A Method to Classify Users on Social Networks Based on Similarity Mesuring

Thi Hoi, Nguyen[1]

[1]*Faculty of Economic Information System and E- Commerce, Thuongmai University, Hanoi, Vietnam; E-mail: hoint@tmu.edu.vn*

Abstracts: With the express growth of social networks, users have joined more and more of these networks and live their lives virtually. Consequently, they create huge amounts of data on these social networks: their profile, interests, and behaviors such as posting, commenting, liking, joining groups or communities, etc. One of the basic issues in these challenges is the problem of estimating the similarity among users on these social networks based on their profile, interests, and behavior. This paper presents a model for estimating the similarity between users based ontheir behavior on social networks. The considered behaviors are activities including posting or sharing entries, liking these entries, commenting and liking the comments in these entries, and joining a group in the social networks. The model is then evaluated with a dataset collected from Facebook users. The results show that the model correctly estimates the similarity among users in the majority of cases.

Keywords: Social Network, User Modeling, Classify User, Similarity.

1. INTRODUCTION

Social networks appeared in the late 20th century, creating favorable conditions for millions of people around the world to connect, establish, and maintain relationships, as well as access and share information. According to Samuel and Shamili [23], Collin et al. [5] the social networks impact on all aspects of social life is increasingly affirming their role in many fields, from education, business, health, tourism, etc., to social issues such as discovery risk, interest, etc. In these digital worlds, according to Zafarani et al. [33], Tang et al. [27] users freely present themselves, share information about their favorites and passions, or share their personal opinion on some issues of economic, social, cultural, etc. Through several activities on social network such as posting entries, sharing video clips, images, or news they read, and then leaving their comments or liking these entries or the comments of others, etc.

Consequently, huge data are created on the social network. This huge data attracts many researchers, businessmen, etc. to mine and exploit it. This tendency also brings some new challenges to researchers: do users having the same profile or interest show the same behavior? One of the basic issues in these challenges is the problem of estimating the similarity among users on these social networks based on their profile, interest, and behavior? The problem of detecting the similarity or the difference between users is not only based on the user profile on the social network, but also based on the data about user behavior such as posting entries, commenting, liking, and etc. This problem has been attracting many researchers.

For instance, Raad et al. [22] and Peled et al. [21] proposed a model to measure the similarity between user profiles. Anderson et al. [1] calculated the similarity between user characteristics. Liu et al. [13] estimated the similarity among preferences of user behavior. Liu et al. [14] and Chen et al. [6] measured the similarity among user mobility behavior. Xu et al. [32] analyzed the user posting behavior on a popular social media website. Singh et al. [24] formulated a metric based on the common words used in social networks to measure the user similarity in textual posting. Sun et al. [25] proposed a mapping method, which integrates text and structure information for similarity computation. Guo et al. [9] developed a model to estimate continuous tie strength between users for friend recommendation with the heterogeneous data from social media community. Nguyen et al. [17] aimed to understand the strategies users employ to make retweet decision. Liu and Terzi [15] approached the privacy issues raised in online social networks from the individual user viewpoint: they proposed a framework to compute the privacy score of a user. Tang et al. [28] adopted a “microeconomics” approach to a model and predicted the individual retweet behavior. Xu et al. [31] introduced several methods to identify online communities with similar sentiments in online

social networks. Zhao et al. [36] proposed to separately model users' topical interests that come from these various behavioral signals to construct better user profiles. Vedula et al. [29] detected pairwise and global trust relations between users in the context of emergent real-world crisis scenarios. Jamali and Ester [10] explored social rating networks, which record not only social relations but also user ratings for items. Bhattacharyya et al. [3] studied the relationship between semantic similarity of user profile entries and the social network topology. In the model of Zhao et al. [36], two social factors, interpersonal rating behavior similarity and interpersonal interest similarity, are fused into a consolidated personalized recommendation model based on probabilistic matrix factorization.

Most of these works try to estimate the similarity among user based on: user profile, user interests or favorites, or user relationship on social network. However, there are not many works which estimate the similarity among social network users based on their activities on social network.

