Prediction of Robot Technology Using Multi-phase Model

Juhyun Lee[1], Junseok Lee[2], Jiho Kang[1], Sangsung Park[3], and Dongsik Jang[1]
[1] Department of Industrial Management Engineering, Korea University, Republic of Korea
[2] MICube Solution, Republic of Korea
[3] Department of Big Data Statistics, CheongJu University, Republic of Korea
Email: {leeju, kangmae, jang}@korea.ac.kr, js.lee@micube.co.kr, hanyul@cju.ac.kr

***Abstract*—Technology changes with the times. It is difficult to predict, as technology develops under the influence of several factors. We analyze the technology by carrying out the patent from a time series perspective. The study consists of two phases. In the first phase, time series models detect the trend, cycle, and seasonality of the technology. Next phase performs to predict the importance of term. In order to confirm the practical applicability of the proposed method, 2,268 industrial robot patents were collected and tested. As a result, it was found that technologies beyond the dual control based on carbon materials among industrial robots will continue to develop.**

***Index Terms*—robot, patent analysis, time series, predictive modeling**

I. Introduction

The global technology market is rapidly changing due to the Fourth Industrial Revolution. Robots are one of the representative technologies that are in the spotlight recently [1]-[4]. Robots are technologies that require expertise in multiple fields such as machinery, information and communication. The initial purpose of robot was to perform human task instead recently, it is used for various industry area. In addition, robots, like drones, are being formed and developed into new industry area. This transforms the industry from labor-intensive to automated smarts. Robots are often used in manufacturing. Industrial robots are specialized in simple repetitive tasks. Industrial robots have the advantages of high precision, low defect rate and sustainable operation. It is therefore preferred as a substitute for human labor.

Determining whether a robot is developed from a national, corporate, research, or school perspective is very important. Because the technology of the robot proceeds in various forms.

Robot development is largely divided into hardware and software. Therefore, it is important to predict the development trend of these various types of robots. This study predicts the development trend of industrial robot through time series analysis. Industrial robot is a technology that requires delicate work with high precision and reliability. Industrial robot needs more hardware development than robots used in other industries. However, since the development of hardware robots is more mature than the software part, it is necessary to predict development demand through predictive modeling. This will determine if there is sustainable development. We analyze the patent document as a time series model [5]-[8]. A patent is a document containing the contents of a developed technology. Therefore, the application characteristics of industrial robot patents are analyzed from a time series perspective to predict development characteristics.

Manuscript received January 9, 2020; revised June 18, 2020.

II. Background

Time series data is divided into trend, cycle, seasonal variations, and random fluctuation. A trend is a change in the observed value that increases of decreases over time. A cycle, also called periodic, refers to fluctuations that are long and not seasonal. Seasonality is fluctuations caused by certain periodic factors, such as daily, weekly, monthly or quarterly. Random fluctuation is factor that is not explained by trend, cycle, or seasonality.

A. Time Series Models

A time series model is that predicts observations over time in time t periods. Representative this includes Binary Variable Model (BVM), Trigonometric Model (TrigM), and AutoRegressive Integrated Moving Average (ARIMA) [9]-[13]. BVM converts time series factors into binary variables and applies them to the regression model. BVM with a monthly cycle contains 12 variables. The first variable is for a time period and has a sequential value of 1, 2, ..., *t*. The remaining 11 variables are binary variables by month. For example, an observation for January has 11 variables {1, 0, 0, ..., 0}. In February it is {0, 1, 0, ..., 0} and in November it is {0, 0, ..., 1}. December is when all values are zero. TrigM constructs time series data using the *sin* and *cos* functions with periods. Equation (1) below shows the model of TrigM with 4 regression coefficients. β is the regression coefficient. t is the time index. $I\pi$ is the factor for the period of the trigonometric function and L is the period of time such as 12 or 24. ε is the error term.

$$y_t = \beta_1 t + \beta_2 \sin(\frac{2\pi t}{L}) + \beta_3 \cos(\frac{2\pi t}{L}) + \varepsilon_t \quad (1)$$

ARIMA is created by combining the AR (AutoRegressive) model and the MA (Moving Average)

doi: 10.12720/jait.11.3.181-185

model. The ARMA (p, q) model that combines the AR model and the MA model is shown in Equation (2). $A\varphi$ and θ are factors for AR and MA. ε is the error term.

$$y_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i y_{t\text{-i}} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t\text{-j}} \tag{2}$$

The ARIMA model is the addition of differencing in ARMA. Differencing is performed by equation (3). That is, diff (d) is a value obtained by subtracting the value at time t and the value before t-d.

$$diff(d): \ Y_t = \nabla^{(d)} X_i \tag{3}$$

In this study, we use the traditional time series models BVM, TrigM, and ARIMA model.

B. Ensemble Learning Model

Machine learning has a simple model with low computational complexity and a complex model with high computational complexity. The ensemble learning model is a machine learning skill that leads to a high generalization performance by repeating a number of simple models [14]-[17]. Random Forest (RF) is one of the representative ensemble learning model. RF learns by repeating many simple decision tree models. RF is generally known to have high generalization performance. RF has the advantage of calculating the relative importance between variables used in modeling.

III. Proposed Methodology

In this paper, we propose a technology analysis framework using a time series model and an ensemble model. As shown in Fig. 1, proposed method is performed.

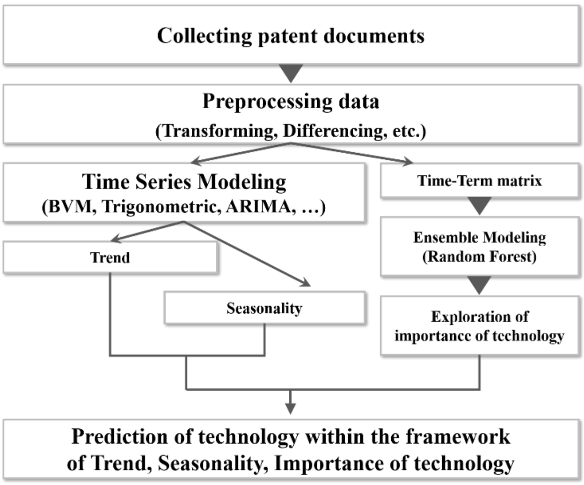


Figure 1. Flowchart of proposed methodology.

First, patent documents are collected from the DB. And data preprocessing is conducted, such as variable conversion and differencing. Modeling proceeds in two phases. In the first phase, time series models such as BVM, TrigM, and ARIMA are used to detect trend and seasonality of the technology. The time series model can then be used to predict the trend and seasonality of a given technology.

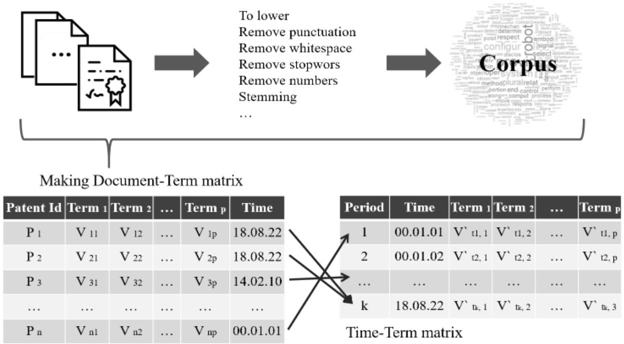


Figure 2. Conceptual diagram of time-term matrix.

Next phase, the technique is analyzed using an ensemble model. For ensemble modeling, the collected patent documents are converted into a time-term matrix for the patent applicant date. Fig. 2 is an example of a Time-Term matrix. The Time-Term matrix covers all documents belonging to a certain time. Therefore, the row of the Time-Term matrix is always less than or equal to that of the Document-Term matrix.

It is possible to detect and predict the trend and seasonality of technology through framework, also to calculate the importance of term.

IV. Experiments

In this chapter, experiments are conducted to confirm the practical applicability of the proposed method. The experiment proceeds with 2,268 patents on industrial robots applicated in 2000 at the United States. Experiments predict the future industrial robots and identify important terms.

