International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 2 Issue: 9

ISSN: 2321-8169
2724 – 2726

A Machine Learning for Rule-Based System

Asma Mehdia[1]
Department Of Computer Science
Greenfort Engineering College
India

Mohd Ayazuddin[2]
Mirza Younus Ali Baig
Department Of Csit
Nawab Shah Alam Khan College Of
Engg & Tech
India

Mirza Younus Ali Baig[3]
Department Of Csit
Nawab Shah Alam Khan College Of
Engg & Tech
India

Abstract-- The main article issue in today On-line Social Networks (OSNs) is to give users the ability to ascendancy the messages posted on their own private space to avoid that superfluous content is displayed. Up to now OSNs provide little support to this prerequisite. To fill the gap, in this paper, we propose a system allowing OSN users to have a direct control on the messages posted on their walls. This is effectuate through a bendsome rule-based system, that allows users to specification the percolate benchmark to be applied to their walls, and a Machine Learning based soft classifier automatically labeling messages in support of content-based filtering.

Keywords*: On-line Social Networks(OSNs), Rule –based system, Prerequiste, Effectuate, Bendsome.*

*****

I. INTRODUCTION

Document routing can be described as a quandary of statistical text apportionment. Documents are to be assigned to one of two departments, relevant or non-relevant, and a large sample of judged documents is available for training. This paper will compare traditional concernment assessment approaches to routing with apportionment based on explicit error minimization. A dominant problem in routing is the gigantic dimensionality of the built in feature space, where there exists one dormant dimension for each unique term found in the collection, typically hundreds of thousands. Standard allotment techniques cannot deal with such a large feature set, since data processing of the solution is not tractable and the results become errorneous due to the lack of sufficient training data. One explication is to reduce dimensionality by using subsets of the authentic features or mutating them in some way. Another avenue does not attack dimensionality reduction, but instead manipulate a learning algorithm without straight forward error minimization. Relevance feedback via Rocchio augmentation , which has been commonly used in IR, is an example of such an access. We will examine two different forms of dimensionality reduction, Latent Semantic Indexing (LSI) and optimal term selection, in order to fact-finding which form of dimensionality devaluation is most perceptual for the routing problem. In routing, the system uses a query and a list of credentials that have been identified as relevant or not relevant to construct a assessment rule that ranks unlabeled credentials according to their presumption of concernment. We go through a number of different methods of generating the credentials classifier: relevance feedback via query expansion (QE), linear discriminate analysis (LDA), logistic regression (LR), linear neural networks (LNN), and non-linear neural networks (NNN). The computative depiction of the assessment rule is extensively emphatic as a target.

II. LITERATURE SURVEY

Thecrucial contribution of this article is the layout of a structure subject to tai lored content-based message filtering for OSNs, based on ML techniques. As we have acuminate out in the introduction, to the best of our knowledge, we are the first affirming such kind of function for OSNs. However, our work has relationships both with the state of the art in content-based filtering, as well as with the field of policy-based personalization for OSNs and, more in general, web contents. Therefore, in what follows, we survey the literature in both these fields.

1. Content-Based Filtering

Illumination filtering systems are designed to allocate a stream of invigorated generated information dispatched all together by an information producer and present to the user those information that are likely to satisfy his/her engrossments. In content-based filtering, each user is be incline to think to operate independently. As a result, a content-based filtering system selects information items based on the irrelation between the content of the items and the user prepossession as contrary to a synergetic filtering system that chooses items based on the correlation between people with similar preferences [1]. While electronic mail was the authentic empire of early work on information filtering, empire including newswire articles, Internet "news" articles, and broader network resources. Credentials processed in content-based filtering are mostly textual in nature and this makes content-based filtering close to text classification. The activity of filtering can be modeled, in fact, as a case of single label, binary classification, partitioning incoming documents into relevant and nonrelevant categories. More complex filtering systems include multilabel text classification automatically labeling messages into martial tthematic categories. Content-based filtering is mainly based on the use of the ML paradigm according to which a classifier is automatically induced by learning from a set of preclassified examples. A remarkable variety of related work has recently appeared, which differ for the adopted feature extraction methods, model learning, The feature uprooting procedure maps text into a appressed representation of its content and is uniformly applied to training and generalization phases. Several experiments prove that Bag-of-Words (BoW) approaches yield good performance and prevail in general over more sophisticated text representation that may have superior semantics but lower statistical quality .As far as the learning model is concerned, there are a number of major approaches in content-based filtering and text classification in general showing mutual advantages and disadvantages in function of application dependent issues. In [3], a abundant balancing analysis has been conducted confirming bulge of Boosting-based classifiers. However, it is worth to note that most of the work related to text filtering by ML has been applied for long-form text and the assessed performance of the text classification methods strictly depends on the nature of textual documents. The application of content-based filtering on messages posted on OSN user walls poses additional challenges given the short length of these messages other than the wide range of topics that can be discussed. Short text classification has received up to now few attention in the scientific community.

