A Proposed Analysis for User Behavior

Tamer Ahmed Abdellatif
Faculty of Computers and Information
Mansoura University

Dr. Aziza S. Asem
Faculty of Computers and Information
Mansoura University

Ass. Prof. Mahmoud M Abd El-latif
Faculty of Computers and Information
Helwan University

Abstract.

This paper presents a model of the "Semantic Web User Behavior Model" (SWUBM). That behavior is an important piece of adaptation and personalization, but it is too complicated to be a part of a generic Model. Semantic Web User behavior differs depending on web type and usages. For example, user behavior model on twitter is not the same on Wikipedia. So, in this thesis the researcher introduces a common general model that fit most of semantic websites.

The researcher uses some modeling features in this paper. Like personal features, content features, and focus features. In the personal features the researcher uses contact information and demographics. "Features reductions using rough sets" is the used method in minimizing the accumulated results from the semantic web sites. With the K-Means clustering algorithm and neural network.

Keywords: *Semantic web, User Behavior, K-Means clustering Modeling.*

1 Introduction.

Online communities form an important part of the web today where a huge portion of the Internet's traffic is driven by and through them. [6] Each single day, people in the Social Web create 1.5 billion pieces of information on Facebook, More than 140 million tweets on Twitter, upload more than 2 million videos on YouTube and about 5 million of images to Flickr site. [7] This huge amount of social data attracts researchers who want to use it to learn more about user preferences and interests, enhance recommendation and personalization systems.

What major current systems have in common is that they use data from a single application and depend on sufficient user information (user behavior or ratings) to produce acceptable results. [8] By using the distributed personal information a single user produces on a daily base, personalization and recommendation quality can be much more enhanced.

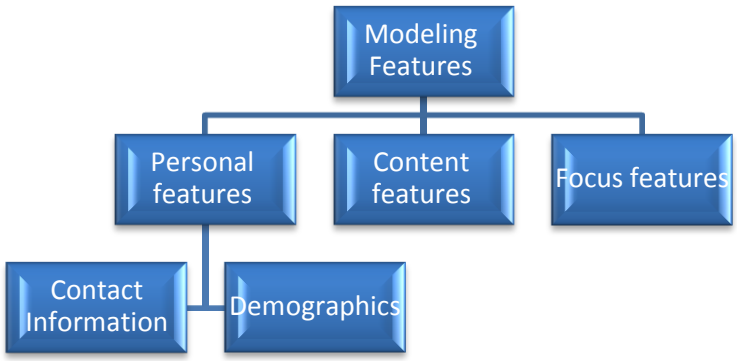

Figure (1): A simple illustration of the modeling features

Nowadays Internet and network related service and content providers try to collect useful service usage data and process it using different ways to know their user's behaviors. Heretofore quality of online contents is measured with number of page impressions (a request to start a single page of a website) or number of hits (refers to a request for a file from an internet web server). But both indicators regard to the quality of incoming links rather than quality of contents. [9].

Table (1): The data of the features

Personal Features		Content Features	Focus Features	
Contact info.	**Demographic**			
User name	Gender	Post length	Topic entropy	Posts Replied Ratio
In-degree	Birthday	Posting delay		Thread Initiation Ratio
First name	In-degree	Polarity		
MobilePhone	Family status	Complexity	Topic Likelihood	
CurrentCity	Education	Readability		
Hometown	Employment	Informativeness		

2. Literature review.

Martin J. Chorley and others in (2015) [1] released a web-based application that examines the personality characteristics and behavior of web users. The study concluded that personality traits help to explain individual differences in web usage and the type of places visited.

Á. Herrero Crespoa and others in (2015) [2] developed an integrative model to present the influence on behavior, as perceived by users of the content sent by other web users on social network sites. And the research concentrated on the hotel sector.

Radovan and others in (2015) [3] studied the possibilities of using the social media networks in the educational processes. Specially to serve in the communications between students and teachers.

Duygu Türkera and others in (2014). [4] aimed to explore the nature of unethical medical web user behavior within a social network perspective and investigated how such behaviors spread among the actors in medical sector.

Leila Weitzel and others in (2014). [5] Proposed how to measure reputation from the social interactions existing in the social networks. The researchers concentrated on the health information.