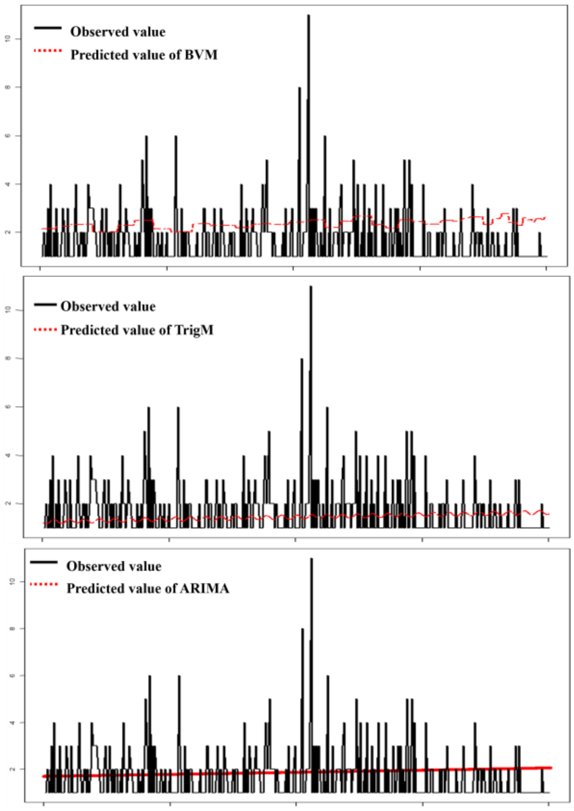


Figure 3. Prediction of time series models.

The experiment proceeds in the order of BVM, TrigM, ARIMA, and RF. As a result of decomposition, industrial robots have a slightly upward trend, and certain seasonality is observed. Now, time series and ensemble models are used to detect and predict each factor. To compare the generalized prediction performance, the data was split as training data until 2014 and the test data afterwards. MAPE (Mean Absolute Percentage Error) was used as the performance index.

As shown in Fig. 3, prediction was performed using BVM, TrigM and ARIMA. The upper direction of the figure is predicted using BVM, the middle position is TrigM, and lower position is ARIMA. BVM and TrigM were able to detect the application trend and cycle of industrial robots. ARIMA performed log transformation. At the bottom of Fig. 3 is the result of ARIMA. ARIMA made it possible to detect the trend of patent applications for industrial robots.

The TF-IDF value is an indicator that weights words that appear a lot in a particular document [18]-[20]. Using the TF-IDF value can reduce the weight of words such as system, which occur frequently in patent documents. We used a TF-IDF based Time-Term matrix for RF.

The following uses words from the patent documents collected to use Time-Term matrix. Text mining techniques were used to preprocess and collect words. Time-Term matrix used for the prediction is 644 dimensions (terms). The frequency of the words used in the Time-Term matrix.

The Fig. 4 below shows the RF predictions. RF predicted the number of patent applications using the Time-Term matrix. As a result of the prediction using RF, it was confirmed that the application cycle and scale of the industrial robot were detected. The number of trees of RF used for prediction is 100. As in the other pictures, the solid black line is an observation. The dotted red line is the predicted result.

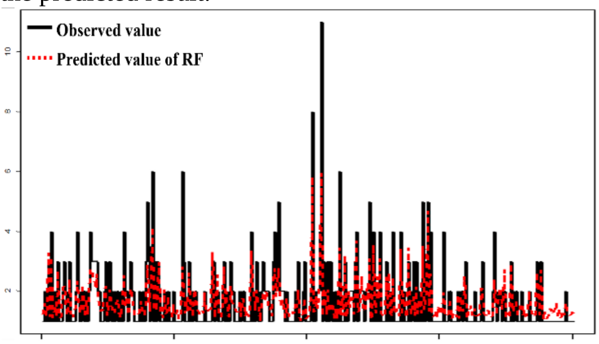


Figure 4. Prediction of ARIMA and Random forest models.

TABLE I. COMPARISON OF THE TEST MAPE BY MODEL

Model	MAPE (%)	Rank
BVM	89.2597	4
Trigonometric	44.0900	2
ARIMA	61.7683	3
Random Forest	24.9740	1

As can be seen from the data in Table I, the best prediction model is RF. MAPE value of RF is 24.9740. The most predictive model of the time series was the TrigM that value is 44.0900. The third model is ARIMA, with a low MAPE of 61.7683. Finally, BVM had the lowest performance, but the MAPE value was 89.2597.

The most influential terms were words commonly used in general robots such as ‘configure’, ‘system’, ‘position’, ‘robot’, ‘portion’, and ‘provide’. The frequency of their words was soaring. However, ‘couple’ and ‘plural’ were high-level words in TF-IDF.

Except for general robot words, dual control related words such as plural and couple had high influence in industrial robot. And functional operators such as hole, select, and clean are mainly derived. In terms of materials, carbon was found to have a high impact [21]-[23].

V. CONCLUSIOINS AND FUTURE WORKS

We proposed a framework for predicting techniques using time series data. The purpose of the proposed model is to detect trend, cycle, and seasonality of time series data. For this, a traditional time series model was used. In addition, an ensemble model was used to improve prediction performance. In order to utilize the ensemble model, a time-term matrix structure is proposed. This is a matrix expressing the frequency of words in a patent over time. It organizes the words in the document in units of hours. In this process, it recognizes documents as one in a single time.

Industrial robot patents were collected and tested to confirm the practical applicability of the proposed method. The experiment proceeds with 2,268 patents on industrial robots applicated in 2000 at the United States. As a result, it was possible to detect the trend and cycle of industrial robot technology with time series models. BVM and TrigM were suitable for detecting trends and cycles in robot patents. In addition, ARIMA was able to detect trends in robot patents. It is predicted that the application of robot patents will continue to increase through ARIMA.

In RF, it was possible to detect the cycle of industrial robot technology. In addition, the RF model was able to derive the relative importance of each term in the Time-Term matrix. Except for general robot terminology, we found that dual control and carbon material terminology are important. Therefore, industrial robots need to develop technology that can control more than dual based on carbon materials.

So far, this study has proposed a model for technology prediction. The proposed model was implemented through a traditional time series model and an ensemble model. As a result of implementation, the predictive performance of the ensemble model was the best. And it was excellent in the order of BVM, TrigM, and ARIMA.

In future research, it is necessary to develop the time series prediction model for each term in the Time-Term matrix and search for the emerging term. By combining the results of this study, it will be possible to establish a more precise technology development strategy.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

JH. L., JS. L., J. K., and S. P. conceived and designed the experiments; D. J analyzed the data to show the validity of this study; JH. L. wrote the paper and performed the entire research step. In addition, all authors have cooperated with each other in revising the paper.

ACKNOWLEDGMENT

This research was supported by the MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under the Fostering Global Talents for Innovative Growth Program (P0008749) supervised by the Korea Institute for Advancement of Technology (KIAT).

REFERENCES

[1] L. Pagliarini and H. Lund, "The future of robotics technology," *Journal of Robotics, Networking and Artificial Lift*, vol. 3, no. 4, pp. 270-273, 2017.

[2] T. Wang, Y. Tao, and H. Liu, "Current researches and future development trend of intelligent robot: A review," *International Journal of Automation and Computing*, vol. 15, no. 5, pp. 525-546, 2018.

[3] L. Yang and M. Zhu, "Review on the status and development trend of AI industry," in *Proc. IEEE 4th International Conference on Cloud Computing and Big Data Analysis*, 2019, pp. 89-93.

[4] A. Rath, D. Parhi, D. H. Das, M. Muni, and P. Kumar, "Analysis and use of fuzzy intelligent technique for navigation of humanoid robot in obstacle prone zone," *Defence Technology*, vol. 14, no. 6, pp. 677-682, 2018.

[5] S. Jun and S. Park, "Examining technological innovation of Apple using patent analysis," *Industrial Management & Data Systems*, vol. 113, no. 6, pp. 890-907, 2013.

[6] J. Kim, J. Lee, G. Kim, S. Park, and D. Jang, "Time series analysis of patent keywords for forecasting emerging technology," *KIPS Transactions on Software and Data Engineering*, vol. 3, no. 9, pp. 355-360, 2014.

[7] A. Supraja, S. Archana, S. Suvetha, and T. Geetha, "Patent search and trend analysis," in *Proc. IEEE International Advance Computing Conference*, 2015, pp. 501-506.

