Development greenhouse environment prediction system using IoT data

Yasunori Takemura, Naoya Nishida, Moeko Tominga
Nishinippon Institute of Technology, 1-11 Aratsu, Kanda-town, Miyako-gun, Fukuoka, 800-0394, Japan

Eiji Mizoe, Tomoyasu Furukawa
Kumamoto Fruits and Vegetables Shipping Association Co. Ltd, 3523, Okimachi, Yatsushiro-city, Kumamoto , 866-0013, Japan
Email: takemura@nishitech.ac.jp

Abstract

Currently, the decrease in the working population is a significant problem in agriculture in Japan. In addition, the prices of imported vegetables and fruits have been rising due to international trade problems such as the weak yen. This study has developed an environmental data storage system using IoT to assist the labor force and to understand the cultivation data of crops that have not been cultivated in Japan. Using this big data, we have constructed and evaluated a forecasting system. In this study, we focus on the domestic cultivation of bananas grown in the tropics as the subject to obtain data and make predictions. For learning, a system was constructed to infer the next day's temperature by time-series analysis, using as input data obtained from the IoT. In this paper, we report a comparison of two time series analysis algorithms.

Keywords: Time-series data analysis, ARIMA model, prophet model, IoT, agriculture

1. Introduction

For In recent years, the Internet of Things (IoT) has become increasingly popular, and various " devices" are now connected to the Internet. The IoT is a system in which sensors are embedded in various " products," connected to the Internet, and a large amount of data acquired via the Internet is stored in the cloud. This accumulated large amount of data is called "Big Data (BD)," and today there are tens of billions of IoT devices in the world, all of which continue to generate new data. BD cannot create value simply by accumulating it, but it can create new value by analyzing the accumulated data. In addition, with the development of computer technology, the computation time of AI (Artificial Intelligence) has been shortened, and many companies have developed home appliances and automobiles that incorporate AI, making the IoT and AI a familiar presence. As evidence of this, the number of IoT devices is on an increasing trend with respect to transition data, and is expected to exceed 32.33 billion units by 2022 [1] . In Japan, the government and companies are actively working toward the realization of Society 5.0, and it is expected that a society that uses information in a wide variety of ways will expand in the future [2].

For example, the utilization of IoT databases is expected in various fields. One example is in the field of agriculture, where the number of farmers is decreasing year by year due to the problem of declining productivity caused by the aging of the farming population and the lack of farmers to take on the workforce. Therefore, as a solution, it is expected to improve productivity with a small number of people, and a necessary method to acquire environmental information on farms, which is a factor that changes the quality and quantity of agricultural products [3]. As a solution to this problem, visualization and prediction functions of environmental information using IoT are expected. One of the challenges in future development is to analyze the acquired IoT data and weather data and utilize them as time-series data in forecasting algorithms in order to improve productivity and realize high-quality and stable crop production [4].

The purpose of this research is to build an AI using time-series data with IoT technology, and to generate and validate prediction data using time-series data. The experiment involves time-series prediction of environmental data in a plastic greenhouse in Yatsushiro City, Kumamoto Prefecture, where tropical crops are harvested. The time-series forecasting algorithms were compared using the ARIMA model [5] and Prophet [6], because of their ability to consider the effects of periodicity and external factors.

2. Methodology

2.1. ARIMA (Auto Regressive Integrated Moving Average) model

The ARIMA model [5], derived by Box & Jenkins, is a model consisting of three components: AR (Auto Regressive) model, MA (Moving Average) model, and summed components (integrated). Time series data may have non-stationary time series processes when a trend analysis is performed, which cannot be handled by the ARMA model. Therefore, non-stationary time series must be converted to stationary time series and then fit into the ARMA model. The ARIMA model, which adds a summation component to the ARMA model, is then applied to the non-stationary time series and aims to make them stationary by taking the "d" th-order difference. The general equation is shown in Equation (1) below.

$$\boldsymbol{y_t - y_{t-d} = c + e_t + \sum_{i=1}^{p} \emptyset_i y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j}} \tag{1}$$

In the equation, t denotes time, d, i and j denote the time at the regression point in time. e indicates the error height, y is the output, and c, ϕ and θ are constants. In recent years, it has been used as a forecasting system for COVID-19 [7] , and has also been used in research such as stock price forecasting [8].

