REVIEW ON ENHANCEMENT IN OSN BY SPAM FILTERING & DETECTION

Prof.S.Y.Raut, Miss. Yewale Pooja B., Mr. Bhagat Vipul G., Miss. Phad Kavita B. , Mr. Suryawanshi Abhishek S.

Dept of Comp Engg., P.R.E.C, Loni.

ABSTRACT

Online social networks' (OSNs) popularity has habitual users to the share information. At the same time, OSNs have been a focus on privacy with respect to the information shared. Therefore, it is important that users have little assurance when sharing on OSNs: popular OSNs give users with mechanisms, to protect shared information access right. However, these mechanisms do not allow collaboration or mixing when given access rights for joint content related to more than one user (e.g., party pictures in which more than one users are being tagged). In fact, the access rights list for these content is defined by the union of the access list represented by each related user, which could result in unwanted leakage or msg. We propose a collaborative access control method, based on secret information sharing, in which sharing of content on OSNs is decided comparative by the number of related users or the user which are connected to each others. We demonstrate that such mechanism is benefits users' privacy on own space.

Keyword: *Online Social Network, Radial Basis Function Network, Social Network Manager, Graphical User Interface, Filtered wall.*

I. Introduction

In the previous years, Online Social Networks, such as Face book or Twitter, have played an important role in changing society, give people the ability to information, and build communities around shared interests. Popular OSNs power users with some or more customizable "privacy settings" to take on access rights decisions based on access list. Limiting access rights or control to a subset of the users, for instance, Friends, friends-of-Friends or everyone. However, those mechanisms are often difficult and may lead to accidental leakage [7], [25]. According to Gurses and Berendt [20] the access control design in current online social Network represent the principal bottleneck of privacy when sharing information. At time, the research community has give solutions to fine tune the access control mechanism that rely on the use of encryption [3], [4], [5], [23] to prevent leakages. However, those solutions aim to protect users information imposing access control by confidentiality, but they do not allow collaboration. The major efforts in building a short text classifier are used for concentrated in the extraction and selection of a set of characterizing and discriminate features.

The solutions investigated or invented in this paper are an extension of those adopted in a previous work by us [5] from which we find the learning model and the a technique used to discreetly gather information method for generating pre-process data. The original set of characteristic, derived from Endogenous properties of short texts, is enlarged here including originating externally knowledge related to the context from which the messages originate. As far as the learning model is concerned, we sure in the current paper the use of deep learning or neural learning which is today recognized as one of the most well organized solutions in text classification [4]. In particular, we base the overall short text classification method on Radial Basis Function Networks (RBFN) for their proven capabilities in acting as soft classifiers, in managing noisy data and congenitally imprecise classes. Moreover, the speed in performing the learning phase creates the premise for an sufficient use in OSN domains, as well as facilitates the experimental evaluation tasks. We insert the neural model or deep model within a hierarchical two level classification strategy. In the first level, the RBFN class short messages as Neutral and Non-Neutral; in the second level, Non-Neutral messages are classified producing moderate estimates of appropriateness to each of the considered category.

III: Related Work:

The issue of spamming over emails and in many other forms is a well studied problem. Spam Detection has been the area of interest of many researchers. Many solutions have been propounded in regard to spam detection. However, spam detection in the social networks, which is a recent phenomenon, has not been studied so widely. Also, the fact that Tweet messages are small in size, restricted to 140 characters only (as opposed to email or web content), the problem of spam detection becomes more difficult. This section summarizes the main contributions of other researchers on spam detection in social networks. Fabricio Benevenuto et al. [2] detected spammers by identifying various user social behaviors and the characteristics of tweet content. These characteristics were used in a machine learning approach to classify the users as spammers and non-spammers. De Wang et al. [3] in their study proposed a general framework to detect spam account across all the OSNs. The main contribution of their work was a new spam detected in any one social networking could be quickly identified across all other OSNs. Alex Hai Wang et al. [5] proposed a model which uses a directed graph that represent the relation between "friends" and "follower" relationship in twitter. Bayesian Classifier was also used in his work, to detect spam accounts. Xin Jin et al. [6] propounded a method for detecting spam accounts in social media network.

