Jestr

Journal of Engineering Science and Technology Review 16 (6) (2023) 90 - 97

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Demand Forecasting for the Full Life Cycle of New Electronic Products Based on KEM-QRGBT Model

Binlong Lin [1], Yi Wu [1], Juanjuan Wu [2,*] and Chenghu Yang [1]

[1]*School of Economics and Management, Fuzhou University, Fuzhou 350108, China*
[2]*School of Mathematics and Statistics, Fuzhou University, Fuzhou 350108, China*

Received 22 July 2023; Accepted 8 December 2023

Abstract

To improve the accuracy of demand forecasting for new electronic products, especially in scenarios with limited historical data, a novel forecasting model was proposed in this study which integrated K-means based on Euclidian distance, Multi-layer perceptron algorithm, and Quantile Regression with Gradient Boosted Trees (KEM-QRGBT). The model also incorporated grid search with K-fold cross-validation to enable the adaptive selection of the optimal parameters for product data. Additionally, the KEM-QRGBT model, which can balance the intricacies of learning parameter patterns with its ability to quantify demand uncertainty, exhibited proficiency in quantifying the uncertainty inherent in demand forecasting. Using a case study from a manufacturing enterprise in Turkey, the effectiveness of the model was validated. Results demonstrate that, for new electronic products with limited historical data, the KEM-QRGBT model with adaptive parameter selection improves demand forecasting accuracy, outperforming benchmark methods, and other machine learning models. The proposed algorithm provides a strong evidence for the demand forecasting of new electronic products, particularly in cases where historical data is limited.

Keywords: Demand forecasting, Time series clustering, Deep learning classification, Ensemble learning, New products

1. Introduction

In the initial stages of product development, accurate demand forecasting for the entire product life cycle becomes crucial. It profoundly influences the development of product positioning strategies and the formulation of intricate market penetration strategies. As technology rapidly progresses and competitive landscapes evolve, many industries are experiencing a trend toward increasingly shorter product life cycles. Prominent sectors such as electronics, automobiles, and fashion are prime examples, often showcasing annual product updates or revisions. This evolving dynamic presents significant challenges for decision-makers (DMs), who often lack sufficient historical data to forecast new product demand accurately before the sales season [1]. The limited availability and incompleteness of historical data skew the forecasting accuracy in the volatile demand [2]. The consequences of inaccurate demand forecasting are not trivial. Major companies such as Microsoft, Samsung, and Lenovo have experienced substantial losses due to miscalculations in demand forecasting for new electronics [3]. These examples underscore the importance of refining forecasting methodologies to better align with the realities of today's fast-changing markets.

Demand forecasting for new products is a classic cold start problem, which has received more attention. Products with brief life cycles face challenges like extended lead times, a dearth of early historical data, and seasonal influences, complicating demand predictions. To address the challenges, a widely adopted method is analogical reasoning, leveraging the historical data of similar products for new product forecasting. However, the traditional demand forecasting models for new products that employ analogical reasoning tend to be subjective. The process of selecting comparable products often hinges on subjective criteria and depends heavily on expert judgment, which results in a larger forecasting error. In addition, point forecasting is frequently-used in demand forecasting for new products. Due to few relevant historical data, traditional point forecasting methods often result in significant discrepancies from actual demand and are ill-suited to environments with short product life cycles [4]. Interval forecasting, by contrast, provides a range of potential outcomes, allowing DMs to more accurately assess the risk associated with future demand for new electronic products. Hence, to address these issues, this study employed analogical learning principles combined with cluster analysis, deep learning classification, and ensemble regression techniques, to develop a novel interval forecasting model for new electronic products before the sales season. We aim to provide DMs with insights into the life cycle trends and demand intervals at different stages for new electronic products, even in the absence of historical data. This model can help manufacturer improve their decision-making processes and decrease marketing costs.

2. State of the art

Demand forecasting for new products often relies on qualitative analysis, yet this approach can be skewed by external factors. There is an urgent need to employ quantitative models for forecasting demand [5]. In this realm, quantitative methodologies are broadly classified into case-based reasoning (CBR) and Bass diffusion model. The CBR method identifies and adapts solutions from similar historical cases to predict new product demand, proving

*E-mail address: 210327009@fzu.edu.cn
ISSN: 1791-2377
doi:10.25103/jestr.166.11

effective in fields like industrial design and printed circuit board production. However, due to the uncertainties in knowledge representation, attribute description, and similarity measurement within CBR, identifying similar cases from a database is challenging. Moreover, the selection process is often subjective. Previous studies have addressed these issues in traditional CBR systems by integrating clustering techniques [6,7]. However, the individual adaptation of models in CBR can be resource-intensive, inefficient, and not fully adaptable to the fluctuating nature of product demand forecasting. The Bass model, on the other hand, considers consumer adoption patterns and word-of-mouth effects to model new product diffusion [8]. Its limitation lies in heavy reliance on historical data for parameter estimation, reducing its effectiveness for products without historical data. To address this issue, some studies have employed analogy-based approaches, and selected parameters for new product diffusion models from similar established products [9]. Lee et al. [10] developed a library of product attributes through expert analysis and created a parameter database from the sales data of older products. They then applied machine learning algorithms to categorize products and construct the Bass model, providing demand forecasting for new electronic products in the US. Yin et al. [2] refined this approach by incorporating fuzzy clustering and sets in the Bass model, forecasting demand for new cellphones and cars in China. Zhou et al. [11] utilized analogical reasoning to assess clothing similarity, and devised an augmented Bass model with consumer preferences and seasonal variations, predicting clothing pre-sale demand. However, the Bass model is mainly applied for durable products and may not be as effective for new electronic products.

