A Data Mining Based Approach for Introducing Products in SaaS (Software as a Service)

R. Kamalraj  
Asst. Professor CSE, SNS College of Technology, Coimbatore, Tamil Nadu, INDIA  
mailtokamalraj@yahoo.com

Dr. A. Rajiv Kannan  
Professor CSE, KSR College of Engineering, Tiruchengode, Tamil Nadu, INDIA  
raji5757@yahoo.com

Mrs. S. Vaishnavi  
I ME / SE, SNS College of Technology, Coimbatore, Tamil Nadu, INDIA

Mrs. V. Suganya  
I ME / SE, SNS College of Technology, Coimbatore, Tamil Nadu, INDIA

Abstract – Introducing new products in cloud based environment that may not be reached related customer. To find exact customers who have interest on new product is a crucial activity in the business. So here we proposed Data Mining Approach ‘Association Rule Learning’ to create different associations from various products’ clusters in that cloud. That may reduce the time or effort for advertising products to the customer side. In this paper we have used Market Basket Analysis also for improving business gain by applying the Data Mining concepts on Cloud Computing.

Key Words – Cloud Computing, Clustering, Association Rule Learning, Data Mining, Market Basket Analysis.

I. INTRODUCTION

Cloud Computing is a technology to control remote servers and maintain data to which consumers will connect to them and use resources whenever they are needed. It delivers the service to the consumers through which they can utilize the shared resources and information. These services are provided to computer and consumers access those services over a network. It allows the consumers to use the applications without installation and access their personal files at any computer with internet access. Cloud computing supports centralized storage. It includes any subscription based or pay-per-use service. Consumers can access cloud based applications at remote location through web browser. The name cloud computing was inspired by the cloud symbol that is to represent the internet in flow charts and diagrams.

Traditional business applications are very complicated and expensive. Customers need of hardware and software is increasing day-to-day life. They need the process of installation, configuration, testing, run and update on software and hardware. This consumes more time and cost. This leads to the tremendous emergence of Cloud Computing. Cloud based applications can run up for days or weeks and they cost less.

Cloud service has three characteristics different from traditional services. They are
1. It is sold on demand-for minutes or hours
2. It is elastic-can consume little or much service needed.
3. Service is fully managed by provider. Consumer just needs to have a personal computer and internet.

A. Different forms of Cloud
Cloud is categorized as two forms as per its service usage. Those are
1. Public Cloud
2. Private Cloud

Public Cloud – Public cloud service providers can attract customers from all over the world. A service provider can have different cloud services type to be with customer for continuing their business. It gives a very good advantage like customers can link or disconnect their business with their cloud service. Suppose a customer is using a service from one Service Provider, it is possible to switch over from one Service Provider to another Service Provider as per his requirement on resources for doing his business process.

Private Cloud – In this form the coverage area is limited like only within an organization the services are utilized by users or persons belong to that organization. So here user can get prescribed set of services to be used for their business. If they need any additional or advanced service they have to spend some amount of money to be ready for their usage. In other words, the breadth of the services list has to be increased dynamically.

A cloud can be public or private. A public cloud is one that gives service to anyone on the internet whereas private cloud is one that gives services only within an organization. When service provider uses public cloud to create private cloud then it is called Virtual private cloud. Goal of cloud computing is to provide scalable and easy access to resources for consumers. A simple example of cloud computing is gmail, yahoo mail etc. for sending mails, they just need a internet connection and doesn't want any server or software.

II. CLOUD SERVICES TYPES

There are many categories of service such as platform, storage, application, security, testing and infrastructure. Some problems arise in need of security-as-a-service and some as platform-as-a-service [4]. Right service should be chosen for appropriate type of problem.
Cloud Computing provides services related to fundamental models are shown in the below figure. Where,

IaaS - Infrastructure-as-a-Service
PaaS - Platform-as-a-Service
SaaS - Software-as-a-Service

![Types of Cloud Services](image)

**A. Infrastructure-as-a-Service**
This is the base for other two layers. In this service, cloud provides computers as physical or virtual. Consumers can have the choice of their virtual computer i.e., they can select a configuration of CPU, memory, storage which is needed for their application. Service providers provide the whole cloud infrastructure which includes servers, firewalls, load balancers and other network components. In Enterprise level, company has to pay to cloud providers as much capacity as is needed. Advantages if IaaS is pay per use, instant scalability, security, reliability and APIs. Example is Amazon Web Services. They provide virtual server to consumers. IaaS service is based on a utility computing basis since amount of resources consumed by consumers will reflect the cost for the service. It is like offering power and storage on demand.

**B. Platform-as-a-Service**
This type of service provides development platform to the cloud upon which new applications can be developed. Developers can develop code for an application and can run that application on the cloud using the hardware or software platform on the cloud. They need not to install those hardware or software on their own PC. Providers have set of software and product development tools in their infrastructure. PaaS layer lies between SaaS and IaaS. They avoid the complexity of buying and managing the hardware and software for the application. Paas offers faster and cost effective service for application development and delivery. Force.com and Google are examples of Paas. Developers create their own code and this service presents it in the web.

**C. Software-as-a-Service**
Instead of installing the hardware and software, this service provides the whole application suite to the consumers [11]. They provide access to applications and also gives security, availability and performance. Cloud providers just install the application software and consumers can directly use it. So consumers need not maintain the platform and infrastructure. Service provider hosts both data and application so end user is free to use the service from anywhere. It is renting the application from service provider and not buying it. It gives greater flexibility and less maintenance. Consumers have no burden of buy, install, maintenance and update. This type of service provides single application to thousands of consumers. Access to application is easy.

**III. PROPOSED TECHNIQUE**
In cloud business, apart from required services of customers, the service provider has to attract the customers by introducing products.

Different ways for attracting customer for new product can be done in some ways. Those are

1. Giving Advertisement of New Product or Service
2. New Product may be provided as a Trial version with some limits on usage.

So reaching more number of customers the new product has to be client side for satisfaction of customer's needs. For this, the products of different services have to be clustered or grouped as per their product or service characteristics. For improving cloud business, Data Mining approaches such as 'Clustering' and 'Association Rule Learning' are introduced on SaaS cloud platform for product classification and introducing new product to the customer side [2], [3].

**A. Data Mining**
Data Processing is a very important task in any business environment for analyzing and taking decisions as per the result coming from 'Data Analysis' phase [7]. So Data Mining is a powerful concept for fulfilling the requirement in any business types. Data Mining is dealt with producing different methods and techniques to process a large volume of data to get expected results within a minimum time gap [8].

For data processing, following steps to be followed one by one. They are,

i) Data has to be collected from various data resources,
ii) It should be organized in a structure,
iii) Applying algorithm to group the data as per their characteristic
iv) Applying algorithms for required result from the collected data.

They are named as 'Data Collection', 'Data Classification' and 'Data Clustering', and 'Decision Tree' for taking decisions as per collected data sets from different data sources.
C. Association Rule Learning

Association Rule Learning provides relationship between variables in the database [1], [9]. It searches for strong rules to determine the occurrence of item in the database. So, it is based on the rule which is “If item A the Item B”.

The rule of finding relationship will help in marketing the product. It must specify minimum thresholds of support and confidence. It aims at finding frequent patterns among sets if items in transaction database. They are rules which implement if/then statements. Using these rules, it is able to uncover relationships in the database or other repository. This has two parts namely: First, minimum support has to be applied to the item set to find frequent items. Second, rules are based on common sets of items and the minimum confidence constraints. All possible subsets are considered to find the frequent item set. Support indicates how frequently the item occurs in database. Confidence discovers number of times the if/then statements have been found to be true. Association rules are used for analyzing customer behavior. They are mainly used in market basket analysis. It is also applied in banking services, purchases using credit and debit cards, supermarkets, and medical field. There are three types of association rules: Actionable Rules, Trivial Rules and Inexplicable Rules.

D. Market Basket Analysis

It is used for recognition of dependencies in data. It is based on the concept that if a collection of items which are bought by certain group of people then they are likely to buy another group of items. The set of items bought by customers is called item set. Market analysis finds the relationship between purchases made by customer [5]. The relationship is in the form:

IF {milk, Coffee powder} THEN {Sugar}

It is that if a customer buys milk and coffee powder then he/she is likely to buy sugar.

It identifies the purchasing behavior of customer. Market analysis finds who are all the customers and discovers why the item is bought. So using this analysis the product which is most liked by customer is found to know how the market of product is. The area where market basket analysis is used are credit card purchase, insurance claims and telephone calling patterns.

Step 1 - Clustering Analysis for Product Classification

It is one of the efficient approaches from Data Mining to create different groups of data items as per the different framed characteristics. This will reduce the effort on identifying a particular data set in a huge data collection. For an example, in SaaS different types of software can be utilized by customers, which is given in the following figure.

B. Clustering

Clustering aims at finding variables which have similar characteristics or behave in similar manner [6]. Cluster analysis is the process of discovering groups of objects with similar or related to one another but differ from other clusters. It finds a structure among unlabeled data. First, raw data is given and clustering algorithm is applied due to which we obtain clusters. It is useful for simplification and pattern detection. It is the process if partitioning the set of objects into meaningful subclasses. A good cluster is one which has high quality characters. Quality depends on measure of similarity and hidden patterns. Similarity can be measured by using distance function d(i,j). Clustering is sometimes defined with classification. In classification there is pre-defined class whereas in clustering classes are also to be defined. In intra-class cluster similarity is high and in inter-class cluster similarity is low.

**Fig. 2. Block Diagram of Proposed Technique**

Identifying new product which have to be marketed

Categorize the new product according to service

Cluster the products by means of category

Determine the frequently moving product

Introduce the new product with frequently moving product

Collect feedback

Decision about the new product in the market
The application such as games can be categorized to different types namely Shooting game, Race game, Mind blowing etc. There are many shooting games available from different organization with new trends. These games are grouped according to their characteristics which forms cluster. Similarly other type of games are clustered.

Procedure for creating Cluster

Start

//reading software attributes
if type[SaaS] == 'Game'
    if type[Game] == 'Shooting'
        Cluster[Shooting] = 'Game Name'
    else if type[Game] == 'Race'
        Cluster[Race] = 'Game Name'
    else
        Cluster[MindBlop] = 'Game Name'
else
    type[SaaS] = 'Business'

End

Step 2 – Association Rule for Introducing new Product

The Association Rule Learning is focusing on relating different data items from available clusters as per the requirement. Multiple associations of related items may reduce the effort on advertising new products. The given below table shows different clusters of Game Software available as a service in SaaS.

Consider three clusters with number of products in it such as X, Y, Z of particular service.

Cluster X have products such as x1, x2, x3, x4, x5. Let x2 be the frequently moving product in the market and x5 be the new product to be introduced for marketing. Cluster Y have products such as y1, y2, y3. Let y3 be the frequently moving product in the market and y1 be the new product. Cluster Z have products such as z1, z2, z3, z4. Let z4 be the frequently moving product in the market and z1 be the new product to be introduced. The clusters are formed using clustering techniques.

By associating a new product with the frequently moving product for Eg y1 with x2, so that y1 will get introduced in the market. The customers using x2 will get a chance to use y1 in their routine life so that y1 can be easily marketed. According to the feedback of the customers about y1 the productivity and sales of the y1 can be increased or decreased.

IV. RESULTS AND DISCUSSIONS

In this paper we proposed Data Mining approaches to make groups of products in SaaS to reach customers with new products for improving gain in business. These approaches are applied for Market Basket Analysis on products in Cloud Computing.

By analyzing each cluster in each game, the game which is frequently used by customers can be identified and while delivering frequently used game the new product is associated with it as a 'Trial Version'. Associating a new product as a free one with a frequently used product may improve the business gain by attracting customers with its features.

<table>
<thead>
<tr>
<th>Race Game (C1)</th>
<th>Shooting Game (C2)</th>
<th>Mindblowing Game (C3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bike Race</td>
<td>Military (F)</td>
<td>Chess (F)</td>
</tr>
<tr>
<td>Boat Race (N)</td>
<td>Die Hard (N)</td>
<td>Sudoku (N)</td>
</tr>
<tr>
<td>Car Race (F)</td>
<td></td>
<td>Treasure Hunt</td>
</tr>
</tbody>
</table>

Table I Sample Input

C – Cluster, F-Frequently used product N- New product in the Cluster.

Different Associations in each cluster of the given example as follows. Format of association pair is (F,N).
1. (Car Race, Boat Race)
2. (Military, Die Hard)
3. (Chess, Sudoku)

The different clusters formed for the given example as follows. The format of association pair is (Cluster (F), Cluster(N)).
1. (C1(Car Race), C2( Die Hard) )
2. (C1(Car Race), C3( Sudoku) )
3. (C2(Military), C3( Sudoku) )
4. (C2(Military), C1(Boat Race) )
5. (C3(Chess), C1( Boat Race) )
6. (C3 (Chess), C2(Die Hard) )

By associating the new product with the frequent moving product, we can market the new product in cost benefit way. According to the customer’s feedback about the new product,
the new product can be marketed or dropped from the corresponding cluster.

The advantages of applying Data Mining Approaches are minimizing time for searching required products for customer expectation and take little effort for reaching customers with new products. Associating new products with existing frequently used products that may increase the product cost in future.

V. CONCLUSION

Cloud computing provides services to consumers in a beneficial way relating to cost, time and performance. Implementation of Clustering and Association procedures in this service will improve the effect of selling product in the cloud. So, new product will be the moving product if we associate it with the frequent product. Marketing new product in this way depends on market basket analysis. This will reveal whether the new product has been purchased by consumers or not. Future work can be performed by implementing any specific algorithms for clustering and association.

REFERENCES

[3] E.W.T. Ngai, Li Xia and D.C.K. Chau, “Application of data mining techniques in customer relationship management: A literature review and classification” Department of Management and Marketing, The Hong Kong Polytechnic University, Hong Kong, PR China, Department of Automation, Tsinghua University, Beijing, PR China.

AUTHOR’S PROFILE

Mr. R. Kamalraj has received M.E degree from Anna University, Chennai, INDIA. He is having overall 86 months of experience in teaching field. He has started his career as a LECTURER at KSR College of Engineering. He is currently working as an Assistant Professor in CSE Department at SNS College of Technology, Coimbatore, Tamil Nadu, India.

He published five papers in International Journals and 1 paper in National Journal. And he has presented papers in International and National level conferences. His areas of Interest are Software Testing, Data Mining and Cloud Computing. He has a life membership in ISTE (Indian Society of Technical Education).

Dr. A. Rajiv Kannan has received his Ph.D. award from Anna University, Chennai, INDIA. He has published many papers in National and International Journals and in Conferences also. He has completed his research work in High Performance Networks and Network Security. Now he is the HOD and Professor of Computer Science and Engineering Department at KSR College of Engineering, Tiruchengode, INDIA. He is guiding 8 Ph.D. registered candidates in different subject domains.

Mrs. S. Vaishnavi has received B.Tech. (IT) degree from Anna University of Technology, Coimbatore. Presently she is pursuing M.E Software Engineering at SNS College of Technology. Her areas of interest are Data Mining and Cloud Computing.

Mrs. V. Suganya has received M.Sc. Software Engineering degree from Anna University, Chennai, INDIA. She is having 18 months of teaching experience in teaching computer subjects. Presently she is pursuing M.E. Software Engineering at SNS College of Technology. Her areas of interest are Data Mining and Cloud Computing.