Resonance-Relationship Network Construction by Information Analysis Based on Microblog Interactions

Meng-Hsuan Fu, Fang-Yu Lin, Yau-Hwang Kuo
Department of Computer Science and Information Engineering
National Cheng Kung University
Tainan, Taiwan, ROC
{mhfu,coffeetrue,kuoyh}@ismp.csie.ncku.edu.tw

Kuan-Rong Lee
Department of Information Engineering,
Kun Shan University Tainan,
Taiwan, ROC
leekr@ ismp.csie.ncku.edu.tw

***Abstract*— In fields of community and relationship analysis for online social medium, there are lots of researches focusing on interest detection and similarity. From those we know people are similar in some interests, but do not know why. Among conventional studies, personal profile information (explicit data) is often the main foundation to analyze. However it may occur inconsistency between a real fact and subjective information written by users. Thus, we think objective information and potential factor are essential to help us to understand the real conditions and progress in the future. So that we proposed a novel model and proposed methods to construct resonance-relationship network with behavioral pattern analysis and coordinate opinion analysis. We leverage interactive and time-varying data to extract resonance-relationship, and model the distribution of interactions. Finally, we showed our observation and result, and explained an actual situation with photography in case study. In summary, we proposed a novel model to analyze and solve potential problems for online social relationships.**

Keywords-resonance-relationship; social relationship; behavior modeling; coordinate opinion.

I. INTRODUCTION

Nowadays, the explicit and implicit data analyses of social community and social relationship are used to mine hidden information, such as the mutual interests between users. The explicit data includes raw text or tag presented in personal profile that is always wrote by users themselves. However, the implicit data was implied in users' behaviors . In the previous modeling studies, behaviors of browse and click on webpage are discussed [3, 4] mostly, and then behaviors are used to analyze user interests. Then, similarity of interests between two nodes are computed. The goal of this kind studies is to improve quality of personal service in networks, there are lots of applications such as recommendation system and matching system.

Online social networks have become more popular because of the blog prevailing. Some traditional methods may have been in deficiency gradually. Firstly, we perceive some important information in each personal profile, which is the main foundation to determine user's affinity. We expect an objective fact instead of personal subjective opinion; however, if the information includes too much subjective opinion, the inconsistency between fact and the information users wrote might occur. In some scenarios, we prefer objective facts, but the results happened via personal profile would not suit our requirements. Therefore, we proposed a method of analyzing interactive information based on online social platform, and expected to provide objective information.

Secondly, past studies in detecting users' interests realized people come together due to similarity of interests [9, 10], but they did not know why people got high similarity score in these fields (interests). In fact, people come together not only the similar interests but also other reasons. For example, people are curious about different characteristics, reliable comments, meeting frequently, and so on. In short, we think there are some key causes to make people come together, so that we made attempt on this study.

In the following section, we will give the detail explanation, definition and comparison. Section 2 reviews the literature on user behavior modeling in OSNs (online social networks), semantic orientation analysis, social relationship discovery and matching systems. Section 3 details the model and the procedure that we proposed method/algorithm. Section 4 discuss the results of our experiment and explain by case study, and finally conclude this paper and give a future work.

II. RELATED WORK

A. User Behavior Modeling

A number of studies observing properties of online social networks recently, and now most people are inseparable from the social networks. So, how do we understand the implicit information in the social medium is an important task. Based on this concept, there are two key issues [1] addressed for the measurements: characterization of user activities and usage patterns. This paper proposed a question: the information of user interaction is really an indicator of analysis on OSNs? Due to the motivation they try to quantify their observing factor and verify their assumption [2]. In order to characterize user behavior in online social network, the methodologies proposed to identify different classes of user behavior by evaluating the feature vector, which they defined [3] and analyze the user workloads in different online social networks to get the usage patterns [4]. Moreover, feature selection is also significant, because concept drift [5] is often caused by poor feature selection. Therefore, major studies chose the key impacting factors

 ISBN: 978-1-61208-220-2

cautiously by modeling extracted features and coefficient of variation.

B. *Social Relationship and Matching*

In the field of community analysis, a part of researches aimed at discovering hidden information and relation. In a study, two algorithms were designed by the original concept of feature extraction to accomplish relation extraction [6]. Regression-based algorithm is suitable when a user provides multiple community examples, but MinCut-based algorithm is suitable when a user provides single community example [6].