In line with our previous works [18, 19, 20], this paper introduces a model for measuring the similarity between users based on their behavior in social network. In this model, the similarity between users is estimated from the similarity of their behaviors such as posting an entry or sharing an existing entry, liking an entry or liking a comment, commenting on a post, and joining a group or a community.

The model is then evaluated with a dataset-collected users from Facebook. The results show that the model estimates correctly the similarity among users in the majority of the cases. The paper is organized as follows: Section 2 presents the similarity model. Section 3 takes some experiments to evaluate the proposed model with empirical data. Section 4 is the conclusion and perspectives

2. USER MODEL BASED ON BEHAVIORS IN SOCIAL NETWORKS

User modeling is a way of representing a user's personal information through the characteristics that users show on social networks. User models, according to studies Benevenuto et al. [2], Gattani et al. [8], Xu el at. [32], are often built based on the following user characteristics: Personal characteristics or demographics; Interests and preferences; Needs and goals; Mental and physical state; Knowledge and background; User behavior; Context; Individual personality traits; etc. According to research, after the user model is built, each user will be represented by a set of personal information called a user profile about the problem being researched. Then, the user model will correspond to the profile containing the corresponding personal information, such as the user's interest profile, mobility profile, special model, etc.

2.1. A Similarity Measure Model for User on Social Networks

The The general model takes the two users as input data, and the output is the estimated similarity between the two entered users. Inside the model, there are four main steps:

- Step 1: Modeling Users
- Step 2: Calculating the value of features for the user
- Step 3: Estimating the similarity between each user's features
- Step 4: Aggregating the similarity between users from their similarities on features. These steps will be described in detail in the next sections.

Social Network

Without loss of generality, we assume that: A social network is a 2-tuples $\mathcal{N} =< U, B >$, in which: $\mathrm{U} = \{\mathrm{U}_1, \mathrm{U}_2, ..., \mathrm{U}_\mathrm{M}\}$ is a set of users, $\mathrm{B} = \{\mathrm{B}_1, \mathrm{B}_2, ..., \mathrm{B}_\mathrm{N}\}$ is a set of behaviors of each user $u \in U$ on the social network $\mathcal{N}$. Each user in a social network could post an entry, join a community or a group, like an entry, comment on an entry, like a set of comments in an entry of a community, share an entry, etc.

User Behaviors in The Social Networks

According to Zafarani et al. [35], Zafarani and Huan [34], Vispute et al. [30] user behavior on social networking sites is how users act and interact with events and phenomena on social networks. User behavior includes actions performed by users on social networking sites such as sharing entry, posting entry, liking posts, commenting on posts, bookmarking, following, creating and joining groups and communities, etc. These behaviors are classified according to individual behavior and collective behavior.

In this model, only five popular behaviors are considered: post an entry, like an entry, comment on an entry, share an entry, and join a group on social network. We assume that a social network has a set of users $\mathrm{U} = \{\mathrm{U}_1, \mathrm{U}_2, ..., \mathrm{U}_\mathrm{N}\}$. Each user $u_i \in U$ posts a set of entries E and acts with a set of behaviors $\mathrm{B} = \{\mathrm{B}_1, \mathrm{B}_2, ..., \mathrm{B}_\mathrm{M}\}$. Each behavior $b_l \in B$ may have a set of features:

- $P = \{post_i\}$ is an entry posted of entry, noted as bl post: the user writes or shares an entry on the user homepage.
- $L = \{like_i\}$: is an entry liked an entry or like a comment, noted as bl like: the user clicks on the like icon of an entry or a comment.
- $C = \{comt_i\}$ is an entry commented, noted as bl comt: the user writes some comments on an entry.
- $J = \{join_i\}$ is a group joined, noted as bl join: the user joins a group or community. A group usually has the name of a group, description of the group and other characters of the group

Each user $u_i \in U$ when represented by behavior will be a set of four as follows:

$$\mathrm{u_i} =< \mathrm{P_i}, \mathrm{L_i}, \mathrm{C_i}, \mathrm{J_i} > \qquad (1)$$

