

Content Based Filtering Via Social Data: Metadata

S.jackline

M.Tech Student, Department of Information Technology, Dr.Sivanthi Aditanar College of Engineering, Tamilnadu.

J.Ganesh

Assistant Professor, Depart. Of Information Technology, Dr.Sivanthi Aditanar College of Engineering, Tamilnadu.

R.Krishnaveni Jeenath

M.Tech Student, Department of Information Technology, Dr.Sivanthi Aditanar College of Engineering, Tamilnadu.

Abstract – Content-based filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile. The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user. When both quantitative and qualitative meta-data is presented, it is the qualitative meta-data that drives selection. The results are consistent with application of the Recognition heuristic, which postulates that when faced with constrained decision-making, humans will tend to exercise judgement based on cues representing familiarity. The information source that content-based filtering systems are mostly used with are text documents. A standard approach for term parsing selects single words from documents.

Index Terms – Decision-making, Metadata, Cues Recognition.

1. INTRODUCTION

Online Social Networks (OSNs) are today one of the most popular interactive medium to communicate, share, and disseminate a considerable amount of human life information. Daily and continuous communications imply the exchange of several types of content, including free text, image, audio, and video data. According to OSN statistics¹ average user creates 90 pieces of content each month, whereas more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) are shared each month. The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data. They are instrumental to provide an active support in complex and sophisticated tasks involved in OSN management, such as for instance access control or information filtering. Information filtering has been greatly explored for what concerns textual documents and, more recently, web content.

Micro-blogging has become a significant channel of communication that is now widely use day-to-day basis around the world of ten through OSN, the current most dominant micro-blogging service. The defining characteristics of this medium

are its informality ,limited message size and the streamed nature of content provision where users opt-in to receive any content published by users of their choice. These online relationships provide a social network structure through which content is mediated, with users being able to republishor received content as they wish. Gives users the ability to control the messages posted on their own private space to avoid that unwanted content is displayed.

2. RELATED WORK

Recommender systems are a special type of information filtering systems. Information filtering deals with the delivery of items selected from a large collection that the user is likely to find interesting or useful and can be seen as a classification task. Based on training data a user model is induced that enables the filtering system The training set consists of the items that the user found interesting. These items form training instances that all have an attribute. This attribute specifies the class of the item based on either the rating of the user or on implicit evidence. One of the reasons for success is the opportunity to post and receive small message updates in real time so as to draw attention to events while they are occurring.

The ability of a learning method to adapt to changes in the user's preferences also plays an important role. The learning method has to be able to evaluate the training data as instances do not last forever but become obsolete as the user's interests change. Another criteria is the number of training instances needed. A learning method that requires many training instances before it is able to make accurate predictions is only useful when the user's interests remain constant for a long period of time. The Bayesian classifier does not do well here. There are many training instances needed before the probabilities will become accurate enough to base a prediction on. Conversely, a relevance feedback method and a nearest neighbor method that uses a notion of distance can start making suggestions with only one training instance.

3. PROPOSED SYSTEM

In this project, to implement The core components of the proposed system are the Content-Based Messages Filtering (CBMF) and the Short Text Classifier modules.

After entering the private wall of one of his/her contacts, the user tries to post a message, which is intercepted by user.A MachineLearning(ML) based text classifier extracts metadata from the content of the message. Metadata provided by the classifier, together with data extracted from the social graph and users' profiles, to enforce the filtering depending on the result of the previous step, the message will be published or filtered. If any user's message content is filtered immediately a pop up window will be displayed to that user saying that your message is filtered do you want to post the message .we have a filtering graph which represents how many bad words are used how many times by each user. The filtering technique we are using here is content based filtering which performs on the basis of content such as text. In existing system there is no content based filtering in our system implementing content based filtering.

