

## Chapter 62

# Observing Customer Segment Stability Using Soft Computing Techniques and Markov Chains Within Data Mining Framework

**Abdulkadir Hiziroglu**  
Yıldırım Beyazıt University, Turkey

### ABSTRACT

*This study proposes a model that utilizes soft computing and Markov Chains within a data mining framework to observe the stability of customer segments. The segmentation process in this study includes clustering of existing consumers and classification-prediction of segments for existing and new customers. Both a combination and an integration of soft computing techniques were used in the proposed model. Segmenting customers was done according to the purchasing behaviours of customers based on RFM (Recency, Frequency, Monetary) values. The model was applied to real-world data that were procured from a UK retail chain covering four periods of shopping transactions of around 300,000 customers. Internal validity was measured by two different clustering validity indices and a classification accuracy test. Some meaningful information associated with segment stability was extracted to provide practitioners a better understanding of segment stability over time and useful managerial implications.*

### MOTIVATION AND BACKGROUND

The main idea of segmentation or is to group similar customers. A segment can be described as a set of customers who have similar characteristics of demography, behaviours, values, and so on (Nairn and Berthon, 2003, Bailey et al., 2009). In fact, one of the most valuable pieces of information based on which a segment can be customers' behavioural characteristics, especially past customer purchases and value-oriented attributes (Bayer, 2010, Kim et al., 2006, Wind and Lerner, 1979). In fact, customer analytics related technological advances have facilitated performing segmentation studies based on those characteristics (Bailey et al., 2009).

DOI: 10.4018/978-1-5225-5643-5.ch062

This study focuses on two main issues regarding customer segmentation. The first issue is related to the difficulty of obtaining homogenous segments. For customer segmentation, a wide variety of data analysis techniques, such as cluster analysis (Ghazanfari et al., 2008; Hruschka et al., 2004, Li et al., 2009, Liu and Shih, 2005, Wang, 2009, Wang et al., 2008; Xia et al., 2010), clusterwise regression (Desarbo et al., 2008), AID/CHAID (Aravamudhan, 2011; Gil-Saura and Ruiz-Molina, 2008, Jonker et al., 2004), multiple regression (Suh et al., 1999), discriminant analysis (Tsotsos, 2006), latent class structure (Wu and Chou, 2011) and sophisticated soft computing techniques including such as fuzzy-oriented approaches (Crespo and Weber, 2005, Hu and Sheu, 2003, Kaymak, 2001, Ozer, 2001, Shin and Sohn, 2004) and neural network algorithms (Bloom, 2005, Chiu et al., 2009, Diez et al., 2008, Ha, 2007, Hsieh, 2004, Hung and Tsai, 2008, Kuo et al., 2006, Lee and Park, 2005, Potharst et al., 2001, Shin and Sohn, 2004) have been used in the related literature. Cluster-based segmentation methods, particularly hierarchical and non-hierarchical methods, have been widely used in the field. But, the hierarchical methods are criticised for non-recovery, while the non-hierarchical methods for their inability to initially determine the number of clusters (Lien, 2005). Hence, the integration of hierarchical and partitioning methods (two-stage methodology) is suggested to make the clustering results powerful for large databases (Kuo et al., 2002). None of those approaches, however, have the ability to establish non-strict customer segments that could play a significant role in today's competitive consumer markets. Although there have been a few studies that utilised fuzzy segments they are not based on the effective two-stage methodology.

