

User Behavior Mining in Software as a Service Environment

B.S. Naveen Kumar, D.R. Kumar Raja and Olive Esther Kumar

Abstract--- Software as a Service (SaaS) provides software application vendors a Web based delivery model to serve large number of clients with multi-tenancy based infrastructure and application sharing architecture. With the growing of the SaaS business, data mining in the environment becomes a challenging area. In this paper, we suggest a new metric along with a few existing metrics for customer analysis in a Software as a Service environment.

Keywords--- Software as a Service, SaaS, Customer Behavior Analysis, Data Mining in SaaS Environment

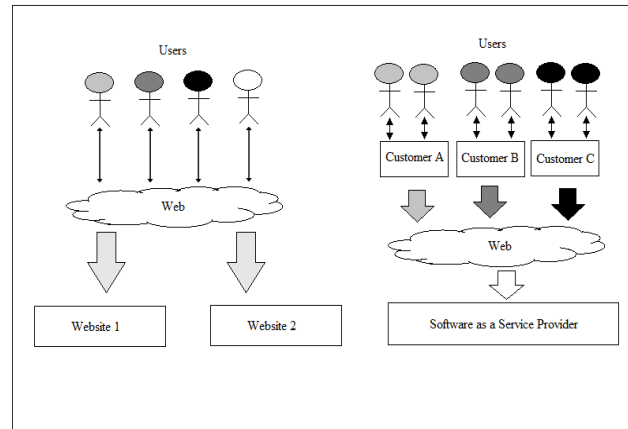


Figure 1

I. INTRODUCTION

With the rapid development of Internet Technology and the application software usage, SaaS (Software as a Service) as a complete innovative model of software application delivery model is attracting more and more customers to use SaaS for reducing the software purchase and maintenance costs as it can provide on-demand application software, and the users can adjust the functions provided by services to meet changes in demand. SaaS is gaining speed with the considerable increase in the number of vendors moving into this space[1]. The SaaS model is different from a regular website model. In a regular website model, users of the software directly interact with the software application. But in the case of a SaaS model, users interact with the application through the service provider. The difference between a regular website model and a SaaS model can be shown in figure 1.

B.S. Naveen Kumar, Software Engineer, HCL Technologies, Chennai, India

D.R. Kumar Raja, Assistant Professor, Sree Vidyanikethan Engineering College, Tirupathi, India

Olive Esther Kumar, Madanapalle Institute of Technology and Science, Madanapalle, India

II. MOTIVATION

Software as a Service (SaaS) is being adopted by more and more software application vendors and enterprises [2]. SaaS is beneficial for the customers in such a way that, a customer can unsubscribe from the services whenever he wants which makes it a challenge to manage customer relationships. One of the characteristics of the SaaS business model is that one SaaS service needs to serve a large number of customers, among which considerable portion are customers for whom services are offered on trial basis. As there is competition in the market, both trial and paying customers may move their business to another service provider based on their requirements. It is essential for a service provider to retain the customers from migrating to another service provider. Previous studies show that a small increase in retention rate would lead to a considerable increment in the new present value of the customers. To withstand the competition in the market, a service provider should satisfy the customers by understanding their current behavior and predicting their next move like if they are having any problems in using the services, how much are the

customers satisfied based on the seriousness and activeness of the customers.

III. RELATED WORKS

A lot of work has been done in the area of analyzing the customers' behavior on website model. Various methodologies are stated by various authors on various processes in mining the web. In [3] Sindhu P Menon and Nagaratna P Hegde, analyzed the views and methodologies stated by various authors on various processes in web mining.

In [4] R. Suguna and D. Sharmila listed out work done by various authors in the web usage mining area.

In [5] the authors Jiehui Ju. Et.al, gives a quick survey on SaaS. It covers key technologies in SaaS, difference between Application Software Provider and Software as a Service Provider, SaaS architecture and SaaS maturity model and the advantages that SaaS offers to small businesses.

In [6], the authors Espadaset. al, presents the analysis of the impart of a set of requirements and proposes guidelines to be applied for application deployment in Software as a Service (SaaS) Environment.

In [7], the authors Ning Duan, et. al, proposed an algorithm and two metrics which work with the collaboration among the users of a customer in a Software as a Service environment.

IV. PROBLEM DEFINITION

In a SaaS Environment, an effective relationship with the customer depends on how much the status of each customer is understood. In order to understand the status of a customer, it is necessary to study the behavior of the customer from time to time. It is necessary to predict the customers' seriousness and activeness in using the service. This prediction may help the service providers in improving their business strategies. In a business to customer website model, the mining is done based on selected metrics like visit frequency, average depth, average stay time etc. In the case of SaaS model, there is another level of users who

actually use the service. So, regular user behavior metrics may not yield accurate results in the case of SaaS model. If individual customer's user's behavior is studied, then the difference between the customers may be identified.