A. Content-based filtering:

Information filtering systems are designed or implement to classify a stream or class of dynamically generated information to send something asynchronously or not with by an information producer and present to the user those information that are like to satisfy requirements [6]. In content-based filtering each user is assumed to operate single-handed. As a result, a content-based filtering system or method selects information items based on the correlation between the gratified of the items and the user preferences as opposed or different to a collaborative filtering system that select items based on the relation between people with same preferences or rights[7], [8]. While electronic mail was the main domain of early work on information filtering, accompanying papers have addressed miscellaneous domains including news wire articles, Internet "news" articles, post ,blog and broader network resources [9], [10], [11]. Documents prepared in content-based filtering are mostly textual or argument in nature and this makes content-based filtering close to text classification.

B. Policy-based personalization of OSN contents:

Recently, there have been some proposals exploiting classification executions for personalizing access in OSNs. For instance, in [27] a classification method has been proposed to categorize short text messages in order to avoid or remove uncontrollable users of micro blogging services by raw data.

The system described in [27] focuses on Twitter2 and members a set of groups with each tweet describing its content. The user can then aspect only certain types of tweets based on user interests. In contrast, Golbeck and Kuter [28] propose an application, called Film depend on that exploits OSN trust relationships and provenance information to personalize approach to the website. However, such systems do not provide a filtering policy layer by which the user can exploit the result of the classification activity to choose how and to which extent filtering out unwanted information. In contrast, our filtering strategy or rule language allows the setting of FRs according to a variety of criteria, that do not consider only the ravages of the classification process or method but also the relationships of the wall owner with other OSN users as well as information on the user profile.

C. Machine Learning-based Classification:

We declaration to short text categorization as a graded two-level classification process or method. The first-step classifier performs a binary hard classifies that labels messages as Neutral and Non-Neutral. The first-level filtering task facilitates the consequent second-level task in which a finer-grained classification is performed. The second-level classifier performs or operates a soft-partition of Non-neutral messages assigning a given message a gradual membership to each of the non neutral classes. Among the collection of multi-class ML models well-suited for text classification, we choose the RBFN model [39] for the adventured competitive behavior with respect to other state of the art classifiers. RFBNs have a single unseen layer of processing units with local, restricted activation domain: a Gaussian function is usually used, but any other normally tunable function can be used. They were introduced as a neural network evolution of exact interpolation [40], and are established to have the universal approximation property [41], [42]. As outlined in [43], RBFN have main advantages are that classification function is non-linear, the model may produce confidence values and it may be robust to outliers; downsides are the potential sensitivity to input parameters, and potential overtraining awareness. The first level classifier is then planned as a regular RBFN. In the second level of the classification stage or level we propose a modification of the normal use of RBFN.

IV. Proposed Approach:

An overview of the complete process of junk mail recognition is shown in the diagram in Figure 1, each of whose steps are explained in this section. The preliminary step for the detection of spammers in any OSN is data collection and necessary preprocessing to convert it into a form, which can be used by the learning algorithms.

A. Data Set Description:

In our work, we have used the dataset received from Fabricio Benevenuto et al. [2] which consists of labeled record of 1064 Twitter users. Dataset comprises of 62 characteristics containing user specific and tweet specific information. The spammer accounts comprised of around 36% of the dataset. Also, as per [2], the users were chosen randomly and not based on any of their characteristics. They [2] have used SVM based machine learning approach as opposed to our work in which we have used other learning methods namely Naive Bayes, Clustering, Decision Trees and finally combined all of them together to achieve a higher junk mail recognition accuracy.

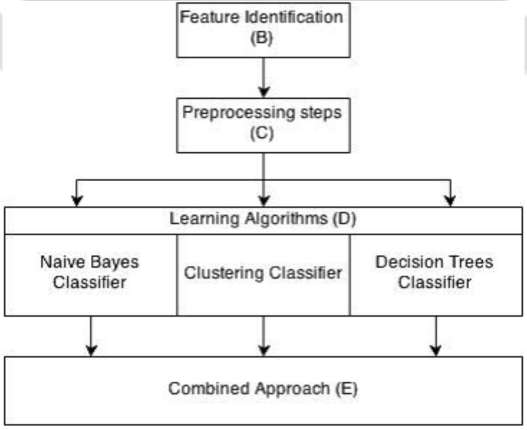


Fig.1: Proposed Spam Detection Approach

B. Feature Identification:

Since, spammers behave differently from non-spammers; therefore we can identify some features or characteristics in which both these categories and class change. Various features which we have used to detect spam accounts include: Number of followers and followers': Followers are the users who follow a particular user, while followers' are the users whom the user follows. Generally speaking, spammers have small number of followers but follow large population with the motive to get noticed by many. Therefore, account with large followers' and small number of followers can potentially be considered as a spam account.

C. Preprocessor:

Twitter user accounts in the dataset [2], labeled as Spammer and Non-Spammers, and were used for training the learning algorithms and also in accuracy calculations.

In preprocessing step, all the repeated structure were converted into separate. The procedure adopted to select the intervals for a particular feature was obtained from [4] according to which all user accounts are arranged in increasing order of their feature principles. Processing start from the first account, if we encounter an account whose category or section is differ from the category of the next summing, and then an interval or space is generate as a mean of both the feature values.

D. Learning Algorithms:

There are various different classification algorithms, which can be used to classify an account as "Spammer". In this method, we have spent Naive Bayes, Clustering and Decision trees as learning algorithms. Although, each of these approaches can be solely used to classify user accounts, but in order to increase the accuracy, we have combined these approaches into an integrated algorithm in our work.

Filtered Wall Architecture

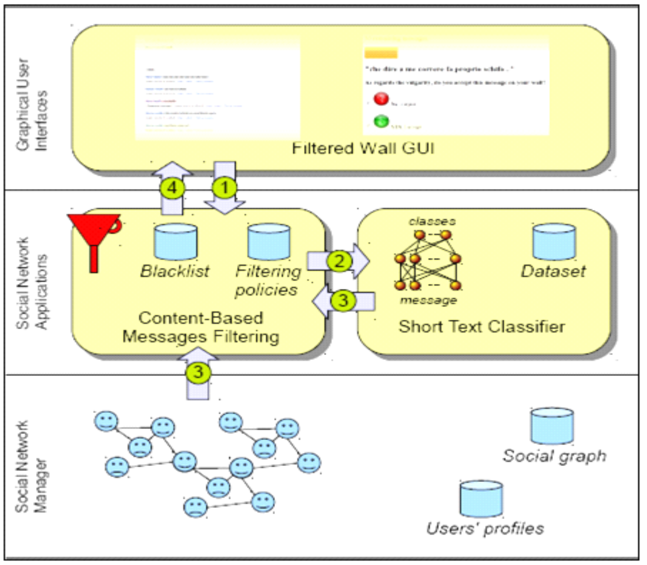


Fig.2: Wall conceptual architecture and filtered stream messages follow, writing for publication.

The first layer or level, called the Social Network Manager (SNM), it provides the basic Functionality OSN (i.e., profile), while the second layer provides stand for external or outer applications Social Network (SCN). SNA situated on in turn may need or required an extra layer to their graphical user interfaces necessary (GUI). To protect a spam message posted on the user wall of the user own space and to protect the user social image is an important issue in the social networking site. To filter or removed unwanted messages, we offer three levels architecture containing a message classifier based on the information and use of machine learning technique. The user is able to customize the filtering rule according to their priorities also set the filter on
different user privileges or level i.e. subsidies to allow the user to insert messages in his / her wall.

Let us explain this by taking an example. Consider an user A who is connect to his friend through social network such as facebook. So this user firstly need to login to his account. For login, the following window opens.

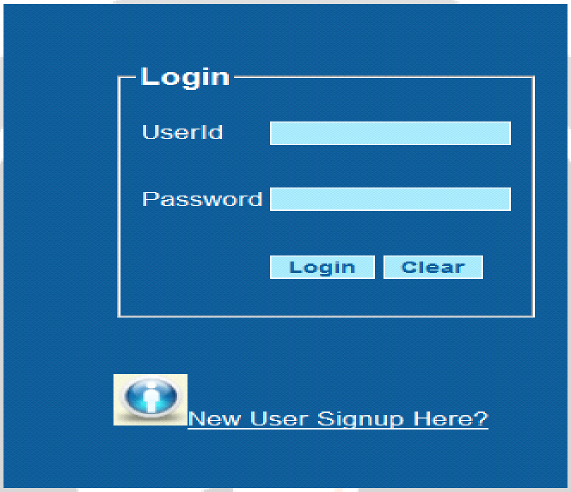


Fig: Login

In this signup process, the user needs to signup with his name, his email ID, new password, gender, date of birth, etc. as shown in fig below:

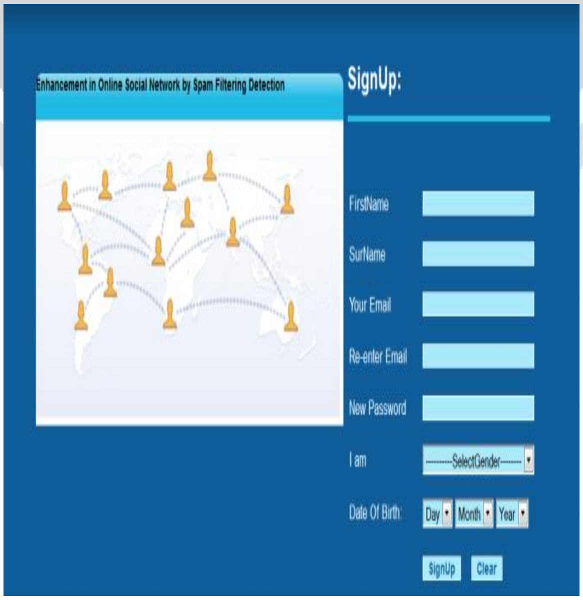
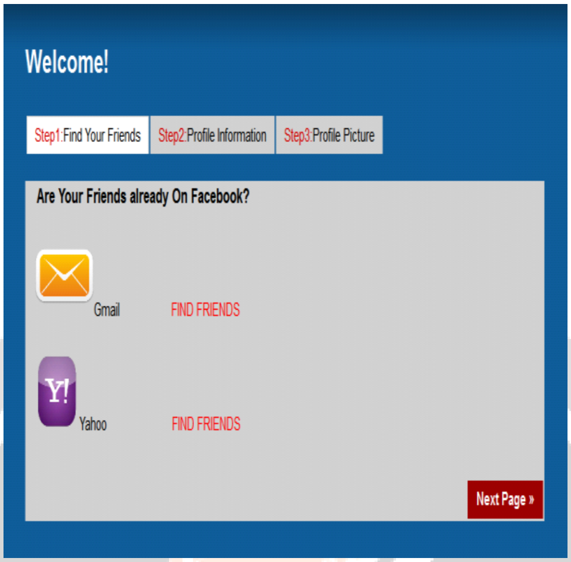


Fig: Signup Page

After this signup process, the user create new account on the social network successfully and he is able to access all data over that sites.



Welcome!

Are Your Friends already On Facebook?

Fig: Send request by user

In which user can send the request to connecting friend or group of friends. Also user can add profile picture on his account also upload the different kind of information.

V. Objective:

1. To provide a filtered current wall (FW) mechanism that is able to filter out unwanted messages wall present in the online social network.
2. Provide classification mechanism to prevent unnecessary data overwhelmed present in Wall position of the user.
3. Improve the quality of classification.

VI. Conclusion & Future Work:

In our work, an algorithm, combining three different learning procedures and algorithms (namely Naive Bayes, Clustering and Decision trees) was implemented. This integrated algorithm categorizes an account as Spammer/Non-Spammer with an overall accuracy of 87.9%. Finally, this algorithm was compared with all the three learning algorithms taken alone. It was observed that the combined approach could give best results in terms of overall accuracy and in detection of non-spammers. Though, Decision Trees alone perform better in detection of spammers but it is poor in detecting non-spam accounts, thus it can't be used solely. As a future work, the integrated approach can be further improved by maintaining high accuracy of Decision Trees approach with respect to detection of Spammer accounts.

References:

[1] A. Adomavicius, G.and Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[2] M. Chau and H. Chen, “A machine learning approach to web page filtering using content and structure analysis,” Decision Support Systems, vol. 44, no. 2, pp. 482–494, 2008.

[3] R. J. Mooney and L. Roy, “Content-based book recommending using learning for text categorization,” in Proceedings of the Fifth ACM Conference on Digital Libraries. New York: ACM Press, 2000, pp. 195–204.

[4] F. Sebastiani, “Machine learning in automated text categorization,” ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

[5] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari,“Content-based filtering in on-line social networks,” in Proceedings of ECML/PKDD Workshop on Privacy and Security issues in Data Mining and Machine Learning (PSDML 2010), 2010.

[6] N. J. Belkin and W. B. Croft, “Information filtering and information retrieval: Two sides of the same coin?” Communications of the ACM, vol. 35, no. 12, pp. 29–38, 1992.

[7] P. J. Denning, “Electronic junk,” Communications of the ACM, vol. 25, no. 3, pp. 163–165, 1982.

[8] P. W. Foltz and S. T. Dumais, “Personalized information delivery: An analysis of information filtering methods,” Communications of the ACM, vol. 35, no. 12, pp. 51–60, 1992.

[9] P. S. Jacobs and L. F. Rau, “Scisor: Extracting information from online news,” Communications of the ACM, vol. 33, no. 11, pp. 88–97, 1990.

[10] S. Pollock, “A rule-based message filtering system,” ACM Transactions on Office Information Systems, vol. 6, no. 3, pp. 232–254, 1988.

[11] P. E. Baclace, “Competitive agents for information filtering,” Communications of the ACM, vol. 35, no. 12, p. 50, 1992.

[12] P. J. Hayes, P. M. Andersen, I. B. Nirenburg, and L. M. Schmandt, “Tcs: a shell for content-based text categorization,” in Proceedings of 6th IEEE Conference on Artificial Intelligence Applications (CAIA-90). IEEE Computer Society Press, Los Alamitos, US, 1990, pp. 320–326.

[13] G. Amati and F. Crestani, “Probabilistic learning for selective dissemination of information,” Information Processing and Management, vol. 35, no. 5, pp. 633–654, 1999.

[14] M. J. Pazzani and D. Billsus, “Learning and revising user profiles: The identification of interesting web sites,” Machine Learning, vol. 27, no. 3, pp. 313–331, 1997.

[15] Y. Zhang and J. Callan, “Maximum likelihood estimation for filtering thresholds,” in Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2001, pp. 294–302.

[16] C. Apte, F. Damerau, S. M. Weiss, D. Sholom, and M. Weiss, “Automated learning of decision rules for text categorization,” Transactions on Information Systems, vol. 12, no. 3, pp. 233–251, 1994.

[17] S. Dumais, J. Platt, D. Heckerman, and M. Sahami, “Inductive learning algorithms and representations for text categorization,” in Proceedings of Seventh International Conference on Information and Knowledge Management (CIKM98), 1998, pp. 148–155.

[18] D. D. Lewis, “An evaluation of phrasal and clustered representations on a text categorization task,” in Proceedings of 15th ACM International Conference on Research and Development in Information Retrieval (SIGIR-92), N. J. Belkin, P. Ingwersen, and A. M. Pejtersen, Eds. ACM Press, New York, US, 1992, pp. 37–50.

[19] R. E. Schapire and Y. Singer, “Boostexter: a boosting-based system for text categorization,” Machine Learning, vol. 39, no. 2/3, pp. 135– 168, 2000.

[20] H. Sch¨utze, D. A. Hull, and J. O. Pedersen, “A comparison of classifiers and document representations for the routing problem,” in Proceedings of the 18th Annual ACM/SIGIR Conference on Resea. Springer Verlag, 1995, pp. 229–237.

[21] E. D. Wiener, J. O. Pedersen, and A. S. Weigend, "A neural network approach to topic spotting," in Proceedings of 4th Annual Symposium on Document Analysis and Information Retrieval (SDAIR-95), Las Vegas, US, 1995, pp. 317–332.

[22] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in Proceedings of the European Conference on Machine Learning. Springer, 1998, pp. 137–142.

[23] ——, "A probabilistic analysis of the rocchio algorithm with tfidf for text categorization," in Proceedings of International Conference on Machine Learning, 1997, pp. 143–151.

[24] S. E. Robertson and K. S. Jones, "Relevance weighting of search terms," Journal of the American Society for Information Science, vol. 27, no. 3, pp. 129–146, 1976.

[25] S. Zelikovitz and H. Hirsh, "Improving short text classification using unlabeled background knowledge," in Proceedings of 17th International Conference on Machine Learning (ICML-00), P. Langley, Ed. Stanford, US: Morgan Kaufmann Publishers, San Francisco, US, 2000, pp. 1183–1190.

[26] V. Bobicev and M. Sokolova, "An effective and robust method for short text classification," in AAAI, D. Fox and C. P. Gomes, Eds. AAAI Press, 2008, pp. 1444–1445.

[27] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," in Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval,
SIGIR 2010, 2010, pp. 841–842.

[28] J. Golbeck, "Combining provenance with trust in social networks for semantic web content filtering," in Provenance and Annotation of Data, ser. Lecture Notes in Computer Science, L. Moreau and I. Foster, Eds. Springer Berlin / Heidelberg, 2006, vol. 4145, pp.101–108.

[29] F. Bonchi and E. Ferrari, Privacy-aware Knowledge Discovery: Novel Applications and New Techniques. Chapman and Hall/CRC Press, 2010.

[30] A. Uszok, J. M. Bradshaw, M. Johnson, R. Jeffers, A. Tate, J. Dalton, and S. Aitken, "Kaos policy management for semantic web services,"IEEE Intelligent Systems, vol. 19, pp. 32–41, 2004.