AI-Based Demand Forecasting for Both Reliable Forecasting and Efficient Operation: Dynamic Ensemble Forecasting

● Isamu Watanabe ● Tooru Yoshibayashi ● Tateki Imaoka

In the distribution industry, supply chains are formed by manufacturers, wholesalers, and retailers. Their business activities, such as product development, production management, logistics, and sales, are based on demand forecasting. Therefore, business reforms and improvements by capturing demand more accurately have become critical for companies. In recent years, however, consumer demand and products have diversified, and product life cycles have become shorter. It is thus increasingly difficult to accurately forecast future demand. Moreover, the aging population and declining birthrate make it difficult to secure sufficient labor. These situations raise expectations for advanced technology, leveraging AI, to automate and enhance demand forecasting. Fujitsu Laboratories has developed the dynamic ensemble forecasting technology, an AI-based demand forecasting. Using a model integration method, this technology flexibly handles various product characteristics and is reliable in executing highly accurate forecasting. It also lessens the operational burden by means of automatic tuning. We have also developed an attribute decomposition model for forecasting demand of brand-new products with no past results data. This paper describes the dynamic ensemble forecasting technology and the attribute decomposition model, a solution based on these technologies, and its application.

1. Introduction

In the distribution industry, where manufacturers, wholesalers, and retailers make up the supply chain, the forecasting of demand is a critical task that is a major determinant of the business success of enterprises. The essence of planning, such as production planning by manufacturers, order acceptance/placement and shipment planning by wholesalers, and order placement and sales planning by retailers, is how to capture demand accurately. To this end, companies use all kinds of information to forecast demand based on experience and know-how.

Unlike in the past, when anything that was produced would sell, it is becoming increasingly difficult to accurately forecast demand because markets are now saturated and consumers can pick and choose from so many choices. Then, there are also the issues of the declining birthrate and aging population, which make it increasingly difficult for companies to secure the workforce required to carry out duties according to demand forecasts. It becomes increasing difficult to keep lines of work going that depend on the experience and intuition of experts as they retire, and ways must be found to address the predicted diversification of demand.

As AI technologies evolve, their use is increasingly being looked to as a solution for demand forecasting. Since 2011, Fujitsu has been engaged in refining demand forecasting through the use of AI. One product of such endeavors is the “dynamic ensemble forecasting technology,” Fujitsu Laboratories’ original AI demand forecasting technology that integrates the know-how in demand forecasting that we have cultivated through demonstration experiments with customers in the fields of industry and distribution.

This technology realizes both stable forecasting accuracy for diversified demand and reduction of operational burden by means of automatic tuning to maintain accuracy during operation as a business system. Yet demand forecasting for new products cannot be solved by the dynamic ensemble forecasting technology alone, because there is no past sales and shipment data that can be used as learning data. This is where the "attribute decomposition model," which solves this problem by utilizing past results data of similar products, comes in.

This paper describes the dynamic ensemble forecasting technology and the attribute decomposition model, the results of demonstration experiments, and our work on demand forecasting solutions based on these technologies.

2. Current state of demand forecasting in distribution industry

The state of work on demand forecasting varies from company to company. In many cases, as the most basic forecast method, moving averages based on past results such as the same day last week or the same month of the previous year, or simple forecasting values obtained by linear regression are used. Although such methods are effective when demand is stable owing to relatively large base demand, they cannot forecast sudden fluctuations in demand. Therefore, while useful for regular products, these methods do not work well for daily foods with a short expiration date and for which demand fluctuates greatly on a daily basis. They also do not meet forecasting of intermittent demand and demand for new products.

Currently, to deal with these types of unstable demand, the production department, the sales department, and the supply and demand adjustment department, among others, formulate planned values. Further, the level of ICT utilization differs among companies. While some companies calculate forecast values manually using spreadsheet software, others construct statistical methods and demand forecasting systems by machine learning, and use these for planning work. As the variety of products and the diversification of demand increase, there are limits to what human labor can accomplish, and work automation and efficiency improvement through ICT are becoming pressing issues at not a few companies.

