Machine Learning Solutions for controlling Cyberbullying and Cyberstalking

Zinnar Ghasem, Ingo Frommholz, Carsten Maple
University of Bedfordshire
Warwick University
United Kingdom
zinnar.ghasem@beds.ac.uk,ingo.frommholz@beds.ac.uk,carsten.maple@warwick.ac.uk

ABSTRACT: *In the last few years cybercrimes take different forms of shapes to propagate unwanted activities. Such email transactions are used by criminals as a means to initiate cybercrimes that inlcudes phishing, spamming, cyberbullying and cyberstalking. Cyberstalking is a relatively new surfacing cybercrime, which is now extensively used by criminals and anti-social elements. Combating email-based cyberstalking is a challenging task that includes the approaches such as -a robust method for filtering and detecting cyberstalking emails and documenting evidence for detecting such cybercrimes and to prevent such actions. We have initiated a research exercise which involves machine learning approach to find and control file evidence. Our mechanisms of research is based on a novel robust feature selection approach to select informative features, aiming to improve the performance.*

Keywords: Cyberstalking, Cyber Crime, Machine Learning, Cyberbullying, Email Crimes

Received: 1 March 2015, Revised 29 March 2015, Accepted 4 April 2015

1. Introduction

Electronic information transfer and file transfers are used effectively to transmit large quantities of data and it becomes a primary mode of professional communication. The benefits of it have been enjoyed and at the same time many cybercriminals are mis-using it for varied anti-social and criminal acts. The availability of pseudonymous and the anonymous nature of email communication, coupled with its protocol vulnerabilities, have contributed in proliferating the number of email crimes rapidly, where email is one of the most common methods utilised in cyberstalking [1]. There is no general agreement on the definition of cyberstalking but In [2] cyberstalking is described as a "*course of actions that involves more than one incident perpetrated through or utilising electronic means that cause distress, fear or alarm*".

Unlike some other email attacks, e.g. spam email sent to a huge number of people, cyberstalking email repeatedly and explicitly targets an individual to harass or threaten a victim, which could have a profound psychological, social and financial consequence on the victim's life. Thus it differentiates itself from most other cybercrimes for being highly personalised, repeatedly targeting individuals, taking many forms; inflame, hate, anger, revenge, etc; it encompasses social, psychological and financial impacts. Cyberstalking, being a unique cybercrime in its own right, requires unique technological solutions not only to filter and detect

cyberstalking emails, but more importantly for filing evidence, which could be used in a victim's initial complaint and help law enforcement in their earliest stage of investigation.

Given the seriousness of cyberstalking, receiving any communication from a cyberstalker is not welcomed, it causes alarm and fear to victims, therefore, unless technical solutions are provided and applied not only to mitigate cyberstalking but also help the victims to document evidence to be used to identify the culprit as a deterrence and prevention measure, the number of cyberstalking might increase and the cyberstalker be emboldened to continue and widen his/her scope of attacks, while victims may fear for their life, cowering in their home and not knowing what will happen next.

To tackle the problem of email-based cyberstalking, we explore and combine the application of machine learning, text mining, statistical analysis and email forensics to detect and mitigate email-based cyberstalking. To this end, this study proposes a hybrid framework to filter, detect and document evidence of one of the most used cyberstalking forms, which is email-based cyberstalking. As machine learning plays a crucial role in this framework, and selecting the right features is important in this respect, we also propose a robust new method for selecting informative features for our work, which has been tested with some well-established feature selection methods such as Chi Square (chi), Information Gain (IG), Odd Ratio (OR), Mutual Information (MI), Deviation from Poisson Distribution (PDM), Class Discriminating Measure (CDM) and Gini index (GI). We discuss related work in the next section. Sections III, IV and V provide an in-depth discussion of email-based cyberstalking. Our ACES framework is introduced in Section VI. Section VII discusses our feature selection approach as part of our framework. Section VIII shows the results of some evaluation, before we finally conclude in the last section.

2. Early Work

Many types of cybercrime have been well researched and documented, but there exists a group of cybercrimes which has yet been somewhat underestimated by both the public and government services; these are online harassment and cyberstalking [3]. Most cyberstalking research has focused on the psychological and social aspect of cyberstalking rather than tackling it through technological solutions. Most research has been descriptive and surveying, such as defining cyberstalking and looking at prevalence aspects [4].

A monitoring system framework is proposed in [5] to record data on a cyberstalking victim's computer. The idea was then developed further, reported in [6], into a prototype called Predator and Prey Alert (PAPA) system as a forensic tool to help law enforcement. However, the PAPA system has some major limitations; the system records every screen and keystrokes of a victim's computer within a session, which raises a privacy concern; it requires a special set-up including software and hardware; furthermore, it neither filters nor detects cyberstalking emails.

