Computers & Industrial Engineering 118 (2018) 33–43

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Forecast of individual customer's demand from a large and noisy dataset

Paul W. Murray[a,*], Bruno Agard[a], Marco A. Barajas[b]

[a] École Polytechnique de Montréal, Montréal, QC H3C 3A7, Canada
[b] Fisheries & Marine Institute of Memorial University of Newfoundland, St. John's, NL, Canada

ARTICLE INFO

Keywords:
Strategic forecasting
Data mining
Intermittent demand
Time-series
Delivery records
Behavior patterns
Market segmentation

ABSTRACT

Optimization of the supply chain relates on data that describes actual or future situation. Besides in many situations available data may not correspond directly to what is expected for the different models because of too large quantity and imprecision of the data that may lead to suboptimal or even bad decisions. Actual problem considers the availability of a large and noisy dataset concerning historical information about each customer that will be used to make improved prediction models, that may fit models to optimize the supply chain.

When dealing with large datasets, market segmentation is frequently employed in business forecasting; many customers are grouped based on some measure of similarity. Segment-level forecasting is then employed to represent the population within each segment. Challenges with successfully applying market segmentation include how to create segments when descriptive customer information is lacking and how to apply the segment-level demand forecasts to individual customers. This research proposes a method to create customer segments based on noisy historical transaction data, create segment-level forecasts, and then apply the forecasts to individual customers. The proposed method utilizes existing data mining and forecasting tools, but applies them in a unique combination that results in a higher level of customer-level forecast accuracy than other traditional methods. The proposed forecasting method has significant management applications in any domain where forecasts are needed for a large population of customers and the only available data is delivery data.

1. Introduction

Supply chains create value by transforming and transporting goods and services that satisfy the demand requirements of downstream partners; understanding demand requirements is necessary for strategic planning (Carbonneau, Laframboise, & Vahidov, 2008). Forecasting methods all share a common prerequisite; they require suitable input information from which demand behaviors can be extracted, interpreted, and predicted. In ideal situations, strategic forecasts incorporate collaborative information shared by downstream partners, however, supply chain partners are not always able or willing to share information (Holweg, Disney, Holmström, & Småros, 2005). When collaborative information is not available, the supplier must rely on other information such as historical data to build its forecasts; however, the available data is not always the correct data to use (Makridakis, 1989) and may not truly represent actual consumption behaviors. Historic data collected at the point of consumption is preferred, but sometimes not available. Data obtained upstream in a supply chain is often distorted due the Bullwhip effect (Forrester, 1958); it can become noisy, lumpy, and intermittent, making it difficult to use for segmentation and forecasting.

In addition to obtaining suitable input data, the forecaster must also consider how to manage large populations of customers. In domains with many customers, it is advantageous to divide the market into analogous segments so that a manageable number of forecasts can be created & maintained (Calvet, Ferrer, Gomes, Juan, & Masip, 2016; Wind, 1978). Market segmentation has the additional benefit of improving forecast accuracy due to forecast errors off-setting each other (Armstrong, 2006). A critical component of market segmentation is to identify customers' distinguishing attributes from which similarities and differences can be measured (Calvet et al., 2016; Huber, Gossmann, & Stuckenschmidt, 2017). When distinguishing attributes are absent, historical data in the form of transaction records is often the only information available—just as forecasts can be built from historical data, markets can be segmented from the same data.

A challenge with forecasting by market segments is applying the segment-level forecasts to individual customers and validating the results. While the literature is rich with methods for creating customer segments, there is little guidance on how to apply segment-related predictions to individual customers and evaluate the results.

In this research, we present a method that produces customer-specific forecasts using noisy delivery data as the only input. This method

* Corresponding author.
E-mail addresses: paul.murray@polymtl.ca (P.W. Murray), bruno.agard@polymtl.ca (B. Agard), marco.barajas@mi.mun.ca (M.A. Barajas).

https://doi.org/10.1016/j.cie.2018.02.007
Received 21 July 2017; Received in revised form 18 December 2017; Accepted 6 February 2018
Available online 08 February 2018
0360-8352/ © 2018 Elsevier Ltd. All rights reserved.

is based on several steps that permit segmenting the customer's consumption behaviors based on the delivery data, generating forecasts for each segment, scaling those forecasts to individual customers, and finally, evaluating the accuracy of the predictions. This new method is a significant contribution in that it produces segment-based forecasts using noisy data and then applies the forecasts to individual customers. Rather than create new tools, we apply proven effective tools in a way that they have not been used before to solve a persistent problem. The proposed method can be effectively applied in many domains and provide significant value in many applications.

The remainder of the paper is organized as follows: Section 2 presents a review of the literature followed by the proposed method in Section 3. Section 4 is a case study in which the method is applied to a real industrial dataset and compared with other methods. The conclusions are presented in Section 5.

2. Literature review

In the context of this research, we propose a new method that combines several sequential steps including treatment of noisy data, market segmentation, forecasting, and forecast evaluation. The state of the art for these steps are summarized in the following sections of the literature review.

