Prediction of customer demands for production planning – Automated selection and configuration of suitable prediction methods

Bernd Scholz-Reiter (1), a, Mirko Kück a, Dennis Lappe b

a University of Bremen, Bremen, Germany
b BIBA – Bremer Institut für Produktion und Logistik GmbH, University of Bremen, Bremen, Germany

ARTICLE INFO

Keywords:
Production planning
Pattern recognition
Predictive models

ABSTRACT

Demand planning is of significant importance for manufacturing companies since subsequent steps of production planning base on demand forecasts. Major tasks of demand planning are the selection of a prediction method and the configuration of its parameters subject to a given demand evolution. This paper introduces a novel method for the automated selection and configuration of suitable prediction methods for time series of customer demands. The research investigates correlations between dynamic time series characteristics and forecasting accuracies of different prediction methods. The evaluation of the method on a database comprising real industry data confirms excellent prediction results.

© 2014 CIRP.

1. Introduction

Production planning is an important task for manufacturing companies. In particular, demand planning, which is premised on forecasts of future customer demands, is the main basis for all following planning steps [1]. Since no single prediction method outperforms all other methods for all cases of time series [2], two major tasks have to be fulfilled. At first, a suitable prediction method must be selected. Subsequently, the parameters of this method have to be configured to match a current demand evolution. Commonly, manufacturing companies select a few prediction methods, configure their parameters and compare the training results to select the best of the considered methods. In an extreme case, only one prediction method is selected and configured with respect to a current demand evolution. The advantage of this approach is the quick establishment of forecasting results. Nevertheless, the probability of poor results in comparison to other prediction methods is high. The other extreme is to configure various prediction methods to the current demand evolution and to select the method with the lowest training error. This approach can lead to more accurate predictions. However, it requires either a high amount of expert knowledge or automated algorithms to find appropriate parameter configurations of the applied prediction methods. Moreover, it takes much time to configure the different prediction methods and to compare the prediction results. Besides, the increasing number of product variants in recent years further complicates demand planning [3].

The paper at hand presents a novel data-driven approach for the automated selection and configuration of suitable prediction methods. Thereby, sophisticated predictions are calculated quickly. By using machine learning methods, correlations between time series characteristics of demand evolutions and the accuracies of different prediction methods are analysed. Here, measures of recurrence quantification analysis are incorporated in addition to common time series characteristics. Furthermore, the study includes several established prediction methods which model time series evolutions locally or globally as well as linear or nonlinear. After setting up a knowledge base, a suitable prediction method for an unknown time series of customer demands is selected and configured automatically. The remainder of this paper is structured as follows. Section 2 gives a theoretical background. The new automated selection and configuration approach is detailed in Section 3. Section 4 describes an evaluation on real time series data of the M3-competition [4]. Conclusions and an outlook on further research directions are given in Section 5.

2. Theoretical background

2.1. Demand planning and time series

Manufacturing companies must solve demand planning problems for every product they sell. In particular cases, additional information can be incorporated to improve forecasts of future customer demands, but in general, univariate time series of past customer demands are the main basis for the predictions. A time series of customer demands for a specific product is an ordered sequence of points \( y = \{y_1, \ldots, y_T\} \) measured in time steps of equal length. The time distance between consecutive points is called a planning period. Each point of the time series represents an amount of customer demands within a period.

Demand planning can be conducted for different prediction horizons. However, the most important and most frequently performed planning step is planning one period into the future, denoted as a one-step-ahead forecast. In this paper, one-step-ahead demand planning problems are considered. These can be defined as follows: For a given time series \( y = \{y_1, \ldots, y_T\} \) of past
customer demands for a specific product, predict the unknown future demand $y_{t+1}$. In order to obtain meaningful results, in this paper, one-step-ahead demand planning is performed for subsequent time steps. This means, after predicting $y_{t+1}$ based on $y_1, \ldots, y_t$, predict $y_{t+2}$ based on $y_1, \ldots, y_{t+1}$ and so on. In this second prediction step, $y_{t+1}$ is known and is included into the prediction of $y_{t+2}$. This procedure is iterated $h$ times. The prediction accuracy is computed as the average error of the predicted values $\hat{y}_{t+1}, \ldots, \hat{y}_{t+h}$. Compared to the true future values $y_{t+1}, \ldots, y_{t+h}$.

