Available online at www.worldscientificnews.com

World Scientific News

WSN 40 (2016) 163-174 EISSN 2392-2192

Survey on Message Filtering Techniques for On-line Social Network

Dr. M. Newlin Rajkumar[1,a], P. Kayathri[2,b], V. Dhurka[2,c]

[1]Assistant Professor, Department of CSE, Anna University Regional Centre, Tamil Nadu, India

[2]PG Scholar, Department of CSE, Anna University Regional Centre, Tamil Nadu, India

[a-c]E-mail address: newlin_rajkumar@yahoo.co.in , kayathri44@gmail.com , dhurkav@gmail.com

ABSTRACT

A social network is a set of people or organizations or other social entities connected by set of social relationships such as friendship, co-working or information exchange. Online Social Networks (OSN) usually not support to the user for message filtering. To solve this issue, which allows OSN users to have a direct control on the messages posted on their walls. The users can control the unwanted messages posted on their own private space .To avoid unwanted messages displayed and they can also block their friend from friends list using filtering rule, content based filtering and short text classification.

Keywords: On-line Social Network; filtering rule; text classification; blacklist

1. INTRODUCTION

Online Social Network is most popular and interesting interactive medium to communicate, share and distribute information among the users. It's used to share the several types of content, that including free text, image, audio and video information. A Social network service consists of each user having her own profile, social link and variety of additional services. In OSNs, information filtering used for a dissimilar, more responsive, purpose. Because, OSNs is the possibility of posting or viewing other posts on public/private regions in common walls. Information filtering is used to customize their wall and it provides the capability to the user for automatically control the messages written on their individual walls, by filtering out unwanted communication. This is not normally present in OSN network. Now, at present OSNs provide very tiny maintenance to prevent unwanted messages on user walls. For an example, in a social network Face book permits the users to edit who is allowed to insert messages in their walls (i.e., friends, mutual friends, defined groups of friends or friends of friends).

In this paper, our main aim is to analyze the classification technique and to design the system to filter the undesirable messages from OSN user wall. Filtered Wall (FW) should be able to filter unwanted messages from OSN user walls. Machine Learning (ML) text categorization techniques are evolved to automatically assign with each short text message based on its content by using a set of categories. In addition, the system will use a flexible language to demonstrate the filtering rules (FRs), with the help the users can decide what contents should be displayed on their walls. The FRs can be personalized according to the users need. Along with it there are user defined blacklists (BLs) which will temporary allows the users to post any type of message on user walls.

2. FILTERED WALL ARCHITECTURE

The architecture of OSN services is a three-tier structure. It consist of three layers

1. Social Network Manager (SNM)
2. Social Network Application (SNA)
3. Graphical User Interface (GUI)

Social Network Manager: It maintains the basic OSN related functionalities (i.e.) user profiles and relationship management and also provide the data to the social network application layer.

Social Network Application: To provide external social network application. That layer applying filtering rules (FR) and blacklists (BL). SNA layer composed of short text classifier and Content Base Message Filtering (CBMF). It is most important layer. Because the classifier categorizes each message according to its content and CBMF filters the message according to filtering criteria and blacklist provided by the user. CBMF is used to select information item based on the correlation between the content of item and the user preferences.

Graphical user interface: It provide the interaction between users and system. Which user provide her input and is able to see published wall messages. Moreover GUI also provides user the facility to apply filtering rules for his wall messages and helps to provide list of BL user who are temporally prevented to publish messages on user’s wall. The GUI also consists of Filtered Wall (FW) where the user is able to see her desirable messages.

1. After entering the private wall of one of users, the user tries to post a message captured by Filtered wall
2. A ML-based text classifier extracts data from the message content.
3. Filtered wall uses data given by the classifier, along with data extracted from the user's profiles, to implement the filtering rules and blacklists techniques.
4. Considering the result of the previous step, message will be filtered.