Apart from spam filtering, social networking harassment and cyberbullying are two other closely related areas, where textual patterns have been used to detect and filter unwanted messages. One of the early approaches was proposed by [7] where the authors try to detect abusive message based on selecting 47 syntax and semantic features, achieving 64% detection rate. Another approach was discussed in [8] to classify cyberbullying based on binary and multi-class text classification, and reportedly binary class classification outperforms multi-class classifiers. However, using and including attackers' characteristics, and their harassing behaviour can improve the performance of cyberbullying detection [9], while in [10] it was shown that performance improves when combining content, sentiment and contextual features to detect harassment on the Web 2.0. A semi-supervised algorithm was proposed in [11] utilising lexical association of profane language to detect offensive tweets.

3. Finding and Recording Email-based Cyberstalking

The key inspirations for our work are twofold; first it is inspired by the work reported in [6]. Their proposed prototype basically focused on capturing data within the duration of a session. Their system has some major limitations. Our proposed approach is clearly different from their system. Unlike their system, our system is able to detect and filter cyberstalking emails, it attempts to address the problem of anonymous email, tries to locate the source of attack and collects and analyses evidence. The second inspiration is the necessity for a system that not only filters email, but also supports manual email evidence collection by victims by automatically collecting evidence. This is done in order to persuade authorities to investigate or prosecute cyberstalkers, where the responsibility is often on the victim to produce such evidence [1]. Additionally, "police do not take a report which can be problematic in terms of victim documentation and they advise victims to document their experiences" [12]. It is therefore vital

that all copies of communication from a cyberstalker, whether email or other communications, is saved and given to law enforcement [4]. Providing evidence and identifying the author of email will not only persuade law enforcement to start an investigation, but it will also help the investigator and law enforcement to prosecute the attacker [13]. Thus, it is imperative that the victim keeps the message and that all headers are readable and exposed for investigation and tracking [3].

Therefore an automated system will not only make the initial complaint process and investigation easier, but will also speedup investigations and make them more efficient. Furthermore it will encourage victims to come forward and complain to prosecute cyberstalkers, and it will ensure that victims receive the required service and support [14].

4. Formulating the Cyberstalking Email Issues

The cyberstalking email problem can be represented by a very well known example of Alice and Bob communicating.The cyberstalker and victim here are depicted by Bob and Alice respectively. *Bob* is cyberstalking Alice through email, and has acquired Alice's email (alice@ alicemailprovider.com). *Bob* composes a cyberstalking email message (From: bob@bobemailprovider.com) on a device,then sends it to Alice. However, Bob might use different technologies and methods including but not limited to remailers, proxy servers, computers in public libraries and internet cafes, pseudonymous accounts, etc to anonymise emails, particularly to penetrate through any prevention mechanism Alice might have employed, as well as to protect himself being identified by law enforcement. As the victim, Alice does not only want to filter and detect emails from Bob, but also wishes to discover whether received anonymous emails are sent by a cyberstalker (in this case *Bob*) for the purpose of prosecution in the court of law. However, *Bob*'s email, both the header and body hold valuable information to detect, identify and establish the link between unwanted received email and its sender, but as mentioned earlier, the problem of identifying a cyberstalker is exacerbated by anonymous technologies and techniques, thus a system needs to be intelligent enough to detect and establish a link between received cyberstalking email and the respective cyberstalker.

As aforementioned, cyberstalking is very personal, thus, Alice does not wish to receive any email whether containing any unwanted words or not. This can be formally represented as

$$\forall e\,(C\,(e, b) \rightarrow P\,(e, cs))$$

Where e is an email, $C\,(e,b)$ is an email from *Bob*, $P\,(e, cs)$ means the email is categorised as a cyberstalking one (cs). This rule says: if any email is sent by Bob to Alice, then classify it as cyberstalking. Therefore, the monitoring system needs to be able to detect, filter and document evidence of any email from *Bob*

$$\forall e\,(C\,(e, b)) \rightarrow (d\,(e) \wedge d_s\,(e))$$

Where $d\,(e)$ means the system needs to detect and $d_s\,(e)$ to document evidence, that is any email send by *Bob*, then the system should be able to detect it and file the evidence. Generalising the problem: suppose x is any person who sends an email to victims, $C\,(x)$ means x is a cyberstalker, $S\,(x)$ means the system needs to detect email from $C\,(x)$ and $S_f(x)$:
the system files evidence from $C\,(x)$; $P\,(x)$ is a person who sends an email to victim. Then

$$\forall x\,(P(x) \wedge C(x)) \rightarrow (S\,(x) \wedge S_f(x)).$$

This can be assumed as a deterrence measure and response defence, in which the response defence is concerned with identifying and penalizing the perpetrator and learning lessons to enable the organization and individual to defend themselves more effectively [15].