In general, many different demand planning problems have to be solved at each period. Thus, usually a high amount of time has to be invested to find individually suitable prediction methods for the demand evolutions of different products. Model selection is an approach to achieve sophisticated predictions in shorter time.

2.2. Selection of prediction methods

In order to obtain sophisticated forecasts of time series, a suitable prediction method has to be selected. For this purpose, different approaches exist which can be divided into expert systems and data-driven approaches [5]. Since expert systems [2] are generally static and inflexible, data-driven approaches in terms of machine learning methods were studied in the recent years [5–7]. These methods achieve an automated selection of a prediction method by linking time series characteristics (features) to the accuracies (labels) of different prediction methods. In this way, a higher flexibility is accomplished [5]. However, most of the proposed model selection methods base either on a small amount of considered features [6] or a small number of prediction methods as possible labels [7]. Hence, Wang et al. developed a data-driven approach which incorporates 13 time series characteristics as features for the selection of one of four prediction methods by a decision tree [5]. This approach showed great potential for a sophisticated model selection. In the paper at hand, a classification approach by linear discriminant analysis is proposed which considers measures of recurrence quantification analysis in addition to the features proposed by Wang et al. [5]. Moreover, prediction methods of nonlinear dynamics are incorporated beside the common prediction methods. Recurrence quantification methods and nonlinear dynamics base on the theory of dynamical systems which is outlined in the next subsection.

2.3. Dynamical systems and phase space reconstruction

In general, customer demands for a specific product depend on various influences, like the amount of substitute products, the success of competitors, changing social or ecological conditions, or general agreements [8]. Here, methods of nonlinear dynamics have shown potential to model complex dependencies in production networks [9,10]. These methods base on dynamical systems which are also applicable to model a customer demand evolution together with its influencing forces. In order to reconstruct the dynamical properties of the system, the method of delay coordinate embedding is applied [11]. This method maps the time series into a so-called phase space. A vector

$$\mathbf{v}_Y^{m, \tau} = [y_{t-(m-1)\tau}, y_{t-(m-2)\tau}, \ldots, y_{t-\tau}, y_t]$$

(1)

is called delay coordinate vector of length $m$ corresponding to time point $t$. The delay time $\tau$ is a multiple of the sampling time of the time series $y$. The length $m$ is called the embedding dimension. By building all delay coordinate vectors $\mathbf{v}_Y^{m, \tau}$ of length $m$ with successive time distance $\tau$, the dynamical properties of the whole dynamical system can be reconstructed if the parameters $m$ and $\tau$ are chosen appropriately. Rules of thumb values for the parameters can be obtained by the $\mathbf{fn}$-algorithm and the first minimum of the average mutual information [11]. In order to build a prediction method based on the dynamical system reconstruction, the parameters have to be chosen more sophisticatedly, like described in Section 3.2.

3. New approach for model selection and configuration

This section describes the new approach for the automated selection and configuration of suitable prediction methods. The process for the creation of the proposed prediction framework is illustrated in Fig. 1. Characteristic measures are computed for a given training set of time series. After performing a principal component analysis (PCA) on these measures, the principal components are stored in a knowledge base. Besides, the time series are predicted by different prediction methods. The rankings of the different prediction methods for each time series are also stored in the knowledge base. Thus, the knowledge base comprises the principal components of the time series characteristics (features) as well as the most appropriate prediction methods for each time series (labels) of the training set. By performing a linear discriminant analysis (LDA), a classifier is created which relates the features to the labels. The proposed automated prediction method consists of this classifier as well as a parameter configurator. If an unknown test time series has to be predicted, firstly, the classifier recommends a prediction method. Secondly, the parameter configurator optimizes the parameters according to the time series evolution. After these two steps, the resulting model can be used to predict the time series into the future. The next subsections describe the three main components of the proposed method, the features, the labels and the classifier.

![Fig. 1. Creation of the proposed automated prediction method.](image-url)
Fig. 2. Considered time series measures.

Marwan et al. [12] propose time series complexity measures of RQA. These measures base on the theory of phase space reconstruction which was outlined in Section 2.3. In general, for performing an RQA, it is recommended to compute the embedding parameters $m$ and $\tau$ by the fnn-algorithm and the first minimum of the average mutual information [12]. However, for specific time series, some measures of RQA may be in calculable, due to different reasons. For instance, RQA was initially introduced to characterise time series of a high length. Though, in this paper, also short time series are considered. Hence, to obtain comparable characteristics, the set of measurements consists of seven measures computed for $m = 1$ and $\tau = 1$. Additionally, three characteristics are measured after calculation of the embedding parameters by the fnn-algorithm and the first minimum of the average mutual information. The set of 12 considered characteristics of RQA is completed by the computed values for $m$ and $\tau$.

