Journal of information and communication convergence engineering
http://jicce.org

J. Inf. Commun. Converg. Eng. 19(3): 142-147, Sep. 2021 Regular paper

Disapproval Judgment System of Research Fund Execution Details Based on Artificial Intelligence

Yongkuk Kim[1], **Tan Juan**[2], **and Hoekyung Jung**[3*], ***Member*, *KIICE***

[1]Senior Principal Researcher, KEIT, Daejeon 35262 Korea
[2]Professor, Weifang University of Science and Technology, Weifang 262700 Shandong, China
[3]Professor, Department of Computer Engineering, Paichai University, Daejeon 35345 Korea

Abstract

In this paper, we propose an intelligent research fund management system that applies artificial intelligence technology to an integrated research fund management system. By defining research fund management rules as work rules, a detection model learned using deep learning is designed, through which the disapproval status is presented for each research fund usage history. The disapproval detection system of the RCMS implemented in this study predicts whether the newly registered usage details are recognized or disapproved using an artificial intelligence model designed based on the use of an 8.87 million research fund registered in the RCMS. In addition, the item-detail recommendation system described herein presents the usage details according to the usage history item newly registered by the artificial intelligence model through a correlation between the research cost usage details and the item itself. The accuracy of the recommendation was shown to be 97.21%.

Index Terms: Artificial intelligence, Detection system, Real-time cash management system, Research fund

I. INTRODUCTION

Despite the detection of abnormal transactions using data business rules, the disapproval of research funds continuously occurs, and the complexity of the system continues to increase with the application of additional business rules owing to the further integration of the research fund management system. In addition, there is a continuing demand for support functions to minimize the possibility of a disapproval of research funds owing to mistakes by researchers unfamiliar with the accounting regulations.

Therefore, through this study, we intend to support the research fund execution of researchers and the research fund settlement of specialized institutions by utilizing big data and the latest artificial intelligence technology possessed by the operating research fund management system. The machine learns the details of a disapproval from the previous execution, allowing the recognition and disapproval of the new execution details to be judged; in addition, if the determined information is provided to the researcher prior to the business being conducted, the cause of the disapproval is resolved by supplementing the execution details. Thus, the efficiency of the research expenditures can be improved. In addition, it will be possible to suggest the details of the research fund execution to which the accounting standard should be executed.

In this study, we designed and implemented an artificial-intelligence-based research fund management system in which disapproval detection and expenditure (item) recommendation functions are added to the existing research fund management system using big data. The implications of this study and future research directions are also presented herein.

Received 29 April 2021, **Revised** 28 May 2021, **Accepted** 28 May 2021
***Corresponding Author** Hoekyung Jung (E-mail: hkjung@pcu.ac.kr, Tel: +82-42-520-5640)
Professor, Department of Computer Engineering, PaiChai University, Daejeon 35345, Korea.

Open Access **https://doi.org/10.6109/jicce.2021.19.3.142** print ISSN: 2234-8255 online ISSN: 2234-8883

II. IMPROVEMENTS TO RESEARCH FUND MANAGEMENT SYSTEM

Excessive administrative work is required in the spending of research funds. The government has therefore introduced a research fund management system and linked it with the university's accounting system, thereby reducing the burden on research fund accounting management. The higher the level of central management, the greater the impact on the research outcomes, such as the number of manuscripts registered and the number of patent applications [1, 2]. Owing to these effects, the research fund management system has continuously improved after being applied to national R&D projects [3, 4].

RCMS, a research fund management system, has established a research fund management system using finance and IT technology, and has reorganized the research fund execution management process based on advanced technologies in the fields of electronic finance and electronic evidence. An objective verification system was established by linking the electronic tax invoice information provided by the National Tax Service, in addition to the card usage history from the previous credit card company based research fund management system [5, 6].

The research fund management system has not only allowed improvements in research, it also applies big data, through which the history of the research expenditures is integrated and managed using the research fund management system.

Through the application of big data, the relationship between the correlations and the difference in research expenses based on the research characteristics was analyzed, and an appropriate method for calculating the expenses of R&D projects was suggested [7, 8].

