

Survey on a Rule Based System to Refine User Walls

Tirgul Aniket Nandkumar

Department of Computer Engineering,
DGOI,FOE,Daund.
Pune Maharashtra India
anikettirgul333@gmail.com

Salunke Shrikant Dadasaheb

Department of Computer Engineering,
DGOI,FOE,Daund.
Pune Maharashtra India
shrikantsalunke25@gmail.com

Rajpure Amol Subhash

Department of Computer Engineering,
DGOI,FOE,Daund.
Pune Maharashtra India
amolrajpure9@gmail.com

Bere Sachin Sukhadeo

Department of Computer Engineering, DGOI,FOE,Daund.
Pune Maharashtra India
sachinbere@gmail.com

Abstract -The core problem in today's Online Social Networks (OSNs) is to allocate users the authority to manage the messages posted on their private space to avert that unwanted content. The unwanted data may contain political, vulgar, non-neural etc. message filtering systems are designed for unstructured or semi-structured data, as opposed to database applications, which use very structured data. In this paper, we proposed a System with the flexible rules to filter the unwanted messages posted on user wall. After passing threshold value, the informing message is sent to that user. This allows users to customize the refining criteria to be applied to their walls, and a Machine Learning-based classifier automatically classifies the messages and labelling messages in support of content-based filtering.

Keywords - flexible rules, message filtering, online social networks, short text classification.

I. INTRODUCTION

Online Social Networks (OSNs) square measure now a day's one amongst the most popular interactive medium to speak, share, and disperse a substantial quantity of human life information. Daily and continuous communications imply the exchange of many sorts of content, as well as free text, image, audio, and video information. Consistent with facebookstatistics1 average user creates ninety items of content every month, whereas over thirty billion items of content (web links, news stories, blog posts, notes, icon albums, etc.) are parceled monthly. The massive and dynamic character of this information creates the premise for the utilization of internet content mining ways aimed to mechanically discover useful info dormant inside the info. They are instrumental to produce a lively support in advanced and sophisticated tasks concerned in OSN management, such as for instance access management or info filtering.

Information filtering has been greatly explored for what considerations textual documents and, additional recently, website (e.g., [1], [2], [3]). However, the aim of the bulk of those proposals is especially to supply users a classification mechanism to avoid they're powerless by useless information. In OSNs, data filtering may be used for a different, additional sensitive, purpose. This can be thanks to the very fact that, in OSNs, there's the likelihood of posting or commenting alternative posts on specific public/private areas, called normally walls. Data filtering will so be used to provide users the flexibility to mechanically management the messages

written on their own walls, by filtering out unwanted messages. We have a tendency to believe that this can be a key OSN service that has not been provided thus far. Indeed, today OSNs offer little support to forestall unwanted messages on users walls. For instance, Facebook permits users to state United Nations agency is allowed to insert messages in their walls (i.e., friends, friends of friends, or outlined teams of friends). However, no content-based preferences area unit supported, and therefore, it's insufferable to forestall sought messages, like political or vulgar ones, regardless of the user United Nations agency posts them. Providing this service isn't solely a matter of victimization antecedently outlined website mining techniques for a special application, rather it needs to design impromptu classification ways. This can be as a result of wall messages area unit official by short text that ancient classification strategies have serious limitations since short texts don't offer sufficient word occurrences. The aim of this work is so to propose and experimentally value an automatic system, referred to as Filtered Wall (FW), able to filter unwanted messages from OSN user walls. We have a tendency to exploit Machine Learning (ML) text categorization techniques [4] to mechanically assign with every short text message a group of classes supported its content. The major efforts in building a strong short text classifier (STC) area unit targeted within the extraction and choice of a set of characterizing and discriminant options. The solutions investigated during this paper area unit A Nextension of these adopted in an exceedingly previous work by US [5] from that we have a tendency to inherit the learning model, and

3511

therefore, the stimulant procedure for generating pre-classified information. The first set of options, derived from endogenous properties of short texts, is enlarged here as well as exogenous information associated with the context from that the messages originate. As so much because the learning model worries, we have a tendency to ensure within the current paper the employment of neural learning that is these days recognized as one of the foremost economical solutions in text classification [4]. In specific, we have 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 are presenting the different methods those are presented to text filtering problems of ONS. A. Adomavicius and G. Tuzhilin, [2] discussed in Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions that, this paper presents an overview of the field of recommender systems and describes the current generation of recommendation methods that are usually classified into the following three main categories: Content based, collaborative, and hybrid recommendation approaches.

This paper also describes various limitations of current recommendation methods and discusses possible extensions that can improve recommendation capabilities and make recommender systems applicable to an even broader range of applications. These extensions include, among others, an improvement of understanding of users and items, incorporation of the contextual information into the recommendation process, support for multi-criteria ratings, and a provision of more flexible and less intrusive types of recommendations. The feature extraction procedure maps text into a compact representation of it's content and are uniformly applied to training and generalization phases.

In Machine Learning Approach to Webpage Filtering using Content and Structure Analysis, M. Chau and H. Chen [3], addressed, 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. This paper reports our research in addressing the second issue. 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 implemented using both a feed forward or 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.

R.J. Mooney and L. Roy, In Content-Based Book Recommending Using Learning for Text Categorization [4], presented 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.

F. Sebastiani [5] represent the automated categorization (or classification) of texts into predefined categories has witnessed a booming interest in the last 10 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.

III. CONCLUSION AND FUTURE WORK

In this paper our proposed system is to send notification as final warning to user who posts unwanted messages. Firstly, apply flexible rules on text document (i. e. message). Secondly, message is classified according to its contents. Finally if particular user sends unwanted message content continuously then notification is send on his/her email id using which he/she got registered.

REFERENCES

- [1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo, "A System to Filter Unwanted Messages from OSN User Walls" VOL. 25, NO. 2, FEBRUARY 2013.
- [2] A. Adomavicius and G. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.
- [3] 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.
- [4] 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.
- [5] F. Sebastiani, "Machine Learning in Automated TextCategorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.