Altogether, 26 characteristics are computed to describe a considered time series. However, some of the measured components may correlate. By performing a PCA, these components are combined to form new linearly uncorrelated components, the so-called principal components. For this purpose, an orthogonal transformation is applied, so that the principal components are arranged according to the possible variances in descending order. In this way, the most important information is identified. The steps of PCA are as follows. Preliminary, all characteristics of the measure matrix $C$ are normalised to $[0,1]$, to indicate the degree of presence of the specific feature. Subsequently, the needed 26-by-2 transformation matrix $L_2$ is computed. This matrix consists of the eigenvectors corresponding to the $z$ largest eigenvalues of the covariance matrix. These are found by applying the singular value decomposition [13]. Finally, the original $n$-by-26 measure matrix $C$ is transformed into a new $n$-by-$z$ measure matrix of principal components $C_{PCA}$ by $C_{PCA} = CL_2$. Here, the number $z$ of incorporated principal components can be 26 or smaller, where a smaller value induces a dimensionality reduction to the $z$ components with the greatest importance.

3.2. Prediction methods (classification labels)

In this paper, six established prediction methods are applied and compared regarding forecasting accuracy. These are random walk (RW), auto-regressive integrated moving average (ARIMA), feed-forward artificial neural networks (ANN) and exponential smoothing (ETS) as well as locally constant (LC) and locally linear (LL) methods depending on phase space reconstruction.

RW is a global constant prediction method. Generally, it is used as benchmark and it is often a strong competitor [5]. A one-step-ahead prediction is the last value of the time series $\hat{y}_{t+1} = y_t$. For the RW method, no parameters need to be optimised.

A global linear prediction method is the ARIMA method. Its forecasts base on an ARIMA $(p, d, q)$/$P, D, Q$, model which is determined by the order of auto-regression $p$, the degree of first differencing $d$ and the order of moving average $q$, their seasonal equivalents $P, D,$ and $Q$ as well as the number of periods per season $s$. The parameters, $p$, $d$, and $Q$ are selected based on KPSS and OCSB tests [14]. The remaining parameters are optimised according to Akaike’s information criterion (AIC) [15].

Feed-forward artificial neural networks (ANN) can be used as global nonlinear prediction methods. In this paper, networks with one hidden layer are applied [15] which build an ANN$(p, P, k)$, model. The network bases on $p$ lagged inputs and $k$ nodes in the hidden layer, $P$ is the seasonal equivalent to $p$ and $s$ represents the number of periods per season. AIC is applied to choose values for the lagged inputs. The nodes in the hidden layer are computed as $k = (p + P + 1)/2$ rounded to the next integer [15].

ETS methods comprise 30 different prediction methods which base on either global linear or global nonlinear models. In order to define an ETS model, three components have to be specified, error (E), trend (T) and seasonality (S). All ETS models are formulated as state space models. In order to select one of these models and to optimise the parameters, the maximum likelihood of the innovations state space models and the AIC are used [14].

In addition to the global prediction methods, two methods are proposed which approximate time series locally. They base on delay coordinate vectors of length $m$ with successive time distance $\tau$ which were described in Section 2.3. For the purpose of prediction, at first, all delay coordinate vectors are sorted by their Euclidean distances to the vector $\mathbf{V}_{y_{T}}^m$ corresponding to the last point $T$ of the time series $y$. Subsequently, the $k$ nearest neighbours $w_{0:n,1}, \ldots, w_{0:n,k}$ of $\mathbf{V}_{y_{T}}^m$ are determined and extrapolated one step into the future, to obtain $P_1(w_{0:n,1}), \ldots, P_k(w_{0:n,k})$. The $q$th nearest neighbour is defined to have the $q$th shortest distance to $\mathbf{V}_{y_{T}}^m$. Now, a prediction of the LC method is defined as $\hat{y}_{t+1} = \frac{1}{k} \sum_{i=1}^{k} P_i(w_{0:n,i})$. The LM method performs a principal component regression on the points $(w_{0:n,1}, P_1(w_{0:n,1}))$ for $i = 1, \ldots, k$ to approximate the time series locally linear. In order to optimise the parameters $m$, $\tau$ and $k$ for the LC and the LL prediction methods, a genetic algorithm is applied [16].