5. Evidences

Our proposed system is assumed to operate on a victim's device. Evidence is stored there, but could equally be saved for instance on a private cloud storage, where law enforcement can have a regular access to the data and monitor the cyberstalking, but this needs to be discussed and agreed between the victim and law enforcement in terms of security, privacy and availability. However, the idea of saving cyberstalking emails on a cloud storage could be an important step to create a national cyberstalking

database, which could be vital to solve some other cyberstalking cases, as well as beneficial to the research community as a whole. However, it is understood that the victim must consent to it and privacy must be protected. Nevertheless, reliable evidence collection is a difficult process and maximum care needs to be taken. Digital evidence is susceptible and easily deleted, altered or damaged, therefore establishing the integrity and authenticity of material in court requires standard techniques and methods for the collection, preservation and presentation of the stored material [16].

The National Security Systems Instruction [17] glossary defines the Information Assurance model as "measures that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation. These measures include providing for restoration of information systems by incorporating protection, detection, and reaction capabilities." While confidentiality and availability are not directly concerned with digital evidence, integrity is essential [18], so it must be ensured that raw data cannot be modified. It is imperative that all communications from cyberstalkers are saved without any modification for the purpose of judicial evidence. Discovering a pattern of harassment is a vital step in the investigation process and to the success of prosecution [19]. Email evidence includes the email body, subject, header information, geolocation information, analysed data results, etc. All this must be carefully saved and be verifiable, thus the integrity of data whether during saving on the storage or in transmission should be preserved and must not be altered. For this purpose the HMAC [20] mechanism would be employed with hash function; SHA-1 or any other later version of the SHA family. However, saving email in a private cloud requires both hashing and encryption. Asymmetric key encryption could be used where the public key resides on the system and the private key is held by law enforcement.

5.1 Email's End-to-End Path

It is vital that we understand the format and working structure of email as both hold valuable information and clues for the detection and identification of cyberstalkers and cyberstalking emails. An email message comprises a header and the body. The header encompasses a number of lines which are called header fields, and each header field consists of two parts; field name and its body holding specific information. RFC2822 [21] defines a number of header fields, which are either added to the email header by the email sender or by nodes along the travel path of the email; these include From, To, Carbon Copy (Cc), Blind Carbon Copy (Bcc), Subject, Date, Message-ID, Reply-To, Received, Return-To, etc.. The email body is the (optional) second part of an email message and is usually formed as text, but could also include multimedia elements in Hyper Text Markup Language (HTML) and attachments encoded as Multi-Purpose Internet Mail Extensions (MIME) [21]. The Mail User Agent (MUA) is an application which operates on the client device and enables users to create, read and manage emails on the server; this includes commercial applications such as Microsoft Outlook, open source ones like Thunderbird and Web mail applications such as Gmail and Yahoo,etc.

A simple email path between sender and recipient includes the MUA, a forwarding Mail Transfer Agent (MTA), zero or more intermediate MTA relays and the receiving MTA (rMTA) then to local receiving Mail Delivery Agent (rMDA), where MUA uses Internet Mail Access Protocol (IMAP) or Post office version 3 (POP3) to download email from the server. Each node along the path consists of required software and hardware to enhance interoperability between all nodes along the path.he Simple Mail Transfer Protocol (SMTP) is mainly used as a transport protocol. Along the path, the forwarding MTA constructs and adds a Message-ID, timestamp and 'received' fields to the header of the email. Then it sends the email to another intermediate MTA relay or to the rMTA. The rMTA takes care of the message and transfers it to the rMDA. The message is then added to a message store or Inbox. Each MTA adds email tracing information by writing a 'received' field to the header. A 'return path' is added to the header as well [22].

Header and path data provides the system with some potentially useful information when it comes to tracing cyberstalkers. However, as a cyberstalker can have several anonymous email accounts, email header information alone is not sufficient to identify a cyberstalker. It is crucial to gather other evidence about a cyberstalker, and text categorization approaches, as they are applied for spam detection, look promising in this respect. We will therefore introduce next our framework and discuss the role of text categorization. Since choosing the right features is crucial in categorization, in particular in nonstandard tasks like the one we have at hand, we will discuss a feature selection approach in the context of the framework.