Various methods of demand forecasting have long been in use. These include forecasting based on past periodicity and trend records, forecasting based on an autoregressive model, and forecasting based on the capture of correlated data and monitoring of changes in external factors. Given this diversity, such methods must be used selectively according to the characteristics of the demand to be forecasted. **Table 1** lists representative forecast methods and their characteristics.

Continuing to individually pick forecast methods from the many that are available for the huge array of

Table 1
Representative forecast methods and characteristics.

Forecast method	Description	Forecast accuracy			Predictor	
		Magnitude of number to be forecasted		When volume of learning data is small	Possibility of incorporation into the model	Ease of understanding of influence of factors and its extent
		Large	Small			
Random walk	The simplest method, which uses the latest actual value as the prediction value	Low	High	High	No	–
Multiple regression	Prediction from correlation between factors (holidays, weather, etc.) and sales figures	Medium	Low	Low	Yes	Easy
Poisson regression	A model in which the error distribution of multiple regression follows the Poisson distribution. This method allows accurate prediction for products with low actual value volume	Medium	High	Low	Yes	Easy
ARIMA	Prediction from correlation with past actual values	Medium	Low	Low	No	Difficult
ARIMAX	A combination of the ARIMA model and multiple regression	High	Low	Low	Yes	Easy
Dynamic linear model	Each time data is added, the internal latent state is updated and prediction is executed accordingly. This model can handle dynamic change.	High	Medium	High	Yes	Easy
Exponential smoothing state space	Prediction focusing on trends and periodicity	Medium	Low	High	No	–
Neural network	Prediction by mathematical model imitating the human brain	High	High	Medium	Yes	Difficult

ARIMA: Autoregressive integrated moving average
ARIMAX: Autoregressive integrated moving average with explanatory variable

increasingly diversified products is simply not realistic. In this situation, an effective approach to achieving forecasting accuracy of a certain level is the automatic selection of an optimum forecast method for each product. This approach is already being put to practical use in some areas.

However, sales trends of products are not necessarily fixed. From the start of sales of a given products, until it becomes well established, and finally it is terminated, sales trends will change according to various factors. When a product has become well established and enjoys stable sales, demand forecasting based on past periodicity and trend records is effective. However, for example, in the case of products that are taken up on a TV show or on social media and suddenly become popular, sales numbers will wildly change. In such situation, any one fixed forecast method will be unable to flexibly deal with these changes.

Another major challenge in demand forecasting is the forecasting of demand of new products. As new products do not have past sales and shipment records, sales of such products cannot be forecasted with conventional statistical methods. Thus demand forecasting for new products must be made with reference to demand trends of similar existing products. However, such approach has not been systemized and humans are left to rely on their limited forecasting abilities as product become increasingly diverse and life cycles grow shorter. In response, Fujitsu developed an attribute decomposition model, which is a proprietary technology for systematizing forecasting demand of new products.

3. Fujitsu Laboratories' original dynamic ensemble forecasting technology

Fujitsu Laboratories has developed the original dynamic ensemble forecasting technology, which improves the accuracy of demand forecasting and implements automatic tuning. This technology is realized by combining conventional time series analysis technology with machine learning, which is a type of AI.

3.1 Conventional technology and its problems

Various demand forecast methods have been proposed, as shown in Table 1. However, there is no universal method that can obtain high accuracy for any case. The optimum method varies depending on the characteristics of the target product and the characteristics of the forecast conditions (magnitude of number to be forecasted, volume of learning data, available predictors, and so on).

Figure 1 shows the relationship between the forecast horizon (how many days later the forecasting applies to) and the forecast error of demand forecasting by various forecast methods. From this graph, we can see that certain methods such as ARIMAX use more complex models than other methods such as multiple regression and exponential smoothing, and are better able to finely capture the features of time series data. These methods make for highly accurate forecasting, even for longer term forecasts that incorporate a greater amount of information (holidays, temperatures, and so on).