Matthew Rowe, and others in (2013). [10] presented an approach to analyze communities based on their role compositions. And presented a behavior ontology that captures user behavior within a given context (i.e. time period and community) and a semantic-rule based methodology to infer the role that a user has within a community based on his/her exhibited behavior.

The researchers described a method to tune roles for a given community-platform through the use of statistical clustering and discretisation of continuous feature values. And demonstrated the utility of their approach through role composition analyses of the SAP Community Network finding differences between communities and predicting community activity.

Michel Buffa, and others in (2008) [11] presented a state-of-the-art of semantic wikis, and they introduced SweetWiki as an example of an application reconciling two trends of the future web: a semantically augmented web and a web of social applications where every user is an active provider as well as a consumer of information. SweetWiki makes heavy use of semantic web concepts and languages, and demonstrates how the use of such paradigms can improve navigation, improve search, and improve usability.

F. Buccafurri, and others in (2006) [12] proposed an agent-based approach relying on a new model, called concept-graph, capable of representing user behavior-dependent relationships among concepts and, importantly, dealing with structural and semantic heterogeneity of Web sources.

Gerd Stumme, and others in (2006) [13] from their point of view that Semantic Web Mining aims at combining the two fast-developing research areas Semantic Web and Web Mining. This survey analyzed the convergence of trends from both areas: More and more researchers are working on improving the results of Web Mining by exploiting semantic structures in the Web, and they make use of Web Mining techniques for building the Semantic Web. these techniques can be used for mining the Semantic Web itself.

The researchers said that The Semantic Web is the second-generation WWW, enriched by machine-processable information which supports the user in his tasks. Given the enormous size even of today's Web, it is

impossible to manually enrich all of these resources. Therefore, automated schemes for learning the relevant information are increasingly being used. Web Mining aims at discovering insights about the meaning of Web resources and their usage.

3. The proposed model.

The proposed in this paper "semantic web user behavior Model" (SWUBM) is depending on the modeling features displayed in Figure (1) which are personal features from the documented contact information and demographics, the content features and the focus features. These modeling features will be reduced using rough sets by two programs Rose and Roseta. After that the K-Means clustering will be used in the proposed model, as illustrated in Figure (2). Which represented the steps of modeling features, features reduction using rough sets and the clustering using k-Means clustering. table (1) display the features data.

The proposed model (SWUBM) is applied with the reduction algorithms on the collected data from some of the most popular web sites Facebook, Twitter, Flicker, and Wiki. The results of the reduction algorithms produced a number of features in reduction and a support of reduction as illustrated in table (2). The reduction algorithms used are: GeneticReducer, JohnsonReducer, and RSES Exhaustive reducer.

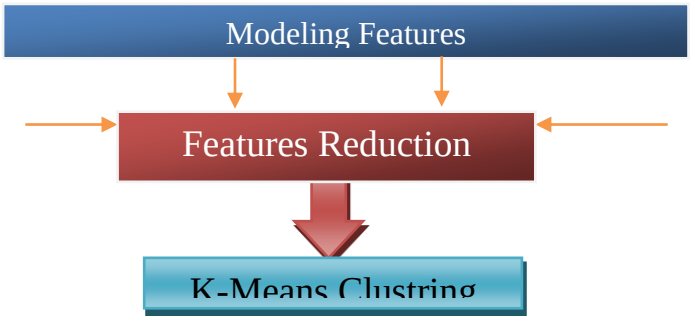

Figure (2): An illustration of the features

4. Discussion and Results

The model was applied using Rosetta and Matlab on Datasets about Facebook , Twitter, Flicker and Wiki .

In order to pick one of each algorithm, it is very important to pick the number of features in reduction to take the lower one in combination with highest one in the support of reduction. Thus in Facebook the GeneticReducer algorithm is picked . Because it resulted number 14 in the number of features in reduction which is the lower result. And in the same time it resulted number 100 in the support of reduction result. And in Twitter JohnsonReducer is picked because it resulted number 6 in the number of features in reduction which is the lower result. And in the same time it resulted number 100 in the support of reduction result

Table (2): The Reduction algorithms, number of features in reduction, and the support of reduction.