2.2. Calculating the Value of Features

The Value of Posting Behavior

The value of user $u_i \in U$ posting behavior on social network $\mathcal{N}$ is determined by the set of posts posted and shared by the user, denoted as $E_i^{post} \in E$ on social network $\boldsymbol{\mathcal{N}}$. Suppose $u_i \in U$ has n posts and shares $E_i^{post} = \{e_{i1}, e_{i2}, .. e_{in}\} \in E$ on social network $\mathcal{N}$, then the value of user u_i posting behavior is calculated the vector $\boldsymbol{p_i}$ with n components, each component is the weight vector of the corresponding posted and shared a post in $E_i^{post} \in E$ calculated according to the formula as follows:

$$\mathrm{u_{ipost}} = \mathrm{post_i} = \mathbf{p_i} = (\mathbf{e_{i1}}, \mathbf{e_{i2}}, .. \mathbf{e_{in}}) \qquad (2)$$

The Value of Like Behavior

The value of user $u_i \in U$ like behavior on social network $\mathcal{N}$ is determined by the set of posts and comment was liked by the user u_i, denoted as $E_i^{like} \in E$ on social network $\boldsymbol{\mathcal{N}}$. Suppose $u_i \in U$ has m posts and comment was liked $E_i^{like} = \{e_{i1}, e_{i2}, .. e_{im}\} \in E$ on social network $\mathcal{N}$, then the value of user u_i like behavior is calculated the vector $\boldsymbol{p_i}$ with m components, each component is the weight vector of the corresponding posts and comment was liked in $E_i^{like} \in E$ calculated according to the formula as follows:

$$\mathrm{u_{ilike}} = \mathrm{like_i} = \mathbf{l_i} = (\mathbf{e_{i1}}, \mathbf{e_{i2}}, .. \mathbf{e_{im}}) \qquad (3)$$

The Value of Comment Behavior

The value of user $u_i \in U$ comment behavior on social network $\mathcal{N}$ is determined by the set of posts and comment was commented by the user u_i, denoted as $E_i^{comt} \in E$ on social network $\boldsymbol{\mathcal{N}}$. Suppose $u_i \in U$ has k posts and

comment was commented $E_i^{comt} = \{e_{i1}, e_{i2}, .. e_{ik}\} \in E$ on social network $\mathcal{N}$, then the value of user u_i like behavior is calculated the vector $\boldsymbol{p_i}$ with k components, each component is the weight vector of the corresponding posts and comment was commented in $E_i^{comt} \in E$ calculated according to the formula as follows:

$$\mathrm{u_{icomt}} = \mathrm{comt_i} = \mathbf{c_i} = (\mathbf{e_{i1}}, \mathbf{e_{i2}}, .. \mathbf{e_{ik}}) \qquad (4)$$

The Value of Join A Group Behavior

The value of user $u_i \in U$ like behavior on social network $\mathcal{N}$ is determined by the set of groups was joined by the user u_i, denoted as $G_i^{join} \in G$ on social network $\boldsymbol{\mathcal{N}}$. Suppose $u_i \in U$ has l group was joined $G_i^{join} = \{g_{i1}, g_{i2}, .. g_{il}\} \in G$ on social network $\mathcal{N}$, then the value of user u_i join a group behavior is calculated the vector $\boldsymbol{p_i}$ with l components, each component is the weight vector of the corresponding groups was joined in$G_i^{join} \in G$ calculated according to the formula as follows:

$$\mathrm{u_{ijoin}} = \mathrm{join_i} = \mathbf{g_i} = (\mathbf{g_{i1}}, \mathbf{g_{i2}}, .. \mathbf{g_{il}}) \qquad (5)$$

2.3. Estimating the Similarity Two Users on Social Networks

The Cosine Similarity Measure

Suppose there are two vectors $\boldsymbol{u} = (u_1, u_2, .. u_n)$ and $\boldsymbol{v} = (v_1, v_2, .. v_n)$ then the cosine similarity of u and v is calculated as:

$$\mathrm{sim(u, v)} = \frac{<\mathbf{u},\mathbf{v}>}{\|\mathbf{u}\| * \|\mathbf{v}\|} \qquad (6)$$

In which, $< \boldsymbol{u}, \boldsymbol{v} >$ is scalar product of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, $\|\boldsymbol{x}\|$ is the Euclidean length of vector $\boldsymbol{x}$