The applications of social matching system are often online social website and applications in some well-known social networks, such as iPartment [7] of Taiwan, Australia online dating website, RSVP [8] and so on. However, the conventional matching processes are almost according to users' profile and some questionnaires, which are all fixed fields for choose; the matching process of applications in some well-known social networks is providing the mutual friends between two users. [9] That study did matching process based on the personal information and their preferable conditions written in RSVP. The method classified users into several groups and matched male groups and female groups according to which male personal conditions are corresponding to female preferences. In addition, a social matching model was proposed [10] based in Twitter, they first detected personal information and interests, and then identify the user and his/her knowledge. Finally, they calculate similarity of interests between users then recommend. Another kind matching is recommending items, and then evolutionary computing method [11] is also used to improve the traditional methods in recommendation system.

Modularity [12] is common in evaluation of community analysis. It is an index to evaluate strength of a community (group), so that it is often used to partition a whole network into several groups. And it is used to understand intensity of one community or group according to the adjacent matrix and expectation value of degrees in a network (graph). Recently, Lu et al. [13] proposed a novel framework to evaluate user' s condition in online social networks. It contains three parts, people rank, social rank, relationship weight individually.

III. OUR MODEL

The core issue we researched for online social medium is to understand the hidden relationship, called resonance-relationship, and then construct a resonance network. We give this phrase a definition: people have a) coordinate opinions and b) enthusiasms for some themes.

Here, we mined referable user's opinion and potential tendentiousness by their posts and comments, and then regarded enthusiasms as the degree of participation. However, in order to target the potential users, we carry out "Active Cluster Detection" first. The overall framework is shown as Figure 1.

We believe this study will be in favor of future applications development.

A. *Environment*

This study is based on online social platform: Facebook [14]. We collected data from fanspages in Facebook, where could obtain enough large amount of raw data. Each fanspage has specific theme, it is a platform for online users to exchange ideas.

We try to solve a problem that the raw data in Facebook is miscellaneous. We assume condition of our environment: "we know there are some hot topics in one theme", because the goal of our research is *resonance-relationship* based on themes and topics. In addition, we need a terminology database in the coordinate opinion model beforehand, here we set "photography" as our query text, the terminology database is obtained from PHOTOGRAPHYTIPS.COM™ [15]. We retrieved data includes theme title, posts, comments and the corresponding behaviors by each user. See the example as Figure 2.

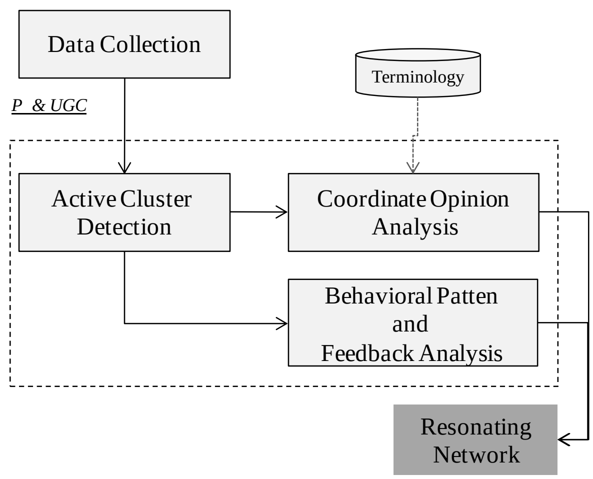

Figure 1. Overall famework of our model.

Figure 2. Example of fanspage in Facebook.

ISBN: 978-1-61208-220-2

B. Active Cluster Detection

After data collection, we proceed to detect active cluster (AC) in a theme. Since there is one kind of people called "flash mob", which means they just come and appear one time and exit rapidly, they do not attend interactions in at all. Here, we leveraged interactive information to achieve our goal that is finding the potential users.