	Reduction Algorithms	No. Of Features In Reduct	Support Of Reduct
Facebook	GeneticReducer	14,18,19,20,21	١٠٠
	JohnsonReducer	15	100
	RSES Exhaustive reducer	19,20,22,23,27	58
Twitter	GeneticReducer	8,9,10	78
	JohnsonReducer	6	100
	RSES Exhaustive reducer	8,9,10,12,14	0
Flickr	GeneticReducer	10,11,14,15,16	88
	JohnsonReducer	10	100
	RSES Exhaustive reducer	10,11,14,15,16	0
Wiki	GeneticReducer	5,6,7,9,10,11,	92
	JohnsonReducer	6	93
	RSES Exhaustive reducer	5,6,7,8,9,10,11,12	1

And for Flicker the JohnsonReducer algorithm also is picked. Because it resulted number 10 in the number of features in reduction which is the lower result. And in the same time it resulted number 100 in the support of reduction result. But, Wiki is a different case. While GeneticReducer is picked. Though it resulted 92 in the support of reduction. but, it resulted number 5 in the number of features in reduction.

Chart (1), (2), (3), and Chart (4) shows an illustration for the reduction algorithms with Facebook, Twitter, Flicker, and Wiki.

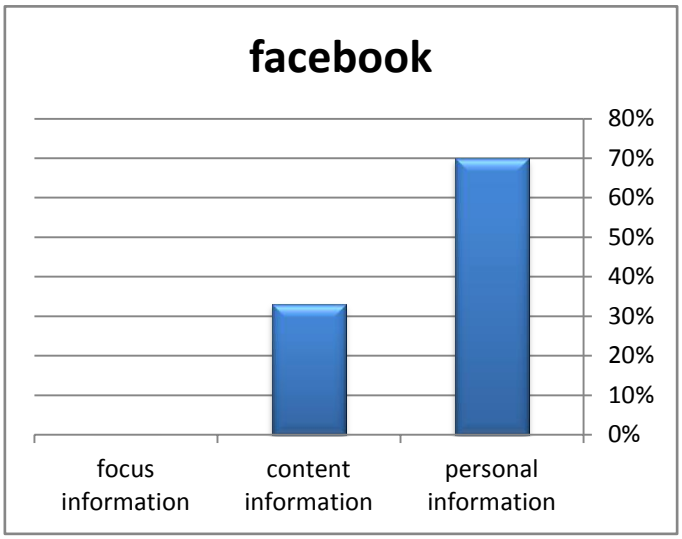

Chart (1): Algorithms used with Facebook

The column chart (5) shows all features reduction ratio for Facebook, Twitter, Flicker, and Wiki. While the pie charts (6), (7), (8), and (9) shows the data related to the mentioned web sites. And the bar chart (10) shows Users clustered by 5 activity features using K-means algorithm through Facebook, Twitter, Flicker, and Wicki.