The Pearson Correlation Measure

The research also using the Pearson correlation to calculate the correlation between two objects, according to the following formula:

$$\mathrm{cor(\mathbf{u}, \mathbf{v})} = \frac{\sum_i (\mathrm{u_i} - \bar{\mathrm{u}})(\mathrm{v_i} - \bar{\mathrm{v}})}{\sqrt{\sum_i (\mathrm{u_i} - \bar{\mathrm{u}})^2} * \sqrt{\sum_i (\mathrm{v_i} - \bar{\mathrm{v}})^2}} \qquad (7)$$

In which, $\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i$ and $\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$ and then the $cor(\boldsymbol{u}, \boldsymbol{v})$ is the correlation measure between $\boldsymbol{u}$ and $\boldsymbol{v}$.

The Entry Similarity

Definition 1*:* Given a set of texts $\boldsymbol{\mathcal{D}} = \{D_1, D_2, \ldots, D_p\}$, each text is represented by a set of terms $D_i = \{d_{i1}, d_{i2}, .., d_{ip_i}\}$. Call $\boldsymbol{\mathcal{V}} = \{v_1, v_2, \ldots, v_q\}$, is a set of different terms, pair by pair. Then, the weight of the term $d \in \boldsymbol{\mathcal{V}}$ with D_i is calculated as follows:

$$\mathrm{w_d} = \mathrm{tf(d, D_i)} \times \mathrm{idf(d, \boldsymbol{\mathcal{D}})} \qquad (8)$$

In there, $tf(d, D_i)$ times of occurrences of the term $d \; in \; D_i$; and $idf(d, \boldsymbol{\mathcal{D}})$ is calculated as follows (2):

$$\mathrm{idf(d, \boldsymbol{\mathcal{D}})} = \log\left(\frac{\|\boldsymbol{\mathcal{D}}\|}{1 + \|\{\mathrm{D_i} | \mathrm{d} \in \mathrm{D_i}\}\|}\right) \qquad (9)$$

After calculating the weights of the terms, each document $D_i \in \boldsymbol{\mathcal{D}}$ is represented by a weight vector; each vector is normalized to the unit interval [0,1]. Then, it is possible to define the text $D_i \in \boldsymbol{\mathcal{D}}$ according to the weight vector as follows:

***Definition 2*:** Given a set of texts $\mathcal{D} = \{D_1, D_2, \dots, D_p\}$, each text is represented by a set of terms $D_i = \{d_{i1}, d_{i2}, .., d_{ip_i}\}$. Call q is number of different terms, pair by pair, in $\mathcal{D}$. Then, each D_i is presented by a q demension vector as follows:

$$\mathbf{w_i} = (\mathrm{w_{i1}}, \mathrm{w_{i2}}, .., \mathrm{w_{iq}}) \text{ in } \mathcal{D} \qquad (10)$$

In there, w_{ik} is calculated follow Definition 1. A user post can be defined as follows according to Definition 1:

***Definition 3**:* Given a set of posts by social media N as $\mathcal{E} = \{\mathrm{E_1}, \mathrm{E_2}, \dots, \mathrm{E_q}\}$, each post $\mathrm{E_i}$ is represented by a set of words $E_i = \{e_{i1}, e_{i2}, .., e_{iq_i}\}$. Let q be the number of words that differ by pair in $\mathcal{E}$. Then, each $\mathrm{E_i}$ is represented by a vector with q dimensions: $\boldsymbol{w_i} = (w_{i1}, w_{i2}, .., w_{iq})$ in $\mathcal{E}$. Where each $\mathrm{w_{ik}}$ is calculated as in Definition 1.

