, and the trend varies widely. They suggest that this kind of forensic linguistic analysis can be valuable evidence.

Very recently Seigfried-Spellar et. al proposed a chat analysis tool in [15] to differentiate between contact-driven vs fantasy driven online child sex offenders. Their tool obtained an accuracy of 87.1%, F1 measure 0.634. SVM was used as the classifier. Our task of classification is a bit different: offensive vs non- offensive. Our offensive class include both the classes of contact and fantasy of [15].

Focuses of the above-mentioned researches are different than the focus of our current research. The above-mentioned researches are focusing on topic detection, authorship characterization, analysing sentiment, distinguishing predators vs victims (not predators vs others) and fantasy vs contact driven offense. Those researches did not mention the use of psychometric information in solving the problems. Though [13] used the same LIWC tool but in a different way. They only analysed the word categories and we focused on psychological information. In this current research, we used psychometric information and analyse its effect in addressing the task of classifying informal online-text. We conducted classification experiments between anti-social offensive texts and non-offensive texts. As an instance of anti-social offensive texts a sample of child-exploiting chat-texts are used.

Before going to the experiment and result section, a brief analyses of online informal text are provided in the following two subsequent sections.

3. SPECIAL CHARACTERISTICS OF ONLINE TEXT COMPARED WITH FORMAL TEXT

The text in the online social media is different from a formal text due to some unique characteristics. This has been corroborated by researchers of text-processing field. For example: Kucukyilmaz et al. in [9] and Rosa and Ellen in [16] provided some good analysis of those special characteristics. Text in online social media is grammatically informal and inherently unstructured. For example, in an online chatting the users are actually trying to "talk" through the text typed in a quickly fashion, and therefore, are reluctant in following grammatical rules and sentence structures. Consequently, unwanted errors occur very frequently and are ignored for correction. Some other distinctive aspects include deliberate misspelling and chat-abbreviations. Some common examples of such abbreviations include: "gr8" instead of "great" and "asl" to ask "age, sex and location". Sometimes clever teenagers use special chat-codes, for example "P911", which stands for "Parent Alert!" or "Parent Emergency" [17]. The text-posts in other social media such as twitter and Facebook also contain similar textual anomalies. Such informal structures, uncommon abbreviations and erroneous texts are difficult to be handled by currently available formal text processing techniques.

To express emotional feelings (such as sadness, happiness and anger) emoticons are also frequently used in those informal kind of texts. "Emoticon" is a chat jargon which is conglomerated from "emotion" and "icon" to become "emoticon". Sometimes emoticons are also called smileys or emoji. Technically these are special string of characters to display emotional feelings through graphical representations. For example, happiness can be expressed by a happy-face imoji such as ''':-)". Emotions can also be expressed by emphasizing a word with repeating characters, for example, "soryyyyyyyyyy". This kind of deliberate misspelling is also frequent in chat text. The emoticons and intentional misspelled words may contain valuable contextual and psychological information in a social media text. For example, in a child-grooming chat the paedophile may reconstruct relation by an emphasized "soryyyyyyyyy" when the child felt threatening by any obtrusive language [2]. In other types of anti-social behaviours such as trolling, stalking or harassment emoticons and emphasized terms also express similar important information of psychological and emotional feelings. However, traditional text processing methods such as stemming and part of speech tagging will find it very difficult for processing such strange sequence of characters when those are preserved.

4. PSYCHOMETRIC FEATURES OF INFORMAL TEXT

Most of the online anti-social behaviours such as child-grooming, cyber-bullying, trolling and harassment have documented psychological trends. For example, in [2], a progressive psychological stages are identified by Rachel O'Connell in online child grooming. The perpetrator does not instantly start exploiting a child online. The adult may start with the child by making an innocent friendship and then gradually through a psychological progression advances towards the stage of exploitation. The model of luring communication theory, proposed by Olson et al. in [3], may be followed by a perpetrator. According to this model a perpetrator builds up a deceptive psychological trust. In [5], Kowalski et al. proposed a general aggression model as a useful theoretical framework for cyber-bullying. The model is based on the relationships between cyberbullying and other meaningful behavioural and psychological variables. The authors found that there are strong associations between cyberbullying perpetration and normative beliefs about aggression and moral disengagement.