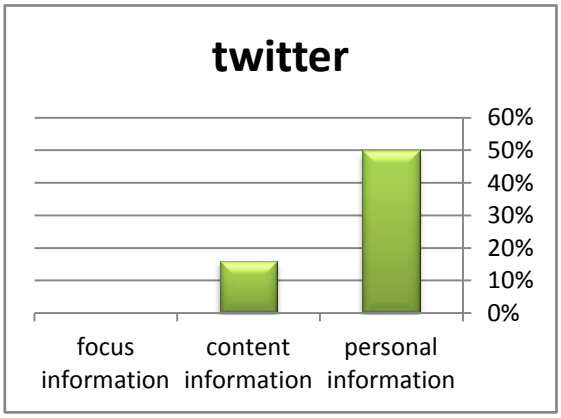

Chart (2): Algorithms used with Twitter

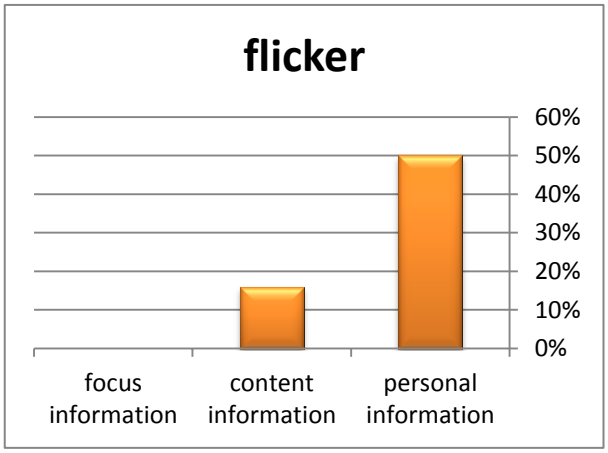

Chart (3): Algorithms used with Flicker

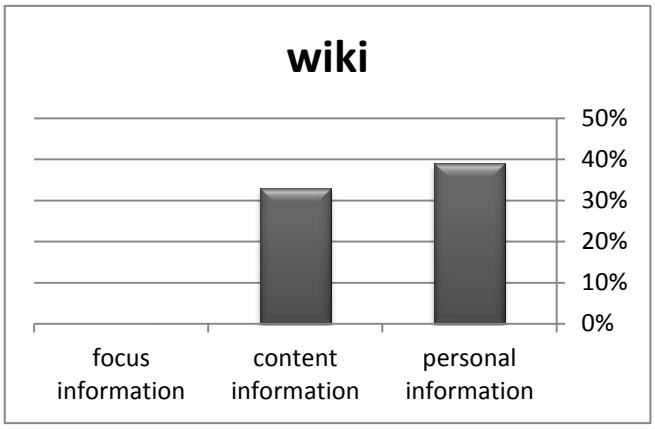

Chart (4): Algorithms used with Wiki

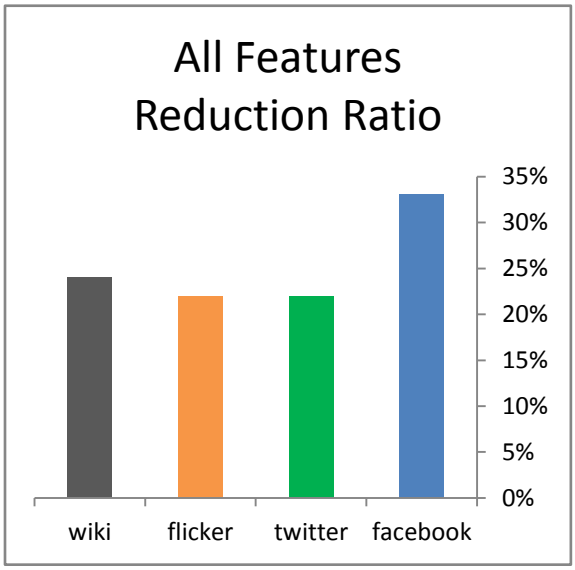

Chart (5): Algorithms used with the four web sites

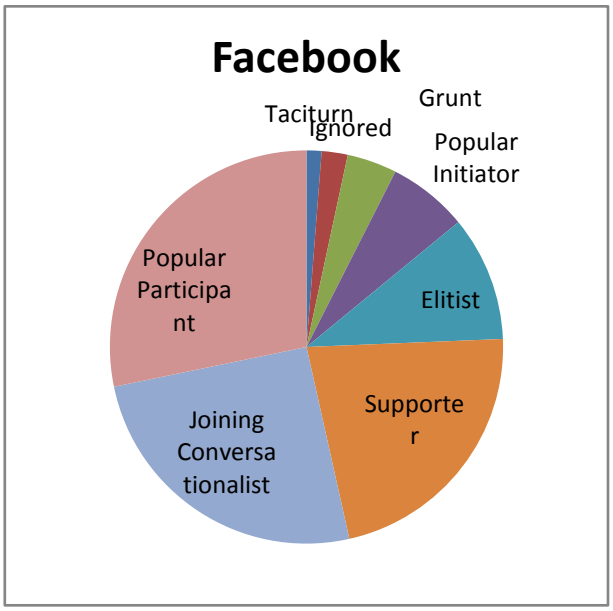

Chart (6): Data collected about Facebook users

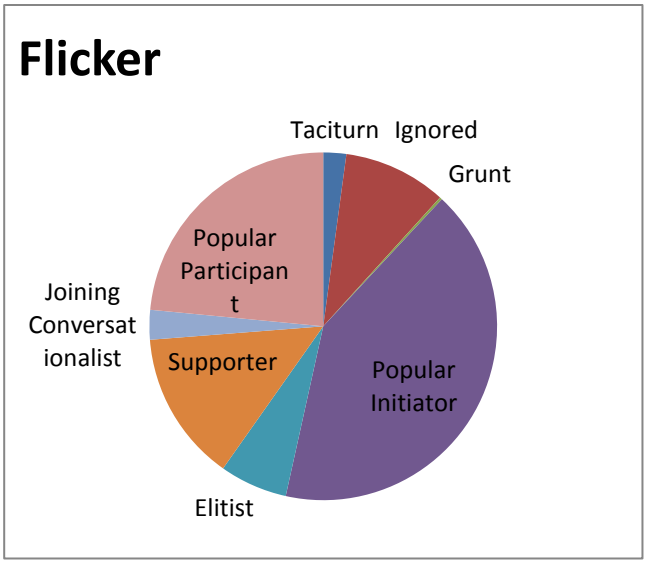

Chart (8): Data collected about Flicker users

Pie charts Numbers (6), (7), (8), and (9) show the collected data from Facebook, Twitter, Flicker, and Wiki users. While each slice presents one of the elements of the data.