A lot of research is done on user behavior analysis in regular website model but those methods used for user behavior analysis may not guarantee accurate predictions. So an extra parameter or metric is to be considered. As in the SaaS model, a tenant is the direct customer of the service provider and the actual users of the service are the users of the customers, one way to study the behavior of the customers may be by summing up the individual user's metrics of a customer to evaluate the customer's behavior. But this way ignore the individual differences of the behaviors of the users of a customer. In addition to these regular web usage mining metrics if collaboration among the users is also considered in the analysis of customer behavior, it may yield better results than just using the regular metrics. But previous works done in user behavior analysis in SaaS uses only collaboration metrics in the analysis which ignores almost half of the analysis data.

The experiment done aims at using collaboration metrics along with another metric which works with the data not considered in the collaboration metric calculation so that all the available data is considered in user behavior prediction.

V. EXPERIMENT

The application we considered here is a mobile College Social Networking application which works on SaaS model. We refer to this application as "InTouch" which is hosted on a local server. The roles in this application are the Service Provider, one who provides this application software as a service, various colleges acting as Tenants who subscribe for this application software service and the users are the faculty, students, and staff of a college.

The experiment is done in two phases, namely Data Collection Phase and Data Processing Phase. In the Data collection phase, the necessary data (like transaction history which is recorded in a table) are collected. In "InTouch", a

separate table is maintained to record all the activities in the application which is used in the second phase of Data processing for analysis. In the second phase i.e. in Data Processing phase, the actual analysis takes place. This phase is further divided into individual modules like preprocessing, pattern discovery, and pattern analysis.

Preprocessing is a process of refining the sever log data and transaction history removing noise in data (if any) and populating database for further use in next modules. It includes data cleaning, user identification, session identification, transaction identification.

Pattern Discovery is the process of discovering the usage patterns from the cleaned raw log data. As in this experiment, it is not regular usage patterns that are to be considered, collaboration patterns are to be considered. Regular usage patterns are the sequences of activities that are performed by the users individually. But, collaboration patterns are those that are performed by users by interaction. Collaboration patterns are not the transaction patterns rather they are the patterns of users that collaborate to perform a transaction.

Definition of Collaboration: Collaboration is defined to happen when different users of a customer work on the same business object during a certain period of time. For example, in a Human Resource management SaaS service, the vacation request is submitted by a regular employee user of a customer and then is approved/rejected by manager user of the same customer. Here two users of a customer are involved in the process of granting a leave. This is called collaboration.

After the raw data is cleansed, the data used for collaboration discovery may contain details of the transactions performed by the users of any tenant with tenant id(tid), user id (uid), transaction id (transaction_id) (may also be called business object id), date, time, service id (sid). In this table more than one user may be involved in the completing of a transaction.

Algorithm: Collaboration User Set Identification

Input: Table 1 that consists of the transaction details

Output: Collaboration_Table with collaboration transaction details

Initially Collaboration_Table is empty

- Get first record from Table 1
- Insert details into Collaboration_Table
- While end of table 1 not reached
- Get next record from table 1
- Search for transaction_id in collaboration_Table
- If found, update collaboration user set and no_of_users
- Else Add details to collaboration_table as new record

Table 1: Sample Table Showing the Contents of Table 1

Transaction Id	Tenant Id	User Id	Date	Time	Activity
101	T001	Stud306	12/01/2010	8:00	Events
102	T001	Fac021	12/01/2010	8:30	Exam
101	T001	Fac05	12/01/2010	8:35	Comment
103	T001	staf06	12/01/2010	9:00	Forward

Table 2: Sample Collaboration Table

Transaction_id	Tenant_id	User set	No. of Users
101	T001	Stud306, Fac05	2
102	T001	Fac021	1
103	T001	staf06	1

Pattern analysis plays vital role in the experiment. This module deals with the behavior analysis based on the collaboration patterns extracted above.

From [7], there are two type of collaboration. They are random collaboration and repeated collaboration by certain group of users. The first type of collaboration can indicate the activeness of the customer no matter which users are involved in the collaboration process. It can be called as Active Collaboration Index (ACI). The second type of collaboration can be described by the usage patterns among the users of a customer. It can be called Patterned Collaboration Index (PCI). A high ACI value tells that a

customer is actively using the SaaS service and if such a customer is still a trial customer, it probably shall be the high priority target to get it converted into paying customer. A high PCI value tells that a tenant is seriously using the SaaS service with relatively strong loyalty, cross-selling or up-selling opportunity can be explored for such a customer. The formula to calculate ACI and PCI are as follows

$$ACI = \frac{\text{total collaboration transactions of a tenant}}{\text{total transaction of a tenant}} \cdot AppCNorm$$

$$AppCNorm = \frac{\text{total collaborations}}{\text{total transactions}}$$

The AppCNorm is the normalizing factor indicating collaboration characteristic of SaaS service. While some SaaS service are rich with collaborations and others may not be. In order to balance the difference among different SaaS services, this normalization factor is employed.

$$PCI_n = \sum_{i=1}^m Supp(P_{ni}) * (len(P_{ni}))^2, n = 1, \dots, N$$

Where P_{ni} denotes the collaboration pattern i of customer n , N is the total number of customers, and m is the total number of patterns in customer n . $supp(p_{ni})$ is the support value of pattern P_{ni} , and $len(P_{ni})$ is the length of the pattern.