6. The ACES Framework

Our proposed framework is called Anti-Cyberstalking Email System (ACES). To the best of our knowledge it is the first system that specialises on the automatic detection and evidence documentation of email-based cyberstalking. A prototypicalimplementation of the framework is under development, and the data collection process is ongoing. ACES will run

on a user's computer and will act as an additional layer to detect unwanted emails. The ACES architecture is depicted in Figure 1. The system combines text categorisation, statistical analysis, and supervised machine learning to detect, filter and file evidence of cyberstalking email. The system comprises four main components: filteringLists, detection module, cyberstalker identifier module, evidence module and result analyser with a cyberstalker's digitometrics database. Digitometrics include writeprints, behaviours, unique vocabularies, and related email header information of an email. As aforementioned, cyberstalking is highly personalised, thus focusing only on offensive language to detect and filter email is certainly inadequate to combat email-based cyberstalking. For this reason, the system combines both detection module and cybestalker identifier to filter and identify the cyberstalking emails. Similar to spam white/black lists, the system filters emails based on a cyberstalking list and a whitelist. Emails whose addresses are in neither lists are analysed by the detection module and cyberstalker identifier.

The aim of the detection module is to detect potential cyberstalking emails based on the contents and the header of email. The email is preprocessed and the corresponding supervised neural network is utilised to detect and filter email, based on three outputs not cyberstalking (00), cyberstalking (10) and grey email (01). The final decision or result of neural network output is represented by β. For developing the detection module component and in order to be effective in real time application, a supervised algorithm needs to be trained.The dataset is preprocessed including tokenization, removal of stop words, stemming and feature selection. Each email is presented as a vector in the feature space, reflecting the respective feature's weight. The performance of the supervised algorithm is bound by the selection of informative features; for this purpose we have developed a new and robust feature selection approach called Informative feature selector (IFS) as described in Section VII.

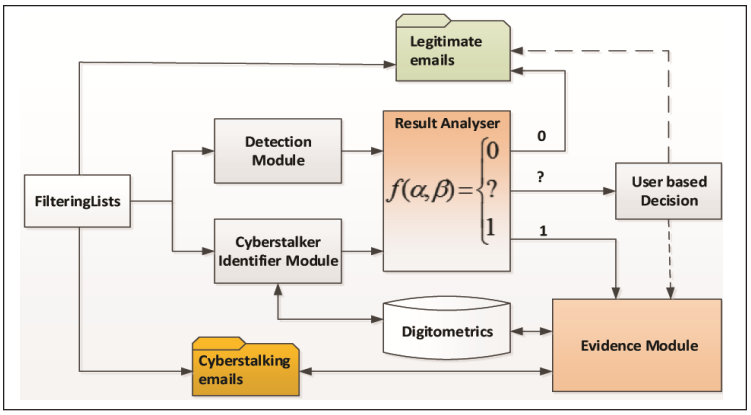

Figure 1. ACES Framework

The cyberstalker identifier module employs a cyberstalker's revious digitometrics history to identify cyberstalking emails including anonymous/pseudonymous emails and cyberstalking emails which are free from any abusive word(s). This is based on the premise that, since by definition a cyberstalking attack must consists of more than one email, there exists a known set of cyberstalking emails $C_E = \{e_1,..., e_2\}$eng where $n \geq 2$. That means there are at least two emails which could be used to identify either afore mentioned types of emails. Obviously the set of emails will increase as the attack continues. Unlike for the detection module component, a supervised algorithm is inapplicable in this module, therefore principal component analysis (PCA) will be utilised in future work.

The result analyser determines whether an email is legitimate (0), grey (?) or cyberstalking (1) based on predefined α and β results from both the supervised algorithm and PCA in the detection module and cyberstalker identifier, respectively. Emails are identified based on $P(\beta, \alpha)$ as follows

$$P(\beta,\alpha)=\begin{cases} 0 \ if \ (\beta=00 \ \wedge \ \alpha \geq r_2), \\ 1 \ if \ (\beta=10 \ \text{v} \ \alpha \leq r_1), \\ ? \ if \ (\beta=01 \ \wedge (r_1 < \alpha < r_2)). \end{cases}$$

r_1 and r_2 are pre-defined threshold values in the cyberstalker identifier (which have to be determined experimentally), and β holds the final neural network's result on the nature of classified email.