To incorporate the psychological trends mentioned in those researches the terms in the text of an online anti-social act should be categorically and psychologically different than the terms used in general text. Therefore, we assume that analysing the psychological and categorical information of the textual-terms used in the offense would be helpful to learn the psychological patterns of the anti-social behaviours.

To find out the categorical and psychological properties of textual-terms LIWC (Linguistic Inquiry and Word Count) dictionary is used in this current research. According to Pennebaker et al, in [14], LIWC is a text analysing tool. It uses a psychometric dictionary to provide an efficient and effective method for studying the various emotional, cognitive, and structural components present in the terms of a text. The LIWC system counts the number of structural and psychologically significant words in the text. For example, it gives the count of the words that contain the information such as: social, family, friend, sexual, positive emotion, negative emotion, sad, anger, and anxiety etc.

5. EXPERIMENTS

This section is divided into two subsections. First subsection describes the data preparation and pre-processing and the second subsection explains experiment procedures.

A. Data preparation and Pre-processing

In the experiments the data set consists of informal text of several online chat-log files. The files include child exploiting (CE) offensive chat-text and non-offensive Non-CE chat-text. The reason to choose the child exploiting chat-text is that, it is a kind of online anti-social act and it adheres to the psychological traits documented by a number of psychological researchers such as in [2], [3], and [7].

The total number of chat-log files is 392. Among the 392 instances there are 200 CE and 192 non-CE chat-files. The non-CE portion of the data set consists of non offensive text-logs collected from different online open forums. The CE chats are collected from pjfi.org [11]. Each of the CE chat-scripts are chat-text between a convicted paedophile and a pseudo-child. By pseudo-child we mean a trained adult posing as a child for sting operations to catch online predators loitering in the internet. The length of the scripts varies from 83 lines to more than 12 thousand lines [7]. Many of the chat scripts were accepted in the American court of law for prosecuting the perpetrators for the charge of online child exploitation.

The chat-scripts went through pre-processing stages of cleansing and feature selection. Any erroneous sequence of unnecessary characters were deleted in the cleansing stage. Also the chat usernames are removed. However, the special sequence of strings that makeup the emoticons are kept with the help of regular expressions and a list of known emoticons. Then the text is transformed into string vectors. We conducted two sets of classification experiments using two types of features. In one set of experiments the term based features are used. The other set of experiments use psychometric categorical information. The psychometric features are produced from the text of chat-logs by using the LIWC dictionary.

B. Experimental setup

Three binary classifiers are used in our experiments from Weka data mining tool [18]. Those classifiers include Naïve Bayes (NB), Decision Tree (DT), and Classification via Regression (CvReg) classifiers. Using those classifiers experiments have been conducted to distinguish between offensive and non offensive chat-texts. In each case stratified 10-fold cross-validation is used. For each fold there were 39 chat-files in the test set and the remaining 353 chat-file in the training set. The test set for each fold is randomly selected keeping the class proportion the same as the class proportion of the entire data set. Each and every chat-log was in the test-set in one of the 10-folds. The evaluation results of a classifier are averaged over the 10-folds.

The results are presented with the metrics of accuracy, precision, recall and F1-measure. These are considered standard metrics in information retrieval. An analysis of the results is given in the following section.

6. RESULT AND ANALYSIS

Table I summarises the result-metrics of our experiments. The leftmost column of the table mentions the titles of the result-metrics. The left side of the table show the results with term based features and the right side shows that with the features of psychometric and word category. From the table it can be seen that, introducing psychometric features have different effect on different classifiers. While the performance of some classifiers are improved, some classifiers performed slightly inferior with the new features.

TABLE I: SUMMARY OF RESULT METRICS;
R = Recall, P = Precision, A = Accuracy, F = F1-Measure

	With Term Based Features			With Psychometric & Word Category Features		
	NB	**DT**	**CvReg**	**NB**	**DT**	**CvReg**
R	0.770	0.915	0.890	**0.940**	0.850	0.910
P	**0.939**	0.906	0.913	0.900	0.895	0.831
A	0.857	0.908	0.901	**0.913**	0.872	0.860
F	0.846	0.910	0.901	**0.920**	0.872	0.869

Table I shows that NB gained the most advantage of psychometric and word category features by improved performance. If we compare the row of F1-measure we can see that NB with psychometric features performed the best among the classifiers with an F1-measure of 92.0%. In the column of NB with psychometric features the accuracy is 91.3% and recall is 94.0% which is also considerably high. The only metric where NB performed slightly lower is the precision, however is still staying within 90% and comparable with the other classifiers.

