

Information Filtering In Automated Text Categorization

Mrs. G. Shoba, A. Amrita, K. Arthi, K. Saranya

Senior Assistant Professor, CSE, Christ College of Engg&Tech, Pondicherry, India

Student, CSE, Christ College of Engg&Tech, Pondicherry, India

ABSTRACT

Online Social Networks (OSNs) are one of the most popular interactive ways to communicate and share information with others. OSN is developing rapidly in the modern world which engages almost of all age's people to be convenient with it. OSN permit any sort of text or content to be posted on user's wall. In order to avoid the unwanted content from being posted on user wall, OSN provide Filtering Rule based system. Machine Learning system is customized for automated filter by means of content- based filtering. This system may remove or filter all such content based on text that has been posted. Since this method will result in automated filtering, the content that has been posted is unaware to the user. Instead of making it completely removed from the user point of view, we facilitate such text to be displayed on user notification. Only the user can view that content and will be aware about the post and can decide whether it can be posed or not and can even simply leave as such.

Keywords: Automated filter, Filtering Rule, Online Social Networks, Short Text Classifier.

I. INTRODUCTION

In the modern world, with the rapid growth of Internet, people start using Internet for various purposes. Initially people used Internet only for Business and commercial purposes. But now Internet has become an indispensable instrument for all sorts of activities. People facilitate themselves by bringing the entire world at the touch of their fingers. Internet provides users to know about all latest news, updates, gossips, invention, shopping, booking, etc. Nowadays Internet even makes people lazy to such an extent that people use to buy home appliances from home via online shopping.

Internet becomes the second home or "home in home" for most of the users. People were used to communicate with others by means of Internet. Internet permits some of the Online Social Networks to communicate and share information among people. If Internet is considered to be a "Home" for the user, then obviously OSN act as a "family" for the user. Relationships between people are widely strengthened by means of this OSN.

OSN is a platform to build social networks or social relationship between people who share interests, activities, backgrounds, or real-life connections. A social network service consists of a representation of each user (often a profile), his/her social links, and a variety of additional services. Most social network services are web-based and provide means for users to interact over the Internet, such

as e-mail and instant messaging. Online community services are sometimes considered as a social network service, though in a broader sense, social network service usually means an individual-centered service whereas online community services are group-centered.

Social networking sites allow users to share ideas, pictures, posts, activities, events, and interests with people in their network. This may allow other user to view our post and provide their point of view over our post. There are possible, in some case, to afford a negative comment or sometime may use certain vulgar text over our comment. In such case it is necessary to remove those comment or text from our post.

In this trendy world, one's character is judged more over via Online Social Network, rather than considering one's personal appearance and behavior. Users are more conscious about their status and desire more security. In such case if a vulgar text is commented on one's post, it may reflect one's personality. It is must that those unwanted messages must be removed from other's point of view.

In order to implement this concept, information filtering method is used. In this method, those unwanted message are assigned a value and frequency is calculated accordingly based on list of word which are predefined in blacklist. And Short Text Classifier is used to filter the phrases or sentences into individual words by extracting the commonly words such as prepositions, adjectives etc.

Interactions among people in which they create, share, and exchange information and ideas in virtual communities and networks. The range of machine methods employed builds on the same principles as those for information extraction. An Information filtering system is a system that removes redundant or unwanted information from an information stream using (semi)automated or computerized methods prior to presentation to a human user.

In content based filtering, when any message is posed it filters those messages by means of Short Text Classifier. The Short Text Classifier simply estimates appropriate value for each text. In the Bag-of-Word, those words are assigned a frequency value. The message which receives the lowest frequency value is been automatically removed from the displaying. The lower valued message is considered to be an unwanted message.

Instead of completing hiding such comments from user, we enhance a methodology to provide those comments in the notification. Since the comment are available in notification user can view those comments and can decide for further blocking or not. We are also creating a separate tool which stores only the filtered unwanted messages.