For the evaluation of the prediction methods regarding forecasting accuracy, the widely used $4$ symmetric mean absolute percentage error ($s$MAPE) is computed as $\frac{100}{h} \sum_{i=1}^{h} \frac{|y_i - \hat{y}_i|}{y_i}$

Rankings of the prediction methods are determined by comparing the $s$MAPE values of the different methods.

3.3. Linear discriminant analysis (classifier)

To achieve recommendations for the selection of a prediction method based on time series characteristics, LDA is applied to relate the characteristics to the prediction accuracies. This is a multivariate statistical method to classify different objects into one of several classes [13]. In this paper, the principal components of the characteristics serve as input features. The best prediction methods constitute the classification labels and thus the output values of the LDA. The LDA obtains classification rules based on prior and posterior probabilities of class affiliations and the assumption that the feature vectors have a Gaussian mixture distribution with different means and equal covariance matrices.

4. Empirical study

4.1. Experimental setup

In order to evaluate the proposed automated prediction method (AP), an empirical study was conducted based on the 3003 time series of the M3 competition [4] which were recorded in yearly, quarterly or monthly periods from different fields of application, namely industry, micro, macro, finance, demographic and other areas. The series were between 14 and 126 time points of length.
Moreover, 8, 12 and 18 future time points were known for the yearly, quarterly and monthly time series which were used for the computation of the smAPE as accuracy measure.

In this paper, the steps of Fig. 1 were accomplished, in order to train the AP method on a training set of time series. Subsequently, a test set of unknown time series was predicted by different methods. The test set of this study consisted of 200 time series which were randomly chosen from the industry data of the M3 competition. The remaining 2803 time series constituted the training set. On the test set, different prediction methods were compared. Firstly, each of the six prediction methods described in Section 3.2 was applied individually to forecast the 200 time series. After that, the AP method was applied. As a benchmark, the method proposed by Wang et al. [5] using decision trees (DT) was applied. Besides, the much more time-consuming method T6 to train all six individual prediction methods for each time series and to select the one with the lowest training error was adapted. In addition, the methods T5–T2 were applied. These methods also selected the best method after training but only considered the best individual methods. In this case, best meant the individual methods which obtained the highest numbers of best rankings like displayed in Table 1. For example, the method T5 considered the five best methods LL, ARIMA, ETS, RW and ANN. The method T2 only trained the two best methods LL and ARIMA and selected the one with lower training error. All the methods were compared to each other and to the optimal predictions (Opt).

4.2. Results

The results of the empirical study are summarised in Tables 1 and 2. The best predictions were obtained by the proposed AP method. Its predictions resulted in the smallest average smAPE of 8.8% and the smallest standard deviation of the smAPEs. In the set of the individual prediction methods, ARIMA, LL and ETS computed better predictions than the other three methods. This was reflected in a higher number of best rankings as well as lower means and standard deviations of the smAPEs. The values of the methods T6–T2 were calculated by training the best 6–2 methods according to Table 1 and selecting the one with the lowest training error. Here, the T2 method, which considered the methods LL and ARIMA, showed the best results. The benchmark method DT performed worse than the proposed AP method as well as the methods T2, T3, T4 and ARIMA but better than the other methods. Fig. 3 illustrates the prediction results for each point considered in the AP prediction results of one time series. It can be seen that applying one of the worse methods led to a higher amount of poor predictions than applying one of the better methods.

5. Conclusion and outlook

In this paper, a new method for the automated selection and configuration of an appropriate prediction method for time series of customer demands was proposed. This method incorporates correlations between 26 time series characteristics and forecasting accuracies of six different prediction methods. An evaluation on a real set of 200 time series delivered excellent prediction results. The proposed method performed better than the six individual methods and a benchmark method using decision trees, less features and less prediction methods. The proposed method also outperformed the more time-consuming method to train all individual methods and to select the one with lowest training error. By using the proposed method, accurate predictions were calculated with low effort of computation time. In further research, other classifiers will be applied and compared to the LDA classifier used in this paper. More individual prediction methods will be included to further improve the predictions.

Acknowledgement

This research has been funded by the German Research Foundation (DFG) under the reference number SCHÖ 540/21–2.

References