International Journal of
Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.03,Issue.23
September-2014,
Pages:4719-4722

www.semargroup.org,
www.ijsetr.com

Removal of Irrelevant Communication using Machine Learning Based Classification in Social Network

CH. SWATHIKEERTHANA[1], N. KESAVA RAO[2]

[1]PG Scholar, Dept of CSE, Narayana Engineering College, Gudur, Andhrapradesh, India.
[2]Asst Prof, Dept of CSE, Narayana Engineering College, Gudur, Andhrapradesh, India.

Abstract: On primary problem in today Online Social Network (OSNs) is to give users the capability to manage the messages posted on their own confidential space to avoid that redundant content is displayed. Up to now OSNs make available a tiny support in this requirement. In this paper, we suggest a system allowing OSN users to have straight control on the messages posted on their walls. This is achieved through a flexible rule based system, that allows users to convert the filtering criteria to be useful to their walls, and machine learning based soft classifier repeatedly labeling messages in support of content –based filtering.

Keywords: Online Social Network, Machine Learning Based Classification, Message Filtering, Machine Learning.

A. INTRODUCTION

They are on-line social networks (OSNs) are nowadays one of the mainly trendy interactive medium to communicate, share and disseminate a significant amount of human life inform-action. On a daily basis and permanent communications involve the exchange of numerous types of content, together with free text, image, and audio and video data. According to facebook statistics1 average user creates 90 pieces of content each month, but more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) are shared each month. The massive and dynamic nature of these data creates the basis for the employment of web content mining strategies meant to automatically determine valuable information dormant within the data. Instrumental to offer an active support in complex and difficult tasks involved in OSN management, such as for instance access control or information filtering. Information filtering has been very much explored for what concerns textual documents and, more recently, web content (e.g., [1], [2], [3]). The intend of the best part of these proposals is primarily to provide users a classification mechanism to avoid they are overwhelmed by useless data. In OSNs, information filtering can also be used for a special, more sensitive. This is due to the reality that in OSNs there is a chance of relocation or commenting other posts on particular public/private areas, called in general walls.

A. Existing System

We consider that this is a key OSNs service that has not been provided so far. Certainly, today OSNs provide extremely small support to avoid unwanted messages on user walls. For instance, face book allows users to state who is permitted to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends). However, no content-based preferences are supported and as a result it is not possible to avoid undesired messages, such as political or vulgar ones, no theme of the user who posts them. Providing this service is not only a matter of using formerly defined web content mining techniques for a dissimilar application, rather it requires to design ad-hoc classification strategies. This is because wall messages are constituted by short text for which conventional classification methods have serious restrictions since short texts do not provide enough word occurrences.

B. Proposed System

The intend of the current work is therefore to suggest and experimentally assess an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSNs user walls. We utilize Machine Learning (ML) text categorization techniques [4] to automatically allocate with each short text message a set of categories based on its content. The main efforts in building a strong short text classifier are determined in the extraction and selection of a set of characterizing and distinguish features. The solutions investigated in this paper are an extension of those adopted in a earlier work by us. From which we take over the learning model and the elicitation process for generating pre-classified data. The original set of features, resulting from endogenous properties of short texts, is enlarged here including exogenous knowledge associated to the context from which the messages begin. As far as the learning model is concerned, we confirm in the current paper the use of neural learning which is nowadays recognized as one of the mainly efficient solutions in text classification. In

particular, we support the overall short text classification strategy on Radial Basis Function Networks (RBFNs) for their verified capabilities in acting as soft classifiers, in managing noisy data and intrinsically vague classes. Moreover, the speed 2 in performing the learning phase creates the basis for an adequate use in OSNs domains, as well as facilitates the tentative evaluation tasks.

II. FILTERED WALL ARCHITECTURE

The architecture in support of OSNs services is a three-tier structure shown in the figure 1. The first layer, called Social Network Manager (SNM), commonly aims to provide the basic OSNs functionalities (i.e., profile and relationship management), whereas the second layer provides the support for external Social Network Applications (SNAs). The supported SNAs may in turn need an additional layer for their desired graphical user interfaces (GUIs). According to this reference architecture, the projected system is located in the second and third layers. In particular, users work together with the system by means of a GUI to set up and manage their FRS/BLS. The core components of the projected system are the Content-Based Messages Filtering (CBMF) and the Short Text Classifier (STC) modules. The final component aims to categorize messages according to a set of categories.