Meanwhile, correctly learning the large number of parameters that constitute a complex model requires a longer learning period (period of the data used for learning). On the other hand, in cases where learning data is insufficient and parameter learning remains incomplete for new products and seasonal goods shortly after their release, and intermittent demand products with few shipments, complex models may yield worse forecasting accuracy than simpler ones that use fewer parameters, such as multiple regression and exponential smoothing.

As described above, the fact that the optimum method differs depending on the forecasting target and the forecast conditions, and the fact that the optimum method changes dynamically according to the life cycle of the product, are major challenges to the continuous maintenance of the demand forecasting accuracy.

3.2 Developed technology

The challenge of selecting the optimum method according to the forecasting target and the forecast condition is not limited to demand forecasting and likewise applies to machine learning and statistical forecasting. "Model selection" is often used as a way to solve this challenge. Model selection can be done, for example, by determining the model selection criteria such as suitability to the data and forecasting accuracy, and selecting the model that has the highest accuracy with past data. In the case of the forecasting target and forecast conditions being fixed, by this technology, the optimum forecast method is selected automatically according to the characteristics of each product.[1)]

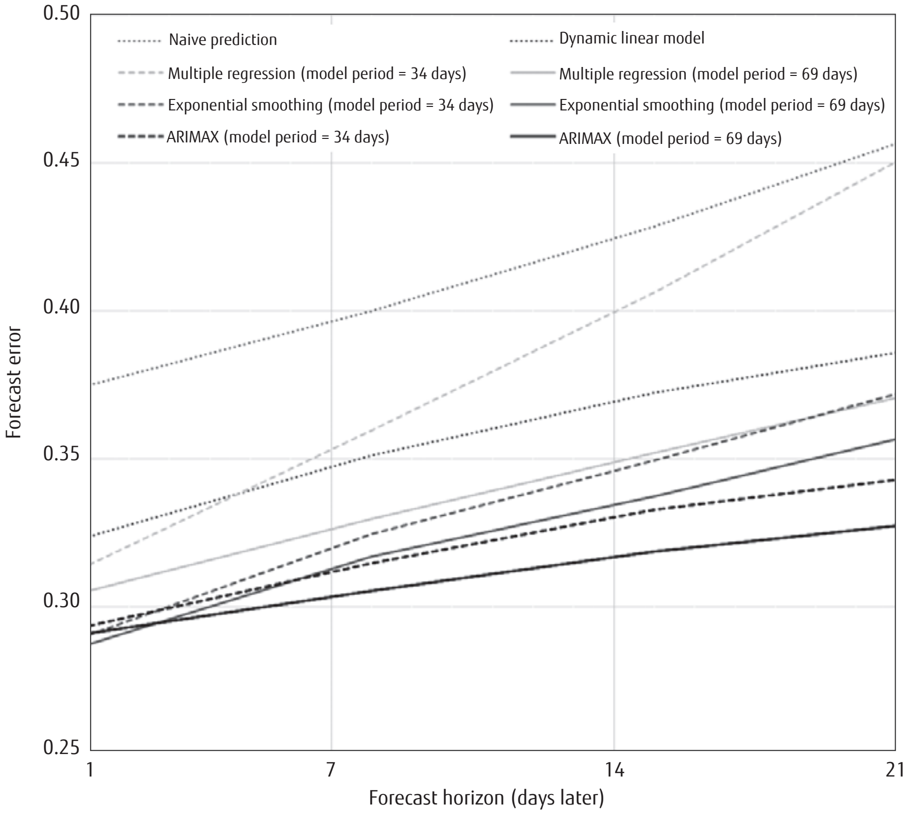

Figure 1
Relationship between forecast horizon and forecast error.