Fig. 1 (next page) shows recall and accuracy through column charts. Those charts give us a closer view of recall and accuracy of the Table I.

In both the recall and accuracy charts we can see that the performance of Naïve Bayes (NB) classifier is greatly improved by using the psychometric features. In case of recall, NB achieves as high measure as 94% with the psychometric features while with term based features the measure is only 77%. That is, NB can catch 17% more offensive type texts when the psychometric features are used. In case of accuracy, the improvement of NB is smaller than the improvement in recall, however considerably high. With psychometric features NB achieves 91.3% accuracy while without those features the accuracy is only 85.7%. This means that, NB distinguishes between offensive and non-offensive texts 5.6% more accurately with psychometric features than without those features.

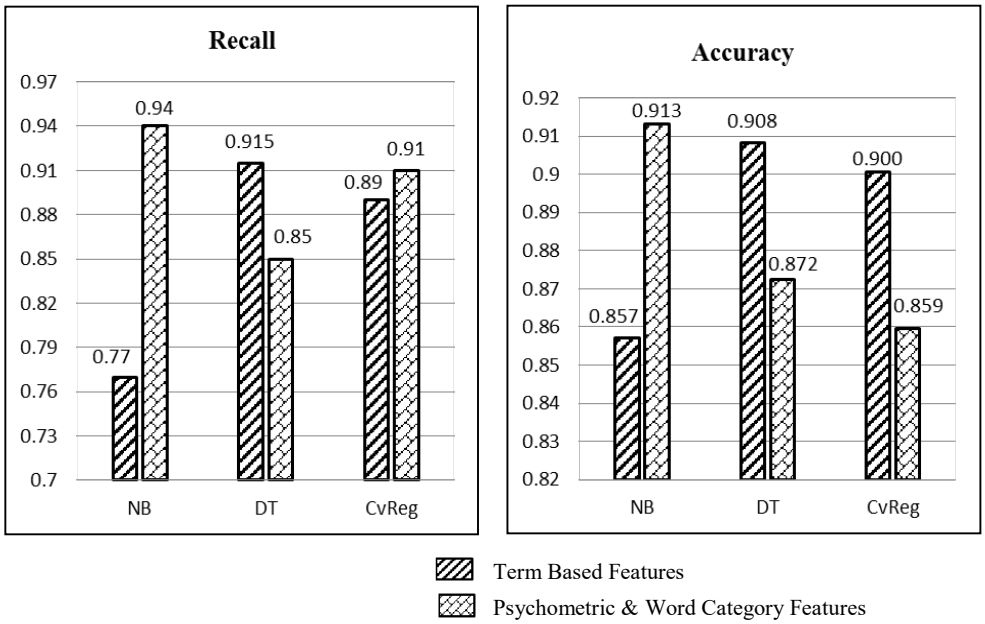

Fig. 1: Recall and Accuracy of different classifiers

For a Decision Tree (DT) classifier the psychometric features do not make any improvement, instead it imposes slightly negative effect.

Interestingly, a Classification via Regression (CvReg) classifier generates a mixed result with the psychometric features. The recall is improved 2%, that is, the recall of CvReg improves from 89% with term features to 91% with psychometric features. However, the precision is dramatically reduced from 90% to 85.9%.

We can understand from the bar-charts and the result metrics table that the NB classifier outweighs other classifiers in the improvement of performances with psychometric features.

7. CONCLUSION AND FUTURE DIRECTION

From the discussion of the previous section of result and analysis it can be concluded that the psychometric and categorical information is useful in some classifiers as a feature set for classification task of unstructured informal text. The new feature set significantly improves the performance of Naïve Bayes (NB) classifiers to detect offensive type texts. In some cases, it also improves the performance of Classification via Regression (CvReg) classifier. It appears that the psychometric and categorical information enrich the features of informal text dataset that improves the classification task.

However, it is interesting to see that while psychometric and categorical information is improving the performance of two classifiers (NB and CvReg), another classifier (DT) the same enriched dataset does not make any performance improvement. It can be a future scope to thoroughly examine the profile of informal text of chats and investigate the interesting behaviour of different classifiers.

In this current research the psychometric features are applied on a limited text data set of offensive child exploiting type and non-offensive general type. Future scopes of this research can be using the new features in a broader informal text data set containing online anti-social problems such as cyber-bullying, stalking and trolling. It will be interesting to see the effect of psychometric features in those data sets.

