

Survey on Filtration of Unwanted Messages from OSN Using Customizable Filtering Rules

Dhivya S, Information Technology, IFET College of engineering, Villupuram, India

Kalaivani J, Assistant Professor, Information Technology, IFET College of engineering, Villupuram, India

Abstract— Now a days On-line Social Networks (OSNs) are one of the most common interactive site to share, communicate and distribute a significant amount of human life information. The difficulty in OSN is to assign users the authority to manage the messages posted on their private space to avoid unwanted content to be displayed. Till now user walls offer little support to this requirement. This paper delivers a survey about, the proposed system allowing OSN users to have a direct control on the messages posted on their walls and block users for particular time limit who has posted unwanted message on wall. This is achieved through a customizable rule-based system, that allows users to adapt the filtering criteria to be applied to their walls, and a Machine Learning approach based soft classifier which automatically labeling messages in support of content-based filtering.

Index Terms— On-line Social Networks, flexible rule-based system, Machine Learning, soft classifier, content-based filtering.

I. INTRODUCTION

Today's modern life is totally based on Internet. Now a day's people cannot imagine life without Internet. Also, OSNs are just a part of modern life. From last few years people share their views, ideas, and information with each other using social networking sites. Such communications may involve different types of contents text, image, audio and video data. The Facebook user creates 90 pieces of content each and every month, while more than 30 billion pieces of content are (web links, news stories, blogs, notes, photo albums, etc.) shared each month. But, in today's OSN, there is a very high chance of posting unwanted content on particular public/private areas, called in general walls. In the online social network such as Facebook, Twitter, etc., it is

possible to display any type of data on the private wall of the user. These data can hold unwanted messages. Posting unwanted content on the user's wall of social network account can corrupt the social image of the user. So, the safety of this wall is an important issue.

In many social networking systems such issue is addressed by providing easy remove content, blocking the installation of the user, which makes post private institution, etc. But if the user's posts on another user's wall and another user not touch social account for so many days so it can accommodate post on his wall for a long time and can cause corruption of the social image and even if the locking scheme of the user. Content built on the content of the wall filter does not yet have any system of online social networking. OSN today offer little support to prevent unwanted messages on the walls of the user by implementing filtering rules (FR) in our system. Also, Black List (BL) will be maintained in this system. For example, Facebook users are only allowed to say who is allowed to insert messages in their walls (i.e. Direct and indirect friends). Therefore the objective of this work is to propose and evaluate experimentally an automated system, called filtered wall (FW), capable of filtering spam walls of OSN user. To automatically assign each short text message a set of categories based on their content; we use machine learning (ML) technical text classification. In precise, it has a tendency to base the short text classification strategy on Radial Basis perform Networks (RBFN) for his or her proven capabilities in acting as soft classifiers, in managing noisy information and as such obscure categories. Moreover, the speed in performing arts the educational part creates the premise or an adequate use in OSN domains, similarly as facilitates the experimental analysis tasks.

II. LITERATURE SURVEY

In this section, we focus on the different methods for text filtering problems of ONS. [1] Describes the content-based filtering is based primarily on the use of one paradigm ML classifier that is automatically induced learning a set of pre-classified examples. A remarkable variety of related work has recently appeared for feature extraction methods adopted, learning style, and the collection of samples. The extraction of text features maps procedure into a compact representation of its content and is uniformly applied to training and generalization phases. Several experiments show that bag of words (Bow) approaches provide good performance and generally outweigh more sophisticated text representation which may have a higher semantics, but the lower statistical quality. As regards the model of learning is concerned, there are a number of approaches depending on the main-wire tiring content and text classification typically showing the mutual disadvantages and advantages according to the issues that rely on the application.

[2] R.J. Mooney and L. Roy describe Recommender systems improve access to relevant products and information by making personalized suggestions based on previous examples of a user's likes and dislikes. Most existing recommender systems use social filtering methods that base recommendations on other users' preferences. By contrast, content-based methods use information about an item itself to make suggestions. This approach has the advantage of being able to recommend previously unrated items to users with unique interests and to provide explanations for its recommendations. We describe a content-based book recommending system that utilizes information extraction and a machine-learning algorithm for text categorization. Initial experimental results demonstrate that this approach can produce accurate recommendations. These experiments are based on ratings from random samplings of items and we discuss problems with previous experiments that employ skewed samples of user-selected examples to evaluate performance.

[3] F. Sebastian describes the automated categorization (or classification) of texts into predefined categories has witnessed a booming interest in the last ten years, due to the increased availability of documents in digital form and the ensuing need to organize them. In the research community the dominant approach to this problem is based on machine learning techniques: a general inductive process automatically builds a classifier by learning, from a set of pre-classified documents, the characteristics of the categories. The advantages of this approach over the knowledge engineering approach (consisting in the manual definition of a classifier by domain experts) are a very good effectiveness, considerable savings in terms of expert labor power, and straightforward portability to different domains. This survey discusses the main approaches to text categorization that fall within the machine learning paradigm. We will discuss in detail issues pertaining to three different problems, namely document representation, classifier construction, and classifier evaluation.