Consequently, various scholars have explored machine learning to address the inherent limitations of the mentioned approaches. For example, Thomassey et al. [12] applied the K-means algorithm to categorize product life curves into distinct groups and utilized decision tree classifiers to describe the life curves of new electronic products. A self-organizing map is introduced into this approach to improve classification accuracy [13]. Lu et al. [14] utilized the K-means algorithm to categorize older products and provided a voting mechanism to categorize new products by using cluster linkage rules. Tehrani et al. [15] used K-means clustering to segment product demand into three tiers and implemented logistic regression and regression tree approaches to forecast future demands of new products. Hu et al. [16] used subjective judgment to classify new products and applied multiple curve-fitting techniques for forecasting life curves. Van Steenbergen et al. [17] expanded the random forest algorithm to predict the total demand for new products, combining the total demand forecasting with the predicted life curve to determine multi-period demand values. Lei et al. [18] developed a dynamic life curve forecasting model using a Bayesian regression, capturing real-time product promotion data. Elalem et al. [19] implemented hierarchical clustering and machine learning techniques to determine the future life curve of new products. However, these models also exhibit limitations. The approaches used in [12], [16], and [18] focus solely on life curve forecasting, hindering precise demand forecasting for different sales periods. The model proposed in [14] ignores product demand heterogeneity, assuming average group demand as the future demand for new products. In [15], subjective segmentation is adopted, leading to greater error margins in classification. The model in [17] lacks objective algorithm selection and hyperparameter optimization, making it less versatile for varied retail scenarios.

The above literature highlights the importance of addressing demand forecasting for new electronic products. Notably, there exists a significant gap in examining the process of parameter selection, especially from a data-driven perspective, and in acknowledging the inherent uncertainties within predictive models. Hence, this study developed a novel demand forecasting model by incorporating K-means based on Euclidian distance (K-means-ED), Multi-layer perceptron (MLP) algorithm, and Quantile Regression with Gradient Boosted Trees (QRGBT). The model calculated the similarity of the life curve of older products based on Euclidian distance, and combined K-means algorithm with a multi-layer perceptron algorithm to forecast the future life curve of new electronic products. Furthermore, the model employed the QRGBT algorithm for forecasting demand intervals at various life stages of new products, predicting the total demand over the product's life cycle. In particular, the model utilized a grid search cross-validation (GSCV) algorithm, which incorporates grid search with K-fold cross-validation, to automatically select the optimal parameters for the collected dataset. Additionally, the QRGBT algorithm effectively quantifies the uncertainty in demand forecasting, balancing the complexity of parameter learning. Using the electronic products of a manufacturing enterprise as an example, the performance of the model was verified on the total demand forecasting of the life cycle of new products and the demand intervals forecasting of different life stages.

3. Methodology

3.1 Model construction

Based on the time series dataset of older products $TS_i^x = \left\{TS_{i,1}^x, TS_{i,2}^x, \ldots, TS_{i,T}^x\right\}$, $i \in \{1,2,\ldots,N\}$, the KEM-QEGBT model can forecast the demand for new electronic products $TS_m^y = \left\{TS_{m,1}^y, TS_{m,2}^y, \ldots, TS_{m,T}^y\right\}$, $m \in \{1,2,\ldots,M\}$. Where the life cycle length of electronic products is T, the number of older products is N, and the number of new electronic products is M. The framework of the KEM-QRGBT model is as follows:

1) Life curve clustering: The life cycles of older products are clustered after normalization using the K-means algorithm based on Euclidean distance. Further, the mean of each cluster is extracted to define a unified trend (prototype) of the life cycle clusters. Finally, the prototype is considered as a potential life curve of new electronic products.

2) Life cycle forecasting: The MLP algorithm is trained by using the characteristics of older products and the corresponding relationship of their life cycles. Based on the features of the new electronic products, the cluster label with the highest probability can be obtained. Finally, the predicted life cycle curve of new electronic products can be learned from the cluster's prototype.

3) Total demand quantile forecasting: The QRGBT algorithm is trained on the features of older products and their total demand over the life cycle. By inputting the features of new electronic products and adjusting the quantiles of the Quantile Regression Tree, the total demand for new electronic products is obtained for quantiles ranging from 0.01 to 0.99.

4) Multi-period demand interval forecasting: Through multiplying the predicted total demand of new electronic products at different quantiles with the predicted life cycle

curve, the multi-period future demand and their respective quantiles for new electronic products are obtained.

The framework of the KEM-QRGBT model is shown in Fig. 1.

