THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Detection of Malicious Accounts on Social Media and Other G-Mail Accounts

Nisha Baste
Student, Department of Computer, K.K. Wagh Institute of Engineering Education and Research Center, Nasik, Maharashtra, India
Tanvi Chaudhari
Student, Department of Computer, K.K. Wagh Institute of Engineering Education and Research Center, Nasik, Maharashtra, India
Rohini Kothavade
Student, Department of Computer, K.K. Wagh Institute of Engineering Education and Research Center, Nasik, Maharashtra, India
Shital Jadhav
Student, Department of Computer, K.K. Wagh Institute of Engineering Education and Research Center, Nasik, Maharashtra, India

Abstract:
Internet is most attractive source for sharing information but security is major issue as far as internet is concern. A majority of the individual users primarily utilize social media and email accounts to stay in touch with their friends and to get information from people or organizations. Most of the people face the problem of getting malicious information such as URLs, emails, posts through adult account on social media. Now a day's Facebook provide facility of spam mail but in case, if user click on it, ultimately users account information sends to the attacker. In existing system an iterative social based classifier (ISC) is used to detect malicious contents on the social media. But the Gmail users also face the problem of getting URLs, posts, mails from adult account. So, in the proposed approach to address this problem ISC is extend to detect and block the adult account on Gmail and to enhance privacy and security on social media and e-mail accounts.

Keywords: *Text analysis, Social media, Security, Facebook, G-mail, Social profiles, Natural Language Processing, Adult content, Graph based classification.*

1. Introduction

The World Wide Web has become the most essential criterion for information communication and knowledge dissemination. Social Networking Sites plays very important role in human life nowadays, it is becoming a main communication media among individuals and organizations. It helps to transact information timely, rapidly & easily. Facebook and G-mail has become an increasingly influential platform for real-time information sharing. With over 200 million monthly active users and half a billion posts sent per day. The Web serves as better medium for large number of malicious activities such as sending adult contents through messages, URL's and images. This type of contents leads to various attacks such as DDoS attacks, Phishing attacks, Spam attacks.

Figure 1: Example of adult content on Facebook

These attacks attract the common users to click links attached in legitimate looking or spam emails and make them to visit the malicious sites. It initiates them to click, urges them to give their personal information. The malicious URLs in Gmail and Facebook leads to the actual Phishing sites which are clones of legitimate websites and force the users into entering sensitive information. The malicious user poses various critical conditions such as account suspension, failed transaction and forcing user to upgrade the newly installed security feature.

However, Facebook at the same time has become an attractive platform for the adult entertainment. A large number of accounts have been created on Facebook for the purposes of promoting services related to adult entertainment, propagating sexually explicit materials, and even recruiting performers for the adult entertainment industry. According to our observation on 5000 accounts, adult accounts are mostly connected with normal accounts. Figure 1 shows example of adult content on Facebook. In addition, many adult accounts post only a few entities related to adult content and much more entities that are not related to adult content. It is difficult for existing graph based classification techniques to identify these adult accounts. In order to identify the adult accounts on Facebook, G-mail one simple solution is to use existing adult content detection techniques including URL blacklisting, text based and image based adult content detection methods.

1.1. Literature Survey

Existing adult content detection techniques that are designed for web pages, however, are ill-suited for adult account detection on Twitter. Over the past decade, a number of techniques have been introduced for detecting adult web content. These techniques leverage different types of information such as text, image, and URL.

1] Hepple et al. proposed to combine statistical text categorization techniques and natural language processing techniques to filter web pages with vulgar languages However, a recent study shows that traditional statistical text categorization techniques do not work well for the short text of micro blog services [i].

2] Furthermore, in order to obtain an effective text based classifier, expensive costs in labelling a large training set are needed. Particularly, given that Twitter is a multi-linguistic OSN which supports multiple languages, we may need to label a large training set for each language that Twitter supports. Lastly, an effective text based classifier fails when accounts just post adult URLs, images or videos. For content-based image analysis, skin colour based detection and the bag of visual word model are proposed for detecting pornographic images [ii], [iii], [iv], [v].