The last component is the evidence module. It basically gathers evidence; extracting domain name, email source IP address or next

server relay's IP in the path, in case the email IP is not available from the newly arriving cyberstalking email's header, and automatically submitted to WHOIS and other IP geolocation website, then returned result, with its date, time and email header fields are saved in the evidence database on the victim's computer. It also utilises statistical methods like PCA and multivariate Gaussian distribution to examine digitometrics of cyberstalkers. Furthermore it extracts similar features, attacker's behaviour, greetings, farewell, etc. to analyse and identify the similarity between anonymous and non-anonymous cyberstalking emails or any other selected emails. Another functionality of this part is regularly updating the database.

7. Informative Feature Selection (IFS)

The selection of suitable features for the task at hand is crucial for the underlying text categorization algorithms used in the detection module. Therefore we have developed a new robust feature selection method called IFS, to select the most discriminative features (for instance words). The principal aim of a feature selection method is to differentiate between infor-mative and uninformative features by giving highest values to the most informative and the lowest value to the least informative features, which subsequently facilitates the process of reducing the size of feature vector-space. However, the following issues must be taken into account by a feature selection method in the process of weighting the features for text classification. A feature that appears in all documents of a class and does not appear in any documents of other class(es) is a good discriminator and shouldbe given the highest value. Consequently, a feature which equally appears in all classes should be given a lower value.The value assigned to a feature should reflect its degree of discrimination and its correlation between the classes.Take a binary classification, for example, if feature t_1 appears in six documents in class one, while it only appears in three documents in class two, but feature t_2 appears in three documents in class one and it does not appear at all in any documents of class two, then, obviously, feature t_2 should be assigned a higher weight than t_1.

Based on these considerations, the IFS method is formulated as follows

$$IFS(t_i,c_j) = log\left(\left| \left(P(t_i \mid c_j)\, P(t_i \mid \bar{c}_j) - P(t_i \mid \bar{c}_j)\, P(\bar{t}_i \mid c_j) \right) \right| \frac{\left| P(t_i \mid c_j) - P(t_i \mid \bar{c}_j) \right|}{min\left(P(t_i \mid c_j), P(t_i \mid \bar{c}_j)\right) + 1} + 1 \right) \quad (1)$$

where $P(t_i|c_j)$ is the probability of t_i appearing in class c_j and $P(\bar{t}_i|c_j)$ is the probability of ti not appearing in class c_j. Similarly $P(t_i|\bar{c}_j)$ is the probability of t_i appearing in another class $\bar{c}_j$ and $P(t_i|\bar{c}_j)$ is probability of ti not occurring in $\bar{c}_j$.

IFS has been formulated in a way in which the overall value given to a feature is sensitive to changes in the number of features which are absent, shared or present in classes. In order for IFS to assign a value which reflects the useful- ness of a feature for classification, and adhere to the above considerations, IFS makes use of the difference between the probability of a feature that appears in both classes, then divides it by the smallest value between $P(t_i|c_j)$ and $P(t_i|\bar{c}_j)$; 1 is added in case the smallest value is zero (this is the 2nd part of the equation). This ensures the feature is assigned an appropriate value not only according to the differences between $P(t_i|c_j)$ and $P(t_i|\bar{c}_j)$ but also according to the probability of ti occurring in intersection between c_j and $\bar{c}_j$. However, the calculation so far does not consider the number of documents which do not contain features in all classes; therefore, both the absence and presence of features in classes $P(t_i|c_j)$, $P(\bar{t}_i|c_j)$, $P(t_i|\bar{c}_j)$ and $P(\bar{t}_i|\bar{c}_j)$ are used in the calculation to reflect the probability of a feature being absent or present in classes. The whole formula makes sure the feature is assigned a value which reflects its usefulness for classification. This is not always the case in some other approaches like OR, they do not take the intersection or number of shared features between classes into account. For example, if $t_1$appears in two documents in $class_1$ and does not appear in $class_2$, while t_2 appears in all four documents in $class_1$ and only in two documents in $class_2$, then both t_1 and t_2 features are assigned the same values in the existing approaches, while our method gives a higher value to t_1. The maximum and minimum values are between zero and one, and a feature which equally appears in both classes is assigned a minimum value, which is zero in case of balanced binary classes. The method adheres to all above mentioned considerations.

8. Evaluation

The aim of our experiments is to evaluate and compare the effectiveness of the proposed IFS method to state-of-the- art feature selection approaches discussed in the introduction. We perform our experiments on datasets that are relevant for our cyberstalking scenario; the chosen scenario, however, was spam detection as we could utilise existing datasets for this. Neural Networks (NN)

and Support Vector Machines (SVM) have been utilised to carry out a 10-fold cross-validation of spam email and SMS classifiction to examine the performance of our method. In this experiment we followed [23] and used binary classification, as the binary classification outcome better reveals the efficiency and usefulness of the used approach than multi-class classification [23]. Moreover, resolving the binary text classification problem also means resolving multi- classes [24].