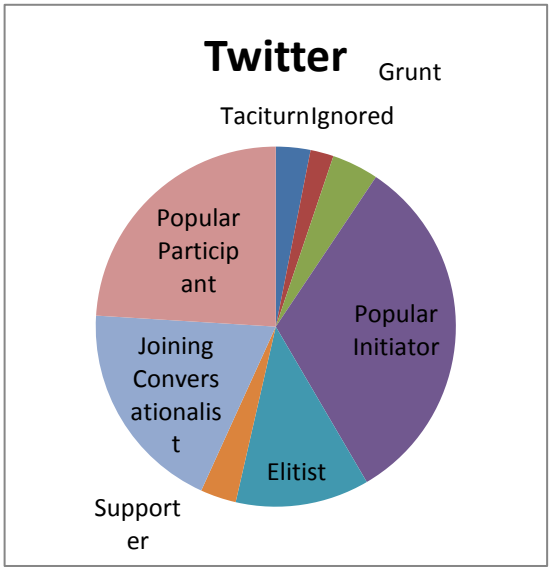

Chart (7): Data collected about Twitter users

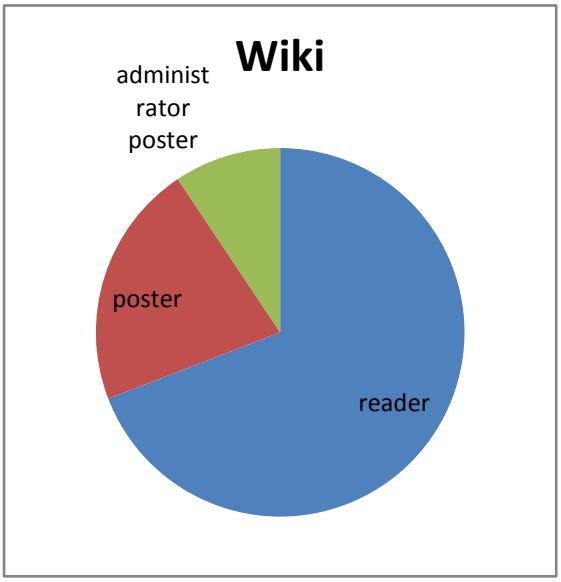

Chart (9): Data collected about Wiki users

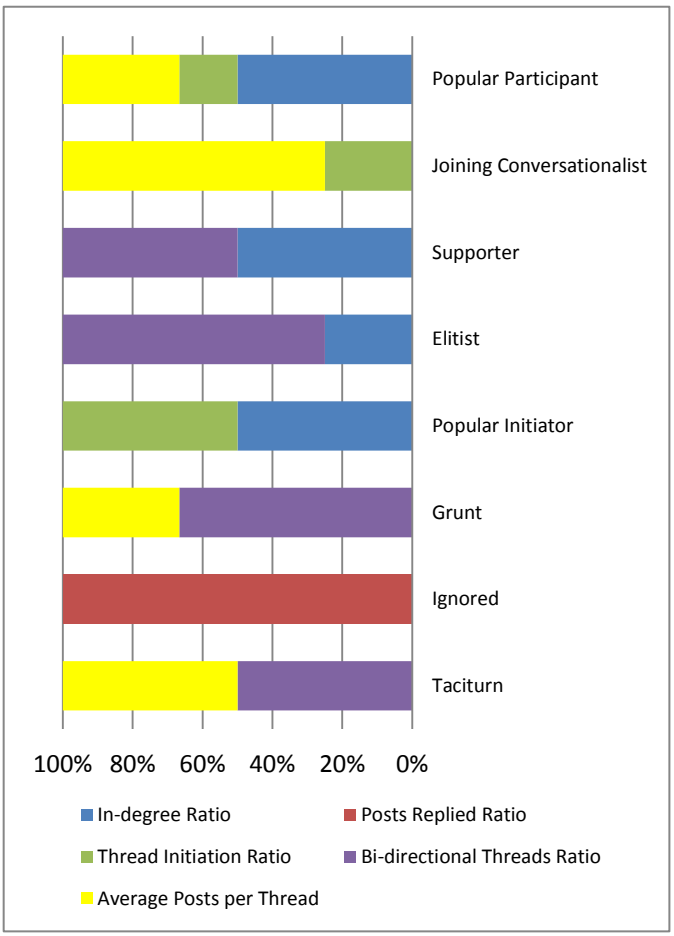

Chart (10): Users clustered by 5 activity features using K-means algorithm through Facebook, Twitter & Flickr

Conclusion and Future Work.

This research paper represented a model "Semantic Web User Behavior Model" (SWUBM). This model aims to help to understand the behavior of the semantic web user in order to offer him better service on the web. That suits each user according to his needs.

How human behave on the web is a very complicated operation. Hence more information should be gathered in order to build more effective tools to understand user behavior. And to give him a better services.

References.

[1] Martin J. Chorley, Roger M. Whitaker, Stuart M. Allen. "Personality and location-based social networks". (2015)

[2] Á. Herrero Crespoa, H. San Martín Gutiérreza, J.M. Hernández Mogollónb. "Perceived influence on behavior of user-generated content on social network sites: An empirical application in the hotel sector" (2015)

[3] Radovan Madle áka, Lucia Madle ákováa, Eva Kiani kováa "Designing a social network to support e-learning activities at the Department of Communications, University of Žilina" (2015).

