Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 79 (2019) 679–683

www.elsevier.com/locate/procedia

12th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 18-20 July 2018, Gulf of Naples, Italy

Demand forecasting in restaurants using machine learning and statistical analysis

Takashi Tanizaki[a,*],Tomohiro Hoshino[a],Takeshi Shimmura[b],Takeshi Takenaka[c]

[a]*Graduate Schoole of Kindai University, 1 Takaya-Umenobe, Higashi-Hiroshima 739-2116, Japan*
[b]*Ritsumeikan University, 1-1-1 Noji-Higashi, Kusatsu 525-8577, Japan*
[c]*Advanced Industrial Science and Technology, 2-3-26 Aomi Koto-Ku, Tokyo 135-0064, Japan*

* Corresponding author. Tel.: +81-82-434-7484; fax: +81-82-434-7890. *E-mail address:* tanizaki@hiro.kindai.ac.jp

Abstract

In this paper, demand forecasting in restaurants using machine learning is proposed. Many researches have been proposed on demand forecasting technology using POS data. However, in order to make demand forecasts at a real store, it is necessary to establish a store-specific demand forecasting model in consideration of various factors such as the store location, the weather, events, etc. Therefore, we constructed a demand forecasting model that functionally combines the above mentioned data using machine learning. In this paper, the demand forecasting model using machine learning and the verification result of the model using real store data is discussed.

Peer-review under responsibility of the scientific committee of the 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering.

Keywords: Demand forecasting, Machine learning, Statistical analysis, Service engineerring, Restaurant management

1. Introduction

The service industry is an important industry accounting for about 70% of Japan's GDP. However, since the labor productivity of the service industry is lower than that of the manufacturing industry, its improvement is an important policy issue of Japan. Especially in the labor-intensive service industry labor productivity is low because service goods are consumed at the same time as they are provided. Among the labor-intensive service industries, restaurant industry, which is integrated production and sales industry, has improved its inventory possibilities by separating service production functions from sales by introducing a central kitchen. On the other hand, in the face-to-face service industry like dinner restaurants, separating service production functions from sales is not applicable because the quality of service will be compromised. In order to solve such problems, we are researching how to advance store management by improving employees' work arrangement and food materials ordering based on accurate forecasting of the number of customers for face-to-face service industries. As part of the research, we are researching demand forecasting methods using internal data such as POS data and external data in the ubiquitous environment such as weather, events, etc. in order to improve the accuracy of demand forecasting. In this paper, we describe the approach to forecasting method based on machine learning and statistical analysis.

2. Forecasting method

In this research, the number of customers is forecasted using machine learning and statistical analysis method with internal data and external data in the ubiquitous environment. Bayesian Linear Regression, Boosted Decision Tree Regression, and Decision Forest Regression are used for machine learning, Stepwise method is used for statistical analysis method. We used Azure Machine Learning as a machine learning tool and SPSS as a statistical analysis tool.

2212-8271
Peer-review under responsibility of the scientific committee of the 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering.
10.1016/j.procir.2019.02.042

2.1. Bayesian Linear Regression

Bayesian Linear Regression (Bayesian) is a method of applying Bayesian network to machine learning. The Bayesian network is a probabilistic model in which conditional dependencies among multiple random variables are expressed using a graph structure and dependency relationships between random variables are expressed by conditional probabilities [1]. The Bayesian network is defined by three variables: random variable, conditional dependency condition between random variables, and conditional probability [2]. By using the Bayesian network, the probability distribution of unobserved variables is calculated using observed some variables and the value with the highest probability value is obtained as the predicted value of that variable.

2.2. Boosted Decision Tree Regression

Boosted Decision Tree Regression (Boosted) is a method of learning using Boosting [3]. Boosting is machine learning using multiple learning devices. In this method, the number of learning times of the incorrectly forecasted case is increased in order to improve learning accuracy by increasing the weight of that case.

2.3. Decision Forest Regression

Decision Forest Regression (Decision) is a method of learning using Random Forest [4]. Random Forest is a method of constructing a forest using multiple decision trees and integrating learning results for each decision tree. Extreme bias in learning of each decision tree can be prevented by incorporating randomness when extracting learning data to be used in each decision tree. As a result, excessive learning can be prevented and high generalization performance can be acquired.

2.4. Stepwise Method

Stepwise method (Stepwise) is a method of constructing a regression model by searching for a combination of objective variables that can most explain the explanatory variable by sequentially increasing or decreasing the objective variable [5]. When adding highly objective variables to regression formulas, there are variables that have already been added, which become useless due to their relevance to objective variables added later. Therefore, in the stepwise method, each time an objective variable is added, the variable which becomes insignificant for the explanatory variable is deleted from the regression formula.

3. Forecasting of customers visiting

3.1. Target data

We got visitor data for 5 stores from the restaurant chain R of the joint researches and forecasted the number of customers by the method described in Chapter 2. Based on the visit record of '14/5/1 to' 15/4/30, we forecasted the number of customers from '15 / 5/1 to '16/4/30 and compared it with the number of actual customers during the same period. Table 1 shows explanatory variables used for forecasting. The forecasting rate α, that is ratio of the number of forecasted customers to that of actual customers, is calculated using the equations (1) and (2).

Table 1. Explanatory variable

Category	Explanatory variable	Definition
Month	January	Jan/1-Jan/31
	February	Feb/1-Feb/28
	March	Mar/1-Mar/31
	April	Apr/1-Apr/30
	May	May/1-May/31
	June	Jun/1-Jun/30
	July	Jul/1-Jul/31
	August	Aug/1-Aug/31
	Septemner	Sep/1-Sep/30
	October	Oct/1-Oct/31
	Novenber	Nov/1-Nov/30
	December	Dec/1-Dec/31
The day of the week	Monday	Weekday and the next day is weekday
	Tuesday	Weekday and the next day is weekday
	Wednesday	Weekday and the next day is weekday
	Thuesday	Weekday and the next day is weekday
	Fryday	Weekday and the next day is weekday
	Saturday	Even if the target day is a holiday it is Saturday.
	Sunday	Sunday and the next day is weekday Even if the target day is a holiday it is Sunday.
	Sunday during holidays	Sunday and the next day is holiday Even if the target day is a holiday it is Sunday.
	Holiday	Holiday and the nextday is weekday
	Holiday during holidays	Holiday and the nextday is holiday
	Before holiday	Weekday and the next day is holiday
	Lastday during holidays	The last day of three or more consecutive holidays
Event	January 1st	January 1st
	January 2nd	January 2nd
	January 3rd	January 3rd
	Year-end	Dec/29-Dec/31
	End of year party	Weekday of December
	Christmas eve	December 24
	Coming-of-age day	Second Monday in January
	Setsubun	February 2nd
	Obon	Aug/13-Aug/15
	New year's party	Weekday till the coming-of-age day except Jan/1-Jan/3
	Farewell party	Weekday in March
	Welcome party	Weekday in April
Weather	Average wind speed	Average wind speed per day (m/s)
	Maximum wind speed	Maximum wind speed per day (m/s)
	Highest temperature	Highest temperature in a day (℃)
	Lowest temperature	Lowest temperature in a day (℃)
	Amount of precipitation	Amount of precipitation in a day (mm)
	Maximum precipitation	Maximum amount of precipitation in ten minutes (mm)
	Maximum instantaneous	Maximum instantaneous
	wind speed	wind speed in a day (m/s)

p_i : Actual number of customers on i-th day
e_i : Forecasting number of customers on i-th day
N : Forecast period
α_i : Forecasting raito on i-th day