2.1. Treatment of noisy data

When transactional data (such as delivery records) is substituted for absent consumption demand information, it must first be transformed into a suitable format for the analysis. Transaction data normally contains a time-stamp that facilitates aggregation into temporal bins, however, selecting an appropriate bin duration is not trivial (Petropoulos & Kourentzes, 2015). High aggregation (longer periods) removes noise and reduces or eliminates zero-quantity periods, however, it also carries the risk of excessive smoothing and loss of information (Spithourakis, Petropoulos, Nikolopoulos, & Assimakopoulos, 2012). Small aggregation reduces smoothing and loss of information, but it tends to create periods of zero-quantity resulting in intermittent time-series. After transformation, the dataset is still difficult to interpret due to noise incorporated from the Bullwhip effect (Forrester, 1958) and logistics decisions.

Croston (1972) proposed a method to update a forecast for intermittent demand where updates only occur when demand occurs; during periods of zero-demand, the forecast remains unchanged. Croston's method is calculated as per formulae (1), (2), and (3):

$$\hat{y}_t = \hat{Z}_t/\hat{X}_t \tag{1}$$

where $\hat{Z}_t$ is the SES forecast for the non-zero periods and $\hat{X}_t$ is the forecast for the number of inter-demand intervals. Both the demand size and the intervals use SES per formulae (2) and (3):

$$\hat{Z}_t = \alpha_z y_t + (1-\alpha_z)\hat{y}_t \tag{2}$$

$$\hat{X}_t = \alpha_x x_t + (1-\alpha_x)\hat{x}_t \tag{3}$$

where y_t is the non-zero demand at time t and x_t is the number of non-zero intervals. The smoothing coefficient α is set to 0.02.

An improved method for resolving noise in time-series data is the aggregate-disaggregate intermittent demand approach (ADIDA), proposed by Nikolopoulos, Syntetos, Boylan, Petropoulos, and Assimakopoulos (2011). ADIDA reduces noise by aggregating into high-level temporal bins and then reverting to the original bin sizes. This approach is appealing due to its effectiveness and simplicity.

2.2. Market segmentation

Market segmentation is comprised of two distinct steps: measuring the difference between customers (known as a distance measure), and then creating analogous segments (a process known as clustering). Many methods exist for dividing a market into segments (Chakraborty, 2013; Han & Kamber, 2006; Le, Agard, & Deveault, 2009), but in general, the methods are based either on descriptive attributes or behavioral attributes (Murray, Agard, & Barajas, 2017). Descriptive attributes (such as size, location, sex, age or transaction frequency) are commonly used since this type of variable is easy to quantify. Success in segmentation based on descriptive attributes relies heavily on two assumptions: that data to support the variables is available, and that the descriptive attributes are truly relevant for creating the segments (Gur Ali & Pinar, 2016). There are many external factors, such as economy, competitors' actions, and social perception, that influence customer behavior (Sevlian & Rajagopal, 2018), however, selecting appropriate factors can be challenging. Segmenting based on behavioral attributes requires historical data that reflects the behavior and a method for extracting and identifying the behavior within the data (Barragan, Fontes, & Embiruçu, 2016; Kashwan & Velu, 2013). In the context of this research, descriptive attributes are not available; the historical delivery transactions, converted into time-series format, are the input.

2.2.1. Time-series distance measures

The advent of data mining and the ability to process vast amounts of data has made it practical to detect differences between multiple time-series. Early efforts to index time-series relationships did not produce meaningful results due to the focus on exact matches rather than similarities (Agrawal, Faloutsos, & Swami, 1993). In order to measure similarity rather than sameness, a method for elastically shifting the time indexes was needed (Keogh & Ratanamahatana, 2005). The ability to identify similar customers rather than exact matches enables the grouping of customers by similar behavior. Planning can then be based on groups rather than on individuals.

Dynamic time warping (DTW) was developed in the 1970s for speech recognition (Sakoe & Chiba, 1978) and is a popular method for indexing time-series data through elastic manipulation of the time axis (Giorgino, 2009). DTW is considered an expensive algorithm with regards to computing time (Zhang, Kaiqi, & Tieniu, 2006), however, faster computer hardware and continual refinement of DTW algorithms have diminished the concern of computational expense. Exhaustive literature reviews have concluded that DTW is the best method for comparing time-series similarities (Ding, Trajcevski, Scheuermann, Wang, & Keogh, 2008; Rakthanmanon et al., 2013).

2.2.2. Clustering time-series

Once customer similarity has been quantified with a suitable distance measure, groups can be established. Clustering is the process of grouping a population such that the inter-group homogeneity is maximized and the intra-group heterogeneity is also maximized (Esling & Agon, 2012). Ideally, group members are similar to each other and dissimilar from members of other groups. Unfortunately, there is a vast collection of clustering algorithms in the literature and no definitive means of selecting a best method (Jain, Murty, & Flynn, 1999). According to Liao (2005), clustering time-series data is most commonly accomplished through partitional clustering, artificial neural networks (ANNs), or hierarchical clustering. Partitional clustering, such as the popular k-means, are generally easy to use and have low computational expense (Kantardzic, 2011). Partitional cluster algorithms commonly create groups by minimizing the total squared error for a giving number of clusters (*K*). Time-series data does not necessarily provide an adequate description of the differences within a population and subsequently, direct comparison through squared error measurement does not always result in useful information (Murray, Agard, & Barajas, 2015). Partitional clustering is also hindered by the non-trivial task of pre-defining the number of clusters (*K*).

Hierarchical clustering, first introduced by Ward (1963), creates a hierarchical structure of grouping based on similarity measures. The hierarchical structure can be graphically displayed as a dendrogram,

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات