2015 International Conference on Computational Science and Computational Intelligence

Personal Recommendation Engine of User Behavior Pattern and Analysis on Social Networks

Cheng-Hung Tsai, Han-Wen Liu, Tsun Ku, Wu-Fan Chien

Institute for Information Industry,
Innovative DigiTech-Enabled Applications & Service Institute,
Taipei, Taiwan, R.O.C.

{jasontsai, tonyhwliu, cujing, juliettechien}@iii.org.tw

Abstract-In the age of information network explosion, Along with the popularity of the Internet, users can link to all kinds of social networking sites anytime and anywhere to interact and discuss with others. This phenomenon indicates that social networking sites have become a platform for interactions between companies and customers so far. Therefore, with the above through social science and technology development trend arising from current social phenomenon, research of this paper, mainly expectations for analysis by the information of interaction between people on the social network, such as: user clicked fans pages, user's graffiti wall message information, friend clicked fans pages etc. three kinds of personal information for personal preference analysis, and from this huge amount of personal data to find out corresponding diverse group for personal preference category. We can by personal preference information for diversify personal advertising, product recommendation and other services. The paper at last through the actual business verification, the research can improve website browsing pages growth 11%, time on site growth 15%, site bounce rate dropped 13.8%, product click through rate growth 43%, more fully represents the results of this research fit the use's preference.

Keywords- Social Networks; Social Persona Analysis; Cost Per Click; Personal of Interest analysis

I. INTRODUCTION

Recently, the flourish of information technology caused people to think how they can promote the chances of interaction with others by information technology. With the progress of information technology and the extension of the concept of interaction, virtual communities (e.g. Facebook, Weibo, Blogspot, PTT, forums, news…) had begun forming and growing. Users have gotten used to interacting and communicating with their families, colleagues and friends over all kinds of social networking sites. Even they will exchange information with others having similar interests or demands. Hence, under the influence of social network, there is a considerable amount of social activities shifting from reality to virtual online platforms. This brings a direct impact to the consumer market, which is too immense to be ignored.

Social networks for business operators, the value of the information can be one of the important sources for understand customer's opinion and behavior analysis which retained by the virtual social group. And the "human" play a very important link in the social network, and also because of the interactive link between people, the social platform has become a very huge information exchange center, spread and scattered that people for any event and personal preferences are derived from the words and social behavior. By collecting and analyzing personal interaction behavior and the content of writing comments on the social platform, we can more understand that the preference for personal interest has more significant results. Therefore, how to effectively analyze the personal preference of the social platform has become a problem to be solved in this paper.

The paper mainly put forward the research of solving social personal preference analysis, through the three modules planning and implementation that the paper mentioned about to collect and analyze data, according to the algorithm of each module for processing, the effective personal preference information analysis for social network platform, through the modules planning in this paper, after the system actual verification, web browsing pages growth 8%, site bounce rate dropped 11%, product click through rate growth 43%, the results can provide extended of follow-up research, the method follow up will be introduced in order.

The framework of this paper is as follows: the first section introduces the motivation and background of this paper. The second section, through the previous research put forward that the relevant literature is discussed and analyzed. The third section introduces the used of the method and technology in the paper. The fourth section puts forward the implementation results and analysis. The fifth section is the conclusion and the future outlook.

978-1-4673-9795-7/15 $31.00 © 2015 Crown Copyright
DOI 10.1109/CSCI.2015.46

II. RELATED WORK

A. Social Network Development Trends

According to the estimation of the international market research authority eMarketer, the number of global users of social network has reached 1.5 billion. The number of active user of Facebook, Twitter, Weibo and Tencent were one billion, 0.5 billion and 0.3 billion at 2012, respectively. Furthermore, the penetration rate of social network services among the global users of all ages has reached 79%. Even those once considered the main composition of digital divide has a rise of penetration rate up to 9.3%. These facts show that the use of social network is very popular. In domestic, the number of user of social network had a breakthrough of ten million at 2011, according to the research of Institute for Information Industry. [3]Among those under 30 and between from 50 to 59, the proportions of using social network were 96% and 70%, respectively, which show an extremely high penetration rate. As for the period of using social network, a research shows that there were more than 70% of domestic people using it over a year. If we take those using over six months but less than a year into account, the proportion is approximately 90%. Learning from the high penetration rate, the application of social network has deeply rooted in our daily life and also changed our lifestyle.

B. Linked Grammar, WordNet and Ontology

Link grammar is an English grammar parser system that was proposed by researchers at the School of Computer Science of Carnegie Mellon University (CMU). Link grammar is a scheme for describing natural language. [26], proposed Link grammar defines a set of words, which are the terminals of grammar, and each has some linking requirements. The linking requirements of each word are gathered in a dictionary. The linkage can be perceived as a graph, and the words can be treated as vertices, which are connected by labeled arcs. Thus, the graph is connected and planar. The labeled arcs that connect the words to other words on either their left or right sides are links. A set of links that proves that a sequence of words is in the language of a link grammar is called a linkage.

WordNet originated from Cognitive Science Laboratory in Princeton University. [27] It is a vocabulary reference system designed by researchers who inspired by psychology theory. WordNet processed the first level classification according to part of speech (POS). Driven by different word meanings and expressions, it forms several Synset. Each Synset symbolizes one vocabulary and takes down other words and expression with the same meaning.

According to the definition of Wikipedia, Ontology is a data model that represents a domain and is used to reason about the objects in that domain and the relations between them. Ontology is very important in various fields, such as knowledge engineering, natural language processing, intelligent information integration, and knowledge management. Ontology provides a shared, common representation of a domain that can be communicated between heterogeneous and widespread application systems. Ontology has been developed in AI to facilitate knowledge sharing and reuse. Ontology provides an explicit conceptualization that describes the semantics of data.

C. Concept Space

As for building ontologies with concept space, if we want to apply concept space as a semantic network into real-world cases, there are problems to be solved: first, in real-world languages are not always monolingual; bilingual or multilingual concept space are frequently used, which means to build a cross-lingual concept space, and previous studies [28] showed that there are still room for improvement. Second, from a macro view, the applications of concept space are much of diversity; as a result, it might require cross-domain knowledge. For example, it might be applied for e-commerce, where user can combine products of the same genre into concept, finding trends for purchase, sales and stock, or it could be used for tracking the topic transition in timelines of news events. [29] proposed a cross-domain medical document retrieval system, and the study revealed the bottleneck of analyzing two or more domains that are not loosely associated. Third, from a micro view, a concept space should be of different semantic distance, even in the same class, each semantic distance between two concepts should be of different values; fourth, a concept space evolves as time shifts, it should update and grows with time. Previous study [30] proposed a dynamic model to deal with real-time data, but the domain is still limited. From the problems mentioned above, it can be inferred that the major problem lies in a dynamic model, which can extract data from different sources to construct concept spaces, and in a real-world data source with features of multilingual languages, cross-domain, different semantic distances and time evolving. To acquire such data resource, it should have been a difficult task; however, with social networks and social software rise, and collective intelligence service such as Wikipedia and social bookmarking, it provides use an opportunity to reach such standards; therefore, to address this problem, it is our research goal to propose a dynamic method that can deal with the data. For this, we proposed an adaptive framework that can automatically generate a dynamic concept space from corpora, and we discuss such framework in a few perspectives in the following section. First, we will illustrate previous works building ontologies, dictionaries, or semantic networks in a corpus-based approach with different formats. Second, we give a formal expression of problem statement and elaborate each of the steps we proposed in the framework;

third, we discuss the evaluation methods and how we measure the experiments.

III. THE REALIZATION OF THE PERSONAL RECOMMENDATION ENGINE OF USER BEHAVIOR PATTERN AND ANALYSIS ON SOCIAL NETWORKS

4.1 The system architecture

The main implementation is set up the system on the Amazon Web Services platform, the system is divided into three modules: Personal Preference Recommendation Interface (PPRI), Data Processing Platform Module (DPPM), Preference Analysis Arithmetic Module (PAAM), to conduct Personal data collection, analysis and temporary storage service, and its architecture planning as shown in Fig. 1.

