Mass Customization
An approach towards personalization of travel packages

[1]Ashish Sharma, [2]Raghav Singh Dhaker
[1]Research Scholar, [2]Assistant Professor
[1]Department of Production and Industrial Engineering,
[1]MBM Engineering College, Jodhpur, India

Abstract: *This paper employs Mass Customization strategy to design best suited travel package for lay customers. Analysis of customer needs is the key step for product and service variety design of Mass Customization. In this paper, popular Data Mining approach is utilized for suggesting a set of most suitable travel package to the customer according to his/her needs and requirements. A model based on the association rule is being used here that minimizes the processing and operation costs for the service providers. A real-world problem is solved using the model and reported results are compared with those from another published study. The outcome establishes the usefulness of the proposed model.*

IndexTerms - Mass Customization, Data Mining, Travel Package, Association Rule

I. INTRODUCTION

The onslaught of web-enabled services and the spread of the net have changed the way business is conducted in the recent ten – fifteen years. Both the seller, as well as the buyer is bombarded with tremendous amount of data, and through that, incalculable growth possibilities exist. It is said that for every seller on the net, there is some customer who may feel satisfied with the bargain and for every buyer, there is some seller or service provider who can fulfill the demand satisfactorily. The whole problem is to get them to find each other. But, it is easier said than done because with this growth in business, the data information within the business has also increased enormously. In such a scenario, making an informed decision for personalized services (or products) is not an easy task for an average customer. Modern technology has brought forward many innovative ways of doing it.

The two recent developments of the modern technological world that have profoundly influenced the way business is conducted. They are:

(1) Mass Customization
(2) Data Mining

Mass Customization is a system that uses information technology, flexible process and organizational structure to deliver a wide range of products and services that meet specific need of individual customer at a cost near that mass-produced products and services' (*M. Kay, 1993*).

The term "data mining" describes the extraction of implicitly available, non-trivial and useful knowledge from large, dynamic databases with relatively complex structures. In here, an attempt is made to fuse the two and to evolve a technique that provides a convenient means of extracting personalized service packages, using the Association Rule.

In the present time, Mass Customization can be seen majorly in every aspect of daily life as there are almost infinite possibilities for customers to customize their preferences in a product. A well-known example is mobile phones, where customers have countless options to choose, within the available range, according to their own individual needs and requirements. Customers can choose a car according to their own requirements. A customer can choose the color of every single section on their Nike sneakers and have their initials printed on it, etc. Such personalized options are now available on several products as well as services in the market.

Web based applications have entirely changed the business scenario, also changed the business functions in competitive market. The power of internet technologies has reduced the complexity in the interface process between customers and retailers; retailers and distributors; distributors and factories; factories and their suppliers. It has enabled the online transaction and also made it imperative to generate large scale real time data. With this big data including the product and all its information, data pertaining to various business transactions and customer requirement database, it is only data mining that can make sense out of these big data. This study explores a way in which both Mass Customization and the Web is of equal importance in finding out the best product or service for an individual customer, using Data Mining.

The area of influence that is chosen here is the business of travel agencies, we suggest suitable and personalized package on the basis of the desire of the customers. In today's competitive market, it is very difficult for travel agencies to sustain their presence, when they are focusing on providing the most appropriate travel package to the customers which are not in front of them. The best

travel package assumes to have the following components: flight tickets, rental cars, hotel reservations, entertainment event tickets, and offers or discounts.

It is proposed to consider here that a standard travel package that may include one or more of the following components:

1. Flight reservations: (a)Round Trip, (b)One-way Trip
2. Hotel reservations: (a)First class hotels, (b)Second class hotels, and (c)Motels
3. Car rental
4. Ticket to entertainment events: (a)Sports event tickets, (b)Performing art tickets, (c)Museum tickets, (d)Cruise tickets, and (e)Fine dining events
5. Offers/discounts

It is clear from the above-mentioned components that there will be too many possible schemes for designing a travel package. For example, for *n* possible components in travel package, the upper limit for possible customized travel packages will be 2^n-1. For the above case, where *n* equals 12, the upper limit will be 4,095 customized travel packages. However, this upper limit will not be reached because a customer could make reservation for only one type of hotel at a time (i.e. either first class or second class). Here, the possible design may have a travel package that includes a round-trip flight ticket, motel reservation, and a fine dining event. Therefore, here it is proposed to develop the best strategy that analyzes the optimality of designing travel packages which is offered by travel agencies and also to develop a mechanism for designing a set of popular travel packages. The motive of the model is to reduce the processing and operation costs also.