8.1 Dataset and Performance Measure

Two data sets have been used in our experiments: the enron1 email collection [25], [26], in total 5172 emails, in which 1500 are spam email and 3672 are legitimate emails, and an SMS collection, which was introduced in [27], [28], and consists of 5574 SMS messages, where 4827 are legitimate SMS and 747 spam SMS. A balanced email dataset was created with the 1500 spam emails and randomly chosen 1500 legitimate emails. Similarly for the SMS dataset, the balanced dataset was created with 747 spam SMS and 747 randomly chosen ones from 4827 legitimate SMS. The top n subset of features with highest weight were used in experiment, and n was set to 10, 15, 25, 50, 100, 200, 300, 400 and 500.

The performance of all methods was assessed using the F-measure (F_1) based on precision and recall as defined in [24]using micro-averaging.

8.2 Data Experimentation

8.2.1 Effectiveness of IFS: The F-measures of the 10-fold cross validation of the email and SMS datasets based on SVM and NN classifiers are shown in Figures 2, 3, 4, and 5 respectively. Based on the results, IFS is performing well with both classifiers. Across all used feature sets, IFS is either superior or comparable to others, while in some instances it is shown as second best. While OR has been regarded by some authors as one of best methods, it is inferior to IFS in almost all cases. However, the performance of OR and CDM slowly increases with the number of features, but still they are inferior compared to all others methods in most cases, except mutual information.