[4] Adomavicius and Tuzhilin discussed an advanced recommendation systems concept, also discussed methods of current recommendation generation are classified into three main categories collaboration, and hybrid

recommendation approaches based on the content. But these approaches do not consider the process of understanding of users and items.

[5] M. Chau and H. Chen describes as the Web continues to grow, it has become increasingly difficult to search for relevant information using traditional search engines. Topic-specific search engines provide an alternative way to support efficient information retrieval on the Web by providing more precise and customized searching in various domains. However, developers of topic-specific search engines need to address two issues: how to locate relevant documents (URLs) on the Web and how to filter out irrelevant documents from a set of documents collected from the Web. We propose a machine-learning-based approach that combines Web content analysis and Web structure analysis. We represent each Web page by a set of content-based and link-based features, which can be used as the input for various machine learning algorithms. The proposed approach was using both a feed forward/back propagation neural network and a support vector machine. Two experiments were designed and conducted to compare the proposed Web-feature approach with two existing Web page filtering methods - a keyword-based approach and a lexicon-based approach. The experimental results showed that the proposed approach in general performed better than the benchmark approaches, especially when the number of training documents was small. The proposed approaches can be applied in topic-specific search engine development and other Web applications such as Web content management.

[6] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari this paper proposes a system enforcing content-based message filtering for On-line Social Networks (OSNs). The system allows OSN users to have a direct control on the messages posted on their walls. This is achieved through a flexible rule-based system, that allows a user to customize the filtering criteria to be applied to their walls, and a Machine Learning based soft classifier automatically labeling messages in support of content-based filtering. This survey discusses the main approaches to text categorization that fall within the machine learning paradigm. Our goal is to design an online message filtering system that is deployed at the OSN service provider side. Once deployed, it inspects every message before rendering the message to the intended recipients and makes immediate decision on whether or not the message under inspection should be dropped. In content-based filtering, each user is assumed to operate independently. As a result, a content-based filtering system selects information items based on the correlation between the content of the items and the user preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences [7], [8]. While electronic mail was the original domain of early work on information filtering, subsequent papers have addressed diversified domains including newswire articles, Internet "news" articles, and broader network resources [9], [10], [11].

III. CONCLUSION

Thus the Usage of Machine Learning has given higher results to the system to trace the messages and the users to distinguish between the good and bad messages, finally the authorized and unauthorized users in the Social Networking User Profiles spontaneously. Thus the Machine Learning approach plays a vital role in this paper in order to generate the blacklist of the bad words and the unauthorized users. In this context, a statistical analysis has been conducted to provide the usage of the good and bad words by the persons in the sites. Overall, the vulgarity of the users has been prevented.

IV. FUTURE WORK

In our future work a system first, apply customizable rules on text document (i. e. message). Secondly, message is classified according to its contents. Finally if particular user has posted the unwanted messages they will be notified through the mail as that their messages were blocked.

REFERENCES

- [1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo, "A System to Filter Unwanted Messages from OSN User Walls Using Filtering Rules", IEEE Transactions On Knowledge And Data Engineering Vol.25 Year 2013
- [2] R.J. Mooney and L. Roy, "Content-Based Book Recommending Using Learning for Text Categorization," Proc. Fifth ACM Conf. Digital Libraries, pp. 195-204, 2000.
- [3] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1- 47, 2002.
- [4] A. Adomavicius, and G. Tuzhilin, "Towards the next generation of recommender systems: A survey of state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering vol.17, no. 6, pp. 734-749, 2005.
- [5] 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.
- [6] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari, "Content-Based Filtering in On-Line Social Networks," Proc. ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML '10), 2010.
- [7] P.J. Denning, "Electronic Junk," Comm. 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," Comm. ACM, vol.35, no. 12, pp. 51-60, 1992.
- [9] P.S. Jacobs and L.F. Rau, "Scisor: Extracting Information from On-Line News," Comm. ACM, vol. 33, no. 11, pp. 88-97, 1990.
- [10] S. Pollock, "A Rule-Based Message Filtering System," ACM Trans. Office Information Systems, vol. 6, no. 3, pp. 232-254, 1988.
- [11] P.E. Baclace, "Competitive Agents for Information Filtering," Comm. 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," Proc. Sixth IEEE Conf. Artificial Intelligence Applications (CAIA '90), pp. 320-326, 1990.
- [13] G. Amati and F. Crestani, "Probabilistic Learning for Selective Dissemination of Information," Information Processing and Management, vol. 35, no. 5, pp. 633- 54, 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," Proc. 24th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 294-302, 2001.

AUTHORS BIOGRAPHY



Ms.S.Dhivya Currently pursuing B.Tech, Information Technology at IFET College of Engineering, Villupuram, India. Her area of interests includes OOPS concepts, Cryptography and Data mining.



Ms.J.Kalaivan received her B.E in CSE from VRS College of Engineering and Technology, Villupuram and M.Tech in CSE from Manonmaniam Sundaranar University. She is currently working as Assistant Professor in Department of Information Technology, IFET College of Engineering, and Villupuram, India. She has published a book on Computer Graphics.

She has published four papers in international journals. Her area of interests includes Computer Networks, Cryptography and Network Security and Computer Graphics.