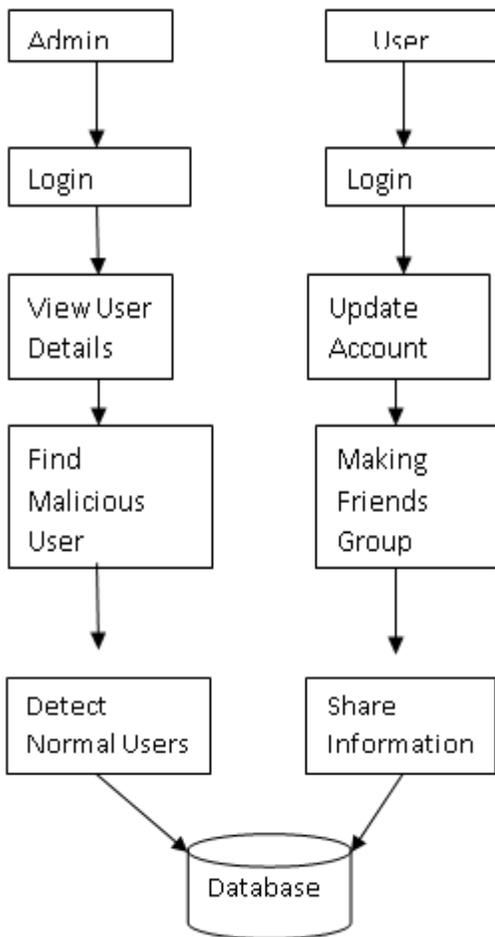


Fig 1 Flow Diagram

4. EXPERIMENTAL RESULTS

In Figure 2, shows that the admin train the abuse keywords for filtering to avoid user profile to display

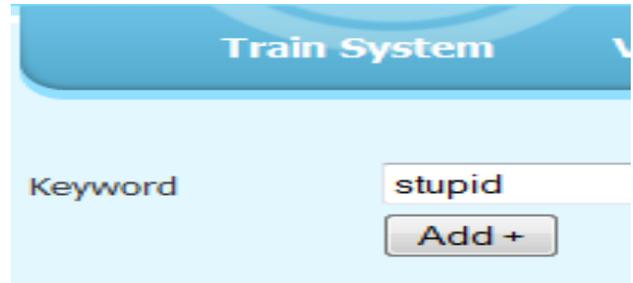


Fig 2 Train Keywords

In Figure 3, shows that the user who post the abuse word to display the other user profile



Fig 3 compose post

In Figure 4, shows the posting word is compare to train keywords.that word is related to abuse word admin block the user

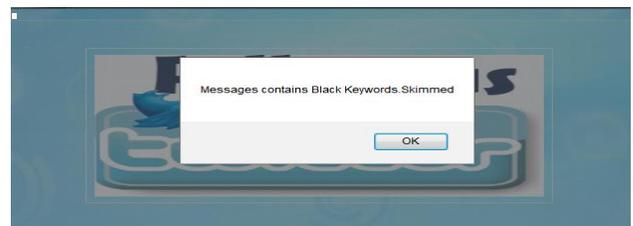


Fig 4 Block Keyword And User

5. CONCLUSION

In this project, Sybil Defender, a scheme that leverages network topologies to defend against Sybil attacks in large social networks. Sybil Defender consists of a Sybil identification algorithm, a Sybil community detection algorithm, and two approaches to limiting the number of attack edges in online social networks. Our evaluation on two large-scale real-world social network samples shows that Sybil Defender can correctly identify Sybil nodes, even when the number of Sybil nodes introduced by each attack edge approaches the theoretically detectable lower bound, and that it

can effectively detect the Sybil community surrounding a Sybil node with different sizes and structures.

REFERENCES

- [1] Abel, F., Gao, Q., Houben, G.-J., Tao, K., 2011. Analyzing user modeling on twitter for personalized news recommendations. In: User Modeling, Adaption and Personalization. Springer, pp.1–12.
- [2] Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G., 2008. Finding high-quality content in social media. In: Proceedings of the International Conference on Web Search and Web Data Mining. ACM, pp.183–194.
- [3] Barchard, K.A., Williams, J., 2008. Practical advice for conducting ethical online experiments and questionnaires for United States psychologists. *Behav. Res. Methods* 40, 1111–1128.
- [4] Cataldi, M., Di Caro, L., Schifanella, C., 2010. Emerging topic detection on twitter based on temporal and social terms evaluation. In: Proceedings of the Tenth International Workshop on Multimedia Data Minin. ACM, p. 4.
- [5] Chen, J.J., Menezes, N.J., Bradley, A.D., North, T., 2011. Opportunities for crowdsourcing research on amazon mechanical turk. *Hum. Factors* 5, 3..
- [6] Chorley, M.J., Colombo, G.B., Allen, S.M., Whitaker, R.M., 2012. Better than you know: social signals on twitter. In: 2012 ASE/IEEE International Conference on Social Computing, pp.277–282.
- [7] DasSarma, A., DasSarma, A., Gollapudi, S., Panigrahy, R., 2010. Ranking mechanisms in twitter-like forums. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM'10, ACM, pp.21–30.
- [8] Dawes, R.M., 1979. The robust beauty of improper linear models in decision making. *Am. Psychol.* 34, 571.
- [9] Duan, Y., Jiang, L., Qin, T., Zhou, M., Shum, H.-Y., 2010. An empirical study on learning to rank soft tweets. In: Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics, pp.295–303