On the other hand, in cases where the characteristics of the forecasting target dynamically change according to the life cycle of the product, such as in the case of demand forecasting in the distribution industry, the optimum method also changes dynamically accordingly. In such cases, it is conceivable to sequentially update the model by selecting the model best suited for the latest data. However, model selection that consists in selecting just one optimum method runs into the problem that the forecast error grows very large when the optimum model is switched (when the model assumed to be optimum is found not to be so).

In the field of machine learning, a method called "ensemble learning" that realizes highly accurate forecasting by combining many learning models is widely used. In the field of time series analysis, a technique called model integration[note)] for calculating forecasted values by weighted averaging of forecasting results produced by multiple methods is known as a related technique of ensemble learning.

Generally, compared with model selection, which selects the best model from multiple forecasting models, model integration, which incorporates models with inferior accuracy for forecasting, tends to be misunderstood as being inferior in terms of accuracy. However, in a situation where the optimal model changes dynamically, model integration, which can reduce risk in case of deviation from the optimum, is known to stably yield forecasting of high accuracy over the long term.[2), 3)]

In model integration, the challenge is how to determine weights for weighted averaging of forecasting results. The easiest way is to take simple averages. This method is known to achieve higher accuracy than model selection. In the case of the dynamic ensemble forecasting technology, forecast error estimation is successively performed using machine learning and dynamically adjusting the weights according to the

note) This technique is also called model averaging or forecast combination.

magnitude of the forecast error, thereby realizing even higher accuracy.

In each method, a forecast error occurs in a certain pattern according to the characteristics of the target product and the forecast method. Thus, using past data, a regression model using the following five characteristic values as the explanatory variables, and forecast error as the objective variable, was built. This allows estimation of the forecast error and determination of the weights for weighted averaging accordingly. This approach is characterized by learning in two stages, namely learning by each method, and learning for integration.

1) Forecast conditions

Conditions determined by the purpose and method of forecasting, such as the forecast horizon, model period, and so on.

2) Static product characteristics

Characteristics such as product category and selling price, which are determined prior to product launch and do not change.

3) Dynamic product characteristics

Changeable characteristics obtained after product launch, such as sale interval, sales volume, and so on.

4) Life cycle characteristics

Characteristics that change according to the life cycle of the product, such as the number of days elapsed since product launch and the product phases (introduction period, growth period, maturity period, decline period, as estimated from sales numbers changes).

5) Most recent forecast error

The most recent average forecast error.

3.3 Assessment of effectiveness using real data

The experimental results and the effectiveness of the dynamic ensemble forecasting technology using real data are shown in **Figure 2**. The forecast error of model integration is smaller than that of each of the

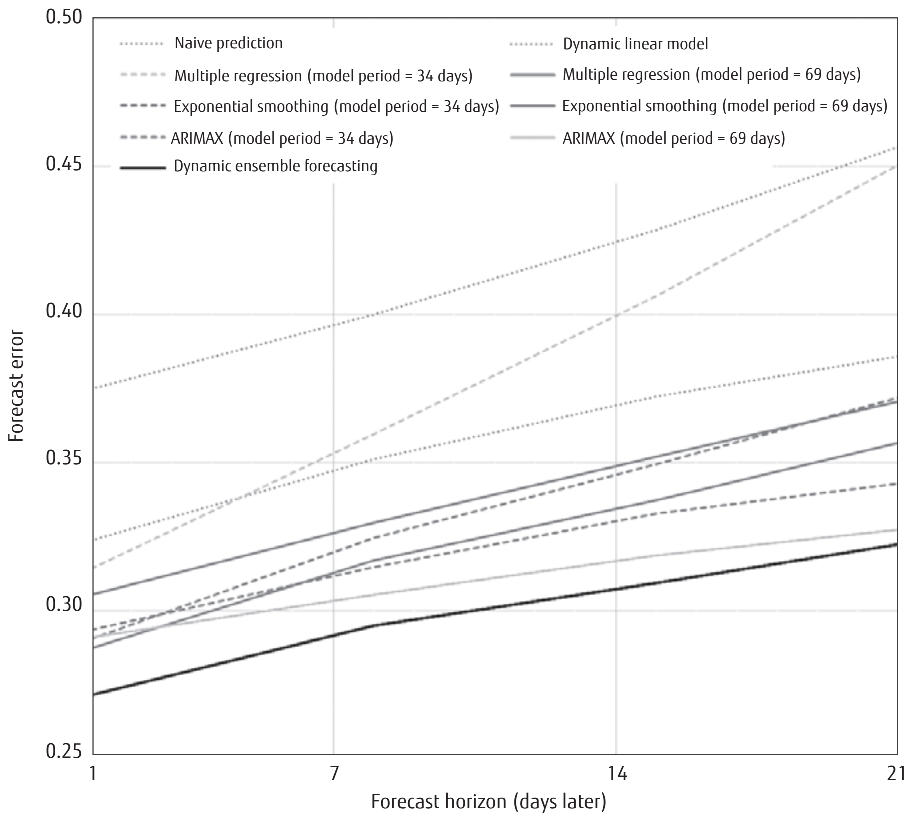

Figure 2
Effect of dynamic ensemble forecasting.

individual forecast methods used for integration, and forecast accuracy is higher.

The aforementioned forecast conditions and static product characteristics do not change with time. By incorporating such characteristics into the regression model, stable forecasting can be realized.

On the other hand, dynamic product characteristics and life cycle characteristics change according to the life cycle of each product. By incorporating such characteristics into the regression model, it is possible to achieve weighting according to changes in product characteristics according to the life cycle. Further, by incorporating the latest average error into the regression model, it is possible to reflect external factors such as trends that cannot be explained by forecast conditions and product characteristics in the weighting. With regard to characteristics other than the most recent forecast error, we can implement highly accurate weighting learning even when there are few data by constructing a unified model for all products. Moreover, even when there is no past data such as in the case of new products, forecasting can be made by utilizing information of products with similar characteristics.

In this way, highly accurate demand forecasting can be continuously performed without manual work through automatic tuning to account for dynamic characteristic changes of various products through the dynamic ensemble forecasting technology that estimates forecast error based on the use of various characteristics mentioned above.

4. New product demand forecasting technology using attribute decomposition model

Compared to about 3,000 items handled at convenience stores, the number of new products reaches about 5,000 a year.[4)] How to forecast demand for new products that are being introduced one after another is a major challenge not only for the distribution industry dealing with such products but also for the manufacturers who develop them.

Fujitsu Laboratories developed an attribute decomposition model that forecasts initial demand from planning information of new products by combining text mining and machine learning.

4.1 Conventional technology and its problems

Table 2 lists data and technologies that can be applied for each demand forecast utilization phase of new products. The dynamic ensemble forecasting technology can be applied shortly after product launch while actual sales and shipment information accumulates for the target product. Meanwhile, due to insufficient data availability in upstream phases, forecasting accuracy using only the past data available for those phases tends to be low.

Therefore, use of a forecast method that references past data of similar products is common. For example, a method of forecasting demand by referring to the pattern of sales of products deemed similar to the target product based on pre-launch orders and/or sales results for the first one or two post-launch weeks has been proposed.[4)]

Table 2
New product demand forecasting pattern.

<table>
<tr><th>Demand forecasting utilization phase</th><th>Major user departments</th><th>Available data</th><th>Applied technology</th></tr>
<tr><td>R&D</td><td rowspan="2">Marketing</td><td>• Planning information
• Product characteristics</td><td>Attribute decomposition model</td></tr>
<tr><td>Planning</td><td>• Planning information
• Product characteristics
• Sales plan</td><td rowspan="2">• Attribute decomposition model
• Regression model</td></tr>
<tr><td>Just before product launch (after sales start)</td><td rowspan="3">• Marketing
• Sales
• Supply and demand</td><td rowspan="3">• Planning information
• Product characteristics
• Sales plan
• Sales information
• Sales results
• Promotion
• Reputation (social media)</td></tr>
<tr><td>Shortly after product launch (1 to 2 weeks after)</td><td>• Attribute decomposition model
• Regression model
• Hazard model</td></tr>
<tr><td>Product launch to end of sales</td><td>• Attribute decomposition model
• Regression model
• Hazard model
• Dynamic ensemble forecasting</td></tr>
</table>

In the R&D and planning phases, which are the most upstream of the demand forecast utilization phases of new products, the information that can be used for demand forecasting is even more limited. Therefore, in most cases, the development side of operations must often rely on subjective assumptions and judgments. For this reason, incorrect estimation of initial demand resulting in inventory shortages or excesses is not uncommon.

4.2 Developed technology: attribute decomposition model

The attribute decomposition model developed by Fujitsu Laboratories applies text mining technology to proposal documents created during the R&D and planning phases, to extract and classify the attribute information such as target layer and product characteristics of each product. Next, using machine learning technology, the sales contribution degree (product characteristics) of each attribute information is quantitatively modeled from past sales data. In forecasting demand for new products, sales volume is estimated by combining the attribute information of the target product with its sales contribution degree.

The conventional method uses past information of similar products as is. On the other hand, in the attribute decomposition model, information of past products is decomposed into product attributes and sales contribution degree for use. Yet, although there may be products that are partially similar to a new product, in many cases no products that are similar overall exist. In this case, with the attribute decomposition model, highly accurate demand forecasting is possible by utilizing partially similar information.

4.3 Assessment of effectiveness using real data

Experiments were conducted to evaluate the accuracy of demand forecasting for new products using data from convenience stores and food manufacturers, yielding the following results.

- Forecast error of ±10% or less for 1/3 of all products
- Forecast error of ±20% or less for 1/2 of all products
- Forecast error of ±30% or less for 2/3 of all products
- Forecast error of ±50% or less for 90% of all products

In all cases, we did not use sales order information or initial sales information at all. Instead, only available planning information was used to forecast the sales volume of the first week of release. This approach yields forecasting patterns that can be implemented at the most upstream levels of the demand forecast utilization phase. Furthermore, combined with sales information and initial sales results, it allows more accurate forecasting.

5. Applications and expected effects

Fujitsu provides FUJITSU Business Application Operational Data Management & Analytics (ODMA) that supports business innovations in the field through the use of information that combines business data such as sales data and inventory data owned by companies and external data such as raw data such as daily work reports and social media data.[5)]

We began offering ODMA Demand Forecasting SaaS (software as a service), a demand forecasting solution utilizing the dynamic ensemble forecasting technology and attribute decomposition model technology described above as a cloud service from April 2018, in addition to the existing ODMA solutions lineup. ODMA Demand Forecasting SaaS is a cloud service that accumulates performance data of companies and external data that affect demand, calculates future demand forecast data, and provides it. This service can be introduced in the form of add-ons to the business systems of companies.

This solution provides a mechanism for the flexible creation and operation of forecasting models by implementing, in addition to the core demand forecasting engine, various functions as standard libraries. There libraries include processing of input and output data, clustering that groups objects based on their attributes, and text mining to extract attribute information from text information. Further, this solution was designed for scalability to accommodate the continuous growth of forecasting models. To realize this, it uses an architecture that can scale out with increases in the number of forecasting targets, a forecast method capable of integration by the dynamic ensemble forecasting technology, and a general-purpose framework that allows the addition of external data.

From among the various demonstration experiments conducted in collaboration with customers to measure the effectiveness of this solution, applications and expected effects at retailers and manufacturers are introduced below.