[4] Duygu Türkera , Ceren Altuntaşb , "Analysis of Unethical Behaviors in Social Networks: An Application in the Medical Sector" (2014).

[5] Leila Weitzel, José Palazzo M. de Oliveira and Paulo Quaresma. "Measuring the reputation in user-generated-content systems based on health information" (2014)

[6] Sofia Angeletou, Matthew Rowe, and Harith Alani. "Modelling and Analysis of User Behaviour in Online Communities". (2011).

[7]http://www.scribbal.com/2011/04/infographic-how-much-daily-content-ispublished-to-twitter-facebook-ickr/

[8] Till Plumbaum, Songxuan Wu, Ernesto William De Luca, and Sahin Albayrak. "User Modeling for the Social Semantic Web". (2012).

[9] Istvan K. Nagy and Csaba Gaspar-Papanek. "User Behaviour Analysis Based on Time Spent on Web Pages". (2009)

[10] Matthew Rowe, Miriam Fernandez, Sofia Angeletou, and Harith Alani. "Community analysis through semantic rules and role composition derivation". (2013).

[11] Michel Buffa, Fabien Gandonb, Guillaume Ereteo, Peter Sander, and Catherine Faron. "SweetWiki: A semantic wiki". (2008).

[12] F. Buccafurri, G. Lax, D. Rosaci, and D. Ursino. "Dealing with semantic heterogeneity for improving Web usage" (2006).

[13] Gerd Stumme, Andreas Hotho, and Bettina Berendt. "Semantic Web Mining State of the art and future directions". (2006).

[14] Shahenda Sarhan, Enas Elharir, and Magdi Zakaria. " A Hybrid Rough-Neuro model For Diagnosing Erythemato-Squamous Diseases" (2014).