II. LITERATURE REVIEW

In manufacturing sector, for implementing mass customization ideas there are three product and process redesign approaches: *standardization, modular design, and process restructuring* (*Lee and Tang, 1997)*. The three important principles suggested by Feitzinger and Lee for achieving customer satisfaction and low cost by mass customization are: (1) Design products to be assembled and customized in a non-expensive way, (2) Design the manufacturing process to support independent models, and (3) The supply network should be designed to perform customization in a cost-effective manner and be flexible and responsive to individual customers' orders.

Earlier research on application of data mining techniques concerning travel agencies focused on establishing methodologies and mechanisms for either only on costing or only on segmenting the market. For example, for estimating the cost of sun-and-beach tour package, a hierarchical regression approach was implemented *(Thrane, 2005)*. The package consists of choice of tour operator, destination, hotel star rating, and many other different attributes. The approach proposed was a decadent price theory which analyzed the indirect effects on cost of each component of the travel package. Also, a cluster analysis was proposed to categorize the adventure travelers with regard to their behavior and the type of travel package *(Sung, 2004)*. This approach was based on segmenting the market according to a specific type of customers. The goal was to increase the likelihood of customer in making reservations. In our approach classification of travel package is done rather than classification of the travelers.

The factors that play important role in the customers' travel agency selection process are pricing and the communication between customers and service provider *(Heung and Chu, 2000)*. Hence, the focus of the study is to minimize the operations and processing cost since, this will increase the chance that the travel agency will be selected by the customer for reservation. Also, the objective of the study is to design travel packages that are popular and have high customer demand.

Another approach of data mining for finding out the best travel package is being done using association rule by (*Al-Salim, 2008*). The approach used many components under consideration of human need but it did not include the major attracting component "Discounts/Offers". This is a major factor for a customer to decide the best package as, in todays' market scenario if any company wants to increase their customer base, it will focus on the offers or discounts they are offering to their customer in addition to the basic components of a product or service. It is proposed to assume that all the customers in the synthetic data set will prefer the component 'discounts' while selecting the travel package.

III. DATA MINING PROCESS

Consider an online store where, the customer can configure a PC according to his/her choice, place an order for it, track its movement, as well as pay for the product and services. Let us understand the whole scenario with the help of flow chart which is reported in Figure 1.

During the last few years, the scholars and researchers have proposed a new area of web mining for all methods that uses data mining to web data. It is to extract information from the web, rather than retrieving it, generally, data-mining is classified into following three categories: (a) Web content mining, in which extraction of useful information from web page contents is done. (b) Web usage mining, it identifies patterns of use by analyzing web data. (c) Web structure mining, it is concerned with application of data mining to the structure of web graph (*Cooley et al., 1997).*

In web usage mining, the outcome is concerned with knowledge discovery for user's behavior. Its aim is to support the process of human decision. Thus, outcome will be a set of data models that tells about data items. Web usage mining consists of the following basic data mining stages: (a) Data collection, (b) Data processing, (c) Pattern Discovery, and (d) Post-processing.

3.1 Data collection

The first step in data mining which is collection of relevant web data, which will be analyzed to provide useful information about consumers' behavior. There are two main data sources (a) data on web server side (e.g. sever log files, cookies, etc.), and (b) data on client side (e.g. data from host that is accessing the web site, etc.).

Now the data which is collected will be personalized, here it is an attempt to identify unique association between data collection methods and the corresponding requirements of the personalized functions. These functions are: (a) memorization, (b) guidance, (c) customization, and (d) task performance support.