5.1 Optimizing ordering at mass merchandisers

Increasing efficiency of order placement work at stores is a challenge in the retail industry. At mass merchandisers with a large number of inexperienced employees, ordering accuracy varies greatly among individuals and order placement takes a long time. As a result, priority merchandise sales area design and product display tend to not receive the attention they need. This situation can be remedied by calculating the optimum order quantities from the demand forecast and inventory situation as recommended values. It can also optimize inventories to reduce sales opportunity losses due to inventory shortages as well as reduce dead stock. Furthermore, the time saved on ordering frees up workers to concentrate on creating attractive sales spaces.

The effect of adopting the recommended order values based on demand forecasting for daily distribution items and deli items was verified at a mass merchandiser by simulation of inventory quantity fluctuations. As a result, a decrease in average inventory quantities and number of inventory shortage occurrences was observed.

5.2 Loss reduction in production planning at food manufacturers

Improving the accuracy of demand forecasting when preparing production plans is a challenge for food manufacturers. In the case of regular products, the range of demand fluctuation is relatively small, so automation of production planning based on demand forecasting can stabilize planning and reduce planning workload. Meanwhile, in the case of new products and seasonal goods, accurate demand forecasting is difficult owing to the variety of products and needs, and production plans are prepared using forecasts based on past shipment records of similar products. However, actual results following product release will often vary significantly from forecasts, resulting in disposal losses or sales opportunity losses.

This can be remedied by systemizing forecasting using planning information of new products and past sales and shipment data of similar products to increase the accuracy of demand forecasting for new products and reduce losses through the optimization of production planning.

In a demonstration experiment with a food manufacturer, this solution was applied to demand forecasting for regular products and new products, and compared with conventional forecasting. For regular products, the average error was on the order of 5%, a level that holds promise for automation supplemented by some conventional manual forecasting correction. As for new products, the inventory reduction effect as measured by total shipping volume at the end of sale of such products, achieved through the adjustment of shipments based on pre-release initial shipment volume and weekly forecasts after shipping was confirmed.

6. Conclusion

This paper introduced the dynamic ensemble forecasting technology and attribute decomposition model technology developed by Fujitsu Laboratories to provide solutions to the challenge of demand forecasting in the distribution industry. It also described solutions based on these technologies and actual application cases.

Looking ahead, we will work on improving demand forecasting for a number of industries and business types, including the distribution industry, based on the demand forecasting technologies and solutions introduced in this paper. In addition to further refinement of the forecasting technologies introduced in this paper, the advancement of demand forecasting requires an approach that expands the data used as predictors to allow more accurate grasp and forecasting of demand. Going forward, we will pursue our work to realize useful forecasting through the accumulation of practical knowledge and the use of the aforementioned scalability of our solutions, as we seek to expand the application of demand forecasting to all the fields of work it could benefit.

References

1) T. Hastie et al.: The Elements of Statistical Learning. Springer New York, 2009.

2) T. Ando: High-Dimensional Data Analysis Methods. Asakura Publishing, 2014. (in Japanese).

3) G. Elliott et al.: Economic Forecasting. Princeton University Press. 2016.

4) H. Nakamura: Marketing of New Products. Chuokeizai-sha, 2001. (in Japanese).

5) Fujitsu: FUJITSU Business Application Operational Data Management & Analytics. (in Japanese). *http://www.fujitsu.com/jp/solutions/business-technology/intelligent-data-services/ba/product/operational-data-management-and-analytics/*

Isamu Watanabe
Fujitsu Laboratories Ltd.
Mr. Watanabe is currently engaged in applied research and practical study for the application of big data, IoT, and AI to business activities such as demand forecasting, marketing, and failure forecasting.

Tooru Yoshibayashi
Fujitsu Ltd.
Mr. Yoshibayashi is currently engaged in planning and development of new information solutions using big data and AI.

Tateki Imaoka
Fujitsu Ltd.
Mr. Imaoka is currently engaged in planning and development of new information solutions using big data and AI.