

A study on classifying the store positioning from the transactional data

***Masakazu Takahashi¹ and Yuji Tanaka²**

¹Graduate School of Innovation and Technology Management,

Yamaguchi University, Japan

²Japan Green Stamp Co., Ltd, Japan

**masakazu@yamaguchi-u.ac.jp*

Abstract

This paper describes a customer analysis for store positioning, using data gathered from supermarkets in Japan. Among the retail industry in Japan, there are many types of reward cards used for customer retention purposes. The rewards cards or “Point Card”, is originally aimed for customer analysis purposes, but at present the full benefits have not been extracted due to issues in data analytics. This reward card has only become a method of simply distributing “virtual money” to the customer. For the efficient use of gathering data, we propose a classification method of the customer based on the objectives of visiting stores. In this study, the customers were classified into their objectives.

Keywords: Data analysis, Machine learning, Retail, Service science, Management engineering

Introduction

There are so many rewards cards distributed in Japan. This card can provide us customer profiles using the gathered data. However, the retail industry has not been able to use the gathered data, efficiently. The data captured includes; Who, When, Where and Which items and Customer spend. This paper attempts to capture the customer profiles by their objectives of visiting the store and then merging with the POS (Point of Sales) data. Hereafter, we define the merged data as ID-POS (Identification Point of Sales) data.

The rest of the paper is organized as follows: Methodology discusses the research backgrounds and related work; Results section briefly summarizes the data on the target sector and describes the analytical results and gives some concluding remarks and future work.

Methodology

ID-POS data was gathered for this study over a four week period. Table 1 shows attributes of gathered data. Then, it was classified into four categories based on the monetary value. a) Upper, over JPY 25,000 spent b) Middle, from JPY 10,000 to JPY 24,999, c) Low, under JPY 9,999 and 4) Non-member, respectively. Table 2 also shows the summary of the data. Further, characteristics of the three levels of customer was analyzed. Cluster analysis was used to assess the data. This method is one of the unsupervised learning in Machine Learning. The primary goals of cluster analysis is to group objects based on their characteristics. The process of analysis is a) grouping object (whether to classify the samples or variables), b) identify classification type (hierarchical or non-hierarchical), c) measure distance between objects for classification (assess data similarity i.e. Euclidean distance, Mahalanobis distance, Cosine distance and so on), d) use merging cluster methods (Ward method, Group average method, Shortest distance method, Longest distance method and so on) (Everitt, 1993; Zume & Mount, 2014).