3.2 Data Processing

The data collected in first step must be assembled into a consistent, integrated and comprehensive view, in order to be used for identifying patterns. Data processing can be done in following steps: (1) data filtering, in this step, the cleaning of raw web data is done, the available data is examined and irrelevant and redundant items are removed. (2) user identification, it is done by allowing the user logging in before using the system. (3) user session identification, a user session is a delimited set of pages visited by some users within the duration of one particular visit to a web site.

Now, data processing for personalization is done, in this removing noise and irrelevant data is first step. Also, user identification is also very important parameter of web personalization. As, a result both of them are required in some form by personalization functions. (a) memorization, (b) guidance, (c) customization, (d) task performance support.

3.3 Pattern Discovery

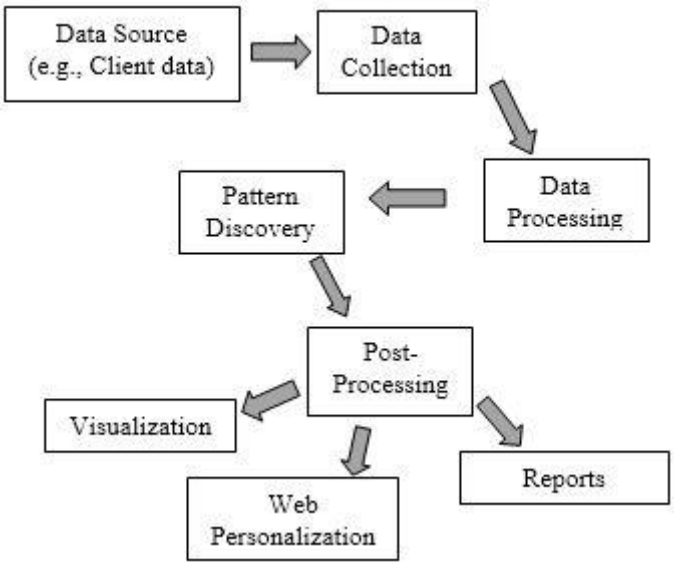
Extracting correct pattern of usage from the processed web data by machine learning and statistical method. This is done by following method: (a) clustering, (b) classification, (c) discovery of association, and (d) sequential pattern discovery.

Now, the pattern discovery for personalization is done, the extraction of usage pattern from web data is essential for construction of efficient user models. This is useful for all personalized functions except memorization function, as user salutation and bookmarking does not require pattern discovery. (a) guidance, (b) customization, (c) task performance support.

3.4 Post-Processing

In this stage, the patterns which are discovered above will be filtered and analyzed to help the decision-making process by human experts. The processes that may be used are: (a) generation of reports, (b) visualization, and (c) query mechanism.

Figure 1: *Generic web based mining process*



IV. MODEL

The method of tackling the above-mentioned travel package selection problem is being done in the following two phases:

Phase I: In this phase, association rule technique is applied on the available data set and best rules are used for further use. The rules are obtained by considering the guidelines which are shown below.

Phase II: In this phase, an optimization model is developed for minimizing operation and processing costs of the remaining customized travel packages in order to propose a set of most suitable travel packages.

4.1 Phase I: Association rule technique

Association Rule Technique or basket market (*Post, 2005*) is one of the most popular data mining technique. In this technique, rules are extracted out of the database and then arranged based on the percentage of times they are correct and how often they occur. If event A (antecedent) occurred, then there is a probability of occurring of event B (consequent) i.e. A⇒B and is read as A implies B. The two main concepts in this technique are: (1) *Accuracy* (*confidence*) – it is the probability that if the event A is true then the event B will be true. High accuracy means that it is a highly dependable rule. (2) *Coverage* (*support*) – it is numbers of records in data set that the rule is applied. High coverage means that the rule can be used very often.

Here, in this paper P_a (set of best travel package), will be determined without affecting the general procedure of association rule, based on sample synthetic data set in Table1.