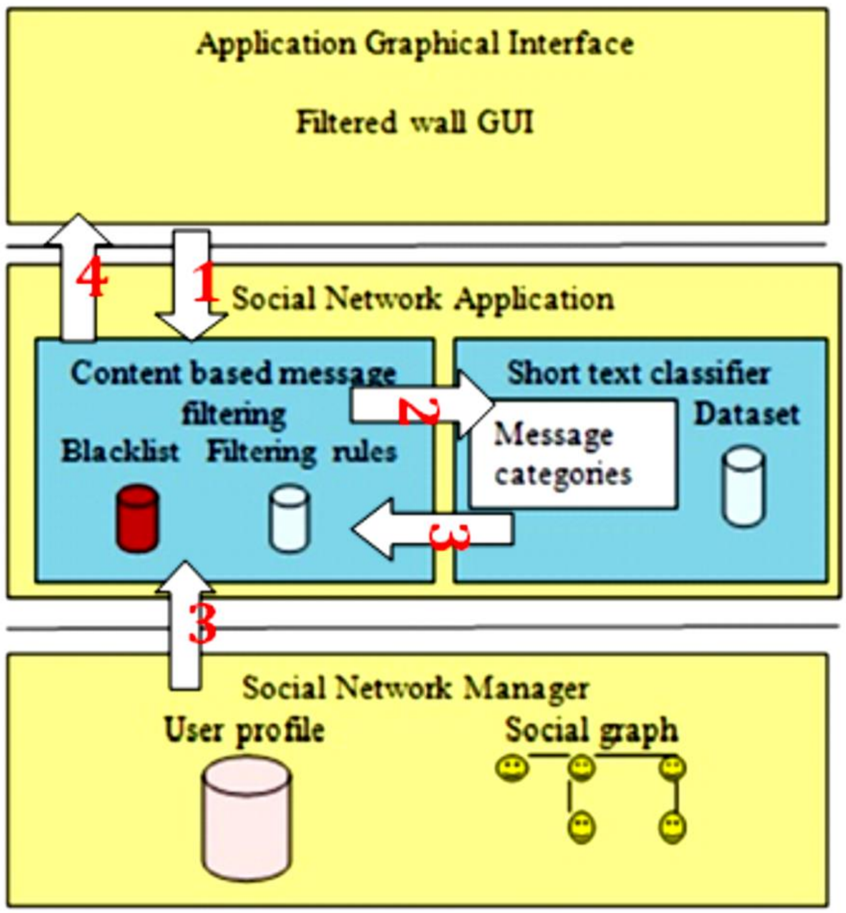

Figure 1. OSN Architecture.

3. TEXT CLASSIFICATION TECHNIQUES

Text Classification is the task to classify documents into predefined classes. Text Classification is also called text categorization, document classification, and document categorization. There are two approaches for classification manual classification and automatic classification. Relevant technologies for Text Classification are:

1. Text Clustering
2. Information Retrieval (IR)

3. Information Filtering
4. Information Extraction (IE)
5. Text Classification.

Text clustering is the application that Create clusters of documents without any external information. Information Retrieval (IR) used to retrieve a set of documents according to the relevant query. Information Filtering is used to filter out irrelevant documents through interactions. Information Extraction (IE) extracts fragments of information, e.g., person names, dates, and places, in documents. In text Classification there is no query, no interactions, no external information, it only decide topics of documents.

Text Classification used in various fields like E-mail spam filtering, categorize newspaper articles and newswires into topics, organize Web pages into hierarchical categories, sort journals and abstracts by subject categories (e.g., MEDLINE, etc.), assigning international clinical codes to patient clinical records etc. Text categorization is the activity of labeling natural language texts with the sematic categories from a predefined set. There are different types feature set and feature representation that are Bag of Words (BoW), Document properties (DP) and Contextual Features (CF).

3. 1. BAG-of-Words Model (BOW)

The bag-of-words model is a simplifying representation used in natural language processing and information retrieval (IR). In this model, a text (such as a sentence or a document) is represented as the bag (multi set) of its words, disregarding grammar and even word order but keeping multiplicity.