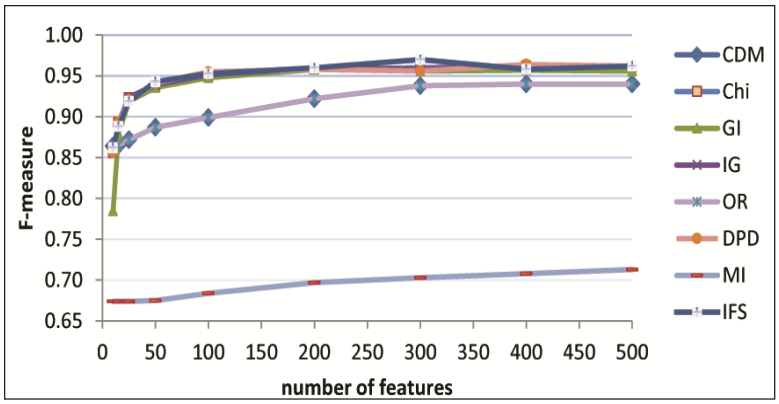

Figure 2. F-measures of SVM with enron1 emails

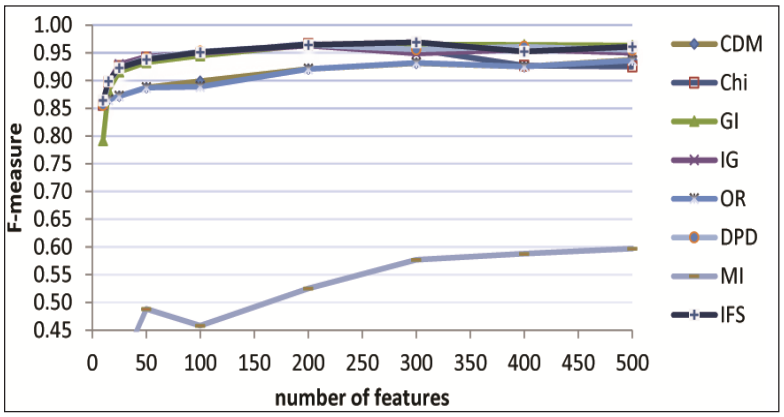

Figure 3. F-measures of NN with balanced enron1 emails

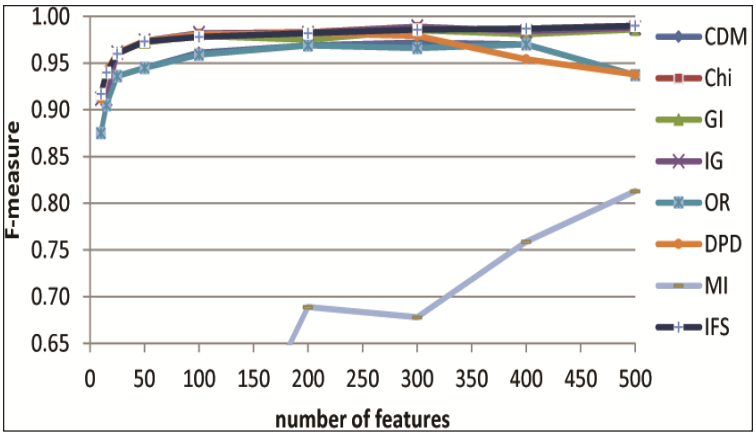

Figure 4. F-measures of NN with balanced SMS collection

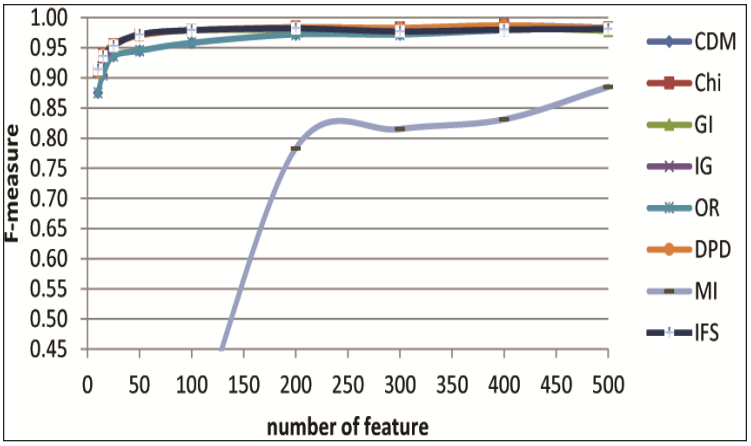

Figure 5. F-measures of SVM with balanced of SMS collection

8.2.2 Processing Time Analysis

While the processing time for selecting features from a small dataset might not be considered very important, it is extremely vital in the case of a big dataset [29]. The processing time of each method including IFS is recorded based on preparing and fetching the top 10 informative features. The processing time is taken on a computer with specification of 4 GB RAM and a 3.00 GHz CPU using Windows 7. Based on the results shown in Figure 6, IFS has less computation time compared to other methods.

	IG	CDM	Chi	GI	OR	DPD	MI	IFS
Processing time(m)	0.042	0.043	0.044	0.039	0.042	0.040	0.041	0.037

Table 1. Computational Processing Time

9. Conclusion And Future Work

Cyberstalking is an emerging type of cybercrime, detecting and mitigating cyberstalking requires new devised methods, techniques and features.

Combating cyberstalking is a challenging task, where technical solutions are a cornerstone in its prevention and mitigation. We therefore presented the ACES framework that filters, detects and documents email-based cyberstalking. In the context of our framework we in particular discussed the potential use of textual machine learning approaches and discussed the difference to classical email categorization. The aim of our solution is not only to mitigate cyberstalking, but also to help victims in the documenting of evidence, which is required for law enforcement. Future work includes the implementation and evaluation of ACES, in particular by measuring the effectiveness of the proposed text categorization methods in this emerging problem area.

References

[1] Roberts. L. (2008). Jurisdictional and definitional concerns with computer- mediated interpersonal crimes : An Analysis on Cyber Stalking, *International Journal of Cyber Criminology*, 2 (1) 271– 285.

[2] Maple.C., Short. E., Brwon. A., Bryden.C., Salter. M. (2012). Cyberstalking in the UK: Analysis and Recommendations, International *Journal of Distributed Systems and Technologies*, 3 (4) 34–51, .

[3] Mccall. R. (2003). Online Harassment and Cyberstalking : Victim Access to Crisis, Referral and Support Services in Canada Concepts and Recommendations, Canada: V*ictim Assistance Online Resources*, 17.

[4]Finn.J. (2004). A survey of online harassment at a university campus. *Journal of interpersonal violence*, 9, 468–483.

[5]Burmester., Henry. P., Kermes. L. S. (2005). Tracking cyberstalkers : a cryptographic approach, *ACM SIGCAS Computers and Society*, 35 (3) 2.

[6]Aggarwal, S., Burmester, M., Henry, P., Kermes, L., Mulholland, J. (2005). Anti-Cyberstalking: The Predator and Prey Alert (PAPA) System, (2005) in *Systematic Approaches to Digital Forensic Engineerin*g, First International Workshop, no. iv. IEEE-CPS, 195-205.

[7] Spertus, E. (1997). Smokey: Automatic Recognition of Hostile Messages, *In*: Proceedings of the Innovative Applications of *Artificial Intelligence*, 1058–1065.

[8] Dinakar, K. (2011). Modeling the Detection of Textual Cyberbullying, in *The Social Mobile Web*, 11–17.