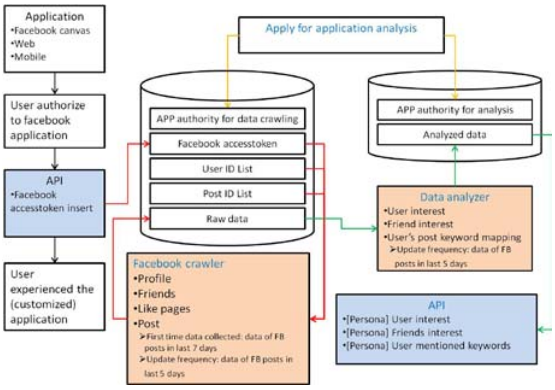

Fig. 1 Field Empirical Chart

The system on this paper collect 30,000 personal preference data with the user's approval through the Personal preference recommendation interface, the WebCrawler of data processing platform module will carry out the personal data collection. Then, the Personal preference analysis calculation module will analyze the personal preference of these data, and finally, through the results of personal preference analysis to convey the product or the article that users interest in.

4.2 The system flow

In this section, we will introduce the details of the process of system flow. First, we set up Facebook Login APP on the web page, when the user login with Facebook Login APP, it will show the user's personal authorization content. And through the user click authorized personal content items, Facebook APP will deliver the user access token to the data processing platform module. Data processing platform module will conduct Facebook Personal page data collection by the user access token (user_profile, user_friends, user_post). By collecting personal data (Raw data), we will carry out three Personal preference analysis calculation modules: 1.user interest, 2.friend interest, 3.user's post keyword mapping. The technical details as described in the third section. Finally, with the results of our Personal analysis, through three different kinds of Personal preference analysis API to fetch with the front-end data, the front-end web site can carry out the recommendation of personal commodities and articles by the result of Personal preference analysis.

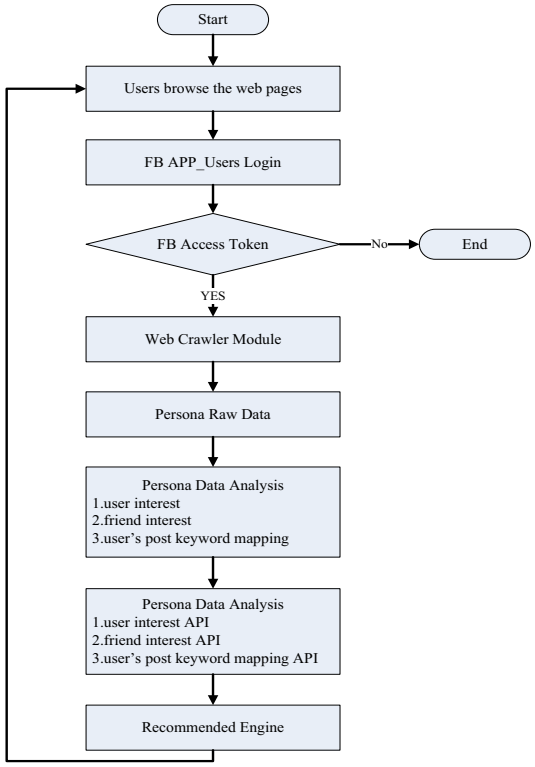

Fig. 2 System flow chart

The research analyzes the user behavior interaction data on social group, further with the way to initiative recommend commodities and articles that the user preferences by the personal preference, and to stimulate the user / the consumer browsing and purchasing the commodities, enhance the owners on click through rate and adhesive capacity of the commodities.