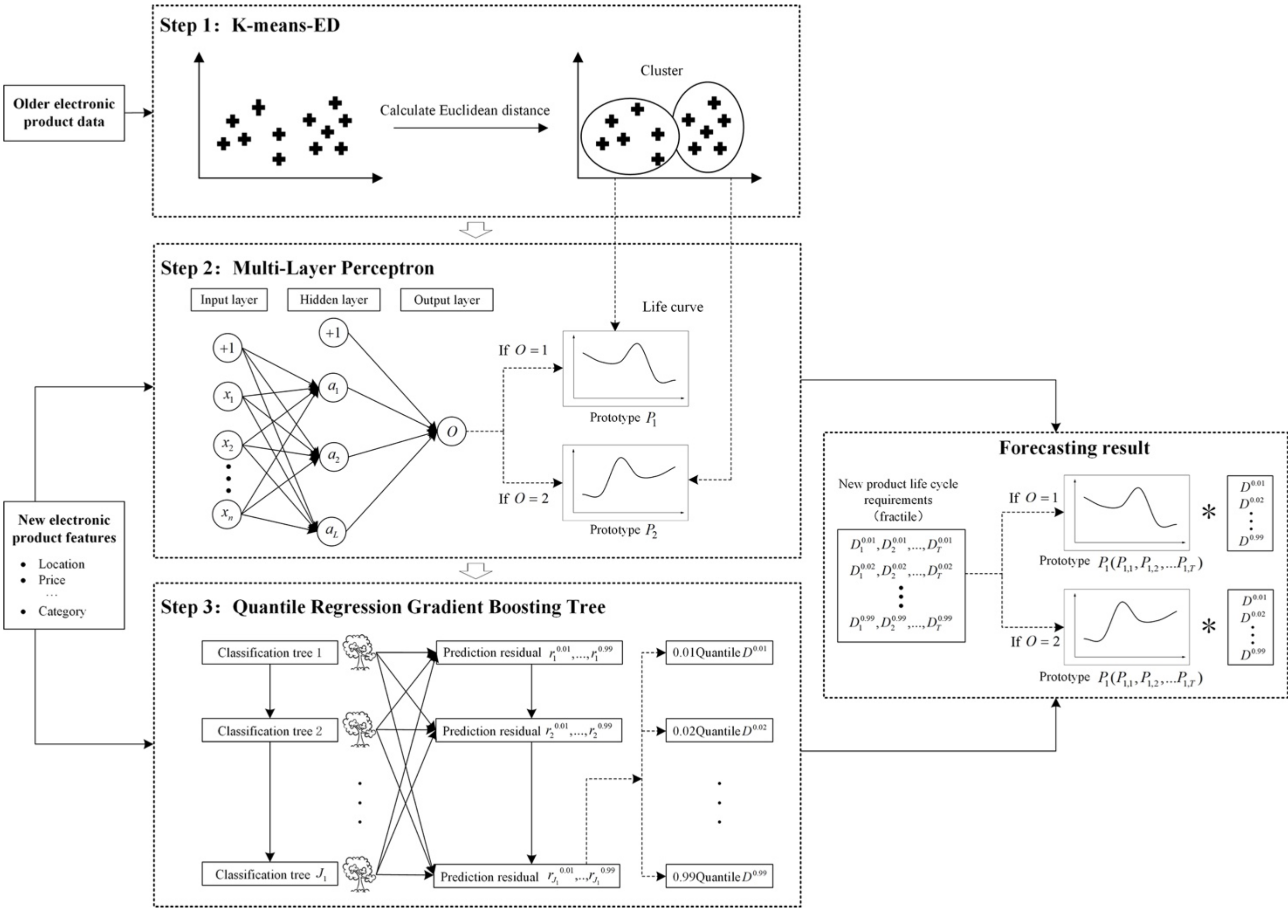

Fig. 1. The framework of the KEM-QRGBT model

3.2 Older electronic product life cycle clustering

Multi-period forecasting methods rely on multi-step forward forecasting based on existing time series. However, new electronic products lack historical demand data for these models to learn. Therefore, the KEM-QRGBT model utilizes the K-means algorithm to group life cycles with high similarity into clusters, and employs centroid extraction based on Euclidean distance to obtain prototypes for potential life cycle curves of new electronic products.

To facilitate the clustering of products together, it is necessary to normalize demand time series to overcome the influence of dimension, as shown in Eq. (1):

$$\overline{TS}_{i,t}^{x} = \frac{TS_{i,t}^{x}}{\sum_{t=1}^{T} TS_{i,t}^{x}} \tag{1}$$

where, $\overline{TS}_{i,t}^{x}$ is the proportion of the demand of the ith older product in period t after normalization. This study adopts the Euclidean distance to measure the similarity of time series, as defined in Eq. (2):

$$dist\left(\overline{TS}_{i}^{x}, \overline{TS}_{i_1}^{x}\right) = \sqrt{\sum_{t=1}^{T}\left(\overline{TS}_{i,t}^{x} - \overline{TS}_{i_1,t}^{x}\right)^2} \tag{2}$$

where $i_1 \in \{1,2,\ldots,N\}$ and $i_1 \neq i$. Considering prototype extraction dependent on the distance measurement method [20], this study selected Euclidean distance gravity center extraction algorithm to obtain the prototype. Assuming that there is n_k life curves in kth cluster C_k, the prototype is obtained as follows:

$$P_{k,t} = \frac{\sum_{i=1}^{n_k} \overline{TS}_{i,t}^{x}}{n_k} \tag{3}$$

where, $P_{k,t}$ refers to the value of the prototype P_k for the tth period. The objective of K-means is to minimize the sum of squared errors, which essentially classifies life cycles with similar trends into one category, as specifically depicted in Eq. (4):

$$SSE_k = \frac{\sum_{k=1}^{K} \sum_{\overline{TS}_i^x \in C_k} dist\left(P_k, \overline{TS}_i^x\right)^2}{n_k} \tag{4}$$

Based on the above, the process of K-means-ED time series clustering algorithm is as follows: (1) the life curve set is generated through normalization, and an arbitrary life curve is selected as the initial prototype. (2) the distance of each life curve is calculated by using Euclidean distance, and each life curve is initially assigned to the nearest cluster. (3) through K-means algorithm, the life curves are reallocated, and the prototypes can be determined. The prototype can summarize the characteristics of all life curves, to avoid the issue of using a single similar old product life curve as the benchmark error of life curve.