Recently, the bag-of-words model has also been used for computer vision. The bag-of-words model is commonly used in methods of document classification, where the (frequency of) occurrence of each word is used as a feature for training a classifier

To represent the image using BOW model, an image can be treated as a document. Similarly "words" in image needed to be defined too. To achieve this, it usually includes the following three steps: feature detection, feature description and codebook generation.

3. 2. Documents Properties (DP)

Document properties as a uniform basis for interaction. Document properties express high-level features of documents that are meaningful to users and usable by systems. Document properties are directly associated with documents, rather than with document storage locations. This means that documents will retain properties even when moved from one place to another, and that property assignment can have a fine granularity.

There are different types of text representation. There are given below

- Good words: if text is not useful and hurt others is called good words.
- Bad words: are computed similarly to the Correct words feature, whereas the set Is a collection of "dirty words" for the domain language?
- Capital words: express the amount of words mostly written with capital letters, Calculated as the percentage of words within the message, having more than half of the characters in capital case. For example the value of the feature for the document "To be OR Not to BE" is 0:5 since the words "OR" "Not" and "BE" are considered as capitalized ("To"

is not uppercase since the number of capital characters should be strictly greater than the characters count).

- Punctuations characters: calculated as the percentage of the punctuation characters over the total number of characters in the message. For example the value of the feature for the document “Hello!!! How’re u doing?” is 5 = 24.
- Exclamation marks: calculated as the percentage of exclamation marks over the Total number of punctuation characters in the message. Referring to the aforementioned document the feature value is 3 = 5.
- Question marks: calculated as the percentage of question marks over the total number of punctuations characters in the message. Referring to the aforementioned document the feature value is 1 = 5.

3. 3. Contextual Features (CF)

A contextualized strategy might allow IR systems to learn and predict what present results relating them to previous information and to the tasks the user has been engaged in and decide who else should get the new information.

3. 4. Vector Space Model *(*VSM*)*

The Vector space model or term vector model is an algebraic model for representing text documents (and any objects, in general) as vectors of identifiers, such as, for example, index terms. It is used in information filtering, information retrieval, indexing and relevancy rankings. The VSM representation of the text that characterizes the environment where messages are posted (topics of the discussion, name of the group or any other relevant text surrounding the messages).

Documental and queries are represented as vectors.

$$d_j = (w_{1,j}, w_{2,j}, \ldots\ldots, w_{t,j})$$

$$q = (w_{1,q}, w_{2,q} \ldots\ldots\ldots, w_{n,q})$$

Each dimension corresponds to a separate term. If a term occurs in the document its values is non zero. Typically the terms are single words, keywords or longer phrases.

If the words are chosen to be the term, the dimensionality of the vector is the number of words in the vocabulary. Vector operation can be used to compare documents and queries.

Advantages of this model is term weights not binary, allows computing a continuous degree of similarity between queries and documents, and allows partial matching.

Advantages of vector space model

- High dimensional input space
- Irrelevant features
- Document vectors are Sparse
- Text classification mostly linearly separable

3. 5. Machine Learning-Based Classification

There are two approaches that you can take that is rule based approach and machine learning-based approach. Rule based approach write a set of rules that classify documents. Rule based classification in which the anticipated request from users influencing how the documents are being classified. Machine learning-based approach using a set of sample documents that are classified into the classes (training data) automatically create classifiers based on the training data.

Machine learning algorithm is three types. There are,

- Supervised learning
- Unsupervised learning
- Semi- supervised learning

Machine Learning is another way of getting computers to classify documents. Machine learning is normally not rule based. Instead, it is normally statistically based. It's the ability of a machine to improve its performance based on previous results so, machine learning document classification is "the ability of a machine to improve its document classification performance based on previous results of document classification. Examples of ML algorithms.

3. 5. 1. Naïve Bayes

This method computes the probability that a document is about a particular topic, T, using a) the words of the document to be classified and b) the estimated probability of each of these words as they appeared in the set of training documents for the topic, T – like the example previously given.