$$\alpha_i = \frac{p_i - |p_i - \alpha_i|}{p_i} \tag{1}$$

$$\alpha = \frac{\sum_{i=1}^{N} \alpha_i}{N} \tag{2}$$

3.2. Forecasting result

Table 2 shows the forecasting results. In the machine learning, the data usage rate, that is ratio of utilized data for machine learning to total data, is changed from 40% to 100%. Since the Stepwise method is a method using all data, the data usage rate is 100%. The columns of yellow background color are the result with the highest forecasting rate among each method. The black box surrounding column shows the highest forecasting rate among the four methods. For store A, the forecasting rate using Bayesian is the highest. For stores B and C, the forecasting rate using Stepwise is the highest. For stores D and E, the forecasting rate using Decision method is the highest. The forecasting rate using Boosted tends to be low. In Bayesian, the higher the data usage rate, the higher the forecasting rate tends to be. In Boosted and Decision, there is no noticeable relationship between data usage rate and forecasting rate. The stores A and B are located in Japan's leading high-class shopping street adjacent to the business area and the store E is located adjacent to the central station in the suburban residential area.

Therefore, many of customers at stores A and B are business customers and tourists, and many of customers at store E are residents in the surrounding area. On the other hand, stores C and D are located in the downtown area adjacent to the central station with many passengers. Therefore, many of customers at stores C and D stores are unspecified customers.

Fig. 1 shows the daily forecasting results using Bayesian at store A with the highest forecasting rate (90%). The forecasting value and the actual value are in the same trend. The difference between the average forecasting value and the average actual value is 15.5, and the difference between the annual total of the forecasting value and that of the actual value is about 5% of the actual value. The difference between the forecasting value and the actual value is large at the Obon , the New Year's Holiday, the beginning of the fiscal year.

Fig. 2 shows a scatter diagram of "the number of customers and the number of reserved customers". Since R^2 value is 0.75, there is correlation between these two variables. Table 3 shows R^2-value of the number of customers and the number of reserved customers and the ratio of the number of reserved customers to the number of customers. Since the R^2 value is 0.6 to 0.7, there is a slight correlation between the number of customers and the number of reserved customers. The ratio of the number of reserved customers in stores excluding store A is about 35% to 40%.

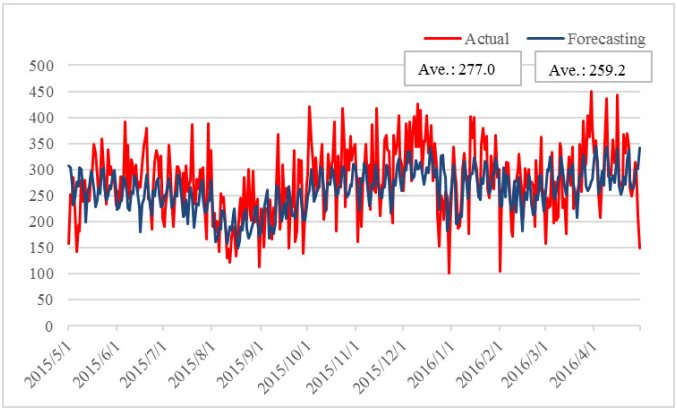

Fig. 1. Daily forecasting results

Table 2. Forecasting results

Store	Data usage rate	Bayesian	Boosted	Decision	Stepwise
A	40%	82.0%	77.4%	80.8%	81.5%
	50%	82.3%	76.7%	80.1%	
	60%	82.6%	75.8%	79.7%	
	70%	82.7%	76.1%	81.4%	
	80%	82.5%	78.2%	80.9%	
	90%	82.7%	76.1%	81.4%	
	100%	82.6%	77.2%	80.7%	
B	40%	80.7%	78.1%	79.2%	81.9%
	50%	78.8%	69.2%	76.4%	
	60%	79.6%	67.2%	78.3%	
	70%	80.0%	69.8%	79.5%	
	80%	80.3%	70.2%	77.8%	
	90%	80.4%	72.4%	79.8%	
	100%	81.0%	73.8%	80.2%	
C	40%	79.0%	70.9%	76.3%	80.2%
	50%	79.1%	68.6%	77.7%	
	60%	79.3%	76.9%	77.1%	
	70%	79.4%	76.6%	79.0%	
	80%	79.4%	77.0%	78.1%	
	90%	79.4%	76.2%	77.8%	
	100%	79.5%	76.7%	78.8%	
D	40%	75.8%	73.0%	78.4%	78.1%
	50%	76.6%	75.0%	75.2%	
	60%	77.7%	72.5%	77.8%	
	70%	77.8%	73.5%	77.4%	
	80%	77.9%	74.9%	75.9%	
	90%	77.8%	74.6%	77.6%	
	100%	78.3%	75.4%	76.7%	
E	40%	82.8%	81.9%	82.1%	83.3%
	50%	82.6%	80.8%	83.9%	
	60%	83.3%	81.1%	83.2%	
	70%	83.4%	80.4%	81.2%	
	80%	83.3%	79.1%	82.0%	
	90%	83.5%	79.7%	82.2%	
	100%	83.9%	80.1%	82.0%	

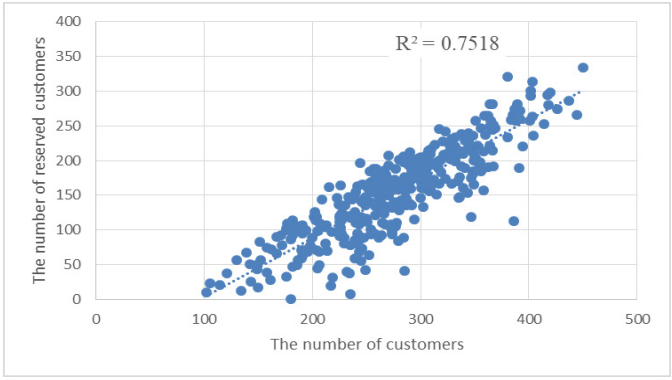

Fig. 2. Scatter diagram of the number of customers and reserved customers

Table 3. R^2-value and ratio of reserved customers

Store	R^2-value	Ratio
A	0.7518	55.5%
B	0.5931	39.3%
C	0.6194	34.1%
D	0.6809	34.6%
E	0.6389	37.2%

3.3. Improvement of forecasting method

From 3.2, it is found that there is slight correlation between "the number of customer" and "the number of reserved customers". The ratio of the number of reserved customers is 35% to 40%, which is not low. The motivation for making reservations at the restaurant is often based on business reasons such as "business trip" and "dinner meeting" and personal reasons such as "family anniversary". Although it is possible to forecast the approximate time when the above events occur, it is difficult to forecast exactly on a daily basis. On the other hand, it is necessary to forecast the number of customers visiting with high accuracy on a daily basis to automate employees' work arrangement and food materials ordering, which is the research goal, Therefore we decide to use the number of reserved customers that can be grasped from several days before in this restaurant chain for forecasting the number of customers. We forecast only the number of customers without reservation and calculate the forecasting value by equation (3).

Z_i : Total number of forecasting customers on i-th day

X_i : The number of forecasting customers without reservation on i-th day

Y_i : The number of reserved customers on i-th day

$$Z_i = X_i + Y_i \quad (3)$$

Table 3 shows the forecasting results. The column of yellow background color and is the result with the highest forecasting rate among each method. The black box surrounding column shows the highest forecasting rate among the four methods. The columns with a background color of pink are the highest forecasting rate in Table 2. The forecasting rate exceeded approximately 85% in all stores.

Fig. 3 shows the daily forecasting results using Bayesian at store A with the highest forecasting rate (100%). The graphs of the forecasting value and the actual value are almost overlapping. The difference between the average forecasting value and the average actual value is 0.2, and the difference between the annual total of the forecasting value and that of the actual value is about 0.1% of the actual value.

We got the evaluation that this method is practically applicable from the restaurant R of the joint researches. In the future, we plan to improve forecasting accuracy and research on the efficiency of store management such as automated food materials ordering and employees' work arrangement based on forecasting results.

Table 4. Forecasting results of improvement method

Store	Data usage rate	Bayesian	Boosted	Decision	Stepwise
A	40%	91.2%	89.9%	90.9%	91.8%
	50%	91.2%	89.5%	90.4%	
	60%	91.0%	89.6%	89.8%	
	70%	91.2%	89.3%	90.6%	
	80%	91.4%	89.3%	91.2%	
	90%	91.5%	90.2%	91.4%	
	100%	91.7%	89.2%	91.0%	
B	40%	87.2%	86.2%	87.2%	88.9%
	50%	87.0%	86.3%	87.2%	
	60%	87.1%	86.1%	86.8%	
	70%	87.3%	86.3%	86.9%	
	80%	87.3%	86.7%	86.9%	
	90%	87.4%	86.7%	86.8%	
	100%	87.6%	87.0%	86.5%	
C	40%	84.6%	83.1%	83.3%	86.0%
	50%	84.5%	83.7%	85.0%	
	60%	84.7%	84.6%	83.9%	
	70%	84.5%	83.8%	84.4%	
	80%	84.7%	83.8%	83.6%	
	90%	84.4%	84.2%	84.8%	
	100%	84.4%	82.9%	84.4%	
D	40%	83.8%	83.3%	84.8%	85.7%
	50%	84.5%	84.7%	84.2%	
	60%	85.1%	83.1%	85.5%	
	70%	85.0%	83.0%	84.9%	
	80%	85.1%	82.7%	84.9%	
	90%	85.5%	83.5%	85.1%	
	100%	85.8%	82.7%	84.2%	
E	40%	85.2%	86.3%	86.3%	84.6%
	50%	84.1%	86.2%	86.0%	
	60%	84.5%	86.1%	85.0%	
	70%	84.8%	86.2%	84.0%	
	80%	84.7%	87.2%	84.5%	
	90%	84.8%	86.8%	84.2%	
	100%	85.0%	87.3%	85.5%	

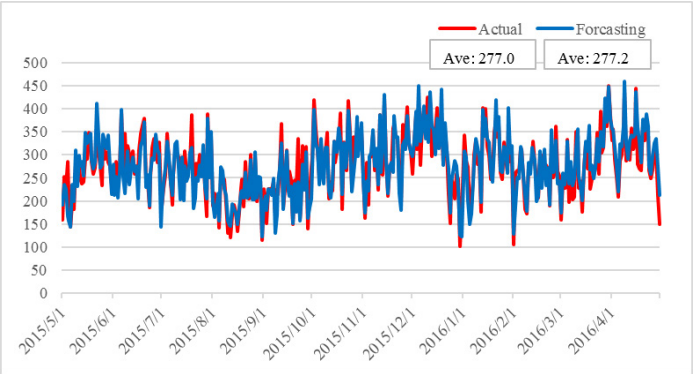

Fig.3. Daily forecasting results

4. Conclusions

In this paper, demand forecasting methods using internal data such as POS data and external data in the ubiquitous environment such as weather, events, etc. are proposed. We use

Bayesian Linear Regression, Boosted Decision Tree Regression, Decision Forest Regression and Stepwise method as the demand forecasting method. There was no big difference in the forecasting rate using the method of Bayesian, Decision, and Stepwise, and the forecasting rate of Boosted was a little low. The forecast rate of any store exceeded approximately 85%.

We got the evaluation that this method is practically applicable from the restaurant R. In the future, we plan to improve forecasting accuracy and research on the efficiency of store management such as automated food materials ordering and employees' work arrangement based on forecasting results.

Acknowledgements

This study is supported by JSPS KAKENHI (16H02909).

References

[1] Motomura Y. Baysian network. Technical Report of IEICE 2003; 103-285: 25-30 (In Japanese).
[2] Motomura Y. Baysian Network Softwares. Journal of the Japanese Society for Artificial Intelligence 2002; 17-5: 1-6 (In Japanese)
[3] Freund Y, Schapire R, Abe N (translation). A short Introduction to Boosting, Journal of the Japanese Society for Artificial Intelligence 1999; 14-5: 771-779 (In Japanese).
[4] Habe H. Random Forests, ISPJ SIG Technical Report 2002; 2012-CVIM-182-31: 1-8 (In Japanese).
[5] Bolch BW, Huang CJ, Nakamura K (translation). Applied Statistics analysis, Tokyo: Morikita Publishing Co., Ltd.; 1968 (In Japanese).