2.2. Prophet model

Prophet is a library for time series analysis developed by Meta in 2017. The advantages of this model are that it can be built without statistical knowledge, it can be trained even with missing values, and it is easy to interpret the prediction results. Prophet is based on the Generalized Additive Model (GAM) [9], which is a model that uses a It predicts future values by modeling and combining four terms: the trend term (growth: $g_{(t)}$), the seasonality term (seasonality: $s_{(t)}$), the holidays effect term (holidays: $h_{(t)}$), and the error term $\varepsilon_{(t)}$). The model equation is shown in Equation (2) below.

$$\boldsymbol{y = g_{(t)} + s_{(t)} + h_{(t)} + \varepsilon_{(t)}} \tag{2}$$

3. Evaluation Methodology

3.1. Correlation coefficient

The correlation coefficient expresses the strength of the relationship and correlation between two items as a positive or negative value, and can be objectively analyzed in the range of 1 to -1. 1 is closer to a positive correlation and -1 to a negative correlation. Equation (3) below shows the formula for obtaining the correlation coefficient. In this research, we aim to use data obtained from IoT data and weather data as explanatory variables in the algorithm in order to increase the accuracy of data prediction, and to select data by examining the correlation coefficient with the output values.

$$\boldsymbol{r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}} \tag{3}$$

3.2. Akaike's Information Criterion (AIC) [10]

It is a statistic that evaluates the predictability of a statistical model using the residuals between observed and theoretical values, with a smaller value indicating a better fit. If the log likelihood of the model is L and the number of parameters in the model is p, the AIC is calculated by the following Equation (4). In this paper, the AIC is used as an evaluation criterion when selecting parameters for the ARIMA model. L is the maximum likelihood and p is the number of free parameters. In the present model, we adopted a low value of AIC to avoid a complex model in determining the AR and MA coefficients in the ARIMA model.

$$AIC = -2\ln(L) + 2p \tag{4}$$

3.3. Coefficient of determination R^2

The coefficient of determination, also known as the contribution ratio, is a measure of the explanatory power of the predicted value of the target variable relative to the observed value of the target variable in a regression analysis. The coefficient of determination is generally expressed as R^2. R^2 takes values between 0 and 1, and the closer to 1, the more valid the regression equation is. When the actual data are (x_i,y_i), the data estimated from the regression equation are $(y_i, \hat{y}_i)$, and the mean value obtained from the entire data is $(\bar{x}, \bar{y})$, the following Equation (5) is shown.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5}$$

4. Data set and Experiment Environment

4.1. Experiment Environment

The field for this experiment was a farmer in Yatsushiro City, Kumamoto Prefecture, who grows bananas and coffee. Since Japan belongs to the temperate zone, farmers across the country are searching for new cultivation methods to make the cultivation of tropical plants such as bananas and coffee possible in temperate Japan. The farmers in Yatsushiro City, which is the focus of this experiment, also grow bananas and coffee in a vast plastic greenhouse, maintaining a tropical environment. However, in Japan, which has four distinct seasons, temperatures vary throughout the year, making it difficult to maintain a constant environment in the vast greenhouses, which affects the growth and production of the plants. For this reason, many companies are introducing agricultural IoT, collecting environmental data in greenhouses and providing visualization services. However, small- and medium-scale private farmers have to spend their budgets on environmental management, such as maintenance and utility costs in greenhouses, and are unable to purchase or rent IoT equipment and monitoring tools, which has hindered the diffusion of data infrastructure construction. Therefore, in this experiment, we developed an agricultural IoT system that can be provided to small and medium farmers, and collected, stored, visualized, and analyzed time-series data acquired from sensors, with the goal of predicting environmental information for the next day based on time-series data.

4.2. *IoT Device and Data set*

Figure 1 shows the IoT device developed. The developed IoT devices were equipped with multiple types of sensors to collect data in order to investigate environmental information in greenhouses from various perspectives. In addition, we conducted a durability test to determine the temperature range at which the device malfunctions, aiming to develop an IoT device that can continue to operate even under harsh environmental conditions. The results showed that the IoT devices failed at temperatures around 50 degrees Celsius, so a fan was installed inside the chamber to prevent failures due to overheating of the IoT devices. In addition, a Watch Dog Timer (WDT) was incorporated into the software to monitor the microcontroller for program stoppages and communication interruptions, and to restart the program in the event of a stoppage. We developed IoT devices that can continue to operate even in harsh environments by implementing countermeasures on both the hardware and software sides.