[9] Dadvar, M., Ordelman, R., Jong, F. D. (2012). Trieschnigg. D. Towards User Modelling in the Combat against Cyberbullying, in Natural Language Processing and Information Systems. *Springer-Verlag Berlin Heidelberg*, 277–283.

[10] Yin, D., Xue, Z., Hong, L., Davison, B. D., Kontostathis, A., Ed- wards, L. (2009). Detection of Harassment on Web 2.0, *In*: Proceedings of the Content Analysis in the WEB 2.0 (CAW2.0) Workshop at WWW2009.

[11]Xiang.G., Fan.B.,Wang, L., Hong, J., Rose, C. (2012). Detecting offensive tweets via topical feature discovery over a large scale twitter corpus, *In*: Proceedings of the 21st ACM International Conference on Information and knowledge management - CIKM '12. New York, New York, USA: ACM Press, 1980–1984.

[12]0Logan. T. K., (2010). Research on Partner Stalking : Putting the Pieces Together, Lexington, KY: Department of Behavioral Science and Center on Drug and Alcohol Research, University of Kentucky, p.1–27.

[13] Pinals, D. A., Stalking. (2007). Psychiatric Prespective and Practical Approach, D. A.Pinals, Ed.

[14] Reyns, B. W., Englebrecht.C. M. (2010). The stalking victim's decision to contact the police: A test of Gottfredson and Gottfredson's theory of criminal justice decision making, *Journal of Criminal Justice*, 38 (5) 998–1005, Sep.

[15] Goodman.S. E., Kirk.J.C., Kirk.M. H. (2007). Cyberspace as a medium for terrorists, *Technological Forecasting and Social Change*, 74, p. 193–210.

[16]Karyda, M., Mitrou, L. (2007). Internet Forensics: Legal and Technical Issues, in Second International Workshop on Digital Forensics and Incident Analysis, no. *Computer Society*. IEEE, August, 3 – 12.

[17] Assurance, I., National Information Assurance. (IA) glos- sary, Tech. Rep. 4009, 2010. [Online]. Available: http://www.cnss.gov/Assets/pdf/cnssi 4009.pdf

[18] Burmester, M., Henry, P., Kermes, L. S. (2004). Tracking cyberstalkers : a cryptographic approach, *ACM SIGCAS Computers and Society*, 35 [3], 2.

[19] Pittaro, M. L. (2007). Cyber stalking : An Analysis of Online Harassment and Intimidation, 1 (2) 180–197.

[20]Krawczyk, H., Bellare, M., Canetti, R. R. (1997). HMAC: Keyed-Hashing for Message Authentication Status, IETF, Tech. Rep, [Online]. Available: https://www.ietf.org/rfc/rfc2104.txt

[21] Resnick.P. W. Internet Message Format, ietf, Tech. Rep., 2008. [Online]. Available: http://tools.ietf.org/html/rfc2822

[22] Banday, M. T., Mir, F. A., Qadri, J. A., Shah, N. A. (2010). Analyzing Internet e-mail date-spoofing, *Digital Investigation*, 7 (3-4) 145–153.

[23] Ogura, H., Amano, H., Kondo, M. (2009). Feature selection with a measure of deviations from Poisson in text categorization, *Expert Systems with Applications*, 36 [3] 6826–6832.

[24] Sebastiani, F. (2002). Machine Learning in Automated Text Categorization, *ACM Computing Surveys* (CSUR), 34 (1) 1–47.

[25] Metsis,V., Androutsopoulos, I., Paliouras, G., (2006). Spam filtering with naive bayes-which naive bayes?, *In*: *CEAS -Third Conference on Email and Anti-Spam*, 27 - 28.

[26] Uysal, A. K., Gunal, S. (2012). A novel probabilistic feature selection method for text classification, *Knowledge-Based Systems*, 36, 226-235.

[27] Cormack, G. V., Hidalgo, J. M. G., Sanz, E. P. (2007). Feature engineering for mobile (SMS) spam filtering, *In*: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and development in Information Retrieval. *ACM*, 871-872.

[28] Almeida, T.A., Hidalgo, J. M. G., Yamakami.A. (2011). Contributions to the study of SMS spam ?ltering: new collection and results, *In*: Proceedings of the 11th *ACM Symposium on Document Engineering. ACM*, 259-262.

[29]Jain, A., Zongker, D. (1997). Feature selection: Evaluation, application, and small sample performance, Pattern Analysis and Machine Intelligence, *IEEE Transactions*, 19 (2) 153–158.