There are 10 records incorporated in the sample here which are being analyzed. Here it is proposed to use '1' when customer prefers the component '0' when customer does not make a request to prefer the component.

From data set table 1, it is clear that it will generate large number of association rules. The basic approach for framing these rules may be explained by an example: if discounts event is preferred, then round trip flight is preferred seven times out of total ten that discounts is preferred thus, accuracy is 7/10 or 70%. Also, this association occurs seven times out of all ten records so, the support is 7/10 or 70%.

Table 1: The synthetic data set as used by *Al-Salim, (2005)*

Record No.	1	2	3	4	5	6	7	8	9	10
Round-trip flight	1	1	1	1	1	0	0	1	1	0
One-way flight	0	0	0	0	0	1	1	0	0	1
Cruise trip	0	0	0	1	0	1	1	0	0	1
First-class hotel	1	1	1	0	0	0	0	1	1	0
Second-class hotel	0	0	0	1	1	0	0	0	0	0
Motel	0	0	0	0	0	1	0	0	0	1
Rental car	1	1	1	0	0	1	0	1	1	0
Sport event	1	0	1	0	0	1	0	1	0	0
Art show event	1	0	0	0	0	0	0	0	0	1
Museum event	0	0	0	1	0	0	0	0	0	1
Fine dining event	1	0	1	0	0	1	0	0	0	0
Discounts	1	1	1	1	1	1	1	1	1	1

In this way all the associations can be made. Since, large number of associations will be made it is suggested to use any available data mining software for it to build the rules easily and efficiently. It is proposed to use WEKA 3.6 © software for obtaining the association rules.

The following guidelines may be suggested for selecting the most efficient association rule, this will result in efficient analysis. (1) accuracy higher than 70% will be significant. (2) coverage higher than 30% will be significant. These guidelines are based on experience and a systematic approach.

Now, using the synthetic data set in Table 1 and applying the above guidelines, the following significant association rules are obtained (A: Accuracy, C: Coverage):

Rule 1: Round trip flight + first class hotel (A=71%, C=50%)
Rule 2: Round trip flight + rental car + first class hotel (A= 71 %, C=50%)
Rule 3: One-way flight + cruise trip (A=75%, C=30%)
Rule 4: Sport event + fine dining event (A=75%, C=30%)
Rule 5: Round trip + rental car (A=71%, C=50%)
Rule 6: Rental car + first class hotel (A=100%, C=50%)
Rule 7: Round trip flight + discounts (A=70%, C=70%)

Now, it is supposed to design cost table by adding above travel package to the initial list of suitable packages, such that each one has all the components of each significant rule, for this we will design a travel package with two components: a round trip and first-class hotel reservation, this is denoted by P_{13}. Table 2 contains the costs list of potential travel packages that will be crunched in the next phase of the model. The first 12 travel packages are single item package and the last 7 are those with two or more items.

4.2 Phase II: Cost optimization model

The cost optimization model, to minimize the costs may be obtained by offering components for a particular suggested travel package. It is proposed that there may be different types of costs which can affect the process of suggesting a travel package to customer. These costs are explained below.

Firstly, set-up and processing costs to process one reservation. Under this category, following activities may be taken under consideration: (1) order fee that may be require to perform the application process. (2) the cost of follow up process which may include human effort (for application review), making assurance of availability of other services by contacting actual service providers, and making of the reservation final. (3) costs of maintaining the database of each transaction for travel agencies. On the other hand, it is to assume that the cost in a travel package that has more than one components may be less because, most of the time it is cheaper to process application that has two components rather than one.

On the other hand, there is the matter of losses in sale costs. The cause of this loss is due to offer made to a customer that sells two reservations together without giving an option to customer that he/she can make a reservation for single event. This is because some customers may have interest in choosing only single event. It is clear from above that, R_a=0 for a travel package that will have only one component.