From this device, temperature, humidity, air pressure, carbon dioxide, solar radiation, and soil moisture in the greenhouse can be obtained. In addition, we also obtained meteorological data supplied by the Japan Meteorological Agency as necessary data for forecasting, such as altar instantaneous wind speed, precipitation, sunshine duration, maximum wind speed, and average wind speed, and examined whether these data could be utilized as explanatory variables for the model.

a. Device

b. Setting Environment

Fig. 1. IoT Device

Table 1. The planning and control components.

Item	Correlation Coefficent
Temperature in plastic greenhouses 1 day ago	0.886252
Average temperature 1 day before [JMA].	0.578104
Lowest temperature 1 day ago [JMA].	0.567764
Highest temperature 1 day ago [JMA].	0.546740
Solar radiation 1 day ago	0.438165
Moisture content 1 day ago (outside)	0.365323
Maximum instantaneous wind speed 1 day ago [JMA].	0.109610
Precipitation 1 day ago [JMA].	0.086130
Sunshine hours 1 day ago [JMA].	0.077532
Moisture content 1 day ago (inside)	0.015999
Maximum wind speed 1 day ago [JMA].	-0.195937
Average wind speed 1 day ago [JMA].	-0.198114
Air pressure 1 day ago	-0.284419
Humidity 1 day before	-0.467905
Carbon dioxide concentration 1 day ago	-0.680186

The correlation coefficients between the output target, temperature in the plastic greenhouse, and the other data sets are shown in Table 1, and the explanatory variables needed to predict the temperature in the greenhouse were extracted from the threshold values. As a result, five items were extracted as positively correlated: the temperature in the greenhouse one day before, the average temperature one day before, the minimum temperature one day before, the maximum temperature one day before, and the amount of solar radiation one day before. Two items, carbon dioxide concentration one day before and humidity one day before, were selected as negatively correlated items, and the temperature in the greenhouse one day before had a high correlation coefficient with the objective variable to be predicted. Therefore, the high correlations were adopted as explanatory variables in the model. As a pre-processing step for time series forecasting, the data set is divided into two types of data sets: training data and validation data. In this experiment, in order to predict the temperature in the greenhouse on the next day, the validation data was set as the last three days, and the other data was used as the training data. The other data were used as training data. Data for training The training data are For the training data, we used 14464 data sets from January 18 to December 25, 2022. We used the temperature inside the greenhouses as input and the temperature, carbon dioxide level, average temperature, humidity, and solar radiation of the previous day with high correlation coefficient as explanatory variables. In addition, 145 pieces of data from December 26 to 28 were utilized as test data for accuracy verification.

5. Experiment Result

Analysis of the ARIMA model chorreogram shows that the data has a period of one year, and the candidate parameters for AR(p) are in the range of 0 to 4, and the candidate parameters for MA(q) should be designed in the range of 0 to 2. Figure 2 and 3 show the prediction results when good parameters are set based on these results. Figure 2 shows the forecasting results without explanatory variables. The black dots indicate the values that are the correct answers, the orange line shows the predictions of the ARIMA model, and the blue line shows the Prophet predictions. The value of R^2 is 0.98 for the ARIMA model, and 0.83 for Prophet's model. The values of R2 for the test data are 0.32 for the ARIMA model and 0.74 for the Prophet model. This indicates that the Prophet model gives better predictions when there are no explanatory variables. Next, Figure 3 shows the results of the model with additional explanatory variables. The black dots indicate the values that are the correct answers, the orange line shows the predictions of the ARIMA model, and the blue line shows the Prophet predictions. The ARIMA model, the value of R^2 in the training data is 0.80, and for the Prophet model, the value is 0.85. The predicted values for the test data were 0.88 for the ARIMA model and 0.85 for the Prophet model. These results indicate that the accuracy improves with the addition of explanatory variables, and that the ARIMA model is more accurate when explanatory variables are added.