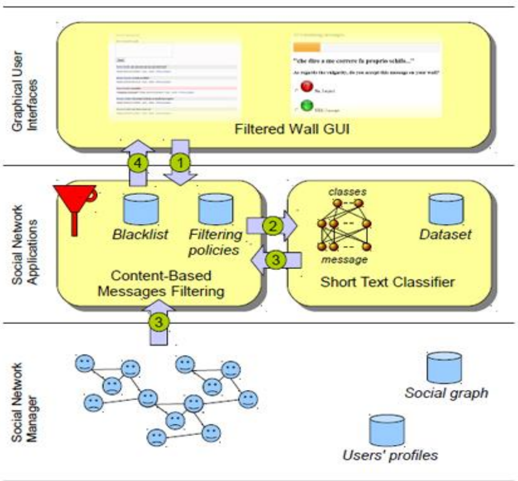
2.**Policy-Based Personalization of OSN Contents**

Latterly, there have been some proposals exploiting classification mechanisms for personalizing access in OSNs. For instance, in a classification method has been proposed to categorize short text messages in order to avoid breathtaking users of more obliging services by raw data. The user can then view only certain types of tweets based on his/her interests. However, such systems do not provide a filtering policy layer by which the user can exploit the result of the classification process to decide how and to which extent filtering out unwanted information. In contrast, our filtering policy language allows the setting of FRs according to a variety of criteria, that do not consider only the results of the classification process but also the relationships of the wall owner with other OSN users as well as information on the user profile. Moreover, our system is complemented by a flexible mechanism for BL management that provides a further opportunity of customization to the filtering procedure.

III. PROBLEM DEFINITION

The aim of the present work is therefore to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. We exploit Machine Learning (ML) text categorization techniques [4] to automatically assign with each short text message a set of categories based on its content. The major efforts in building a robust short text classifier are concentrated in the extraction and selection of a set of characterizing and discriminate features. The solutions investigated in this paper are an extension of those adopted in a previous work by us from which we inherit the learning model and the elicitation procedure for generating pre-classified data.

IV. SYSTEM ARCHITECTURE



V. METHODOLOGIES

1. Filtering rules

In defining the language for FRs specification, we consider three main issues that, in our opinion, should affect a message filtering decision. First of all, in OSNs like in everyday life, the same message may have different meanings and relevance based on who writes it. As a consequence, FRs should allow users to state constraints on message creators. Creators on which a FR applies can be selected on the basis of several different criteria; one of the most relevant is by imposing conditions on their profile's attributes. In such a way it is, for instance, possible to define rules applying only to young creators or to creators with a given religious/political view. Given the social network scenario, creators may also be identified by exploiting information on their social graph. This implies to state conditions on type, depth and trust values of the relationship(s) creators should be involved in order to apply them the specified rules. All these options are formalized by the notion of creator specification,

2. Online setup assistant for FRs thresholds:

As mentioned in the previous section, we address the problem of setting thresholds to filter rules, by conceiving and implementing within FW, an Online Setup Assistant (OSA) procedure. OSA presents the user with a set of messages selected from the dataset discussed in Section VI-A. For each message, the user tells the system the decision to accept or reject the message. The collection and processing of user decisions on an adequate set of messages distributed over all the classes allows to compute customized thresholds representing the user attitude in accepting or rejecting certain contents. Such messages are selected according to the following process. A certain amount of non neutral messages taken from a fraction of the dataset and not belonging to the training/test sets, are classified by the ML in order to have, for each message, the second level class membership values.

3. Blacklists:

A further component of our system is a BL mechanism to avoid messages from undesired creators, independent from their contents. BLs are directly managed by the system, which should be able to determine who are the users to be inserted in the BL and decide when users retention in the BL is finished. To enhance flexibility, such information are given to the system through a set of rules, hereafter called BL rules. Such rules are not defined by the SNM, therefore they are not meant as general high level directives to be applied to the whole community. Rather, we decide to let the users themselves, i.e., the wall's owners to specify BL rules regulating who has to be banned from their walls and for how long. Therefore, a user might be banned from a wall, by, at the same time, being able to post in other walls. Similar to FRs, our BL rules make the wall owner able to identify users to be blocked according to their profiles as well as their relationships in the OSN. Therefore, by means of a BL rule, wall owners are for example able to ban from their walls users they do not directly know (i.e., with which they have only indirect relationships), or users that are friend of a given person as they may have a bad opinion of this person. This banning can be adopted for an undetermined time period or for a specific time window. Moreover, banning criteria may also take into account users' behavior in the OSN. More precisely, among possible information denoting users' bad behavior we have focused on two main measures. The first is related to the principle that if within a given time interval a user has been inserted into a BL for several times, say greater than a given threshold, he/she might deserve to stay in the BL for another while, as his/her behavior is not improved. This principle works for those users that have been already inserted in the considered BL at least one time. In contrast, to catch new bad behaviors, we use the Relative Frequency (RF) that let the system be able to detect those users whose messages continue to fail the FRs. The two measures can be computed either locally, that is, by considering only the messages and/or the BL of the user specifying the BL rule or globally, that is, by considering all OSN users walls and/or BLs.

VI. CONCLUSION AND FUTURE WORK

In this article, we have presented a system to filter blackballed messages from OSN walls. The system abuse a ML soft classifier to carry out customizable content-clinging FRs. Likewise, the extensibility of the system in terms of filtering options is magnify through the mainframe of BLs. This work is the first step of a wider project. The early encouraging results we have obtained on the allotment procedure prompt us to continue with other work that will aim to improve the quality of classification. In future work, we plan to address this problem by investigating the use of online learning paradigms able to include label feedbacks from users. Additionally, we plan to enhance our system with a more sophisticated approach to decide when a user should be inserted into a BL.

VII. REFERENCES

[1] 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.

[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," Proc. Fifth ACM Conf. Digital Libraries, pp. 195-204, 2000.

[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," Proc. ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML '10), 2010.

[6] N.J. Belkin and W.B. Croft, "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" Comm. ACM, vol. 35, no. 12, pp. 29-38, 1992.