[8] H. You, M. Li, K. Hipel, J. Jiang, B. Ge, and H. Duan, "Development trend forecasting for coherent light generator technology based on patent citation network analysis," *Scientometrics*, vol. 111, no. 1, pp. 297-315, 2017.

[9] S. Chae and J. Gim, "A study on trend analysis of applicants based on patent classification systems," *Information*, vol. 10, no. 12, pp. 364, 2019.

[10] J. Long and J. Freese, *Regression Models for Categorical Dependent Variables Using Stata*, Stata Press, 2006.

[11] A. Agresti and M. Kateri, *Categorical Data Analysis*, Springer Berlin Heidelberg, 2011, pp. 206-208.

[12] A. Livera, R. Hyndman, and R. Snyder, "Forecasting time series with complex seasonal patterns using exponential smoothing," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1513-1527, 2011.

[13] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, 2015.

[14] T. Dietterich, "Ensemble learning," *The Handbook of Brain Theory and Neural Networks*, vol. 2, pp. 110-125, 2002.

[15] C. Zhang and Y. Ma, *Ensemble Machine Learning: Methods and Applications*, Springer Science & Business Media, 2012.

[16] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18-22, 2002.

[17] J. Speiser, M. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Systems with Applications*, 2019.

[18] J. Ramos, "Using tf-idf to determine word relevance in document queries," in *Proc. 1st Machine Learning Conference*, 2003, pp. 133-142.

[19] H. Noh, Y. Jo, and S. Lee, "Keyword selection and processing strategy for applying text mining to patent analysis," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4348-4360, 2015.

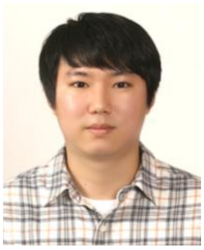
[20] J. Joung and K. Kim, "Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data," *Technological Forecasting and Social Change*, vol. 114, pp. 281-292, 2017.

[21] J. Behrens, K. Stepanova, R. Lange, and R. Skoviera, "Specifying dual-arm robot planning problems through natural language and demonstration," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2622-2629, 2019.

[22] C. Schmidt, P. Weber, T. Hocke, and B. Denkena, "Influence of prepreg material quality on carbon fiber reinforced plastic laminates processed by automated fiber placement," *Procedia CIRP*, vol. 67, pp. 422-427, 2018.

[23] Z. Kálai and T. Schüppstuhl, "Robot-guided, feature-based 3D printing strategies for carbon fiber reinforced plastic," in *Proc. Tagungsband des 4. Kongresses Montage Handhabung Industrieroboter*, 2019, pp. 178-186.

Juhyun Lee was born in Korea. He received his Bachelor of Science degree in statistics from CheongJu University. He is currently pursuing his PhD degree while researching as master's and doctor's integration process in Korea University. His major is industrial management engineering and his research interests are Patent Analysis, Statistical Learning, and Natural Language Processing.



Junseok Lee was born in Korea. He received his Bachelor of Engineering degree in mechanical engineering and automation from Zhejiang University in China. He received his Master and PhD of Engineering degree in industrial engineering from Korea University. His research interests are Patent Analysis, Data Mining, and Management of Technology.

Jiho Kang was born in Korea. He received his Bachelor and Master of Engineering degrees in industrial information system from Ajou University in Korea. He is currently pursuing his PhD degree while researching in Korea University. His major is industrial management engineering and his research interests are Management of Technology, Innovation Strategy, Data Science, and Text Mining.

Sangsung Park was born in Korea. He received his Master and PhD of Engineering degree in industrial engineering from Korea University. He was research and assistant Professor at Korea University from 2006 to 2018. He is an assistant Professor at the CheongJu University. His research interests are Patent Analysis, Data Mining, Management of Technology, and Technology Evaluation.

Dongsik Jang was born in Korea. He received his Bachelor of Engineering degree in industrial engineering from Korea University. He received his Master of Engineering degree in industrial engineering from Texas State University. He received his PhD of Engineering degree in industrial engineering from Texas A&M University. He is a Professor at the Korea University in Korea. His research interest is Pattern Recognition.