Computers & Industrial Engineering 60 (2011) 89–98

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Using logistic regression formulation to monitor heterogeneous usage rate for subscription-based services ☆

Y. Samimi, A. Aghaie *

Department of Industrial Engineering, K.N. Toosi University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 9 August 2009
Received in revised form 26 August 2010
Accepted 7 October 2010
Available online 12 October 2010

Keywords:
Customer relationship management
Statistical process control chart
Logistic regression
Generalized linear models
Deviance statistic

ABSTRACT

This paper explores the effect of heterogeneity across different classes of customers as well as their time dependent usage behavior on the purchase rate of multiple services supplied by a subscription-based service provider. It is shown that a suitable model for customer usage pattern based on the logistic regression can effectively be employed to represent both the cross correlation and serially correlation of purchase rate for different kinds of services. Then the deviance statistic is proposed as an appropriate control statistic to simultaneously monitor the usage of multiple services. On the basis of three comparative scenarios, simulation results indicate that the power of the deviance-based control chart is considerably greater than some traditional counterparts like Hotteling's T^2 control chart. Research findings provide promising results for marketing managers and practitioners in terms of both better understanding of the behavior of different classes of customers as well as timely evaluation of investment opportunities that can lead to enhancement of the firm's relationship with customers.

1. Introduction

This paper deals with employing an appropriate statistical method for monitoring usage pattern of various services supplied to multiple classes of customers by a subscription-based service provider. This type of enterprise may involve a broad spectrum of service industries such as telecommunications, insurance services, internet service providers, as well as many internet-based service firms like online content providers. Such corporations can operate on either contractual (i.e., subscription-based model) or non-contractual basis (i.e., pay per use) (Dover & Murthi, 2006). This paper investigates the contractual business model or the so-called subscription-based services. Customers of such a firm usually purchase a quarterly, an annual, or an "annual billed monthly" subscription.

Momentarily changes of customer needs along with the rapid alterations of market conditions demand to have a prompt reaction against unusual shifts in customer usage pattern. Such flexibility can be gained through a responsive monitoring mechanism, which can detect significant changes in customer behavior via exploration of usage data. Statistical process control (SPC), extensively used to monitor quality characteristics in manufacturing processes, consists of problem-solving tools suitable for monitoring purposes in service industries. Nonetheless, little research has been done on the application of SPC methods to the monitoring of customer activities so that appropriate marketing campaigns and service customizations can be developed (Jiang, Au, & Tsui, 2007). An overview of literature, especially studies that investigate the application of SPC charts for service usage monitoring, is provided in the rest of Section 1.

To reduce customer churn, Pettersson employs multivariate T^2 control chart to develop an early warning system, which monitors the proportion of likely churned customers in telecommunication industry (Pettersson, 2004). Qian, Jiang, and Tsui (2006) employ functional mixture models to derive the profile of customer transactions in a given period of time. Based on the estimated profile, a monitoring method, which resembles the profile monitoring in SPC has been proposed to predict customer churn. Recently, a SPC framework has been introduced for business activity monitoring which can be applied for the purpose of prevention of customer churn and detection of fraud in financial transactions (Jiang et al., 2007). Multi-way principle component analysis (MPCA), a statistical process monitoring technique in batch production systems, has been also used for fraud detection in the service sector (Tsung, Zhou, & Jiang, 2007).

Using probability models to model customer time dependent non-homogenous usage behavior, makes it possible to employ inferential statistics in order to develop a parametric monitoring model. Mixture probability models (Fader & Hardie, 2007; Johnson, Moe, Fader, Bellman, & Lohse, 2004) and Markov chain models

☆ This manuscript was processed by Area Editor E.A. Elsayed.

* Corresponding author. Address: Department of Industrial Engineering, K. N. Toosi University of Technology, P.O. Box 19395-1999, Tehran, Iran. Tel.: +98 21 88678580; fax: +98 21 88674858.

E-mail addresses: Y_Samimi@dena.kntu.ac.ir (Y. Samimi), AAghaie@kntu.ac.ir (A. Aghaie).

0360-8352/$ - see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.cie.2010.10.002

(Chen & Cooper, 2002; Ha, Bae, & Park, 2002; Pfeifer & Carraway, 2000) are prominent methods that have been successfully used to capture heterogeneity of customer behavior as well as its time dependency.

Logistic regression model is employed in this research to construct a predictive model of the probability of usage for multiple services proposed by a service provider. This approach provides the possibility to represent heterogeneity across customers regarding their intention toward various services. It is shown that the cross correlation across multiple services can be represented using this approach, though it is assumed that purchases from different types of services are performed independently. Logistic regression has also been widely used and investigated in predictive data mining to predict customer churn in retail industry (Buckinx & Van den Poel, 2005), financial services (Larivie & Van den Poel, 2005), telecommunications (Ahn, Han, & Lee, 2006), as well as subscription-based services (Burez & Van den Poel, 2007; Coussement & Van den Poel, 2008). The idea of employing SPC charts using logistic regression model to monitor usage behavior originates from the rich literature on model-based control charts commonly used in multistage production processes (Loredo, Jearkpaporn, & Borror, 2002). Recently, generalized linear regression models including Poisson and logistic regression models have been employed to derive suitable monitoring methods in situation where key output characteristics of process or product quality are multivariate discrete random variables and are affected by upstream process stages. Jearkpaporn, Borror, Runger, and Montgomery (2007) and Skinner, Montgomery, and Runger (2003, 2004) are recommended to be referred in this respect. Integrating SPC charts with predictive models of customer usage in order to facilitate monitoring customer behavior is the main objective of our research.

The remainder of this paper is organized as follows. Next section provides precise definition of the problem. Section 3 deals with how to use logistic regression to model heterogeneous usage behavior across multiple classes of customers. Monitoring statistics are also derived in this section. To compare the performance of the proposed method with that of the other competing approaches, three illustrative cases are studied in Section 4. Section 5 outlines final remarks and conclusions.

2. Problem description

The proper understanding of customer behavioral patterns is accounted as an integral part of analytical customer relationship management. To obtain such valuable knowledge, organizations use segmentation techniques to differentiate customers based on their usage patterns. To this end, direct marketing companies commonly use the recency, frequency, and monetary value (RFM) framework that captures the main characteristics of customer usage behavior. "Recency tells how long it has been since each customer made the last purchase; frequency tells how many times each customer has purchased; and monetary tells how much each customer has spent in total" (Ha et al., 2002). Cluster analysis algorithms, which are currently available in numerous software packages, have been also attracted a great deal of attention among practitioners to determine well-differentiated segments by examining customers' usage pattern (Berry & Linoff, 2004; Giudici & Figini, 2009). Classification techniques are then used to discriminate customers by associating their characterizing peculiarities like household and socio-demographic features with the identified usage patterns. This is a prevalent procedure when an organization intends to develop product differentiation strategies or marketing mix allocation strategies that can satisfy the particular preferences of each segment. In this study, we assume that customer segments have been identified using the appropriate methods.

Besides heterogeneity across customers, time dependency is another important feature of customer behavior. Considering time dependent non-homogenous usage behavior for the majority of subscription-based services, this research provides answers to the following questions:

- How can a subscription-based service provider differentiate between usual and unusual usage behavior of multiple groups of customers who use several kinds of services?
- In case of an unusual usage behavior, how can the firm identify the particular class or classes of customers whose intention toward proposed services has changed?

To clarify the problem, we continue with an illustrative case. Suppose a telecommunication firm which offers six major types of products including voice mail, wireless services, toll free, long distance calling, internet and three-way calling. Company's customer base contains 1000 customers categorized into multiple classes based on their profitability level. Recency of customer's last transaction along with his buying frequency and monetary value over tenure are commonly used measures to segment customers. We assume that the firm has already categorized its customer base into a given number of clusters. By allocating a zero-one variable to each service, we suppose that the telecommunication company is able to record customer usage in each subscription period as a binary vector. The elements of this vector, which is specified more precisely in the next section, indicate which services a given customer has chosen in a specific subscription period. Moreover, suppose that in order to find association across products, analysis of customers' usage data by hierarchical clustering method has resulted in a dendrogram shown in Fig. 1. In this case, it can be inferred that customers choose three different groups of services. Correlations between pairs of services are shown as a matrix in Table 1. Admittedly, cross correlations verify the result of clustering.

Moreover, the result of chi-square independence test in Table 2 indicates the mild but significant correlation for usage of services during two consecutive periods. Six contingency tables that are placed inside Table 2 contain percentage values where rows and columns respectively denote the usage of the corresponding services associated with time periods $t-1$ and t.

Recall the first objective of the paper. As assumed before, customer usage from a given service may be represented by a Bernoulli random variable. Hence, the total number of customers who use the same service might be modeled by a binomial random variable, were the heterogeneity across customers overlooked. Therefore, the total usage of different products in each subscription period would form a multivariate binomial distribution suffering

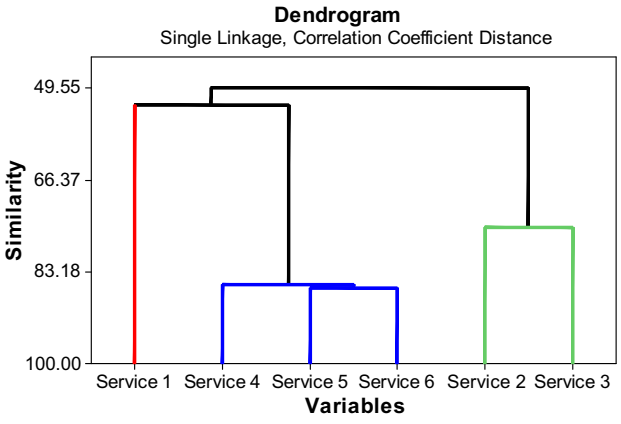

Fig. 1. A dendrogram as the result of hierarchical clustering for the problem discussed here.