Another future work may include sentiment markers from SentiWordNet [19] along with the psychometric features for addressing the above-mentioned problems.

REFERENCES

[1] K. Young. "Profiling Online Sex Offenders, Cyber Predators, and Pedophiles." *Journal of Behavioral Profiling*, vol. 5 (1), pp. 1-18, ABP, 2005.

[2] R. O'Connell. "A Typology of Child Cybersexploitation and Online Grooming Practices." *Cyberspace Research Unit*, University of Central Lancashire. 2003.

[3] L. N. Olson, J. L. Daggs, B. L. Ellevold, and T. K. K. Rogers. "Entrapping the Innocent: Toward a Theory of Child Sexual Predators' Luring Communication."

Communication Theory, vol. 17 (3), pp. 231-251, Wiley-Blackwell, 2007.

[4] T. Krone. "Queensland Police Stings in Online Chat Rooms." *Trends & Issues in Crime and Criminal Justice Series*, Australian Institute of Criminology, 2005.

[5] R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner. "Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth." *Psychological Bulletin*, vol. 140(4), pp. 1073-1137, 2014.

[6] A. Bifet, and E. Frank. "Sentiment Knowledge Discovery in Twitter Streaming Data." in *Proc. of the 13th International Conference on Discovery Science (DS10), Canberra, Australia, 2010*. pp. 1-15.

[7] I. McGhee, J. Bayzick, A. Kontostathis, L. Edwards, A. McBride, and E. Jakubowski. "Learning to identify internet sexual predation." *International Journal of Electronic Commerce*, vol.15(3), pp.103–122, 2011.

[8] J. Bengel, S. Gauch, E. Mittur, and R. Vijayaraghavan. "ChatTrack: Chat Room Topic Detection using Classification." *Intelligence and Security Informatics, in the series of Lecture Notes in Computer Science*, vol. 3073, pp. 266-277, Springer. 2004.

[9] T. Kucukyilmaz, B. B. Cambazoglu, C. Aykanat, and F. Can. "Chat Mining: Predicting User and Message Attributes in Computer-Mediated Communication." *Information Processing & Management*, vol. 44 (4), pp. 1448-1466, Elsevier, 2008.

[10] N. Pendar. "Toward spotting the pedophile telling victim from predator in text chats." in *Proc. of the First IEEE International Conference on Semantic Computing,* pp. 235–241. Irvine, CA: IEEE PRESS, 2007.

[11] pjfi.org. *Perverted Justice Foundation Incorporated.* Available at http://pjfi.org/ (accessed October 2019).

[12] E. Villatoro-Tello, A. Juárez-González, H. J. Escalante, M. Montes-y-Gómez, and L. V. Pineda. "A Two-Step Approach for Effective Detection of Misbehaving Users in Chats." *CLEF2012 (Online Working Notes/Labs/PAN Workshop),* Rome, Italy, 2012.

[13] M. Drouin, R. L. Boyd, J. T. Hancock, A. James, "Linguistic analysis of chat transcripts from child predator undercover sex stings," *The Journal of Forensic Psychiatry & Psychology*, vol-28, pp. 437–457, 2017.

[14] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth. "The Development and Psychometric Properties of LIWC2007." Published in LIWC.Net. 2007.

[15] K. C. Seigfried-Spellar, M. K. Rogers, J. T. Rayz, S.-F. Yang, K. Misra,T. Ringenberg, "Chat analysis triage tool: Differentiating contact-driven vs. fantasy-driven child sex offenders," *Forensic science international*, vol. 297, pp. e8–e10, 2019.

[16] Rosa, K. D., and Ellen, J. 2009. "Text Classification Methodologies Applied to Micro-Text in Military Chat." in *Proc. of the Eighth IEEE International Conference on Machine Learning and Applications (ICMLA '09)*, Miami Beach, Florida, USA, 2009. pp.710-714.

[17] teenchatdecoder. Available at https://www.zoobuh.com /tools/chatdecoder/. (accessed October 2019).

[18] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. "The WEKA Data Mining Software: An Update." In ACM SIGKDD Explorations Newsletter, vol. 11 (1), p. 10-18, ACM, 2009.

[19] S. Baccianella, A. Esuli, and F. Sebastiani. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." *LREC-2010*, pp. 2200-2204. 2010.