The challenge is that some users are aggressive to express their views but some users are relative passive to interact, so that we probe the problem for two parts. We assume the users are positivity if they post actively, users just comment or do other actions occasionally passively, so that we think they are relative inactive and they always need someone to lead their opinion. We sought out active users preferentially, and next discovered passive users by setting a threshold. The model is presented as (1), where *PT* is the function to gather statistics in theme k, *surf* is the function to calculate attending ratio in topics of theme k, and $AC(u_i)$ is the function to determine whether u_i active is. We will compute post-times (*PT*) and visiting ratio (*surf*) by raw data. Here, α is a Boolean value, then β and γ are control parameters. We set β value 0.5. γ is equal to m multiplied by acting times.

$$AC(u_i) = \alpha(PT * \beta) + (1-\alpha)(surf * \gamma) . \qquad (1)$$

C. Coordinate opinion Analysis

In level of opinion orientation analysis, our goal is to analysis and determine the potential users' level of comment orientation in specific topics. The algorithm is shown as Figure 3. Firstly, word and phrase are mentioned by a user in the specific topic then we look them up in the terminology database. In Figure 3, Line5 to Line8 are claclulated the factoring value of a specific theme. Line10 is the process to decide users' temporary orientation by words. SO function is a method of semantic orientation [16]. Then, we calculate the value of importance weighting for a user by line13 (Z is set as a normalized factor). Recursively, we could conclude users' personal orientation and levels in line15. Consequenctly, the user who has strong personal orientation is regarding as strong concious level to some topics. We do that algorithm for every posts under a topic until their orientation in convergence. So that we could know users' conscious orientation in each attending topic.

In the following, we obtained OPL value via the algorithm in Figure 3, and then compute the coordinate value between pairwise users by (2). Define $\delta(o_i, o_j)$ to be 1 if user *i* and user j belong to the same orientation and zero otherwise.

$$Co\_op_{a,b} = \left| opl_i - opl_j \right| * \delta(o_i, o_j) . \qquad (2)$$

Input: P: post, C: comment, u : user.
w: the words
Th: theme
1. repeat
2. repeat
3. do segmentation;
4. if w_i find in **terminology** Database
5. if (everTalk(u , Th, w_i))
6. w_i.value ++;
7. else
8. insert node w_i, set value = 1;
9. end if
10. tempOp := SO(w_i);
11. until there is no C in P
12. until there is no P
13. $IM(u,Th) = \frac{1}{z}\sum_{i=0}^{n} w_i.value \cdot$
14. get user orientation: OP;
15. $OPL = IM(u,Th) * OP$

Figure 3. Algorithm of Level of opinion.

D. Degree of Participation Analysis

In behavioral feedback analysis, we would analyze different types of behaviors and attending level in some topics to discover the degree of preferences. (Assume each user's personal behavior patten is changed in different topics.) Here, we regarded user behaviors as three states. s_1 is like button reply or simple text reply; s_2 is pure text reply; s_3 is additional remark such as outer link reply. S=[s_1,s_2,s_3]. And each state has a feedback set value.

We analyze each user's behavioral patterns in different topics and themes they attended by (3), which calculates the maxmum likelihood of behavioral patterns via markov model (The algorithm is shown in Figure 4.). The behavioral patterns are regared as behavioral sequences under one of the topics that user attended, and then we leverage (4) to calculate the participation score. $Y^{i,z}$ is represented the probability of *user i* who has actions on *topic z*; $Y_T^{i,m}$ is represented the probability of *user i* who has actions on *post m*, *T* is the length of behavioral sequence. And it is subject to

$$\begin{cases} \text{len} = 1 : \text{pre} = 0, \text{now} = 1,2,3(\text{initial}) \\ \text{len} > 1 : \text{pre} = 123, \text{now} = 1,2,3,\text{end} \end{cases}$$

For example, a user do actions (maximum likelihood): comment it and then click like button, then we know his/her behavioral sequence is <s_2, s_1>.