Finally, the marketing costs, marketing theories shows that it is more likely that advertising for a group of items is more expensive than promoting a single item. However, estimating it should be treated carefully, especially with regards to using rights units. To obtain the marketing costs per unit of sold travel package, the one-time advertising costs should be divided by the number of the travel package that are sold after placing the advertisement and until the next advertisement. The number could be obtained from previous records that are stored in information system of the travel agency or based on developing a forecasting model for the future customers' demands.

To solve the problem of obtaining a set of most suitable travel packages, the following notations will be used in this paper:

a: travel package index
b: component index
m: number of potential travel package
n: number of available component
P_a: travel package a
Q_a: cost for processing the travel package P_a
R_a: sales cost losses due to having travel package P_a
S_a: marketing costs for travel package P_a
T_a: total costs for providing the travel package P_a
Y_{ab}: indicator function having component b in P_a.
Mathematically;

$$Y_{ab} = \begin{cases} 1, \text{ If component } b \text{ is an item in } P_a \\ 0, \text{ If component } b \text{ is not an item in } P_a \end{cases}$$

And finally, the decision variable will be,
X_{ab}: indicator function for offering or not offering travel package P_a.
Mathematically;

$$X_{ab} = \begin{cases} 1, \text{ If component } b \text{ is an item in } P_a \\ 0, \text{ If component } b \text{ is not an item in } P_a \end{cases}$$

Mathematically, the total cost of "*a"* (travel package) will be:

$$T_a = Q_a + R_a + S_a$$

$$\text{Minimize:} \quad \sum_{a=1}^{m} T_a \times X_a \qquad (1)$$

$$\text{Subject to:} \quad \sum_{a=1}^{m} (Y_{ab} \times X_a) \geq 1 \quad b=1,2,\ldots n \qquad (2)$$

$$X_a = 0 \text{ or } 1 \qquad (3)$$

Here, equation 1 represents the objective function and equation 2 and equation 3 represents the constraints.

It is assumed that each component b is represented in at least one travel package. If this is not the case, we do not concern ourselves with adding this component in the first place. The first constraint is to ensure that this assumption is satisfied.

Solution of the above model can now be easily calculated using any of the following method:

(1) Modeling Language for Mathematical Programming (AMPL) technique.
(2) Linear programming using LINGO © solver.
(3) Heuristic Algorithm Approach

On solving the problem with LINGO ©, a famous linear programming solver software, an optimal set of suitable travel packages found out are:

P_5, P_6, P_9, P_{10}, P_{14}, P_{15}, P_{16}, and P_{19}

V. CONCLUSION

With a view to optimizing the best deals available to the customer a model has been proposed that assists a lay customer in making informed choices regarding the best travel package with separate focus on major components such as flight tickets, hotel reservation, rental car, entertainment event tickets, and specially, discounts.

An attempt has been made to apply web based service personalization to the fast growing traveling industry. This approach uses the synthetic data set from *Al-Salim,2005* and the authors extend the algorithm by using association rule. The results obtained for the problem given in *Al-Salim,2005* are in agreement with the results reported by other researchers and proves that the proposed model is working efficiently. The comparison of both approaches is reported in Table 3A and Table 3B. The results clearly show that by adding a component the sensitivity of the model changes and thus discounts may play important role while suggesting the best travel package to customer.

Table 3A: Comparison of travel package having single component

Approach	Set of Best suited travel packages
Earlier approach by Al-Salim, (2005).	P_5, P_6, P_9, and P_{10}
Proposed approach by adding a component "Discounts"	P_5, P_6, P_9, and P_{10}

Table 3B: Comparison of travel package having more than one component

Approach	Rank	Set of Best suited travel packages
Earlier approach by Al-Salim, 2005.	1	*Round trip + Rental Car + First Class hotel*
	2	*one-way trip + Cruise trip*
	3	*Sport event + Fine Dining event*
Proposed approach by adding a component "Discounts"	1	*Round trip + Rental Car + First Class hotel*
	2	*One-way trip + Cruise trip*
	3	*Sport event + Fine Dining event*
	4	***Round trip + Offers/Discounts***

As, evidenced by the last package (i.e. ***Round trip + Offers/Discounts***) selected through the proposed approach, the package that gives some discounts/offers is also chosen. This investigation shows that in times of mass customization in the service sector the proposed model proves to be a powerful tool.