Suppose there are two posts e_{il} and e_{jk} by two users u_i and u_j respectively on social network $\mathcal{N}$. Then, the similarity between two entries e_{il} and e_{jk} is calculated by the similarity between the two respective weight vectors as follows:

$$\mathrm{sim}(\mathbf{e_{il}}, \mathbf{e_{jk}}) = \frac{<\mathbf{e_{il}}, \mathbf{e_{jk}}>}{\|\mathbf{e_{il}}\| \times \|\mathbf{e_{jk}}\|} \qquad (11)$$

Then, the similarity between two sets of entries E_i and E_j is calculated by the similarity between two sets of corresponding weight vectors of users u_i and u_j denoted as follows:

$$\mathrm{sim}(\mathbf{E_i}, \mathbf{E_j}) = \max_{\mathrm{ik,il}}(\mathrm{sim}(\mathbf{e_{il}}, \mathbf{e_{jk}})) \qquad (12)$$

In which, $sim(\boldsymbol{e_{il}}, \boldsymbol{e_{jk}})$ is calculated formula (11).

The User Similarity Measure Based in Behaviors

Suppose there are two users $u_i, u_k \in U$ on the social network $\mathcal{N}$, the similarity measure of the two users according to the behavior calculated by the integration weighted similarity measure on the user's behaviors on the social network according to formula as follows:

$$\mathbf{sim_{beha}}(\mathrm{u_i}, \mathrm{u_k}) = \mathrm{w_{post}} * \mathrm{s_{post}}(\mathrm{u_i}, \mathrm{u_k}) + \mathrm{w_{like}} * \mathrm{s_{like}}(\mathrm{u_i}, \mathrm{u_k}) + \mathrm{w_{comt}} * \mathrm{s_{comt}}(\mathrm{u_i}, \mathrm{u_k}) + \mathrm{w_{join}} * \mathrm{s_{join}}(\mathrm{u_i}, \mathrm{u_k}) \qquad (13)$$

In which, $w_{post}, w_{like}, w_{comt}, w_{join}$, are respectively the weights of the behavior of posting or sharing an post, the behavior of liking an post the behavior of commenting on an post, and the behavior of joining a group on social networks, and they satisfy the condition: $w_{post} + w_{like} + w_{comt} + w_{join} = 1$. The $s_x(u_i, u_k)$ is the similarity of each behavior of two users u_i, u_k.

- The similarity on posted/shared behavior calculated by as formula as follows:

$$\mathrm{s_{post}}(\mathrm{u_i}, \mathrm{u_k}) = \mathrm{sim}(\mathrm{E_i^{post}}, \mathrm{E_k^{post}}) = \mathrm{sim}(\mathbf{p_i}, \mathbf{p_k}) \qquad (14)$$

- The similarity on liked behavior calculated by as formula as follows:

$$\mathrm{s_{like}}(\mathrm{u_i}, \mathrm{u_k}) = \mathrm{sim}(\mathrm{E_i^{like}}, \mathrm{E_k^{like}}) = \mathrm{sim}(\mathbf{l_i}, \mathbf{l_k}) \qquad (15)$$

- The similarity on commented behavior calculated by as formula as follows:

$$\mathrm{s_{comt}}(\mathrm{u_i}, \mathrm{u_k}) = \mathrm{sim}(\mathrm{E_i^{comt}}, \mathrm{E_k^{comt}}) = \mathrm{sim}(\mathbf{c_i}, \mathbf{c_k}) \qquad (16)$$

- The similarity on joined a group behavior calculated by as formula as follows:

$$\mathrm{sim}_{\mathrm{join}}(u_i, u_k) = \mathrm{sim}\left(G_i^{\mathrm{join}}, G_k^{\mathrm{join}}\right) = \mathrm{sim}(\mathbf{j_i}, \mathbf{j_k}) \quad (17)$$

3. METHOD EXPERIMENTS

3.1. Collection of Data

The study performed the collection of real data from the Facebook site; after removing the posts containing no text and noise type, the study obtained a set of 500 users, of which 100 posts, 100 posts or comments were liked, 100 comments in entries, and 20 groups joined in the social network were valid. The experimental data set parameters are in Table 1.

Table 1. Data set sample

Feature	Experimental data set
Users	500
Post an entry	50.000
Like an entry	50.000
Comment in an entry	50.000
Group (joined)	10.000
Weighted	TF.IDF
Presented	Weighted vector

3.2. Construction of Sample Set

Each sample is constructed as follows: Each sample contains three users collected from Facbook.com. These users are called as user A, user B, and user C, respectively. We ask a number of selected volunteers to answer the question: Which user, user B or user C, is more similar to user A than the other?