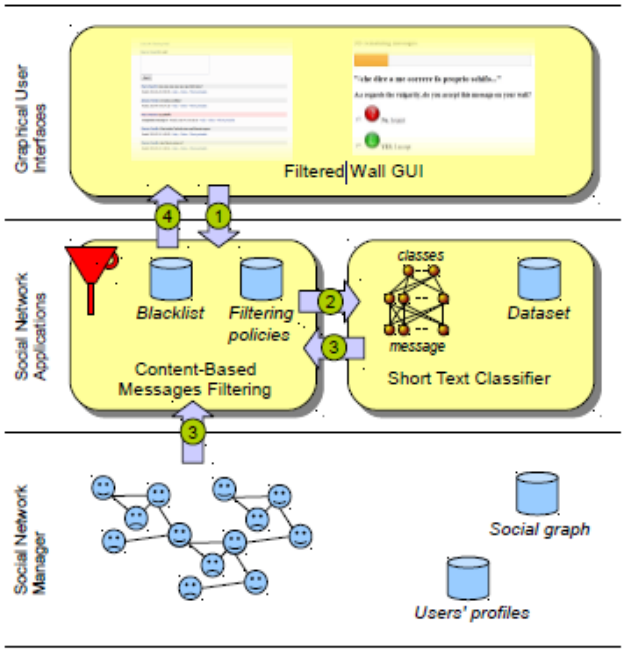

Figure 1 Filtered wall architecture.

As graphically depicted in figure 1, the path followed by a message, from its writing to the feasible final publication can be summarized as follows:

1. Behind entering the private wall of one of his/her contacts, the user tries to post a message, which is interrupted by FW.
2. A ML-based text classifier extracts metadata from the content of the message.
3. FW uses metadata provided by the classifier, jointly with data extracted from the social graph and users' profiles, to put into effect the filtering and BL rules.
4. Depending on the result of the earlier step, the message will be published or filtered by FW.

III. SHORT TEXT CLASSIFIER

Our lessons are intended at designing and evaluating various representation techniques in grouping with a neural learning strategy to semantically classify short texts. From a ML point of view, we approach the task by defining a hierarchical two level strategy assuming that it is improved to recognize and get rid of "NEUTRAL" sentences, then classify "NON NEUTRAL" sentences by the class of interest instead of doing everything in one step. This alternative is motivated by associated work showing advantages in classifying text and/or short texts using a Hierarchical strategy [1]. The first level task is conceived as a hard classification in which short texts are labeled with crisp neutral and non-neutral labels. The second level soft classifier acts on the crisp set of non-neutral short texts and, for each of them, it simply produces expected suitability or "GRADUAL MEMBERSHIP" for each of the conceived classes, without taking any "HARD" decision on any of them.

A. Text Representation

The extraction of a suitable set of features by which on behalf of the text of a given document is a crucial task strongly affecting the performance of the on the whole classification strategy. The most suitable feature set and feature representation for short text messages have not yet been adequately investigated. We consider three types of features, Document properties (DP) and Contextual Features (CF). The first two types of features, already used in [5], are endogenous, they are entirely resulting from the information contained within the text of the message. Text representation using endogenous knowledge has a good general applicability, however in prepared settings it is valid to use also exogenous knowledge.

B. Machine Learning Based Classification

We deal with short text categorization as a hierarchical two-level classification process. The first-level classifier performs a binary hard categorization that labels messages as neutral and non-neutral. The first-level filtering task facilitates the following second-level task in which a finer-grained classification is performed. The second-level classifier performs a soft-partition of non-neutral messages transmission a given message a continuing membership to each of the non neutral classes. Among the range of multi-class ML models well-suited for text categorization, we choose the RBFNs model for the experimented competitive behavior with esteem to other state of the art classifiers. RBFNs have a single hidden layer of processing units with local, restricted activation domain: a GAUSSIAN function is usually used, but any other nearby tunable function can be used. RBFNs major advantages are that classification function is non-linear, the model may create confidence values and it may be robust to outliers drawbacks are the potential sensitivity to input parameters, and potential overtraining sensitivity.