Navie bayes classifier are a family of simple probabilistic classifiers based on applying bayes's theorem with strong (naive) independence assumptions between the features. Naïve Bayesian classification is used for anti-spam filtering technique. It has two different phases. The first phase has been applied for training set of data and the second phase employs the classification phase. Bayesian filter can also be used for classification of text.

A Bayesian network is a probabilistic graphic model that represent a set of random variables and their conditional independencies via directed acyclic group.

3. 5. 2. Neural Networks

During training, a neural network looks at the patterns of features (e.g. words, phrases, or N-grams) that appear in a document of the training set and attempts to produce classifications for the document. If its attempt doesn't match the set of desired classifications, it adjusts the weights of the connections between neurons. It repeats this process until the attempted classifications match the desired classifications.

It is an interconnected group of nodes, aki to the vast network of neurons in abrain.

3. 5. 3. Instance Based

The Saves documents of the training set and compares new documents to be classified with the saved documents. The document to be classified gets tagged with the highest scoring classifications. One way to do this is to implement a search engine using the documents of the training set as the document collection. A document to be classified becomes a query/search. A classification, C, is picked if a large number of its training set documents are at the top of the returned answer set. There are varieties of multi-class ML models well-suited for text classification. Whereas RBFN model is the best model.

3. 5. 4. Radial Basis Function Network (RBFN)

The In the field of mathematical modeling, a RBFN is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks have many uses, including function approximation, time series prediction, classification, and system control. RBFN main advantages are that classification function is non-linear, the model may produce confidence values and it may be robust to outliers; drawbacks of RBFN are the potential sensitivity to input parameters, and potential overtraining sensitivity.

3. 5. 5. Cluster analysis

Cluster analysis or clustering is the task of grouping a set of subjects in such a way that object in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in the other groups (cluster). Types of cluster analysis,

- Connectivity models
- Centroid models
- Distribution models
- Density models
- Subspace model
- Group model
- Graph based model

Clustering is the process of division of vertices into groups, with a higher density of edges within groups than between them. Clustering is an important task for the discovery of community structures in networks. Its goal is to sort cases (people, things, events, etc.) into clusters so that the degree of association is relatively strong between members of the same cluster and relatively weak between members of different clusters.

3. 5. 6. Decision tree

Decision tree learning uses a decision tree as a predictive model which maps observation about an item to conclusions about an item's target value. Tree models where the target variable can take a finite set of values are called classification trees. In these trees structures leaves represent class labels and branches represent conjunctions of features that leads to those class labels.

Association rule learning is the method for discovering interesting relation between variables in large database. There are several different kinds of splits in the decision trees are available. The listed splits are

- Single attribute split
- Similarity-based multi-attribute split

- Dimensional- based multi-attribute split

A social structure made of individuals (or organizations) who are connected, By one or more specific types of relationship i.e friendship, kinship, shared interest, business relationship.

3. 6. Content based filtering

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.

Content based filtering is based on a description of the item and profile of the user's preference. In this filtering keywords used to describe the items and a user profile is built to indicate the type of items user likes.

3. 7. Collaborative filtering

Collaborative filtering methods are based on collecting and analyzing a large amount of information on user's behaviors, activities or preferences and predicting what users will like based on their similarity to other users.

When building a model from a user's behavior, a distinction is often made between explicit and implicit forms of data collection.

Examples of explicit data collection include the following:

- Asking the user to rate an item on a sliding scale.
- Asking a user to search.
- Asking a use to rank a collection of items from favorite to least favorite.
- Presenting two items to a user and asking him/her to choose the better one of them.
- Asking a use to create a list of items that he/she likes.

Examples of implicit data collection include the following:

- Observing the items that a user views in online store.
- Analyzing items/users viewing items
- Keeping a record of the items that a user purchases online.
- Analyzing the user's social network and discovering similar likes and dislikes.

Collaborative filtering approaches often suffer from three problems: cold start, scalability and sparsity.

- Cold start: The systems often require a large amount of existing data on a user in order to make accurate recommendation.
- Scalability: In many of the environments in which these systems make recommendations that are millions of users and products.
- Sparsity: The number of items sold on major e-commerce sites is extremely large. The most active users will only have rated a small subset of the overall database.