Then, we compare the number of people who chooses user B, and that of people who chooses user C. If the number of answer user B is greater than that of user C, then the value of this sample is 1. It means that user B is more similar to user A than user C. On the contrary, if the number of answer user C is greater than that of user B, then the value of this sample is 2. It means that user C is more similar to user A than user B. If the number of the answers user B and user C are not significantly different, this sample will be removed from the sample set.

After this step, we have a set of samples. We use the samples and save them in a set of samples. In experiments, we calculated that the convolution 3 of 500 users is 20.708.500 sample sets, but we only used 20.000 sets to experiment and compare with Buscaldi et al. [7] and our research precedence in Nguyen et al. [20] regarding the given sample set.

3.3. Scenario

The experiment is performed as follows: For each sample, we use the model proposed in this paper to estimate the similarity between user B and user A, and that between user C and user A. If user B is more similar to user A than user C is, then the result of this sample is 1. On the contrary, if user C is more similar to user A than user B is, then the result of this sample is 2. We then compare the result and the value of each sample. If they are identical, we increase the variable number of correct samples by 1.

3.4. Output Parameters

The correct ratio (CR) of the model over the given sample set is calculated as follows:

$$\mathrm{CR} = \frac{\text{number of correct sample}}{\text{total of sample}} \times 100\%. \quad (18)$$

The more the CR value is close to 100%, the more is the model correct. We expect that the obtained value of CR would be as high as possible.

4. RESULTS AND DISCUSSION

4.1. Results Experiment

The results are presented in Table 2. In total, the correct ratio of the model over all samples is about 18522/20000, reaching 92.61%.

Table 2. Correct ratio CR of the sample set

Sample set	Number of correct samples	Correct ratio CR
Facebook	18522	92.61 %

For more details, we run experiments with several combinations of weights from criteria of an entry, and weights from behavior of user with the following detailed scenario in Table 3:

- The same principle is applied at the level of behavior: we run the experiment with 1/4, 2/4, 3/4, and 4/4 behaviors. Each combination is also applied in the same manner as the previous level.

- For each combination, we run the experiment with different weights for each selected criterias. The changing step for each weight is 0.05. Therefore, each criteria weight runs from 0.05 to 1.00 as long as the sum of all criteria weights in the experiment is equal to 1.

Table 3. Weight of behavior

Behavioral combination	w_{post}	w_{like}	w_{comt}	w_{join}	Total of sample correct in 20000	Accuracy (%
¼ behavior				1.00	8304	41.52
		1.00			11288	56.44
			1.00		13142	65.71
	1.00				15828	78.14
2/4 behavior		0.65		0.35	15866	79.33
		0.45	0.55		16232	81.16
			0.65	0.35	16542	82.71
	0.75			0.25	16848	84.24
	0.70	0.30			17074	85.37
	0.75		0.25		17290	86.45
3/4 behavior		0.30	0.45	0.25	17396	86.98
	0.60	0.25		0.15	17492	87.46
	0.60		0.30	0.10	17628	88.14
	0.60	0.25	0.35		18070	90.35
4/4 behavior	0.35	0.25	0.30	0.10	18522	92.61
Weight	0.35	0.25	0.30	0.10	18522	92.61

4.2. Discussion

They indicate that our model, which reaches the correct ratio of 92.61%, is significantly better than the models of Buscaldi et al. [15] (with CR = 69.51%) and Nguyen et al. [24] (with CR = 79.14%), regarding the given sample set in Table 4.

Table 4. Correct ratio CR of the sample set

Model	CR %	Best weight combination			
		w_{post}	w_{like}	w_{comt}	w_{join}
Buscaldi et al. [7]	69.51%	1			
Nguyen et al. [20]	87.60%	1			
Our model	92.61%	0.35	0.25	0.30	0.10

The results also determined the best combination of feature weights for each model. Meanwhile, the model of Buscaldi et al. [7] concentrates 100% on the content, so there is no option to choose the best. The model of Nguyen et al. [20] considered not only content but also category, tag, sentiment, and emotion. The best combination of four weights corresponding to the four behaviors post, like, comment, and join a group in our model is 0.35: 0.25: 0.30: 0.10, respectively. These research results said that user modeling based on behavior could be used to classify users in social networks more than entry or text.

