Expert Systems with Applications 41 (2014) 521–534

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization

Yong Wang [a], Xiaolei Ma [b,*], Yunteng Lao [b], Yinhai Wang [b,*]

[a] School of Management, Chongqing Jiaotong University, Chongqing 400074, China
[b] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195 2700, USA

ARTICLE INFO

Keywords:
Customer clustering
Fuzzy set theory
Clustering algorithm
Clustering validity index
Axiomatic fuzzy set

ABSTRACT

Customer clustering is an essential step to reduce the complexity of large-scale logistics network optimization. By properly grouping those customers with similar characteristics, logistics operators are able to reduce operational costs and improve customer satisfaction levels. However, due to the heterogeneity and high-dimension of customers' characteristics, the customer clustering problem has not been widely studied. This paper presents a fuzzy-based customer clustering algorithm with a hierarchical analysis structure to address this issue. Customers' characteristics are represented using linguistic variables under major and minor criteria, and then, fuzzy integration method is used to map the sub-criteria into the higher hierarchical criteria based on the trapezoidal fuzzy numbers. A fuzzy clustering algorithm based on Axiomatic Fuzzy Set is developed to group the customers into multiple clusters. The clustering validity index is designed to evaluate the effectiveness of the proposed algorithm and find the optimal clustering solution. Results from a case study in Anshun, China reveal that the proposed approach outperforms the other three prevailing algorithms to resolve the customer clustering problem. The proposed approach also demonstrates its capability of capturing the similarity and distinguishing the difference among customers. The tentative clustered regions, determined by five decision makers in Anshun City, are used to evaluate the effectiveness of the proposed approach. The validation results indicate that the clustered results from the proposed method match the actual clustered regions from the real world well. The proposed algorithm can be readily implemented in practice to help the logistics operators reduce operational costs and improve customer satisfaction levels. In addition, the proposed algorithm is potential to apply in other research domains.

1. Introduction

Logistics network optimization plays a critical role in contemporary logistics planning and supply chain network designs (Altiparmak, Gen, Lin, & Paksoy, 2006; Bidhandi, Yusuff, Ahmad, & Bakar, 2009). With a well-designed transportation and logistics network, logistics operators can significantly improve the entire freight system efficiency, and customers' needs are better accommodated in a timely manner (Wang, Ma, Wang, Mao, & Zhang, 2012). Under this circumstance, logistics operators and customers will achieve a win–win situation: logistics operators gain more revenues and customers are better served with a lower price due to the reduced operational and transportation costs (Wang, Ma, Lao, Wang, & Mao, 2013). Therefore, properly optimizing the logistics network has become a vital objective for logistics operators.

The typical logistics network optimization includes distribution center location selection, customer clustering and vehicle routing problem (VRP) (Lau, Jiang, Ip, & Wang, 2010; Manzini & Bindi, 2009; Sadjady & Davoudpour, 2012). The aim of logistics network optimizations is to design and allocate a set of logistics facilities for better satisfying the demands of customers (Taaffe, Geunes, & Romeijn, 2010). However, in reality, when the number of customers increases, the logistics network optimization problem become very challenging, and thus, customer clustering should be undertaken before conducting the vehicular dispatching (Jarrah & Bard, 2012). Customer clustering approach groups the customers with similar characteristics into the same category (Anzanello & Fogliatto, 2011). It not only improves the logistics system efficiency, but also reduces the operational costs. For instance, by categories these customers who require cold chain services, logistics companies can dispatch several refrigerated trucks to store and deliver temperature-sensitive goods with the same area. With the proper customer clustering strategy, a large logistics zone can be decomposed into smaller zones where customers share certain common features (i.e. geospatial location, demand etc.). Then, VRP is further

* Corresponding authors.
E-mail addresses: yongwx6@gmail.com (Y. Wang), xiaolm@uw.edu (X. Ma), laoy@uw.edu (Y. Lao), yinhai@uw.edu (Y. Wang).

0957-4174/$ - see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.eswa.2013.07.078

simplified within each small zone. Thereby, it is necessary to understand customers' characteristics and conduct the customer clustering analysis before vehicle routing optimization in a large-scale network.

To reasonably cluster customers in the logistics network, multiple factors should be taken into account. These factors include customers' geospatial location, travel risk, accessibility, goods compatibility, transportation conditions, etc. However, certain customer attributes cannot be directly measured quantitatively. Considering the high dimensions and ambiguousness of customers' characteristics, traditional clustering algorithms may not be functional well. Therefore, it is desired to develop an innovative clustering approach incorporating both the customer's quantitative and qualitative attributes for logistics network optimization.

The remaining paper is organized as follows: relevant studies are firstly discussed, and then a hierarchical analysis structure for customer clustering is established based on the fuzzy comprehensive evaluation theory. With various definitions and notations documented in the clustering algorithm procedure, a proposed framework including fuzzy integration method, fuzzy clustering algorithm procedure, and clustering validity index is detailed. To evaluate the effectiveness of the proposed algorithm, a case study of customers clustering problem in Anshun, China is presented, followed by a thorough comparison with different prevailing approaches. In addition, the proposed algorithm is further validated using the actual clustered regions determined by local decision makers. Finally, conclusions are summarized at the end of this paper.