3.3 Forecasting of new electronic product life cycle

The KEM-QRGBT model utilizes K-means clustering to convert the subjective forecasting of new electronic product life cycles into a data-driven classification problem. Considering the uncertainty and complex relationship between new electronic product life cycles and their features, the MLP algorithm is selected to predict the life cycles of new electronic products. The MLP algorithm enhances the adaptive learning and inferencing abilities of models and describes the nonlinear relationship by weighting neuron inputs and utilizing activation functions. This effectively mines the deep intrinsic connections between multi-dimensional life cycles and product features.

$$k = softmax\left(o\left(W^{(2)}a\left(W^{(1)}x + b^{(1)}\right) + b^{(2)}\right)\right) \tag{5}$$

$$\overline{TS}^{fore} = P_k \tag{6}$$

Eq. (5) describes a fully connected MLP algorithm, where *k* is the number of the predicted life cycle cluster, and $\overline{TS}^{fore}$ is the predicted life cycle of new electronic products. The hidden layer can be denoted as the weight vector $W^{(1)}$. The output of the hidden layer, after being mapped by the activation function *a*(·), connects to the nodes $W^{(2)}$ of the output layer. The output of the layer is then obtained through another activation function *o*(·), with $b^{(1)}$ and $b^{(2)}$ representing the biases of the nodes at each layer. Since the forecasting of $\overline{TS}^{fore}$ is a classification problem, the KEM-QRGBT model brings the output of MLP into *softmax* function, transforms it into the probability distribution, and selects the prototype of cluster number with the largest probability value as the forecasting life curve.

3.4 Forecasting of total life cycle demand for new electronic products

The life curve of new electronic products describes the proportion of demand in various periods. However, it does not consider the significant differences in demand scales among different products. To accurately assess this heterogeneity, the model introduces the QRGBT algorithm to predict the total demand for products, which combines with the predicted life curve to determine the specific demand for various periods.

Assuming that the QRGBT model comprises J_1 quantile regression trees, each with *L* leaf nodes $R_1, R_2, \ldots R_L \in R$ and response variables $y_l = \{y \mid x_i \in R_l\}$, *l*=1,2,…,*L*. Given a training set $(x, D) = \{(x_1, D_1), (x_2, D_2), \ldots (x_N, D_N)\}$, where $D_i (D_i \in R)$ represents the total demand corresponding to the older product feature vector $x_i (x_i \in R_l)$, the output values of each leaf node with τ quantile are calculated as follows:

$$\begin{aligned}\hat{Q}_y(\tau \mid x_i \in R_l) &= F^{-1}(y \mid x_i \in R_l) \\ &= \inf\{y_l : F(y \mid x_i \in R_l) \geq \tau\}\end{aligned} \tag{7}$$

where $F(y \mid x_i \in R_l) = 1/n_l \sum_{x_i \in R_l} I(y_l \leq y)$ represents the empirical conditional distribution function, *I*(·) is the indicator function, and n_l is the number of occurrences of y_l.

The QRGBT algorithm uses the steepest descent method to produce multiple quantile regression trees iteratively. The results of all regression trees are linearly combined to determine the QRGBT algorithm. The specific steps for generating the QRGBT algorithm are as follows:

1) Initialize the QRGBT algorithm as $G_0(x)$;

2) When the QRGBT algorithm iterates to the *j*th tree, for feature x_i, define the residual r_i as follows:

$$\begin{aligned}r_i &= -\frac{\partial L_\tau(D_i - G)}{\partial G}\bigg|_{G = G_{j-1}(x_i)} \\ &= I\left(\left(D_i - G_{j-1}(x_i)\right) \geq 0\right) - (1 - \tau)\end{aligned} \tag{8}$$

where, $L_\tau(D_i - G)$ represents the loss function. Through fitting r_i, we can obtain the quantile regression tree $T_j(\hat{\theta}_j(\tau))$. Where $\hat{\theta}_j(\tau) = \{R_l^j, \hat{Q}_r^j(\tau \mid x_i \in R_l^j)\}_{l=1}^L$ is an independently and identically distributed variable controlling the growth of the quantile regression tree.

3) Update the quantile regression with gradient boosting tree algorithm as follows:

$$G_j(x_i) = G_{j-1}(x_i) + T_j(\hat{\theta}_j(\tau)) \tag{9}$$

4) Repeat the first two steps until the QRGBT algorithm reaches the convergence or the maximum number of iterations, resulting in the following model:

$$G_j(\tau \mid x_i) = \sum_{j=1}^{J_1} G_j(x_i) \tag{10}$$

Through utilizing the feature vector x of new electronic products, quantiles of the conditional distribution of the total demand forecasting are calculated as follows:

$$\hat{Q}_y(\tau \mid X = x) = \sum_{j=1}^{J_1} \hat{Q}_r^j(\tau \mid x \in R_l^j) \tag{11}$$

4. Result Analysis and Discussion

4.1 Data processing

The data selected in this study is from the product data of a manufacturing enterprise. To verify the effectiveness of the KEM-QRGBT model, two sample sets from different product lines are utilized. Sample set A contains 697 samples, and sample set B contains 570 samples. The split ratio of the training set and testing set is 8:2.

Considering the heterogeneity of identical products from different suppliers, the data should be processed into the same format, including demand time series and features. This study extracts products and suppliers as primary keys, aggregates product demand monthly, and filters out products with life cycle lengths greater than or equal to 12 months. Before applying the MLP algorithm, this study performs one-hot encoding on non-numeric features, achieving a mapping from non-numeric to numeric features. Additionally, considering the sensitivity of regression algorithms to the scale of numeric features, numeric features are standardized before applying the gradient boosting tree algorithm. The mean of the standardized feature columns is 0, with a variance of 1.