The two main benefits which may be obtained from the above suggested model are:
(1) reduced processing and operation costs, and
(2) designing popular travel packages.

The small set of viable packages obtained as an outcome of this model is very helpful to a lay customer because it reduces the number of preferred options to a manageable few. Otherwise, the option available on the web are so large in number that it becomes physically impossible to compare and evaluate all of them. The final choice of any one (best) package can now be made by the customer either in the order of availability, or some specific preference such as, 'view from the balcony' or 'floor of the premises' (top floor/ground floor).

REFERENCES

[1] Al-Salim B., "Data mining application in designing popular and optimal travel packages", In: Proceedings of the IEEE International Conference on Electro Information Technology, Lincln, NE, 2005.

[2] Al-Salim B, "Mass customization of travel packages: data mining approach", Int. J. Flex. Manuf. Syst., 2008, 19: 612-624.

[3] Cooley, R., Srivastava, J. and Mobashaer, B., "Web Mining: Information and pattern discovery on the world wide web", In Proceeding of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), 1997, pp. 558-567.

[4] Feitzinger E., Lee HL, "Mass customization and Hewlett-Packard: the power of postponement". Harvard Bus Rev, January/February, 1997, 116-121.

[5] Heung V. S., Chu R., "Important factors affecting Hong Kong consumers' choice of a travel agency for all-inclusive package tours". J Travel Res, 2000, 39 (1): 52-59.

[6] Lee H. L., Tang C., "Modeling the costs and benefits of delayed product differentiation. Manag. Sci, 1997, 43 (1): 40-53.

[7] M. Kay, "Making mass customization happen: Lesson for implementation, planning review",1993, 21 (4), 14-18.

[8] N. R. S. Raghavan, "Data mining in E-commerce: A survey", April /June, 2005, pp. 275-289.

[9] Nicolas B., Jurgen H., "Data Mining", Wirtschaftsinformatik, 1993, 35(5), 481–487.

[10] Post G. V., "Database management systems: Designing and building business applications". McGraw-Hill, 2005, Irwin.

[11] Sung H. H., "Classification of adventure travelers: behavior, decision making, and target markets". J. Travel Res., 2004, 42 (4): 343-356.

[12] Thrane C., "Hedonic price models and sun-and-beach package tours: The Norwegian case". J. Travel Res., 2005, 43 (3): 302-308.

[13] Xia Shi-Sheng, Wang L., "Customer requirements mapping method based on association rule mining for mass customization", 2008, 13 (3): 291-296.

Table 2: The obtained travel package and their associated cost

Round trip flight + Offers	P_{19}	2	10	5	3	18
Rental car + First class hotel	P_{18}	2	9	3	3	15
Round trip flight + Rental car	P_{17}	2	12	4	3	19
Sport event + Fine dining event	P_{16}	2	9	2	3	14
One-way flight + Cruise trip	P_{15}	2	15	3	3	2
Round trip flight + Rental car + First class hotel	P_{14}	3	10	6	3	19
Round trip flight + First class hotel	P_{13}	2	15	8	3	26
Offers	P_{12}	1	5	0	2	7
Fine dining event	P_{11}	1	6	0	2	8

Component →	Travel Package component (s)	Round trip flight	One-way flight	Cruise trip	First class hotel	Second class hotel	Motel	Rental car	Sport event	Art show event	Mus-eum event
Package index	P_a	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
Modulus	|Pa|	1	1	1	1	1	1	1	1	1	1
Processing cost	Q_a	10	9	12	6	5	5	6	8	3	2
Sale loss cost	R_a	0	0	0	0	0	0	0	0	0	0
Marketing cost	S_a	2	2	2	2	2	2	2	2	2	2
Total cost	T_a	12	11	14	8	7	7	8	10	5	4