Collaborative filtering are classified as memory-based and model based Collaborative filtering. A well know example of memory-based approaches is user-based algorithm and that model based approaches is kernel mapping recommendation.

3. 8. Spam filtering

Social spam is unwanted spam content appearing on social network and any website with user-generated content (comments, chats, etc.). It can be manifested in many ways, including bulk message, profanity, insults, hate speech, malicious links, fake friends and personally identifiable information. These spammers can utilize the social network's search tool to target certain demographic segments, or use common fan pages or group pages to send notes from fraudulent accounts. Many social networks have include a "report spam/abuse" button or address from one throw-away to another.

Bulk submission are a set of comments repeated multiple times with the same or very similar text. These message also called as spam-bombs. User-submitted comments that contain swear or slurs are classified as profanity. User-submitted insults are comments that contain mildly or strongly insulating language against a specific person or person.

Threat is user-submitted threats of violence are comments that contain mid or strong threats of physical violence against a persons or group. Fraudulent reviews are reviews of a product or service from users that never actually used it and therefore insincere or misleading. Fake friends occurs when several fake accounts connect or become "friends". Person identifiable information is user-submitted comments that inappropriately display full names, physical address or credit card numbers are considered leaks of Person identifiable information.

3. 9. Policy based personalization filtering

Policy based personalization is useful in various context. Policy defines the rules according to which authorization is regulated. System specified policies, with respect to who defines the policies. System-specified policies (SP) are system wide general rules enforced by the OSN system; while user-specified policies are applied to specific users and resources.

OSN trust relationships and provenance information to personalize access to the website. 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.

4. FILTERING RULES AND BLACKLIST

4. 1. Filtering Rules

User can state what contents should be blocked or displayed on filtered wall by means of Filtering rules. Filtering rules are specified on the basis of user profile as well as user social relationship.

FR {Author, creator Spec, content Spec, action}

FR is dependent on following factors

1. Author
2. Creator Spec

3. Content Spec

4. Action

- Author is a person who defines the rules.
- Creator Spec denotes the set of OSN user
- Content Spec is a Boolean expression defined on content
- Action denotes the action to be performed by the system on the messages matching content Spec and created by users identified by creator Spec.

4. 2. Online Setup Assistant for FRS Thresholds

There are problem of setting thresholds to filter rule. Online Setup Assistant (OSA) procedure allows user to select a set of messages from dataset of messages. Such messages are selected by 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 value. On the basis of this the user tells the system the decision to accept or reject the message. The collection and processing of user decisions on an sufficient set of messages circulated over all the classes allows to compute customized thresholds representing the user attitude in accepting or rejecting certain contents.

4. 3. BLACKLIST

BL helps to the users whose messages are prevented independent from their contents. BL rules enable the wall owner to determine users to be blocked on the basis of their profiles and relationship with wall owner. Like BL is also dependent on author, creator specification and creator behavior.
BL tuple {Author, creator Spec, creator Behaviour, T}

- Author is the OSN user who specifies the rule.
- Creator Spec denotes the set of OSN users to which rule applies.
- Creator Behaviour = RF blocked V minBlanned
    - RFBlocked = (RF, mode, window)

  *RF = #bMessages/#tMessages, where #tMessages is the total number of messages that each OSN user identified by creatorSpec has tried to publish in the author wall (i.e., mode = myWall) or in all the OSN walls (i.e., mode = SN); whereas #bMessages is the number of messages among those in#tMessages that have been blocked.

  *mode 2 fmyWall; SNg specifies if the messages to be considered for the RF Computation have to be gathered from the author's wall only (i.e., mode = myWall) or from the whole community walls (i.e., if mode = SN).

  *window is the time interval of creation of those messages that have to be considered for RF computation; minBanned = (min, mode, window) is defined such that min is the minimum number of times in the time interval specified in window that OSN users identified by creatorSpec have to be inserted into the BL due to BL rules specified by author wall (i.e., mode = me) or other OSN users (i.e., mode = SN) in order to satisfy the constraint.
- T denotes frequency of occurrence between the creator Spec and creator Behaviour

For denoting users' bad behavior considered two main measures.

1. If any user has already into BL and again inserted many times in it for given time interval, say greater than a given threshold, he/she might deserve to stay in the BL.
2. 2. Relative Frequency (RF), it detect those users whose messages continue to fail the FR.

5. CONCLUSION

Online social networking applications are mostly used by all the people in and around the world. Most of the time of a day is spent in online social networking applications. However, the users of online social networks are unaware of the security issues do exist in OSN platform. A system to filter unwanted messages or posts from Online Social Network walls using the above techniques. The future direction of the research will be modelling effective security algorithms to defend the security issues exist in online social networks.

References

[1] 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.

[2] R. Prashant Tomer, "On Line Social Network Content and Image Filtering Classifications," ISSN 2319-5991 Vol. 2, No. 4, © 2013 IJERST. November 2013.

[3] M. Carullo, E. Binaghi, I. Gallo, and N. Lamberti, "Clustering of short commercial documents for the web," in Proceedings of 19th International Conference on Pattern Recognition (ICPR 2008), 2008.

[4] F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, Vol. 34, no. 1, pp. 1-47, 2002.

[5] A. Adomavicius, G.and Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-theart and possible extensions," IEEE Transaction on Knowledge and Data Engineering, Vol. 17, no. 6, pp. 734-749, 2005.

[6] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li, "Rcv1: A new benchmark collection for text categorization research," *Journal of Machine Learning Research*, 2004.

[7] Bonchi and E. Ferrari, Privacy-aware Knowledge Discovery: Novel Applications and New Techniques. *Chapman and Hall/CRC Press*, 2010.

[8] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, Short text classification in twitter to improve information filtering, in Proceeding of the 33rd *International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR 2010, 2010, pp. 841-842.

[9] DIK L. LEE "Document Ranking and the Vector-Space Model" Hong Kong University of Science and Technology HUEI CHUANG, Information Dimensions KENT SEAMONS, Transarc.

[10] Carla Teixeira Lopes "Context Features and their use in Information Retrieval" Doctoral Program in Informatics Engineering of Faculdade de Engenharia da Universidade, 2007.

[11] 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.

[12] S. Dumais, J. Platt, D. Heckerman, and M. Sahami, “Inductive learning algorithms and representations for text categorization,” in Proceedings of Seventh International Conference on Information and Knowledge Management (CIKM98), 1998, pp. 148-155.

[13] D. D. Lewis, “An evaluation of phrasal and clustered representations on a text categorization task,” in Proceedings of 15th ACM International Conference on Research and Development in Information Retrieval (SIGIR-92), N. J.

[14] Belkin, P. Ingwersen, and A. M. Pejtersen, Eds. ACM Press, New York, US, 1992, pp. 37-50.

[15] R. E. Schapire and Y. Singer, “Boostexter: a boosting-based system for text categorization,” *Machine Learning,* Vol.39, no. 2/3, pp. 135-168, 2000.

[16] E. D. Wiener, J. O. Pedersen, and A. S. Weigend, “A neural network approach to topic spotting,” in Proceedings of the Annual Symposium on Document Analysis and Information Retrieval (SDAIR-95), Las Vegas, US, 1995, pp. 317-332.

[17] S. Zelikovitz and H. Hirsh, “Improving short text classification using unlabeled background knowledge,” in Proceedings of 17th International Conference on Machine Learning (ICML-00), P. Langley, Ed. Stanford, US: Morgan Kaufmann Publishers, San Francisco, US, 2000, pp. 1183-1190.

[18] www.en.m.wikipedia.org.

[19] T. M. Mitchell. Machine Learning. WCB/McGraw-Hill, 1997.

[20] Y. Li, A. Jain. Classification of text documents. *The Computer Journal,* 41(8), pp. 537-546, 1998.

(Received 24 January 2016; accepted 07 February 2016)