Clickstream Analysis on WMC Platform

[1]Heng-Li Yang, [2]Qing-Feng Lin
[1] *Dept. of MIS, National Cheng-Chi University, yanh@nccu.edu.tw*
[2] *Dept. of MIS, National Cheng-Chi University, qflin@nccu.edu.tw*
doi:10.4156/jcit.vol6.issue1.25

Abstract

With the rapid development of information technology and Internet, more and more new applications emerged to challenge the way we live. Microsoft introduced a digital media entertainment platform called Windows Media Center (WMC). On this platform, users can easily access media services, such like online movies, radio, videos, games and photos from registered service providers by a special personal computer. In this paper we introduce an approach to discover user behavior on this platform. Our approach not only focuses on the rules between pages viewed by user but also focuses on what they have done in each page.

Keywords*: User Behavior, Clickstream, WMC, Data Mining*

1. Introduction

In current era, digital home technology integration is more and more important. It can bring a lot of new application. In old days, users' behavior in a living room such like "how a user control TV by his remote controller" is not easy to be collected. However, collecting behavior data in digital home technology integration application is relatively easy. In this study, we collaborated with a Taiwan WMC (Widows Media Center) service provider and mined some user behavior rules for helping their decision making.

2. Windows Media Center

Nowadays using information technology to support both work and daily life becomes very common. On one side, your personal computer can be easily used as a video player, music player, television or radio to support personal entertainment by installing some software. On the other side, your traditional entertainment equipments are also changing. Television can connect with a personal computer as a monitor. CD player can support MP3 format files, not just traditional music track. Computer industry and home entertainment equipment industry had a dramatic digital convergence and many of new applications emerged in this context.

In 2002, Microsoft introduced "Widows Media Center" platform. This platform includes a new media center edition operation system integrated home theater PC (HTPC), media service providers, network service provider as shown in Figure 1.

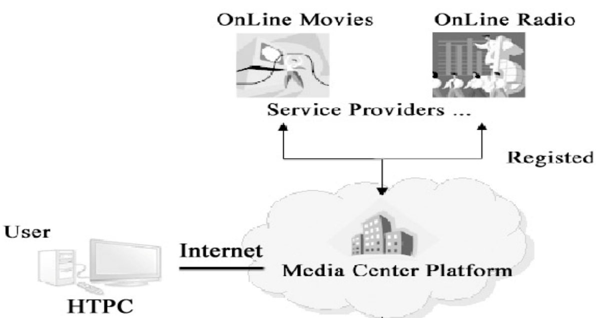


Figure 1. Windows Media Center Platform

With a simple hardware configuration (Figure 2), users can easily connect the platform. Users can sit in sofa and use remote controller to access this online multimedia service. WMC provides a simple and convenient interface to connect an integrated multimedia content (Figure 3).

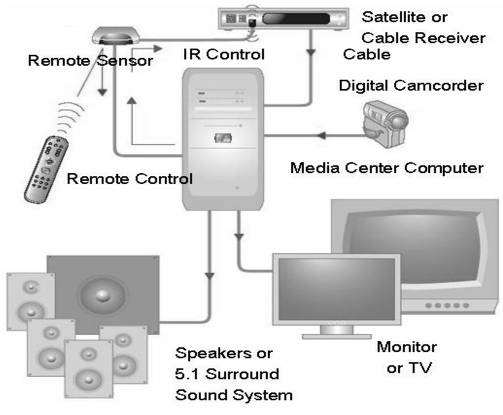


Figure 2. Hardware configuration of Widows Media Center platform

Figure 3. User interface of Widows Media Center platform

3. Data Warehousing, Data Mining and Clickstream Analysis

With the rapid development of information technology, log recording technology and data management system, data analysis approach obtained a great number of attentions. A data warehouse is a repository of all required business information or the single logical storehouse of all the information used to report on the business [1]. Data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision making process [2]. Data mining in data warehouse, also called knowledge discover in data warehouse (KDD), is a process of nontrivial extraction of implicit, previous, unknown and potentially useful knowledge from a large amount, incomplete of noisy, fuzzy and random data [3]. There are many previous data mining studies [4,5,6,7,8,9,10].Web mining is a sub-branch of data mining. There are three types of web mining: web content mining on page contents or search results, web structure mining on website reference hyperlinks, and web usage mining on web logs [11]. Clickstream analysis is a web usage mining. A clickstream is the recording of what a user clicks on while browsing or using another software application [12], in which users can only use very limited types of keys (mainly up, down, left, right, back, enter and home) on the remote controller to move cursor on the system. A sample of clickstream logs can be shown as Table 1.