The first level classifier is then prepared as regular RBFNs. In the second level of the classification stage we establish a alteration of the standard use of RBFN. Its usual use in classification includes a hard decision on the output values: according to the winner-take-all rule, a given input pattern is assigned with the class corresponding to the winner output neuron which has the highest value. To work well, a ML-based classifier needs to be trained with a set of adequately complete and reliable pre-classified data. The complexity of satisfying this constraint is fundamentally related to the subjective character of the interpretation process with which an specialist decides whether to classify a document under a given category.

IV. FILTERING RULES AND BLACKLIST MANAGEMENT

We model a social network as a directed graph, where each node corresponds to a network user and edges indicate relationships between two different users. In particular each edge is labeled by the type of the recognized relationship (e.g., friend of, colleague of, parent of) and probably, the corresponding trust level, which represents how much a given user considers trustworthy with respect to that specific kind of relationship the user with whom he/she is establishing the relationship. Without loss of generality, we assume that trust levels are rational numbers in the range [0, 1].

A. Filtering Rules

In defining the language for FRS specification, we consider three main issues should influence a message filtering decision. First of all, in OSNs like in daily life, the same message may have different meanings and relevance based on who writes. It as a significance, FRS should allow users to state constraints on message creators. Creators on which a FRS applies can be selected on the basis of several unusual criteria, one of the most related is by impressive conditions on their profile's attributes. A Filtering Rule FR is a Tuple (author, creator spec, content spec, action), where:

1. Author is the user who specifies the rule.
2. Creator spec is a creator specification, specified according to definition 1.
3. Content spec is a boolean expression defined on content constraints of the form (c; ml), where c is a class of the first or second level and ml is the minimum membership level threshold necessary for class c to make the constraint satisfied.
4. Action 2 f block, notify denotes the action to be performed by the system on the messages matching content spec and formed by users identified by creator spec.

B. Online Assistant for FRs Thresholds

We deal with the problem of setting thresholds to filter rules, by conceiving and implementing within FW, an Online Setup Assistant (OSA) procedure. For each message, the user tells the system the decision to accept or reject the message. The collection and processing of decisions adequate of messages spread over all the classes allows computing customized thresholds representing the user attitude in accepting or rejecting certain contents. Such messages are chosen according to the following process. A definite 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.

C. Blacklist

BL mechanism to avoid messages from undesired creators, independent from their contents. BLs is openly managed by the system, which should be able to determine who are the users to be inserted in the BL and make a decision when user's preservation in the BL is finished. To improve flexibility, such information is given to the system through a set of rules, here after called BL rules. Similar to FRs, our BL rules make the wall owner able to recognize users to be blocked according to their profiles as well as their relationships in the OSNs.

V. DISCUSSION

Overall estimation of how efficiently the system applies a FR, we look once more at table ii. this table allows us to guess the precision and recall of our FRs, since values reported in table ii have been computed for FRS with content specification component set to (*c, 0.5*), where *c 2 Ω*. Results achieved by the content-based specification component, on the first level classification, can be measured good enough and practically aligned with those obtained by well-known information filtering techniques. Results obtained for the content-based specification component on the second level are somewhat less brilliant than those obtained for the first, but we should understand this in view of the intrinsic difficulties in assigning to a message a semantically most specific category.

VI. CONCLUSION

The system exploits a ML soft classifier to impose customizable content-dependent FR furthermore, the flexibility of the system in terms of filtering options is improved through the management of BLs. The development of a GUI and a set of associated tools to make easier BL and FR specification is also a direction we plan to examine, since usability is a key requirement for such kind of applications.

VII. 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.

Author's Profile:

CH Swathikeerthana has received her B.Tech in Information Technology from Narayana Engineering College (NECG), Gudur affiliated JNTU, Anantapur in 2012 and Pursuing M.Tech degree in Computer science at Narayana Engineering College (NECG), Gudur, AP, India.

N Kesava Rao has received his B.Tech in Computer science and Engineering and M.Tech degree in Computer science and Engineering from KLCE, vaddeswaram, Nagarjuna University in 1995 and Sathyabama University in 2008 respectively. He is dedicated to teaching field from the last 14 years. He has guided 4 P.G and 32 U.G students. At present he is working as Assistant Professor in Narayana Engineering College (NECG), Gudur, AP, India.