CONCLUSIONS

This paper presented user modeling in social networks by behaviors to estimate the similarity among users in the networks. The model is then validated with empirical data collected from Facebook. The experimental results indicate that the proposed model could reach a higher value in accuracy than some recent related models. With this result, it can be applied to classify users on many different social networks or in suggestion systems based on users' search histories. These research results will be presented in our future work.

FUNDING

This research is funded by Thuongmai University, Hanoi, Vietnam.

REFERENCES

[1]. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J.: Effects of user similarity in social media. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, WSDM ’12, pp. 703–712. ACM, New York, NY, USA (2012)

[2]. Benevenuto, F., Rodrigues, T., Cha, M., Almeida, V.: Characterizing user behavior in online social networks. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, IMC ’09, pp. 49–62. ACM, New York, NY, USA (2009)

[3]. Bhattacharya, P., Zafar, M. B., Ganguly, N., Ghosh, S., & Gummadi, K. P. (2014). Inferring user interests in the Twitter social network. In Proceedings of the 8th ACM Conference on Recommender Systems (pp. 357–360), 6–10 October, 2014, Foster City, Silicon Valley, California, USA.

[4]. Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13(1), 210–230.

[5]. Collin, P., Rahilly, K., Richardson, I., & Third, A. (2011). The benefits of social networking services: A literature review. Melbourne: Cooperative Research Centre for Young People, Technology and Wellbeing.

[6]. Chen, X., Pang, J., Xue, R.: Constructing and comparing user mobility profiles for location-based services. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing, SAC ’13, pp. 261–266. ACM, New York, NY, USA (2013)

[7]. D. Buscaldi, P. Rosso, J. M. Gomez-Soriano, and E. Sanchis, “Answering questions with an n-gram based passage retrieval engine,” Journal of Intelligent Information Systems, vol. 34, no. 2, pp. 113–134, 2010.

[8]. Gattani, A., Lamba, D. S., Garera, N., Tiwari, M., Chai, X., Das, S., & Rajaraman, A. (2013). Entity extraction, linking, classification, and tagging for social media: A Wikipedia-based approach. Proceedings of the VLDB Endowment, 6(1), 1126–1137.

[9]. Guo, C., Tian, X., Mei, T.: User specific friend recommendation in social media community. In: 2014 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6 (2014)

[10]. Jamali, M., Ester, M.: Modeling and comparing the influence of neighbors on the behavior of users in social and similarity networks. In: 2010 IEEE International Conference on Data Mining Workshops, pp. 336–343 (2010)

[11]. Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. Information, 10(4), 150.

[12]. Li, L., Peng, W., Kataria, S., Sun, T., & Li, T. (2015). Recommending users and communities in social media. ACM Transactions on Knowledge Discovery from Data, 10(2), 1–27.

[13]. Liu, H., Hu, Z., Mian, A., Tian, H., Zhu, X.: A new user similarity model to improve the accuracy of collaborative filtering. Knowl. Based Syst. 56, 156–166 (2014)

[14]. Liu, H., Schneider, M.: Similarity measurement of moving object trajectories. In: Proceedings of the Third ACM SIGSPATIAL International Workshop on GeoStreaming, IWGS ’12, pp. 19–22. ACM, New York, NY, USA (2012)

[15]. Liu, K., Terzi, E.: A framework for computing the privacy scores of users in online social networks. ACM Trans. Knowl. Discov. Data 5(1), 6:1–6:30 (2010)

[16]. Manning, C. D., Raghavan, P., & Schütze, H. (2009). Introduction to Information Retrieval. New York, USA: Cambridge University Press.

[17]. Nguyen, D.A., Tan, S., Ramanathan, R., Yan, X.: Analyzing information sharing strategies of users in online social networks. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 247–254 (2016)

[18]. Nguyen, M.H., Nguyen, T.H.: A general model for similarity measurement between objects. Int. J. Adv. Comput. Sci. Appl. 6(2), 235–239 (2015)

[19]. Nguyen, T.H., Tran, D.Q., Dam, G.M., Nguyen, M.H.: Integrated sentiment and emotion into estimating the similarity among entries on social network. In: Chen, Y., Duong, T.Q. (eds.) Industrial Networks and Intelligent Systems, pp. 242–253. Springer, Cham (2018)

[20]. Nguyen, T.H., Tran, D.Q., Dam, G.M., Nguyen, M.H.: Multifeature based similarity among entries on media portals. In: Akagi, M., Nguyen, T.T., Vu, D.T., Phung, T.N., Huynh, V.N. (eds.) Advances in Information and Communication Technology. Proceedings of the International Conference on Advances in Information and Communication Technology (ICTA 2016), pp. 373–382. Springer, Thai Nguyen, Viet Nam (2016)

[21]. Peled, O., Fire, M., Rokach, L., Elovici, Y.: Entity Matching in Online Social Networks. Social Computing/IEEE International Conference on Privacy, Security, Risk and Trust, 2010 IEEE International Conference on 0, pp. 339–344 (2013)

[22]. Raad, E., Chbeir, R., Dipanda, A.: User profile matching in social networks. In: Proceedings of the 2010 13th International Conference on Network-Based Information Systems, NBIS '10, pp. 297–304. IEEE Computer Society, Washington, DC, USA (2010) 123 Vietnam Journal of Computer Science

[23]. Samuel, C. J., & Shamili, S. (2017). A study on impact of social media on education, business and society. International Journal of Research in Management & Business Studies, 4(3), 52–56.

[24]. Singh, K., Shakya, H.K., Biswas, B.: Clustering of people in social network based on textual similarity. Recent Trends in engineering and material sciences. Perspect. Sci. 8(Supplement C), 570–573 (2016)

[25]. Sun, S., Li, Q., Yan, P., Zeng, D.D.: Mapping users across social media platforms by integrating text and structure information. In: 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 113–118 (2017)

[26]. Takale, S. A., & Nandgaonkar, S. S. (2010). Measuring semantic similarity between words using web documents. International Journal of Advanced Computer Science and Applications, 1(4), 78–85

[27]. Tang, J., Chang, Y., & Liu, H. (2014). Mining social media with social theories: A survey. ACM SIGKDD Explorations Newsletter, 15(2), 20–29. doi: 10.1145/2641190.2641195

[28]. Tang, X., Miao, Q., Quan, Y., Tang, J., Deng, K.: Predicting individual retweet behavior by user similarity. Know. Based Syst. 89(C), 681–688 (2015)

[29]. Vedula, N., Parthasarathy, S., Shalin, V.L.: Predicting trust relations within a social network: A case study on emergency response. In: Proceedings of the 2017 ACM on Web Science Conference, WebSci '17, pp. 53–62. ACM, New York, NY, USA (2017)

[30]. Vispute, A., Jadhav, P., & Kharat, P. V. (2014). Collective behavior of social networking sites. Journal of Computer Engineering (IOSR-JCE), 16(2), 75–79.

[31]. Xu, J., & Lu, T.-C. (2015). Toward precise user-topic alignment in online social media. International Conference on Big Data (pp. 767–775), 29 October, 2015–01 November, 2015, Santa Clara, CA, USA.

[32]. Xu, Z., Zhang, Y., Wu, Y., Yang, Q.: Modeling user posting behavior on social media. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '12, pp. 545–554. ACM, New York, NY, USA (2012)

[33]. Yin, D., Hong, L., & Davison, B. D. (2011). Exploiting session-like behaviors in tag prediction. Proceedings of the 20th International Conference on World Wide Web (pp. 167–168), March 28–April 1, 2011, India.

[34]. Zafarani, R., & Huan, L. (2014). Behavior analysis in social media. IEEE Intelligent Systems, 29(4), 69–71.

[35]. Zafarani, R., Abbasi, M. A., & Liu, H. (2014). Social Media Mining: An Introduction. New York, USA: Cambridge University Press.

[36]. Zhao, G., Qian, X., Feng, H.: Personalized Recommendation by Exploring Social Users' Behaviors, pp. 181–191. Springer, Cham (2014)

DOI: https://doi.org/10.15379/ijmst.v11i1.3523