4.2 Parameter setting

In parameter optimization, this study integrates grid search with K-fold cross-validation. The KEM-QRGBT model pairs up the parameters to be set for each algorithm. For the parameter selection process, the forecasting system randomly splits the training set into K_1 parts. Where $K_1 - 1$ subsets are used for training and the remaining subset for testing. Each data point is only allocated once to the training set or the testing set. In the K_1th cross-validation, a grid search is conducted to traverse the two-dimensional grid formed by candidate parameter value combinations. After k_1 th $\left(k_1 = 1, 2, \ldots, K_1\right)$ cross-validation, the K_1 evaluation metric values of each combined parameter in the two-dimensional grid are averaged, and the combination with the best performance is selected as the model parameters. Considering the complexity of combining cross-validation with grid search, this study set $K_1 = 5$.

4.3 Benchmark comparison analysis

To assess the effectiveness of the KEM-QRGBT model, this study analyses the predicted results of the life curve, total demand interval, and multi-cycle demand for new electronic products predicted from the KEM-QRGBT model.

1) Accuracy of full life curve characterization: A comparative analysis between the K-means-ED algorithm and common clustering methods (self-organizing maps, K-shape, K-means with Dynamic Time Warping (K-means-DTW)) is conducted to verify that the K-means-ED algorithm can better describe the product similarity. Based on the results of the K-means-ED algorithm, a comparative analysis between the MLP algorithm and common classification methods, including Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and Gradient Boosting Regressor Tree (GBRT), is carried out to demonstrate that the integration of K-means-ED and MLP algorithms can more accurately determine the life curve of new electronic products.

2) Demand interval forecasting: By comparison among KEM-QRGBT, KEM with Quantile Regression Forest (KEM-QRF), and KEM with Lasso Penalized Quantile Regression (KEM-LassoQR) models to verify whether the integration of QRGBT can reduce demand uncertainty.

4.4 Evaluation index

Due to the unsupervised nature of K-means-ED, the size of cluster number K value needs to be set in advance. The KEM-QRGBT model utilizes the Calinski-Harabasz (CH) index to assess the effectiveness of the clustering algorithm under different K values. The CH index is calculated as follows:

$$CH(K) = \frac{Tr\left(B_K\right)}{Tr\left(W_K\right)} \times \frac{N-K}{K-1} \tag{12}$$

where K is the number of clusters, $Tr(X)$ depicts the trace of the matrix X, N describes the total number of samples, B_K is the inter-group covariance, and W_K is the intra-group covariance. The larger the CH index, the better the clustering effect. In this study, the contour coefficient and Davies-Bouldin (DB) index are not used to measure clustering effect. The silhouette coefficient is calculated by measuring the distances between each life curve and all other life curves within the same cluster. Due to the high dimensionality of life curve data, the computational complexity of the algorithm for calculating the silhouette coefficient grows exponentially as the sample size increases. The distance metric for the DB index is the Euclidean distance, which may not effectively assess clustering algorithms based on alternative distances (i.e., dynamic time warping).

The classification accuracy is used to assess the effectiveness of the classification procedure. It represents the proportion of correct predicted samples, which reflects the probability of correctly assigning the new electronic product to the right life curve. For the mth (m=1,2,…,M) new electronic products, its real life curve belongs to cluster l_1 and predicted life curve belongs to cluster l_2. The classification accuracy is calculated as follows:

$$accuracy = \sum_{m=1}^{M} \frac{c}{M} \times 100\% \tag{13}$$

where c is a 0-1 variable. If l_1 equals l_2, c=1, otherwise c=0. The larger the accuracy, the better the classification effect.

For demand interval forecasting, Prediction Interval Coverage Probability (PICP) and Prediction Interval Normalized Average Width (PINAW) are used to comprehensively assess the effectiveness of interval forecasting. PICP represents the proportion of actual values that fall within the predicted interval bounds. A higher PICP value indicates better interval forecasting performance, as detailed in Eq. (14):

$$PICP = \frac{1}{M} \sum_{m=1}^{M} c_m \times 100\% \tag{14}$$

where c_m is a binary variable: it is 1 if the actual value falls within the forecasting interval, and 0 otherwise. The PICP metric does not consider the width of the interval forecasting. Therefore, this study introduces the PINAW metric to measure the confidence level of interval forecasting, calculated as follows:

$$PINAW = \frac{1}{MR} \sum_{m=1}^{M} \left(U_m - L_m\right) \tag{15}$$

where U_m and L_m denote the upper and lower bounds of the predicted demand for new electronic products, R denotes the range of forecasting target values. As a narrower forecasting interval provides more precise information, a smaller PINAW value is preferable.

4.5 Forecasting results of life curve of new electronic products

Due to the unsupervised nature of clustering, the most critical issue is the selection of the number of clusters (i.e., K). As shown in Table 1, with the same K value, the K-means-ED algorithm chosen in this study outperforms others like Self-Organizing Maps, K-shape, and K-means-DTW in clustering performance. It performs optimally at K=2, respectively averaging improvements of 3.47%, 59.71%, and 63.57% over the other models.