3] However, recent studies [vi], [vii], [viii] indicate that these techniques are sensitive to background noise, illumination variance and image quality, and thus do not perform well for images captured by low-quality web cameras equipped on mobile devices. In addition, the high computation cost of image based methods may restrict their practical use for online social networks where a huge volume of images and videos are posted every day.

4] URL blacklisting is one of the widely used techniques in commercial content filtering software like K9. Hammami et al. also proposed a solution that combines text, image and URL for adult content detection on the Internet. However, URL blacklisting faces more challenges for adult content detection on social networks as an increasing amount of adult content moves to blog sites (e.g., tumbler.com and blogspot.com) and cloud based image hosting services (e.g., instagram.com). It is almost impossible for URL blacklisting to respond quickly to the fast-growing adult content in these blog sites and cloud based image sharing services [viii].

5] Skillicorn and David [ix], has used matrix decomposition techniques, where they applied to message-word and message rank matrices. This technique can be used to filter out interesting subset from the set of all messages. However, they have shown results only for artificial small dataset and particular modifications to it.

6] Wu et al. [x] proposed an algorithm that initially identifies a set of bad pages based on the common link set between incoming and outgoing links of WebPages, and then expands this set by marking a page as bad if it links to more than a certain number of other bad pages. Wu et al. also proposed methods of combining the trust and dis-Trust scores of pages to demote spam pages in the Web [xi]

7] Gao et al. present a study on detecting and characterizing social spam campaigns in Facebook [xii].

Meanwhile, most Twitter criminal account detection work can be classified into two categories. The first category of work, such as [xiii, xiv, xv], utilizes machine learning techniques to classify legitimate accounts and criminal accounts according to their collected training data and their selections of classification features. The second category of work (e.g., [xvi]) detects and analyzes malicious accounts by examining whether URLs or domains posted in the tweets are labelled as malicious by public URL blacklists or domain blacklists.

2. Design model

By considering all the drawbacks of existing methods, Iterative social based classifier is feasible to implement as a solution for adult content detection on Facebook and G-mail. Iterative social based classifier (ISC), an effective classification algorithm resistant to noisy links. ISC consists of three key components: a collective correlation model (CCM), a social driven classifier (SDC), and an iterative classification procedure. CCM is an effective model designed for extracting discriminative collective correlation features

from a noisy graph. SDC is an effective classifier specifically designed for the collective correlation features which can further reduce the impacts of noisy links on the classification performance.

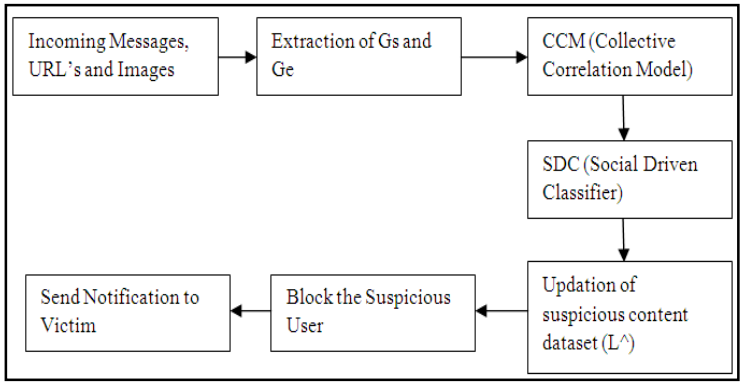

Figure 2: Block Diagram of ISC classifier

As shown in figure, Input to the system is the incoming contents on the G-mail and Facebook which in contains Text, Images and URL's. These contents are given as an input to the ISC (Iterative social based classifier). The Initial requirement of the system is to have dataset of suspicious Images, Text, and URL's. Based on the incoming contents Gs and Ge are extracted, where Gs denotes the "following" relationships between all accounts. Ge denotes the "posting" relationships between accounts and entities. This Gs and Ge are given as an input to the CCM (Collective Correlation Model). CCM generates one collective correlation feature for each type of relationships. It calculates the correlation score (CS) for each node based on the labels of its neighbouring nodes. CCM updates the extracted features of Gs, Ge based on the link based collective correlation (LCC) and entity based collective correlation (ECC) in Xa and Xb respectively. This updated values are given as an input to the next block that is SDC (Social Driven Classifier). A social driven classifier based on the newly updated features X A and X B. Two decision functions, one for LCC and the other for ECC, are jointly learned in SDC. The prediction values of these two decision functions for all accounts are calculated by SDC in yA and yB, respectively. We update dataset L ^ based on the combined prediction of y A and y B. Each time L ^ is expanded by adding a certain number of accounts that have not been placed in this set before. In particular, we add those accounts which have the highest probability of being either adult or normal accounts. In our implementation, we first divide all accounts into adult class and normal class in terms of their prediction values, then we calculate the ranking scores for accounts in each class. We double the size of L ^ in each iteration by adding those accounts which are ranked on the top. Once the accounts are classified into two classes that is Normal class and adult class, the accounts in the adult class are blocked and in the last step the information about the suspicious user is send to the user.

2.1. Iterative Social Based Classifier

Based on the link-entity graph, we design iterative social based classifier. Algorithm 1 describes the overall procedure of ISC. At a high level, ISC uses an iterative classification procedure. Before presenting details about ISC, we first define some notations. Gs denote the "following" relationships between all accounts. Ge denotes the "posting" relationships between accounts and entities. Two types of collective correlation features, link based collective correlation (LCC) and entity based collective correlation (ECC), are extracted from Gs and Ge, respectively. We use XA and XB to denote these two types of features. L denotes the training set. ^ L denotes the automatically labeled set generated by ISC during the iterative classification procedure. _ is a constant between 0 and 1, which is used to combine two decision functions learned in ISC.

Algorithm 1. Iterative Social Based Classifier

Input: G_s, G_e, L, λ
Output: $\mathbf{y}$
1: $MA = 0$;
2: $\hat{L} = L$;
3: **while** *MA increases* **do**
4: $X^A \leftarrow \texttt{CCM}(G_s, \hat{L})$;
5: $X^B \leftarrow \texttt{CCM}(G_e, \hat{L})$;
6: $\mathbf{y}^A, \mathbf{y}^B \leftarrow \texttt{SDC}(X^A, X^B, L, \lambda)$;
7: $\mathbf{y} = \lambda \mathbf{y}^A + (1-\lambda)\mathbf{y}^B$;
8: $\hat{L} \leftarrow \texttt{UpdateLabelSet}(\mathbf{y}, \lambda, \hat{L})$;
9: $MA \leftarrow Jaccard(\mathbf{y}^A_+, \mathbf{y}^B_+)$;
10: **end**

Figure 3

In each iteration, we perform the following operations. First, we update both LCC and ECC based on Gs, Ge and the automatically labeled set ^ L using the collective correlation Model. Initially, ^ L is set to L. Second, we train a social driven classifier based on the newly updated features XA and XB. Two decision functions, one for LCC and the other for ECC, are jointly learned in SDC. The prediction values of these two decision functions for all accounts are denoted by yA and yB, respectively. Third, we update the automatically labeled set ^ L based on the combined prediction of yA and yB. Each time ^ L is expanded by adding a certain number of accounts that have not been placed in this set before. In particular, we add those accounts which have the highest likelihood of being either adult or normal accounts. In our implementation, we first divide all accounts into adult class and normal class in terms of their prediction values, then we calculate the ranking scores for accounts in each class. For example, if a Twitter account is ranked as 10th out of 1,000 accounts, then its ranking score will be 1 –(10/1000). Based on the ranking scores of all accounts, we double the size of ^ L in each iteration by adding those accounts which are ranked on the top. Lastly, we calculate the mutual agreement (MA) on adult account prediction between the two decision functions learned in SDC. Suppose yA and yB represent adult accounts predicted by the two decision functions of SDC, respectively. The mutual agreement is defined as the Jaccard similarity between yA and yB, which can be calculated as || yA intersection yB| | / || yA union yB|| . Here ||.|| denotes the size of a set. If the mutual agreement starts to decrease, we stop the iterative classification procedure.