$$Y^{i,m}_{T=len} = [\prod_{len} \Pr(X_{T=len} = s_{now} \mid X_{T=len-1} = s_{pre})] * Y^{i,m}_{T=1} \text{ . (3)}$$

$$\text{AttendS} = \sum BehSeq(w_{state}) \text{ .} \quad (4)$$

AttendS is the score that *user i* got in topic *z*. It is represented the score users prefered the topic. According to the personal behavior sequence in a topic, we sum the value of weight of elements, and w is the average value calculating by all actions that people did in the topic.

```
Behavior Pattern Recognition
Input: ui: user i;
   Topic k;
   doc(ui , topick) the behavioral record in DB;
   pm : post m;
   s0,s1,s2,s3,send: define user state;
   index, j, g : counter;
tr = getRawdata(doc(ui , topick ));
While(tr){
   mes = getContent(pm);
   while(mes){
         determineType(pm .mes[index]);
   }
   saveTempSeq(ui ,bsq[j++]);
len = bsq.length;
if(length==1) Y^{i,m}_{T=1} = Pr( X_{T=1} = s_{initial} );
else
Y^{i,m}_{T=len} = [∏_len Pr(X_{T=len} = s_now | X_{T=len-1} = s_pre)] * Y^{i,m}_{T=1} ;
}
for(g=1 ; g<=t ; g++){
 Y^{i,z} = Max(Y^{i,m}_T);
}
getPattern();
```

Figure 4. Algorithm of obtaining behavioral pattern.

E. Resonance-Relationship Netwok

Eventually, we would like to construct a resonating network using the hidden features we analyzed in part C and D. This network is represented as a graph $G = (V, E)$. Each node indicates an user that has a feature vector to stand for himself/herself. And each edge is the resonance-relationship between users, if the score of resonance-relationship is larger than the threshold, then the edge will be set up. We leverage (5) to compute resonance score between user *s* and user *t*. w_j is weight of each feature. We found some conditions changed according to the whole circumstance, thus we adjust weight value appropriately.

$$RS_{s,t} = \frac{\sum_{j=1}^{m} w_j * \frac{1}{dist(s_j - t_j)}}{dist(s,t)} \text{ .} \quad (5)$$

In (5), *dist* function is to compute the distance between users and users' feature vector.

IV. EXPERIMENT AND CASE STUDY

In this paragraph, we set the experiments based on Facebook raw data (explained in Environment and Figure 2.). Our preliminary experiments focus on observing active cluster and acquiring behavioral pattern. Moreover, we find some worth discussing phenomenon and then explain in discussion.

A. Experiment Results

The number of attending user and interaction are 83 and 100 in dataset 1. Likewise, the number of attending user and interaction are 175 and 205 in dataset 2.

We query "Photography" to get related themes (fanspages) then obtain hot topics via posts. The difference between dataset1 and dataset2 is the type. This is due to human factors and inherent properties in target platform. When we query target text, lots of related fanspages appear, and how shall we choose? If the type of dataset is cross-theme, it means we may filter some people we don't know whether they are important.

We know there is 9.8% active rate in dataset 1 and there is 5.1% active rate in dataset 2 by (1). Here, we set the threshold value 0.2. See TABLE I., we understand the real active cluster is minority. While a dataset is huge, the active cluster is also colossal but manageable. Re-visit indicates how many re-interactions are in whole datasets.

TABLE I. DATASET 1 AND 2.

	Active Rate	**Re-visit(act)**	**Type**
Dataset 1	9.8%	13.3%	Cross-theme
Dataset 2	5.1%	15.3%	Cross-topic

After active cluster detection, we do the algorithm of behavioral pattern by user generated contents (UGCs). We picked up part of attending users with topics and bsq (behavior sequence) shown in TABLE II. If an user has several bsq in one topic and then we get the maximum likelihood bsq to be his/her behavior pattern. Then we could use (4) to calculate participation score. We adjust w_{state} according to the conditions in topic. (An example is in Figure 5.)

TABLE II. PART RESULT OF BEHAVIOR PATTERN IN DATASET 2.

User	**Topic (multiple posts)**	**Behavior pattern**
Amit Mohod	4	(s1,s2,s1)
Dhruvell Dave	4	(s1,s2,s2)
Kristy Lopp Smith	3	(s1,s1,s2)
Nore Sanada Tozh	3	(s2,s2)
Photography Tips	1,2,4	(s1),(s3),(s2,s3)
Sribha Jain	4	(s2,s2)
Suhasini Gotmare	2,4	(s1),(s2)
Vasu Devan	3	(s2,s2)
Vikram Singh Grewal	4	(s1,s1)
…	…	…

TABLE III. COMPARISON.