2. Literature review

With the advent of new technology such as electronic commerce and new data collection devices, customer behavior information becomes more and more available. Based on the customer's information, valuable knowledge can be extracted using appropriate data mining techniques (Chen, Chiang, Wu, & Chu, 2013; Ngai, Xiu, & Chau, 2009;Pishvaee, Rabbani, & Torabi, 2011). For example, Customer Relationship Management (CRM) has been recognized as a critical component in the business strategy development for companies (Hiziroglu, 2013; Ho, Ip, Lee, & Mou, 2012). Customer clustering categorizes the customers into multiple clusters. Within each cluster, customers share common behaviors. In this way, a company can develop the corresponding business strategy to retain the existing customers. Instead of taking care of each individual customer, the company can allocate their limited resources and efforts into certain clusters for cost savings. In the past decades, numerous business-related customer-clustering approaches were conducted. Wu and Chou (2011) established good customer relations and refined their marketing strategies to match customer expectations. They developed a latent mixed-class membership clustering approach to classify online customers based on the purchasing data across categories. Ren, Zheng, and Wu (2009) presented a clustering method based on genetic algorithm (GA) for telecommunication customer subdivision. Similarly, Ho et al. (2012) proposed a K-means clustering approach based on a robust GA to classify the existing customers. Different from the traditional K-means algorithm, their proposed algorithm is able to detect the optimal number of clusters. Huang, Chen, and Khoo (2012) took customers' voices and opinions into account, and developed a genetic clustering method for emotional design using a combined design structure matrix. Carpaneto, Chicco, Napoli, and Scutariu (2006) made special efforts to cluster electricity customers' representational load patterns on the frequency domain.

In the domain of logistics operations, clustering customers by their characteristics in a large-scale network is not an easy task. Customer similarity is affected by various factors, such as the customer demand, local traffic condition, market environment, and the time window requirement, etc. Most of these attributes are difficult to measure in a quantitative form. This is because the majority of the above attributes are discrete, and traditionally obtained by human perception. Fuzzy set theory is considered as an appropriate countermeasure to tackle this vagueness and ambiguity. Many researchers have utilized fuzzy set theory to handle ambiguous scenarios in the decision-making procedure (D'Urso, Giovanni, Disegna, & Massari, 2013; Golmohammadi, 2011; Hu & Sheu, 2003; Selim, Araz, & Ozkarahan, 2008; Sheu, 2006; Wang, 2010; Zadeh, 1965). In Fuzzy set theory, linguistic terms are used to evaluate different subjective attributes (Wong & Lai, 2011; Chan, Kwong, & Hu, 2012), for example, "Very Low", "Low", "Medium", "High", "Very High", "Very Poor", "Poor", "Fair", "Good", "Very Good", etc. (Jacobsen, 2002; Lao, Wu, Wang, & McAllister, 2012; Li, Dai, & Tseng, 2011a; Liu & Jin, 2012; Wang et al., 2012). These linguistic variables are well suitable to transform into fuzzy numbers. There exist a variety of typical forms for fuzzy numbers, including trapezoidal fuzzy numbers, triangular fuzzy numbers and interval fuzzy numbers. Trapezoidal fuzzy numbers are considered as the general form of fuzzy numbers, and they are easy and accurate to process the linguistic variables (Liu & Jin, 2012).

Due to the inherent advantages of fuzzy set theory, many fuzzy systematic analysis methods are widely adopted into the logistics network operations and customer clustering process in different research fields. Sheu (2004) proposed a hybrid fuzzy-based method that combined fuzzy-AHP with fuzzy-MADM approaches for determining global logistics strategies. Sheu (2008) presented a hybrid neuro-fuzzy approach to choose appropriate global logistics operational modes for global supply chain management. Qin and Ji (2010) utilized the fuzzy programming tool to design a product recovery logistics network based on different criteria. Vahdani, Moghaddam, and Jolai (2013) presented a new solution approach combined fuzzy probabilistic programming and fuzzy multi-objective programming to address the bi-objective model for designing a reliable logistics network. Shin and Sohn (2004) proposed a fuzzy K-means clustering algorithm to group the stock trading customers into three tiers (Normal, Best and VIP level). The major criterion is the total trade amount over a 3-month period for each customer. Wang (2010) developed a clustering algorithm based on the fuzzy equivalence relation. The linguistic data sequences were firstly interpreted by fuzzy data sequences, and then the sequence with similar preference was classified into one cluster. The proposed clustering algorithm was successfully applied to mine the customer relationship. Chan et al. (2012) recently proposed a new methodology to perform market segmentation based on customers' requirements. In their paper, fuzzy compression technique was firstly used to reduce the high dimensions of customer requirements to two dimensions, and then, a hierarchical fuzzy clustering algorithm was applied to grouping customers with similar characteristics into same cluster based on the compressed data.

Very few relevant studies were conducted to address the customer clustering issues for logistics operations. As mentioned in the introduction, customer clustering is an intermediate stage during the optimization procedure for logistics distribution networks, and it is the critical premise on the vehicle routing planning issue. Hu and Sheu (2003) initially utilized fuzzy clustering to classify potential logistics customers into multiple groups based on the five major attributes: safety, transit time, transportation cost, accessibility and service quality. Based on the research by Hu and Sheu (2003), Sheu (2007) presented an integrated fuzzy-optimization framework to identify the customers with similar characteristics considering multiple attributes of customer demands. However, both of the above studies may suffer from the following issues: (1) The traditional clustering algorithm is not able to handle the