3. Collective Correlation

In this section, we first introduce the collective correlation model, an effective model for extracting noise-resistant collective correlation features from a graph full of noisy links. Then we apply CCM to the link-entity graph for extracting link based and entity based collective correlation features based on the "following" relationships and "posting" relationships in the link-entity graph, respectively.

3.1. Collective Correlation Model

Following a two-step procedure, CCM generates one collective correlation feature for each type of relationships. In the first step, we calculate the correlation score (CS) for each node based on the labels of its neighboring nodes. For undirected relationship, the neighboring nodes of a node refer to all the nodes which are directly connected to the node by one edge. For directed relationship, the neighboring nodes of a node could refer to its outgoing neighboring nodes, or its incoming neighboring nodes, or both, varying from applications. In the second step, CCM aggregates the correlation scores of the neighboring nodes for each node. Here we first introduce the two steps, and then analyze CCM.

3.1.1. Analysis of CCM

CCM is robust to noisy links for two reasons. First, the correlation score estimation is confidence-aware. Consider two nodes u1 and u2, both have 10 neighboring nodes. Suppose each of u1's neighbors have only one neighbor labeled as adult and one labeled as normal, each of u2's neighbors has 50 neighbors labeled as adult and 50 neighbors labeled as normal. Based on CCM, we can infer u2 is more likely to be an adult account than u1. Second, using the distribution also makes CCM more robust. For example, an adult account may post many normal entities not related to adult content as well as one or two entities which are correlated with adult label. By using the distribution of neighboring entities' correlation scores, we can infer this account is likely to be an adult account.

CCM can generate good collective correlation features by labeling a small number of nodes. Although a certain number of labeled neighboring nodes are important to estimate the correlation score of a single node, we can obtain reliable correlation score estimation for most nodes by labeling a small number of high-degree nodes. Here the high-degree node is defined as a node which has a large number of neighboring nodes. For many real-world graphs, there exist some such high-degree nodes. For example, in the link entity graph there are 1,734 accounts out of 1:07 million that are followed by more than 10,000 accounts. CCM is a scalable model for large graphs. The major computation cost of CCM includes three parts: calculating, sorting and aggregating correlation scores for all nodes. The computational complexity for the second part is O (||E||+nlog(n)) using popular sorting algorithms (e.g., heap sort), where n is the number of nodes on the graph. For the other two parts, since both of them need to iterate each edge of the graph once, their computational complexity is O(||E||), where ||E|| represents the number of edges on the graph. Therefore, the overall computational complexity of CCM is O (||E||+nlog (n))+ O(n).

4. Social Driven Classifier

In order to improve the classification performance on the Link-entity graph full of noisy links; we take advantages of Two social properties of the collective correlation features in our SDC design. We first introduce the properties of the collective correlation features used in SDC design, and then present the details of SDC design.

4.1. Social Properties of Collective Correlation Features

In SDC, the following two social properties of the collective Correlation features are exploited.

- Property I. The collective correlation features for different types of relationships are correlated. For example, link based collective correlation is correlated with entity based collective correlation, because if the followers of a Twitter account are interested in adult content, then some of the entities posted by this Twitter account are probably related to adult content.
- Property II. As a distribution of the correlation scores of neighboring nodes, the portions of collective correlation feature that correspond to the neighboring nodes with strong correlation scores are stronger clues for classification than other portions. For example, if a Twitter accounts posts an entity that is strongly correlated with adult content, this Twitter account is

probably an adult account even though this account also posts many entities that are not related to adult content. Similarly, if one follower of a Twitter account is strongly interested in adult content, then this Twitter account is probably an adult account even though all other followers demonstrate low interests in adult content.

5. Computational Complexity

The computation cost of ISC depends on the number of iterations and the computation cost for each iteration. The computation cost of each iteration mainly consists of two parts: (1) calculating LCC and ECC features; and (2) training SDC and classifying all accounts using SDC. The computational complexity of the first part is O(||E||+nlog(n)), where ||E|| denotes the number of edges and n denotes the number of accounts in the link-entity graph. The second part includes training SDC and calculating the predicted labels of all accounts using SDC. The computation cost for the second part is O (n), where n denotes the number of accounts in the link-entity graph. Both of these two parts can be computed efficiently. Both of these parts can be computed efficiently, so we can say that ISC is feasible to implement as a solution for adult account detection on Facebook and Gmail. Hence, ISC is P class problem because work can be completed in polynomial time.