Table 1. Clustering methods CH index table

Sample Set	Number of Clusters	K-means-ED	Self-organizing Mapping	K-Shape	K-means-DTW
A	K=2	115.51	110.69	75.04	74.43
	K=3	102.45	72.97	42.92	42.98
	K=4	93.34	64.40	40.93	43.93
	K=5	83.23	57.99	31.58	37.43
B	K=2	83.11	81.02	50.22	48.80
	K=3	67.62	64.12	29.66	44.83
	K=4	71.62	53.07	27.90	47.62
	K=5	70.28	48.39	19.64	41.93

This study categorizes data sets A and B into two clusters, as shown in Fig. 2. Note that the full lines $\left(U_m - L_m\right)$ represent the life cycles, and the dotted lines denote the prototypes. Fig. 2 shows that the prototypes effectively reflect the general trend of most product life cycles. However, due to the information loss in the clustering center extraction process, the prototypes do not accurately represent the cases of extremely high demand for a few products. Furthermore, as observed in Fig. 2, due to the clustering approach's emphasis on maximizing differences between clusters, the life cycle trends within the two clusters exhibit strong complementarity. For set A, the first cluster has demand troughs concentrated in the middle of the life cycle, whereas the second cluster shows demand peaks in the middle. For set B, the first cluster's product demand peaks occur in the latter half of the life cycle, while the second cluster's demand peaks are in the first half.

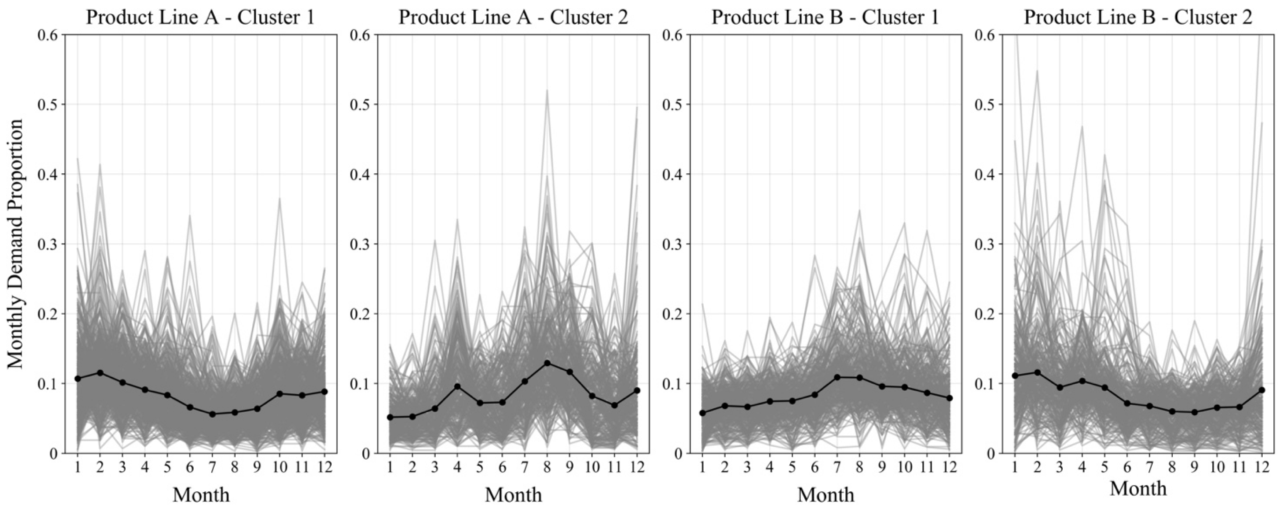

Fig. 2. K-means-ED clustering

After the K-means clustering, the KEM-QRGBT model identifies potential values for the future life cycle curves of new electronic products. The KEM-QRGBT model further employs the MLP algorithm to predict the life cycle curves of these products. As shown in Table 2, in experiments with two sets, the MLP algorithm achieved the most accurate life cycle curve forecasting, outperforming other models by an average of 6.14%, 7.15%, 9.28%, and 18.44%, respectively.

Table 2. Forecasting accuracy of classification method

Sample Set	MLP	RF	SVM	LR
A	80.71%	74.29%	80.71%	76.42%
B	70.18%	63.16%	61.40%	65.79%

4.6 Forecasting results of total demand for new electronic product life cycle

This study assesses the accuracy of demand interval estimates for new electronic products by using PICP and PINAW, assessing the KEM-QRGBT, KEM-QRF, and KEM-LassoQR models. This further measures the model's ability to depict demand for new electronic products in uncertain environments. The performance of the three algorithms is assessed within a 90% forecasting interval (with lower and upper quantiles at 5% and 95%, respectively), as shown in Table 3. Regarding the PICP metric, the KEM-QRF model performed the worst, lagging behind the QRGBT algorithm by 7.32% and 18.00% in sets A and B, respectively. In terms of the PINAW metric, the KEM-LassoQR model performed the worst, underperforming the KEM-QRGBT model by 105.47% and 55.50% in sets A and B, respectively. Thus, incorporating the QRGBT algorithm into the forecasting model can improve the accuracy of demand forecasting for new electronic products.

Table 3. Quantile regression algorithm total demand 90% interval forecasting performance

Index	Sample Set	KEM-QRGBT	KEM-QRF	KEM-LassoQR
PICP↑	A	87.86%	81.43%	88.57%
	B	87.72%	71.93%	86.84%
PINAW↓	A	403.46	376.10	829.00
	B	620.31	610.33	964.56

4.7 Multi-cycle demand forecasting results for new electronic products

Based on the KEM-QRGBT model, the forecasting interval for the multi-period demand of new electronic products can be determined by multiplying the forecasted life curve by the upper and lower bounds of the total demand interval forecasting. Considering the rationality of interval selection, this study investigates the impact of quantile interval size on the PICP and PINAW metrics for multi-period demand interval forecasting, as illustrated in Fig. 3.