	Analyzing information	Methodology base	Effect in Case Study
Our model	User time-varying comment and several behaviors to get hidden information.	Resonance-relationship	Knowing coordinate opinion and high participation users
Slah's	User explicit information and two type behaviors.	Clustering then similarity	Knowing users by your setting conditions
# twintera	Profile information and interests in microblog.	Similarity	Knowing similar users with near knowledge level
Recommendation in FB	Explicit and friend relation.	Mutual friends	Knowing mutual friends

B. *Case Study*

We leverage our model to apply for "**opportunistic social matching**". This idea we proposed is different from the existing social matching. Previous studies matched users according to the explicit information and past records in their profiles or even the questionnaire of personal requirement. Thus, the differences are explained in our novel model, it leveraged user behaviors and comments in a short period and discovered preferences and resonance-relationships among users. Therefore, oppotunistic social matching could well reflect real user's conditions and socialization.

The environment is designed based on facebook platform. Users participate in activities on *Photography* page in publishing media category that we regard it as a specific theme. Then, there were two topics with ten posts in three days (2012 Jan. tenth to twenty-fourth): "Canon lover", "single-lens-reflex cameras/SLR" and "Quiz Time". The number of attending users is 175, and the number of interactions is 205. Therefore we could know a section of users have higher frequency of attending interactions (Dataset 2).

Firstly, attending conditions of all attending users' are calculated by (1) in order to obtain the AC cluster. Secondly, if the outputs of all comments and posts for each topic are looked up in the terminology database the ever talked function is triggered. Then, check whether user talks about the terminology again. For instance, a user talked "DSLR" in this terminology database repeatedly, we accumulate the times, and vice versa. In case that a user just acts "good" or "I love it" frequently, we thought it is not useful information for a specific theme. We obtain users' opinion after getting value of temporary orientation iteratively. In other words, the orientation is convergent in an interval value. When the value of user's $IM(u,Th)$ is large enough, and the interval value of OP is positive, we could say that an user tends to post referable comments on this theme. After obtaining user's OPL, we leverage (2) to understand coordinate opinion between users.

Thirdly, we obtain a user's behavioral patterns to understand user's attending score. For example an user commented a sentence "I did not understand, can you explain with another…". Then we found they usually did actions in this topic, so that we set their pattern $<s_2>$ in this topic. If an user always clicks like button or says "wow", we set their pattern $<s_1>$. Each user has a feature vector including AttendS, Co_op and characteristic decided by potential parameters, and these features are all corresponding to topics of themes. In the end, we get RS score by (5) to construct the resonance-relationship network.

C. *Discussion*

We give another example about photography; there are two topics: "some photo shots about 2012 New Year" and "the shots about weather in Taiwan" in a period of time we crawled. (Topic3 is dataset 2.) The number of attending users in the first topic is nine, and we could obtain the each number of states of behaviors. Simultaneously, we also obtained the number of attending users and each state of behaviors in the second topic is 211. The simple observation is shown as Figure 5. X-axis is represented each state. Y-axis is represented the corresponding occupying ratio with the topic. We calculate the average weight of each state according to the observation Figure 5. The number of attending users in topic one is few to refer to; therefore, we would select datasets carefully to make them in real condition. (The tendency may be changed in different social medium, because each social media owns its limitations.)

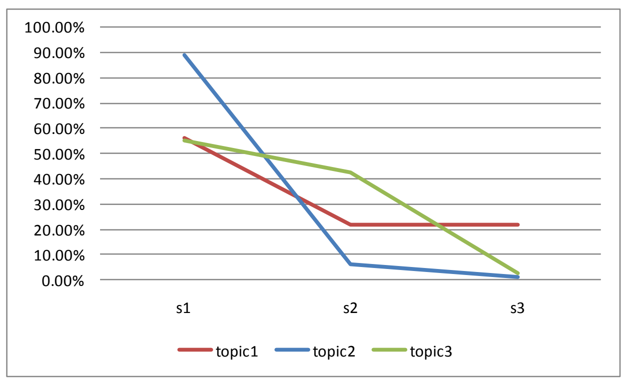

Figure 5. The observation with behavior distribution.

In Table III, we compare our model on social matching scenario with others.