These collaboration metrics works only with the collaboration data and neglects the remaining data which is almost half of the data. Hence another metric can be added along with the above metrics which considers the non-collaboration transactions. As the new metric is for non-collaboration transactions of a tenant, it can be called Average Usage Index (AUI). This can be calculated using the formula

$$AUI = \frac{\text{total non collaboration transactions of a tenant}}{\text{total number of transactions of a tenant}} \cdot AppUNorm$$

$$AppUNorm = \frac{\text{total non collaboration transactions}}{\text{total number of transactions}}$$

This AUI increases the accuracy of prediction of activeness of the customer along with ACI.

VI. RESULTS

For this experiment, the data created is for 100 customers of a Software as a Service provider who is providing 6 different components of an application as different services. Among these 100 customers, first 50 are taken as paid customers and the other 50 are taken as trial customers.

Table 3: Summary of Transactions

Customer type	Collaboration user set length distribution			
Length of pattern	1	2	3	4
Paid Customer	43158	23102	14411	8647
Trial Customer	34241	17940	11357	6789

Table 4: Sample Pattern List

Type of customer	Customer id	Collaboration pattern		
		Pattern	Support	Length
Paid	T0	Staf	0.31873200	1
		Stud, Fac	0.04553314	2
		Stud, Fac, Stud	0.06167147	3
		Fac, Stud, Staf, Fac	0.01095100	4
	T01	Fac	0.07826087	1
		Fac, Staf	0.04347826	2
		Staf, Stud, Staf	0.00489130	3
		Fac, Staf, Stud, Staf	0.00815217	4
Trial	T50	Stud	0.09480676	1
		Stud, Stud	0.01449275	2
		Fac, Staf, Fac	0.01570048	3
		Stud, Fac, Stud, Stud	0.00120772	4
	T51	Stud	0.09698276	1
		Fac, Fac	0.02101293	2
		Stud, Fac, Stud	0.01131465	3
		Stud, Stud, Fac, Stud	0.00215517	4

Table 5: Sample Calculated Metrics

Type of customer	Customer id	ACI	AUI	PCI
Paid	T0	0.00150	1.99625	0.33214
	T45	2.96354	1.80213	4.55162
Trial	T50	1.04676	0.95008	2.56944
	T89	0.0	0.03214	0.00231

From the above calculated values, we can observe that though T0 is a paid customer, less ACI and PCI values indicate that this customer is not using the services to the full and hence revenue generated from this particular customer is not appreciable. Rather, this customer may be planning to unsubscribe from the service and hence is an important target for the service provider to retain the customer. In the case of T45, it has high ACI value, high AUI value

indicating active usage of the services and high PCI indicating that this customer is completely migrating his business onto the SaaS service generating the service provider more revenue. Among the sample trial calculated values, customer T50 is active and serious and hence, there is a high probability for this customer to convert into paid customer. On the other hand, customer T89 is not very active and is not serious indicating that he may be facing technical difficulties in using the services and hence should be helped with or is thinking to unsubscribe from the services.

Table 6: Summary of Calculated Metrics

Type of custr	ACI			AUI			PCI		
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
Paid	0	3.665	1.346	0	3.015	1.765	0	4.878	2.66
Trial	0	1.325	0.661	0	1.865	0.394	0	3.774	1.26

From the above table, for any tenant to be considered active in using the services, minimum ACI and AUI values needed are 1 and 1 respectively and minimum PCI value needed is 2.

VII. CONCLUSION

The metrics ACI and PCI are introduced in previous works done by NingDaun, et. al in [7] which works with collaboration data and leaving the non collaboration data. In our work, a new metric is introduced AUI which considers the non collaboration data also in customer behavior analysis. Still further, frequent pattern analysis can be applied on this non collaboration data to get usage patterns and so the analysis can be further improved.

REFERENCES

[1] Wei Sun, Xing Zhang, Chang Jie Gou, Pei Sun, Hui Su, IBM China Research Lab, Beijing 100094, "Software as a Service: Configuration and Customization Perspective" IEEE Congress on Services Part II, IEEE 2008.
 [2] E. Knorr, " Software as a Service: The Next Big Thing", http://www.infoworld.com/article/06/03/20/76103_12FEsaas_1.html

[3] Sindhu P Menon, Nagaratna P Hegde, "Requisite for Web Usage Mining – A Survey", Special Issue of International Journal of Computer Science & Informatics: 2231-5292, Vol-II, Issue-1, 2, pp. 209-215.
 [4] R. Suguna, D. Sharmila "An Overview of Web Usage Mining", International Conference of Computer Applications (0975 – 8887), Vol. 39, No, 13, February 2012, pp. 11 – 13.
 [5] Jiehui, et. al, "Research on Key Technologi=ies in SaaS", International Conference on Intelligent Computing and Cognitive Informatics, 2010, pp. 384-387.
 [6] Espadas et. al, "Application Development over Software-as-a-Service platforms", The Third International Conference on Software Engineering Advances, 2008, pp. 87-104.
 [7] Ning Duan, et. al, "Tenant Behavior analysis in Software as a Service Environment": Service Operations, Logistics and Informatics (SOLI), 2011 IEEE International Conference, pp 132-137, July 2011.