For the PICP metric, the KEM-LassoQR model performs better in set A, while the KEM-QRGBT model excels in set B. However, for the PINAW metric, incorporating the Lasso quantile regression performs the worst in both sets, with the other two algorithms showing similar results. Notably, at a 98% forecasting interval, the PINAW values for the KEM-QRF and KEM-LassoQR models are significantly higher than the KEM-QRGBT model. From Fig. 3, the PINAW metric shows exponential growth after a 90% forecasting interval, while the PICP curve remains linear. Thus, this study evaluates the performance of each algorithm in multi-period demand interval forecasting at a 90% forecasting interval.

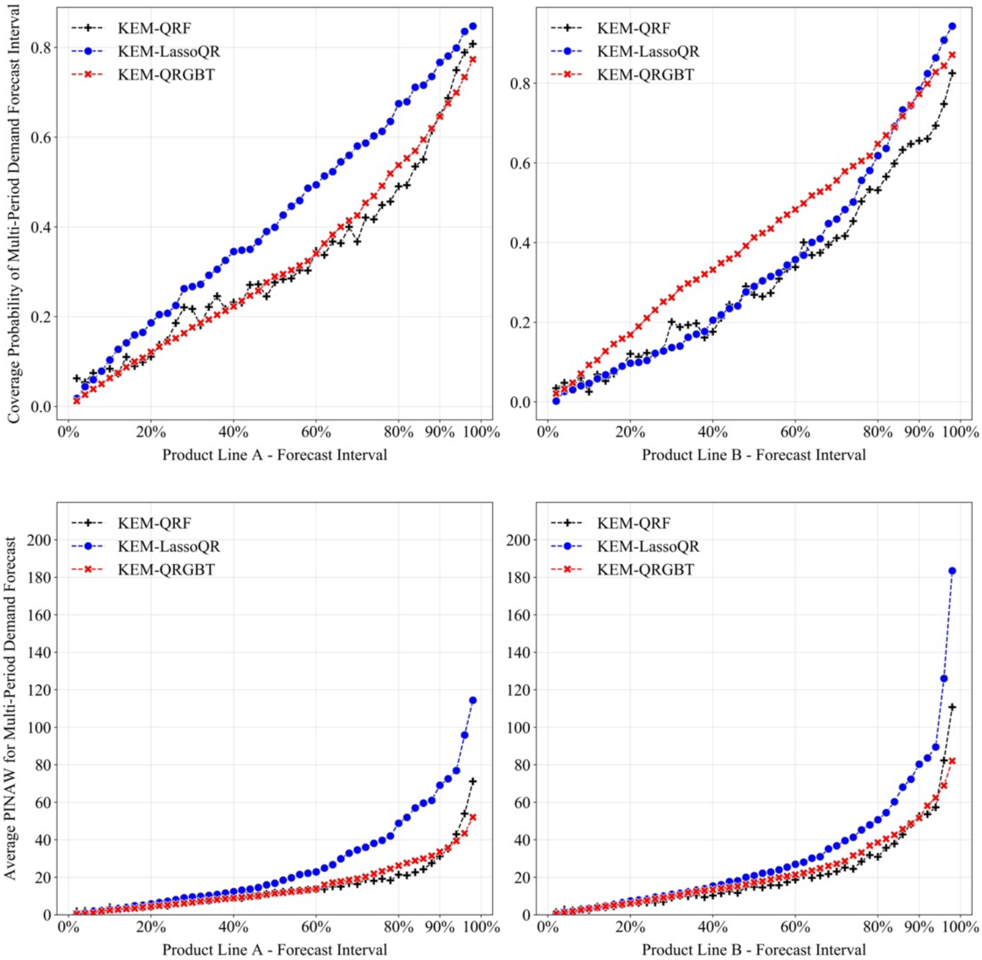

Fig. 3. Multi-cycle demand PICP and PINAW change with quantile interval

Table 4 shows the results of multi-period demand interval forecasting at a 90% forecasting interval (with lower and upper quantiles at 5% and 95%, respectively). From Table 4, it is evident that the KEM-QRF model is the worst in terms of the PICP metric, particularly worse than the KEM-QRGBT model by 16.45% in set B. The KEM-LassoQR model performs poorly on the PINAW metric, underperforming the KEM-QRGBT model by 105.47% and 55.50% in sets A and B, respectively. Consequently, the KEM-QRGBT model demonstrates the best overall performance, providing accurate demand information within the smallest feasible confidence interval.

Table 4. Multi-cycle demand forecasting results under 90% interval

Index	Sample Set	KEM-QRGBT	KEM-QRF	KEM-LassoQR
PICP↑	A	70.95%	70.35%	80.59%
	B	77.33%	64.62%	78.36%
PINAW↓	A	33.62	31.34	69.08
	B	51.69	50.86	80.38

5. Conclusions

An integrated demand forecasting model was proposed to analyze the full life cycle demand of new electronic products with limited historical data. The algorithm performance on demand forecasting of new electronic products was analyzed in the context of an actual case. The main conclusions are as follows:

(1) The KEM-QRGBT model can effectively capture the demand uncertainty and adaptively select optimal parameters, to more accurately predict the life cycle of new electric products with limited historical data. In particular, the GSCV algorithm with grid search and K-fold cross-validation was introduced into the KEM-QRGBT model, which can learn inherent correlations between products and avoid the risk of model overfitting.

(2) The effectiveness of the proposed demand forecasting model was verified through a real case study. Compared with the KEM-QRF model and the KEM-LassoQR model, introducing the QRGBT algorithm was employed to effectively balance the parameter learning complexity with demand forecasting accuracy.