6. Conclusion

Social networking sites are modern popular platform of communication. But now a day's Social Networking sites as well as E-mail accounts are became an attractive platform for illegal activities like planning a terrorist activities or spreading rumours or hate message, do activities which is harmful to the society. In this article, we present a novel solution to effectively classify Facebook, G-mail accounts that contain adult content. We first formulate the adult account detection as a graph based classification problem and construct a graph based on social links and entities. Since adult Facebook, G-mail accounts are mostly connected with normal accounts and an entity not related to adult content, the constructed graph is intrinsically full of noisy links that connect nodes with different labels. Our major contribution in this work is the design of an iterative social based classifier which can accurately classify nodes on the graph full of noisy links by labelling a small number of nodes for Facebook as well as G-mail.
So, the proposed system can be used by crime investigation agencies.

7. References

i. M. Hepple, N. Ireson, P. Allegrini, S. Marchi, S. Montemagni, and J. G_omez, "Nlp-enhanced content filtering within the poesia project," in Proc. 4th Int. Conf. Lang. Resources Eval., 2004, p. 1.
ii. A. Lopes, S. de Avila, A. Peixoto, R. Oliveira, and A. Ara_ujo, "A bag-of-features approach based on hue-sift descriptor for nude detection," in Proc. 17th Eur. Signal Process. Conf., Glasgow, Scotland, 2009, pp. 224–231.
iii. J. Ze Wang, J. Li, G. Wiederhold, and O. Firschein, "System for screening objectionable images," Comput. Commun., vol. 21, no. 15, pp. 1355–1360, 1998.
iv. Fleck, D. Forsyth, and C. Bregler, "Finding naked people," in Proc. 10th Eur. Conf. Comput. Vis., 1996, pp. 593–602.
v. T. Deselaers, L. Pimenidis, and H. Ney, "Bag-of-visual-words models for adult image classification and filtering," in Proc. IEEE Conf. Pattern Recognit., Dec. 2008, pp. 1–4.
vi. H. Cheng, Y. Liang, X. Xing, X. Liu, R. Han, Q. Lv, and S. Mishra, "Efficient misbehaving user detection in online video chat services," in Proc. 5th ACM Int. Conf. Web Search Data Mining, 2012, pp. 23–32.
vii. P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," Pattern Recognit., vol. 40, no. 3, pp. 1106–1122, 2007.
viii. K9 web protection [Online]. Available: http://www.k9webprotection.com
ix. Skillicorn and David, "Keyword filtering for message and conversation detection," Queen's University. http://www.cs.queensu.ca/home/skill/beyondkeywords.pdf, 2005.
x. B. Wu and B. D. Davison. "Identifying link farm spam pages". In ACM Int'l Conference on World Wide Web (WWW), 2005.
xi. B. Wu, V. Goel, and B. D. Davison. "Propagating trust and distrust to demote web spam". In Workshop on Models of Trust for the Web, 2006.
xii. F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. "Detecting Spammers on Twitter". In Collaboration, Electronic messaging, Anti-Abuse and Spam Confference (CEAS), 2010.
xiii. H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao. "Detecting and Characterizing Social Spam Campaigns". In Proceeding of ACM SIGCOMM IMC (ICM'10), 2010.
xiv. C. Grier, K. Thomas, V. Paxsony, and M. Zhangy. @spam: "The Underground on 140 Characters or Less". In ACM Conference on Computer and Communications Security (CCS), 2010.
xv. K. Lee, J. Caverlee, and S. Webb. "Uncovering Social Spammers: Social Honeypots + Machine Learning". In ACM SIGIR Conference (SIGIR), 2010.
xvi. G. Stringhini, S. Barbara, C. Kruegel, and G. Vigna. "Detecting Spammers on Social Networks". In Annual Computer Security Applications Conference, 2010.