Comparing to ours, the affinities could be modified with time goes by, so as to reflect the real users' conditions. Alsaleh's research [9] matched similar condition users after clustering attributes of male and preferences of female and considering the target user and "yes or no" of email and kiss actions. In addition, #twintera social matching model [10] collected profile information and user's interest then proposed knowledge indicators from several aspects by Twitter API. On the contrary, ours analyzed the time-varying hybrid data so the result will be changed according to user's interaction in social medium.

In addition, we realized behavioral distribution varies obviously. The distribution is not regular, we know the more trivial and intuitive action the more popular. Of course, it depends on the contents of posts. If the issue of post is suitable for discussion, the state-2 type will become more.

V. Conclusion and Prospect

This research mainly proposed and dicussed a novel model and methods. Instead of the conventional methods, which almost analyzed by personal profile information, and then calculate similar interests relationship. In order to solve the problem of inconsistency between subjective opinion and objective fact, and understanding hidden causes, we perceive that people know each others in several factors and not at all by similarity. So that we leverage users' interactive data: posts, comments and behaviors to develop our method, instead of static personal profile information (even explicit links). This model is useful to construct resonance-relationship network, we could understand the relationship factoring by some hidden causes. Thus it could be applied to several scenarios such as social matching, marketing. And we expect it will be in favor of future applications development.

Acknowledgment

This work was supported in part by the National Science Council of Taiwan under grant NSC-97-2221-E-006-144-MY3 and NSC-100-2221-E-006-251-MY3.

References

[1] Gyarmati, L. and T. Trinh, "Measuring user behavior in online social networks," *Network, IEEE*, vol. 24(5), pp. 26-31, 2010.

[2] Wilson, C., Boe, B., Sala, A., Puttaswamy, K.P.N., and Zhao, B.Y., "User interactions in social networks and their implications," *Proceedings of the 4th ACM European conference on Computer systems*, Nuremberg, Germany, 2009, pp. 205-218.

[3] Maia, M., J. Almeida and V. Almeida, "Identifying user behavior in online social networks," *Proceedings of the 1st Workshop on Social Network Systems*, Glasgow, Scotland, 2008, pp. 1-6.

[4] Benevenuto, F., Rodrigues, T., Cha, M., and Almeida V., "Characterizing user behavior in online social networks," *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, Chicago, USA, 2009, pp. 49-62.

[5] Concept drift, http://en.wikipedia.org/wiki/Concept_drift, *Wikipedia*. [retrieved: April, 2012]

[6] Cai, D., Shao, Z. He, X. Yan, X., and Han, J., "Mining hidden community in heterogeneous social networks," *Proccedings on the 3rd international workshop on Link discovery*, Chicago 2005, pp. 58-65.

[7] iPartment: http://www.i-part.com.tw/ [retrieved: April, 2012].

[8] RSVP: http://www.rsvp.com.au/ [retrieved: April, 2012].

[9] Alsaleh, S., Nayak, R., Xu, Y., and Chen, L., "Improving matching process in social network using implicit and explicit user information," *Web Technologies and Applications*, 2011, pp. 313-320.

[10] Vrabl, S., J. Oliveira, and C.L.R. Motta, "#twintera!: A social matching environment based on microblogging," *Computer Supported Cooperative Work in Design (CSCWD), 15th International Conference on*, 2011, pp.556-561.

[11] Hyun-Tae Kim, Jong-Hyun Lee, and Chang Wook Ahn, "A Recommender System based on interactive evolutionary computation with data grouping," *Procedia Computer Science*, vol. 3, pp.611-616, 2011.

[12] Modularity, http://en.wikipedia.org/wiki/Modularity_(networks), *Wikipedia*. [retrieved: May, 2012]

[13] Lu, Z., Yanlong, Wen, Haiwei, Zhang, Ying Zhang, and Xiaojie, Yuan, "User relationship index based on social network community analysis," *Business Management and Electronic Information (BMEI), International Conference on*. 2011, vol. 4, pp.66-69.

[14] Facebook: http://www.facebook.com/ [retrieved: May, 2012].

[15] PHOTOGRAPHYTIPS.COM™: http://www.photographytips.com/page.cfm/1587 [retrieved: May, 2012].

[16] Turney, P. and M.L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Transactions on Information System*, vol. 21, pp. 315-346, 2003.