4.3 System Verification

4.3.1 Verify field setting

The research is to verify the effectiveness of this system. The effectiveness will be actual verification to the official website that built by the cooperation owners of this research. On the website, we will plan seven recommended fields (Test_A: article below / you might like, Test_B: article below / extended reading, Test_C: Blog below / you might like, Test_D: video below / you might like, Test_E: Health wall, Test_F: recipes below / you might like, Test_G: body encyclopedia) to carried out two recommended service verification (1. Personal

Preference Recommend Analysis Service 2. Rule of thumb recommendation services)

Table 1、The resource of recommended articles

Recommended classification	Latest in today	Finance and Economics	Hot issue	Data chart
Article quantity	35	15	20	17
Recommended classification	Business Manage ment	Fashion life	Magazine	Character istics Channels
Article quantity	20	20	15	20

Fig. 3 Recommended service verification field

Our research actually verified on the web site, the verification period is a month, between 2015/06/01 ~ 2015/06/30, the eight classifications totally have 162 popular articles, which carried out the user recommended verification, as shown in Table 1.

4.3.2 System statistics

The research experimentally verified on the web site during a month, based on the back-end data which provide by the co-operative enterprise to verify the result of this service, such as: Page View (PV), time on site, site bounce rate, the commodities click through rate. The results of verification are shown as follows:

Table 2、Verification of Personal preference recommended analysis services

	Personal Preference recommend analysis services			
	PV	Time on site	Bounce rate	Click through rate
Test_A	189,371	189	5.23%	2.07%
Test_B	44,382	155	23.16%	0.75%
Test_C	31,151	298	5.76%	2.29%
Test_D	10,392	120	5.08%	2.06%
Test_E	5,112	209	4.43%	1.46%
Test_F	1090	357	6.71%	2%
Test_G	913	151	37.43%	0.67%

The verification of this research verified through the data of the seven recommended field which collected through the Personal preference recommend analysis services by the co-operative enterprise. The results of data as shown in Table 2, the total browsing rate of PV reading data is 282411, the total average of page views is 40,344.4, the average of total time on site is 1,479 seconds, the average of bounce rate is 1.6%, and the average of click through rate is 12.5%. Then, we verified and analyzed through the rule of thumb recommended method (ROT) at follow up.

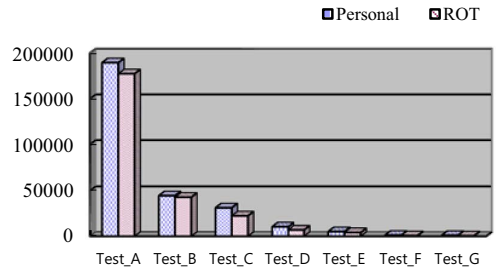

Fig. 4-1 PV empirical data

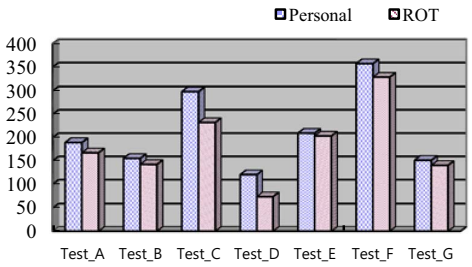

Fig. 4-2 Time on site empirical data

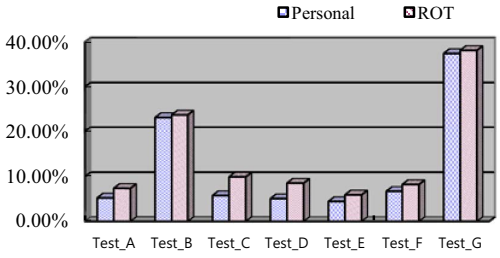

Fig. 4-3 Bounce rate empirical data

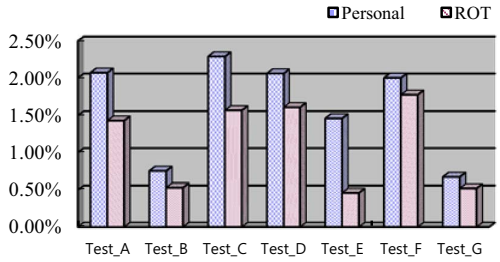

Fig. 4-4 Click through rate empirical data

Table 3、The rule of thumb recommended services data

	The rule of thumb recommended services (ROT)			
	PV	Time on site	Bounce rate	Click through rate
Test_A	177,223	167	7.37%	1.43%
Test_B	42,728	142	23.73%	0.53%
Test_C	22,357	232	9.91%	1.57%
Test_D	6,879	74	8.53%	1.61%
Test_E	4,097	203	5.91%	0.46%
Test_F	633	328	8.23%	1.78%
Test_G	507	140	38.12%	0.52%

The verification of this research verified through the data of the seven recommended field which collected through the rule of thumb recommended services by the co-operative enterprise. The results of data as shown in Table 3, the total browsing rate of PV reading data is 254424, total average of page views is 36,346.2, the average of total time on site is 1,286 seconds, the average of bounce rate is 14.5%, and the average of click through rate is 1.1%.

After experimentally verify on the web site, we compared and analyzed through the Personal preference recommend analysis services and the rule of thumb recommended services, as shown in Fig. 4, we found that in the process of the user's interaction of social behavior, we can get the data of social interaction to further understand the user's preferences. And the preference recommend method is more superior than the traditional rule of thumb recommended method, to achieve its overall effectiveness: website page views growth 11%, time on site growth 15%, website bounce rate dropped by 13.8%, commodities click through rate growth 43%.

We found out that its recommended effect of the growth rate is relatively poor among test_B, test_G, by the results after analysis, we believe that due to the test_B, test_G mainly by article "issue" that the user clicked and read, once more and then recommend the same "issue" of the article, so that the user due to the "issue" of the article reading reaches a certain freshness, and makes the user has met saturated reading the issue of the article, therefore, test_B, test_G growth rate is more limited. Through the actual verification to verify the system of Personal preference recommended analysis and it has reached certain accuracy.

IV. CONCLUSION

This research is a project of Research on Intelligence Techniques and Service Modes of Social Media conducted by Institute for Information Industry. By the design and implementation of our research: Personal preference analysis technology based on the social networks, with 30,000 user authorization acquired the Facebook personal data, further analyze 30,000 personal user preference data to recommend services to achieve the overall efficiency as: website browsing pages growth 11%, time on site growth 15%, site bounce rate dropped 13.8%, commodities click through rate growth 43%. Through the analysis structure of the system that planning and developing by our research, after experimentally verified and analyzed, the effectiveness of the system already has certain accuracy, and it can effectively achieve precise recommendation to enhance the contribution of the future research issue for social analysis. This is certainly one of our goals.

The purpose of this paper is how to quickly understand the personal preferences on social behavior of the user, and then carry out all the recommendation and other innovative application services. In the future, due to the rapid development of social media, through the analysis of the interest of personal preference, to reach a variety of innovative applications services, will be one of the next big issues of the social analysis. Therefore, personal preference analysis data conducted by the research can be used as reference of follow-up research, expected the follow-up research will be put forward further industrial application, population analysis, social policy ... etc, this will be the future of our concerned.

V. ACKNOWLEDGEMENT

This study is conducted under the "Social Intelligence Analysis Service Platform(3/4)" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China .

References

[1] Zhang, C. ; Fang, Y.A "Trust-Based Privacy-Preserving Friend Recommendation Scheme for Online Social Networks" IEEE Transactions on Dependable and Secure Computing, 2015, 12(4), 413-427.

[2] Louta, M.D. ; Varlamis, I. "A Trust-Aware System for Personalized User Recommendations in Social Networks", IEEE Transactions on Systems, Man, and Cybernetics, 2014, 44(4), 409-421.

[3] Cheng-Hung Tsai, Tsun Ku, and Wu-Fan Chien, "Object Architected Design and Efficient Dynamic Adjustment Mechanism of Distributed Web Crawlers", International Journal of Interdisciplinary Telecommunications and Networking, 2015, 7(2), 57-71.