Table 1. Clickstream log in WMC platform

Clickstream Log			
Time	**Page**	**Tag**	**Action**
2010-03-04 12:00:512	Purchase	-	Home
2010-03-04 12:13:343	Movie List	MC_p1_p2	→
2010-03-04 12:21:128	Movie List	MC_p2_p6	↓
2010-03-04 12:22:032	Movie List	MC_p6_p7	→
2010-03-04 12:23:321	Movie List	MC_p7_p8	→
2010-03-04 12:32:214	Movie List	SM_p8	Enter

In Table 1, the column of time and page can be recorded by web system automatically. The "action" column records what buttons had been pressed by user and the "tag" column records more details on user behavior in this action. For example, the second record indicates "at 2010-03-04 12:13:343, on Movie List page, this user performs right (→) key, which moves the cursor from p2 to p6 place". These two types of data can be recorded by a customized tagging system.

4. Building a analysis system on WMC platform

In this paper, we had cooperated with a WMC service provider in Taiwan and helped them establish a system to find behaviors of their users. The system architecture in our work is shown in Figure 4. In the architecture, ETL (extract, transform, and load) is a process involves extracting data from outside sources, transforming it to fit analysis needs, and loading it into database. It needs to refer to outside database (OSDB), which stores some code tables (e.g. page codes).

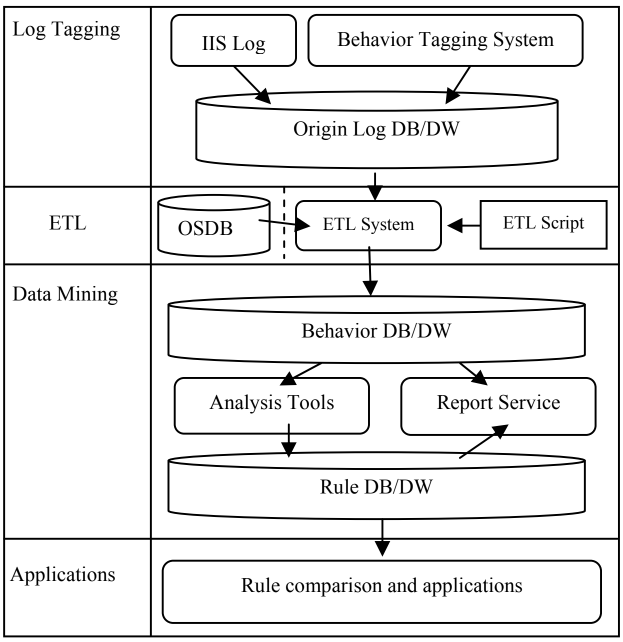


Figure 4. System architecture

The core components of this architecture are behavior tagging system and the analysis tools. Because WMC platform is used in the living room and controlled by a remote controller, the user interface of WMC platform is more simple and intuitive than regular web. If we record only user action such like "→ → ↓ → Enter", we will never know what happen on the screen. So, we introduce a behavior tagging system to tag more behavior data on this platform. This tagging system is compatible with IIS web log system. Tag logs processed by this tagging system will be saved in the same system log file with web system logs which processed by IIS web server. We can use a parser tool to save both log data at one time. Because we face a quite interesting log format, we need to consider not only the patterns in page and actions but also the waiting time and the products user had passed or selected. We adopted our previously proposed algorithm [13] to analyze the data.

5. Findings

After implementation, we had found some interesting results. Firstly, like Table 2, we can use this system to find user behavior rules on WMC platform. Rule #1 in Table 2 told us "if a user moving his cursor on c220 and wait over 20 seconds in the Movie List page, then he gets 90% possibility will to play p1902 movie in the next Play Movie page over an hour". Rules such like rule #1 can help developer customize their system make it more effective. Rule #2 in Table 2 told us "if a user moving his cursor on c230 in the Movie List page and he plays movie p1201 on the next Play Movie page, then he gets 87% possibility that he came from Promo page and he had moved his cursor on C01". Rules such like rule #2 can help developer make a short cut in their system for user, make it more efficient.

Table 2. User behaviour rules on WMC platform

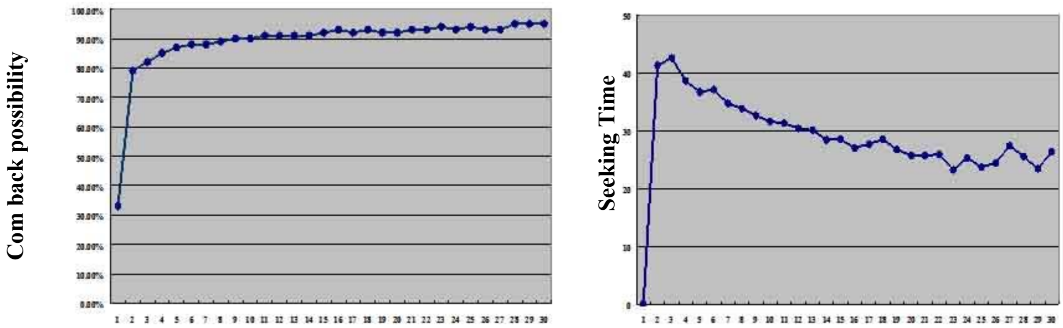
#	LHS	RHS	Type	%
1	[Movie List][MC] on c220 over [20 sec].	**[Movie List][MC] on c220 over [20 sec]** → [Play Movie][Play] on p1902 over [1 hour]	→	90%
2	[Movie List][MC] on c230→[Play Movie][Play] on p1201	[Promo][MC] on C01→**[Movie List][MC] on c230→[Play Movie][Play] on p1201**	←	87%

Secondly, our system can help user segmentation. We applied user behavior segmentation analysis to find clustering rules like Table 3. Not applying demographic data, we used three behavior variables and four indexes to segment users in our system. There are (a) recency — number of hours from last login time till now, (b) frequency— number of login times in a month, (c) Monetary — number of kinds of product used by a user and number of times of product used by a user. Managers can use this segmentation rules to understand how their diamond (Lv5) customers act and they can also use this rules to organize some promotion for transforming Lv4 user into Lv5.

Table 3. User segmentation rules

	Low <<<<<< Importance <<<<<< High					
	Lv1	**Lv2**	**Lv3**	**Lv4**	**Lv5**	**Unit**
last login time till now	>2503	937~2502	386~936	119~385	1~118	Hours
login times in a month	1~7	8~23	24~59	60~175	>176	Times
kinds of product used	1~3	4~8	9~22	23~69	>70	Kinds
times of product used	0~41	42~155	156~427	428~1008	>1009	Times

Thirdly, our systems can analyze user comeback possibility. By analyzing our data, we can find a comeback possibility diagram like Figure.5 (a). The x-axis represents for "number of user had login", the y-axis represents for "the possibility of next login". In Figure.4, if a user already used our system for 9 times, the possibility he will have 10th or more login are more than 90%. This rule has become a golden rule now for our service provider.



number of login times. (a) User login times and his possibility to come back

number of login times. (b) User login times and his seeking time

Figure 5. Rules of others findings

Finally, our systems can analyze the time between a user login and use his first product in a session; we call the time as "seeking time". By analyzing our data, we can find a seeking time diagram like Figure.5 (b). The x-axis represents for "number of user had login", the y-axis represents for "the seeking time". In Figure.5, if a user already used our system for 12 times, the average seeking time will be less than 30 seconds .This rule has been also used as a key rule for our service provider.

6. Conclusion

In this paper we introduce a digital media entertainment platform called Windows Media Center (WMC) and try to establish a platform to collect and analysis user behavior on this WMC service. We also show what we can find in this system. There are several kinds of clickstream behavior rules, segmentation rules, association between login times and come back possibility, and association between login times and his seeking time.

7. Acknowledgments

The authors would thank the National Science Council, Taiwan, for financially supporting this research under contract NSC 99-2410-H-004-102-MY2.

8. References

[1] B. A. Devlin, P. T. Murphy, "An architecture for a business and information system", IBM Systems Journal, vol. 27, no.1, pp.60-80, 1998.
[2] W. H. Inmon, Build the Data Warehouse, John Wiley & Sons, USA, 1993.
[3] W. Frawley, G. Piatetsky-Shapiro, and C. Matheus, "Knowledge discovery in databases: An Overview", AI Magazine, vol. 13, no.3, pp.57-70. 1992.
[4] R. Agrawal, S. Ghosh, T. Imielinski, B. Iyer, and A. Swami, "An interval classifier for database mining applications", In Proceeding of the 18th VLDB Conference, Vancouver, Canada, August, pp.560-573, 1992.
[5] R. Agrawal, T. Imielinski, and A. Swami, "Database mining: A performance perspective", IEEE Transactions on Knowledge, vol.5, no.6, pp.914-925, 1993.
[6] R. Agrawal, R. Srikant, "Fast algorithms for mining association rules", In Proceedings of the 20th International Conference on Very Large Data Bases, Santiago, Chile, pp.487-499, 1994,

[7] J. Han, Y. Cai, and N. Cercone, "Data-driven discovery of quantitative rules in relational databases", IEEE Trans. on Knowledge and Data Eng., vol.5, pp.29-40, 1993.
[8] J. Han, K. Koperski, and J. Adhikary, "Spatial data mining and challenges", In 1996 SIGMOD'96 Workshop on Research Issues on Data Mining and Knowledge Discovery(DMKD'96), Montreal, Canada, June, 1996.
[9] J. Han, Q. Yang, and E. Kim, "Plan mining by divide-and-conquer", In Proc. 1999 SIGMOD'99 Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'99) , Philadelphia, PA, pp.8:1-8:6, May, 1999.
[10] H. Imai, M. Inaba, and N. Katoh, "Experimental results of randomized clustering algorithm", In Proceedings of the twelfth annual symposium on Computational geometry, pp.20-25, 1996.
[11] R. Cooley, B. Mobasher, and J. Srivastava, "Web mining: Information and pattern discovery on the World Wide Web", In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), November 1997.
[12] W. W. Moe, P. S. Fader, "Capturing evolving visit behavior in clickstream data", Journal of Interactive Marketing, vol. 18, no. 1, pp.5-19, 2004.
[13] H. L. Yang, Q. F. Lin, "A new algorithm to discover page-action rules on web", Lecture Notes in Artificial Intelligence, no. 3683, Part III, pp.1331-1337, 2005.