The second problem is pertaining to the issue of segment stability, which is one of the most ignored concepts in the customer segmentation area (Plank, 1985). This problem is also highlighted in the study of Dibb and Simkin (2001) in which they investigated diagnosing and treating the segmentation barriers. Stability is one the most important segmentation effectiveness criteria that reflects whether the segments are stable over time (Wedel and Kamakura, 2000). According to Wind (1978), the stability of segments depends on three sets of factors. Those factors are the bases for segmentation, the volatility of the marketplace, and consumer characteristics. The competitive characteristics of the marketplace (i.e., Political and economic conditions) increase the switching possibility among segments. Also, the characteristics of consumers are supposed to be changeable over time and this influences segment stability. As opposed to these two factors, the first factor (segmentation bases) is more observable and easy to obtain by means of the advanced data collection methods particularly companies' data warehouses. Therefore, it is very important to have longitudinal data for more applicable segmentation studies. However, the more specific bases produce less stable segments compared to general bases, such as actual shopping behaviour and benefits- or needs-related attributes. Some recent research carried out by Myers (1971) and Nicosia (1970), which was based on the expectation of changes, assumed that consumers' shopping behaviours and perceptions may change over time. Calantone and Sawyer (1978) analysed benefit segments based on the relative importance of product attributes in two distinctive time periods and found that some consumers were likely to be grouped into other segments, which had previously been classified differently.

Extracting meaningful information related to segment stability from corporate data warehouses using soft computing techniques, a member of the data mining techniques family, has become of great importance as these techniques were proved to produce robust and effective solutions. Over the last decades, there has been a rapid growth in the applications of Artificial Intelligence-based soft computing technologies, which involve Fuzzy Logic (FL), Artificial Neural Networks (ANNs), Evolutionary Algorithms and Rough Set Theory in the marketing field. These technologies were utilized in various industrial and commercial applications (Bonissone et al., 1999, Dote and Ovaska, 2001, Mohanty et al., 2010; Widrow et al., 1994). Marketing researchers are more interested in the application of these

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/observing-customer-segment-stability-using-soft-computing-techniques-and-markov-chains-within-data-mining-framework/205841](http://www.igi-global.com/chapter/observing-customer-segment-stability-using-soft-computing-techniques-and-markov-chains-within-data-mining-framework/205841)

## Related Content

---

### An Extensive Evaluation of New Federated Learning Approaches for COVID-19 Identification

Ranjit Barua and Sudipto Datta (2024). *Pioneering Smart Healthcare 5.0 with IoT, Federated Learning, and Cloud Security* (pp. 246-257).

[www.irma-international.org/chapter/an-extensive-evaluation-of-new-federated-learning-approaches-for-covid-19-identification/339436](http://www.irma-international.org/chapter/an-extensive-evaluation-of-new-federated-learning-approaches-for-covid-19-identification/339436)

### On Solving Linguistic Bi-Level Programming Problem Using Dynamic Programming

Vishnu Pratap Singh (2021). *International Journal of Fuzzy System Applications* (pp. 43-63).

[www.irma-international.org/article/on-solving-linguistic-bi-level-programming-problem-using-dynamic-programming/274885](http://www.irma-international.org/article/on-solving-linguistic-bi-level-programming-problem-using-dynamic-programming/274885)

### 3D Gesture Recognition Based on Handheld Smart Terminals

Yunhe Li, Yi Xie and Qinyu Zhang (2018). *International Journal of Ambient Computing and Intelligence* (pp. 96-111).

[www.irma-international.org/article/3d-gesture-recognition-based-on-handheld-smart-terminals/211174](http://www.irma-international.org/article/3d-gesture-recognition-based-on-handheld-smart-terminals/211174)

### Statistical Evaluation of Power-Aware Routing Protocols for Wireless Networks: An Empirical Study

Bhupesh Lonkar and Swapnili Karmore (2022). *International Journal of Intelligent Information Technologies* (pp. 1-14).

[www.irma-international.org/article/statistical-evaluation-of-power-aware-routing-protocols-for-wireless-networks/309589](http://www.irma-international.org/article/statistical-evaluation-of-power-aware-routing-protocols-for-wireless-networks/309589)

### Improved Data-Driven Root Cause Analysis in a Fog Computing Environment

Chetan M. Bulla and Mahantesh N. Birje (2022). *International Journal of Intelligent Information Technologies* (pp. 1-28).

[www.irma-international.org/article/improved-data-driven-root-cause-analysis-in-a-fog-computing-environment/296238](http://www.irma-international.org/article/improved-data-driven-root-cause-analysis-in-a-fog-computing-environment/296238)