[4] Rashid, A. “Who Am I? Analyzing Digital Personas in Cybercrime Investigations” IEEE Computer Magazines 46(4).
[5] Sheth, A., Thomas, C., & Mehra, P. “Continuous semantics to analyze real-time data.” IEEE Internet Computing, 2010, 14(6), 0084-89.
[6] Cheng-Hung, Tsai, Tsun Kua, Liang-Pu Chena, Ping-che, Yang, “Constructing Concept Space from Social Collaborative Editing,” Proceedings of AHFE 2015.
[7] Hui Zheng ; Ji-kun Yan ; Ye Jin “Persona Analysis with Text Topic Modelling” Proceedings of IEEE ACAI 2012.
[8] Shiga, A. , “A Support System for Making Persona Using Bayesian Network Analysis” Proceedings of IEEE ICBAKE 2013.
[9] Clark, J.W. , “Correlating a Persona to a Person,” Proceedings of IEEE SocialCom 2012.
[10] Cunha, B. ; Pereira, J.P. ; Gomes, S. ; Pereira, I. ; Santos, J.M. ; Abraham, A., “Using personas for supporting user modeling on scheduling systems”, Proceedings of IEEE HIS 2014.
[11] Letier, E. ; Sasse, M.A., “Building a National E-Service using Sentire experience report on the use of Sentire: A volere-based requirements framework driven by calibrated personas and simulated user feedback” Proceedings of IEEE RE 2014
[12] Bo Wu ; Qun Jin ; Nishimura, S. ; Julong Pan ; Wenbin Zheng ; Jianhua Ma, “Social Stream Organization Based on User Role Analysis for Participatory Information Recommendation” Proceedings of IEEE UMEDIA 2014
[13] Chengjie Mao ; Zhenxiong Yang ; HanJiang Lai, “A new team recommendation model with applications in social network”, Proceedings of IEEE CSCWDM 2014.
[14] Piedra, N. ; Chicaiza, J. ; Tovar, E., “Recommendation of OERs shared in social media based-on social networks analysis approach”, Proceedings of IEEE FIE 2014.
[15] Djauhari, M.A., “A recommendation on PLUS highway development: A social network analysis approach”, Proceedings of IEEE IEEM 2014.
[16] Mishra, S., “An analysis of positivity and negativity attributes of users in twitter”, Proceedings of IEEE ASONAM 2014.
[17] Lu Yang, “A collaborative filtering recommendation based on user profile and user behavior in online social networks”, Proceedings of IEEE ICSEC 2014.
[18] Huaimin Wang ; Gang Yin ; Ling, C.X. “Who Should Review this Pull-Request: Reviewer Recommendation to Expedite Crowd Collaboration” Proceedings of IEEE APSEC 2014.
[19] Castiglione, A. ; De Santis, A. “Friendship Recommendations in Online Social Networks” Proceedings of IEEE INCoS 2014.
[20] Shaffer, C.A. ; Weiguo Fan ; Fox, E.A., “Recommendation based on Deduced Social Networks in an educational digital library”, Proceedings of IEEE JCDL 2014
[21] Gerogiannis, V. ; Anthopoulos, L., “Project team selection based on social networks”, Proceedings of IEEE ITMC 2014.
[22] Diaby, M. ; Cataldi, M. ; Viennet, E. ; Aufaure, M.-A., “Field selection for job categorization and recommendation to social network users”, Proceedings of IEEE ASONAM 2014.
[23] Zegarra Rodriguez, D. ; Bressan, G. “Music recommendation system based on user's sentiments extracted from social networks”, Proceedings of IEEE ICCE 2014.
[24] Sun Yanshen ; Yu Xiaomei, “Personalized recommendation based on link prediction in dynamic super-networks”, Proceedings of IEEE ICCCNT 2014.
[25] Xueming Qian ; Dan Lu ; Xingsong Hou ; Liejun Wang, “Personalized tag recommendation for Flickr users”, Proceedings of IEEE ICME 2014.