(3) From the manager's perspective, the KEM-QRGBT model can provide practical guidance regarding how to obtain a more accurate life curve and multi-cycle demand for new electric products.

This study introduces time series clustering, neural network classification, and ensemble tree model algorithms to construct the combined demand forecasting model for new electronic products. The performance of demand forecasting model relies on the precision of cluster-based extraction of life curves. The KEM-QRGBT model employs a multilayer perceptron algorithm to predict the life curve of new electronic products. A customized deep learning framework can be developed to extend the applicability of the KEM-QRGBT model.

Acknowledgements
The authors are grateful for the support provided by the National Social Science Foundation of China (Grant No. 20BGL112).

References

[1] C. H. Yang, X. L. Su, and P. Wu, "A data-driven distributionally newsvendor problem for edge-cloud collaboration in intelligent manufacturing systems," *Eng. Appl. Artif. Intel.*, vol. 126, pp. 106995, Nov. 2023.
[2] P. Yin, G. Dou, X. Lin, and L. Liu, "A hybrid method for forecasting new product sales based on fuzzy clustering and deep learning," *Kybernetes*, vol. 49, no. 12, pp. 3099-3118, Nov. 2020.
[3] K. B. Kahn, "Solving the problems of new product forecasting," *Bus. Horizons*, vol. 57, no. 5, pp. 2787-2793, Oct. 2011.
[4] X. Li, Y. Yin, D. V. Manrique, and T. Bäck, "Lifecycle forecast for consumer technology products with limited sales data," *Int. J. Prod. Econ.*, vol. 239, pp. 108206, Sep. 2021.
[5] P. Goodwin, K. Dyussekeneva, and S. Meeran, "The use of analogies in forecasting the annual sales of new electronics products," *IMA J. Manag. Math.*, vol. 24, no. 4, pp. 407-422, Oct. 2013.
[6] P. C. Chang and C. Y. Lai, "A hybrid system combining self-organizing maps with case-based reasoning in wholesaler's new-release book forecasting," *Expert Syst. Appl.*, vol. 29, no. 1, pp. 183-192, Jul. 2005.
[7] C. H. Liu and Y. W. Wang, "Establish a cluster based evolutionary adaptive Weighted Fuzzy CBR for PCB sales forecasting," in *7th Int. Conf. Comput. Convergence Technol.*, Seoul, Korea (South), Dec. 2012, pp. 1417-1422.
[8] C. L. K. Yamamura, J. C. C. Santana, B. S. Masiero, J. A. Quintanilha, and F. T. Berssaneti, "Forecasting New Product Demand Using Domain Knowledge and Machine Learning," *Res. Technol. Manage.*, vol. 65, no. 4, pp. 27-36, Jun. 2022.
[9] R. J. Thomas, "Estimating market growth for new products: An analogical diffusion model approach," *J. Prod. Innovat. Manag.*, vol. 2, no. 1, pp. 45-55, Mar. 1985.
[10] H. Lee, S. G. Kim, H. W. Park, and P. Kang, "Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach," *Technol. Forecast. Soc.*, vol. 86, pp. 49-64, Jul. 2014.
[11] X. Zhou, J. Meng, G. Wang, and X. Qin, "A demand forecasting model based on the improved Bass model for fast fashion clothing," *Int. J. Cloth. Sci. Tech.*, vol. 33, no. 1, pp. 106-121, Feb. 2021.
[12] S. Thomassey and A. Fiordaliso, "A hybrid sales forecasting system based on clustering and decision trees. Decision Support Systems," *Decis. Support Syst.*, vol. 42, no. 1, pp. 408-421, Oct. 2006.
[13] S. Thomassey and M. Happiette, "A neural clustering and classification system for sales forecasting of new apparel items," *Appl. Soft Comput.*, vol. 7, no. 4, pp. 1177-1187, Aug. 2007.
[14] C. J. Lu and L. J. Kao, "A clustering-based sales forecasting scheme by using extreme learning machine and ensembling linkage methods with applications to computer server," *Eng. Appl. Artif. Intel.*, vol. 55, pp. 231-238, Oct. 2016.
[15] A. F. Tehrani and D. Ahrens, "Enhanced Predictive Models for Purchasing in the Fashion Field by Applying Regression Trees Equipped with Ordinal Logistic Regression," *Artif. Intell. Fashion Industry Big Data Era*, vol. 32, pp. 131-138, Sep. 2016.
[16] K. Hu, J. Acimovic, F. Erize, D. J. Thomas, and J. A. van Mieghem, "Forecasting New Product Life Cycle Curves: Practical Approach and Empirical Analysis," *M&som-Manuf. Serv. Op.*, vol. 21, no. 1, pp. 66-85, May 2018.
[17] R. M. van Steenbergen and M. R. Mes, "Forecasting demand profiles of new products," *Decis. Support Syst.*, vol. 139, pp. 113401, Dec. 2020.
[18] D. Lei *et al.*, "New product life cycle curve modeling and forecasting with product attributes and promotion: A Bayesian functional approach," *Prod. Oper. Manag.*, vol. 32, no. 2, pp. 655-673, Feb. 2023.
[19] Y. K. Elalem, S. Maier, and R. W. Seifert, "A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks," *Int. J. Forecasting*, vol. 39, no. 4, pp. 1874-1894, Oct. 2023.
[20] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *Proc. 2015 ACM SIGMOD Int. Conf. Manage. Data*, Victoria, Australia, May 2015, pp. 1855-1870.