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. The terms are assigned automatically a method has to be chosen that can extract these terms from items. The terms have to be represented such that both the user profile and the items can be compared in a meaningful way

Information filtering can therefore be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages. An automated system will able to filter unwanted messages from OSN user walls. Short Text categorization technique is used to automatically assign with each short text message a set of categories based on its content. A system to automatically filter unwanted messages from OSN user walls on the basis of both message content and the message creator relationships and characteristics the section.

In this method, the unwanted messages are removed automatically from the user's wall and obtainable in the user's notification. From the notification only the user can view personally and can decide to bring them into light or remove them permanently. In certain case even the unwanted messages can provide positive value to the sentence. So in such case the user can decide whether the text

typed it in positive manner or in negative form. Based on user's option the decision is made.

II. EXISTING SYSTEM

In Online Social Network, human life started communicating their option and share information widely. When any particular issue happened they post those matters over OSN in their wall and trying to spread it worldwide. Different users view those posts and comment their point of view. Certain comment may hurt the user or the user may prefer some comment not to be posted. Content based preferences are supported to prevent those unwanted or undesired messages. The STC concentrate to extract the selection of msg predefined by the user in blacklist, this is possible by means of content based filtering.

In content based filtering, when any message is posed it filters those messages by means of Short Text Classifier. The Short Text Classifier simply estimates appropriate value for each text. In the Bag-of-Word, those words are assigned a frequency value. The message which receives the lowest frequency value is been automatically removed from the displaying. The lower valued message is considered to be an unwanted message.

Drawback

1. Automatically filter the unwanted message.
2. User is not aware of message, since content based is operated independently.
3. STC may understand message in a different manner and eliminate even a good comment.

III. PROPOSED SYSTEM

The Online Social Network becoming one of the easiest modes of communication to spread any information very quickly. People create account on their own and share their view over worldwide via this OSN. When anything is posted on the user wall, any other users who are permitted to view the post can comment their views. OSN so far provide facility to remove the unwanted message automatically by means of filtering rules. This system will hide or remove the unwanted comments. The user will not be aware of those comments.

Instead of completing hiding such comments from user, we enhance a methodology to provide those comments in the notification. Since the comment are available in notification user can view those comments and can decide for further blocking or not. We are also creating a separate tool which stores only the filtered unwanted messages.

Advantages

1. Privacy space in OSN is secure
2. No comment will be unaware to user.
3. User can decide to allow the comment or not.

IV. CONCLUSION

With the help of Information Filtering System in Online Social Network, one can provide security to one's profile by avoiding the unwanted messages. This system will not completely destroy such messages, instead it afford those messages or comments in user's notification. The user will be aware of those messages at least once and can decide whether the messages can be posted or avoided.

References

1. Macro Vanetti, Elisabetta Bingaghi, "A System to filter Unwanted Message from OSN user walls", IEEE Transaction on knowledge and Data engineering, February 2013.
2. P.E. Baclace, "Competitive Agents for Information Filtering," Comm.ACM, vol.35, no.12, p.50, 1992.
3. U.Hanani,B.Shapira, and P.Shoval, "Information Filtering : Overview of issues, Research and Systems," User Modeling and User-Adapted Interaction, vol,pp.203-259, 2001.
4. M.Carullo, E.Binaghi, I.Gallo, and N.Lamberti, "Clustering of Short Commercial Documents for the Web," Proc.19th Int'l Conf. Pattern Recognition (ICPR'08), 2008.
5. L.Fang and K.LeFevre, "Privacy Wizards for Social Networking Sites," Proc.19th Int'l Conf. World Wide Web (WWW'10), pp.351-360, 2010.
6. V.Bobicev and M.Sokolova, "An Effective and Robust Method for Short Text Classification," Proc.23rd Nat'l Conf. Artificial Intelligence (AAAI), D.Fox and C.P.Gomesm, eds., pp.1444-1445, 2008