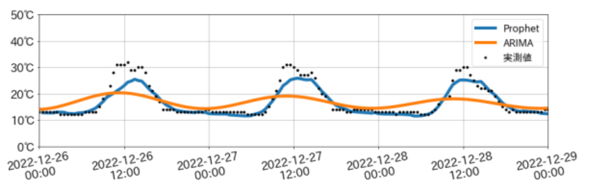


Fig. 2 The experiment data using ARIMA and Prophet model (No explanatory variables)

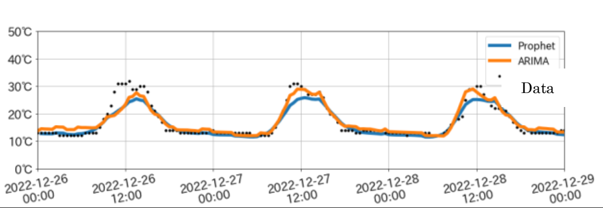


Fig. 3 The experiment data using ARIMA and Prophet model (No explanatory variables)

6. Conclusion

Then paper, we developed a model that acquires data in agriculture using IoT devices and utilizes the acquired data to make predictions. The ARIMA model and the Prophet model were used for development and comparison. In the experiment, we examined the correlation between the output temperature and the data acquired to use the explanatory variables, and compared the models with and without the variables with high correlation values added as explanatory variables. The results showed that the models with explanatory variables were more accurate and the ARIMA model had the best performance. Future issues include repeated experiments to improve accuracy and evaluation through long-term operation of the forecasting system.

Acknowledgements

This work was supported in part by Kumamoto Fruits and Vegetables Shipping Association Co. Ltd.

References

1. Ministry of Internal Affairs and Communications, 2022 White Paper on Information and Communications, Part 2, Chapter., ICT Market Trends, Section 9, Trends in the Postal and Letter Service Businesses., p.111, 2021, in Japanese
2. Ministry of Economy, Trade and Industry of Japan, "Future Investment Strategy 2017 -Reforms towards the realization of Society 5.0-", pp. 39-43, 2017, in Japanese
3. Prime Minister's Office: The Fifth Science and Technology Basic Plan, Chapter 3: Measures to Address Economic and Social Challenges, pp. 16-18, 2016, in Japanese
4. Shingo Otsuka, "Acquisition of Environment Data at Farmers" Vol.30, No.1, p.35-41, 2018, in Japanese
5. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung : Time Series Analysis: Forecasting and Control, 5th Edition, 2015.
6. Sean J. Taylor, Benjamin Letham : Forecasting at Scale, 2017.
7. Domenico, Benvenuto, at el. al, "Application of the ARIMA model on the COVID-2019 epidemic dataset, Data in brief, Vol29, pp.1-4, 2020
8. A. A. Ariyo, A. O. Adewumi and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, 2014, pp. 106-112, doi: 10.1109/UKSim.2014.67.
9. Trevor Hastie and Robert Tibshirani : Generalized Additive Models, Statistical Science, vol.1, No.3, pp.297-318, 1986
10. Akaike, H., "Information theory and an extension of the maximum likelihood principle", Proceedings of the 2nd International Symposium on Information Theory, Petrov, B. N., and Caski, F. (eds.), Akadimiai Kiado, Budapest: 267-281 (1973)

Authors Introduction

Dr. Yasunori Takemura

He received his M.S., and Ph.D. degrees from Kyushu Institute of Technology, Japan, in 2006 and 2009, respectively. In 2010, he joined Nippon Bunri University. In 2015 he joined Nishinippon Institute of Technology, where he is currently an Professor of the Department of Engineering. His research interests include robotics, IoT and AI Technology.

Mr. Naoya Nishida

He received his Bachelor's degree and M.S. in Engineering in 2021 and 2023 from the Faculty of Engineering, Nishinippon Institute of technology in Japan. He is currently working in Information Technology company.

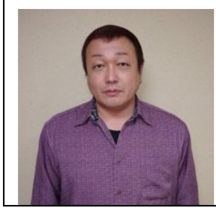
Dr. Moeko Tominaga

She is an Assistant Professor at the Department of Integrated System Engineering, Nishinippon Institute of Technology, Japan. Her research interests are in multi-agent systems and field robotics.

Mr. Eiji Mizoe

He is currenntry working in Kumamoto Fruits and Vegitables shipping Association Co. Ltd.. (Kumamoto seikabutsu syukka kumiai).

Mr. Hiroaki Wagatsuma



He is currenntry CEO in Kumamoto Fruits and Vegitables shipping Association Co. Ltd.. (Kumamoto seikabutsu syukka kumiai).