III. INTELLIGENT RESEARCH FUND MANAGEMENT SYSTEM

In this chapter, to improve the method for checking research funds based on the application of rules, which has been a limitation of the existing research fund management system, we implemented an artificial-intelligence-based disapproval detection model for the use of research funds.

A. Scope of Improvement of Intelligent Research Fund Management System

In this study, the entire system was designed according to the components of an intelligent research fund management system based on artificial intelligence. User convenience has been improved by adding an artificial-intelligence-based disapproval detection model and an expenditure (item) recommendation function to the existing research fund management system.

As an improvement, the disapproval detection model presents whether the individual usage details have been rejected in the reconciliation task menu of the existing research fund management system. When checking the entire usage history, the management agency and the person in charge of the accounting firm conducting the settlement should carefully check whether the usage history has been marked as disapproved.

B. Target System Composition of Intelligent Research Fund Management System

The target system composition of the improvement area of the existing research fund management system consists of a detailed disapproval detection system and an expenditure (item) recommendation.

The disapproval detection system consists of a part generating a disapproved item detection model through data collection, data preprocessing, training data generation, and unrecognized item detection, by which the detection model determines the disapproved items and provides them in the form of an API. The expenditure (item) recommendation system is composed of an expenditure (item) recommendation model generation part for generating learning- and rule-based items, as well as an expenditure (item) recommendation service part that provides the recommendation results in the form of an API.

C. Research Fund Rejection Detection System

Fig. 1 shows a system configuration diagram for detecting the disapproval of research funds. A model building server is placed separate from the existing RCMS system, and service is provided in the form of an API. The procedure for modeling the disapproval detection is as follows:

1) Problem Definition

A list of data related to the problem is derived, the possi-

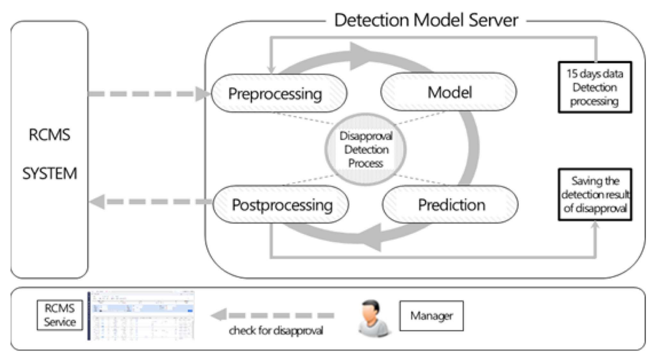

Fig. 1. Configuration diagram of disapproval detection system.

bility of collecting such data is determined, and the first provisional decision on which model to apply based on the problem and available data is made.

2) Data Collection and Analysis

Data were obtained from the RCMS system currently in operation. The columns of data to be analyzed numbered 656, with 8,867,981 records, and 8,358,314 data were derived by removing duplicates. This process of identifying the final modeling target data is applied by checking the target data. Through this process, duplicates, null values, and unnecessary columns were removed from the 656 columns, and a final 118 columns were used for the model design. The ratio of the data to be recognized and disapproved of the target data was 98.9% for recognition and 1.1% for disapproval, and among the reasons for disapproval, non-recovery of levied tax accounted for the largest proportion.

Individual data were analyzed by separating them into coded, informational, and monetary data. Codification data, such as the research fund tax code, were recognized and disapproved for each code.

3) Data Preprocessing

To purify the data in the data preprocessing stage, data were searched for normalization, outliers included in the data were removed, missing values were filled in, and rows/columns were removed. First, the missing values were replaced with mean values, modes, and predicted values through machine learning.

4) Feature Extraction

From among the 620 variables in which the features were generated, 226 were used, from which the features were extracted using a feature selection technique. The 620 variables created through the combination of variables were first extracted through an ×2 test, Pearson test, and t-test, and finally, using XGBoost, a recursive feature estimator, random forest, and light GBM, the 226 variables were extracted.

5) Model Learning

The structure of the disapproval detection model is an ensemble of a deep neural net (DNN) model and a recurrent neural net model, as shown in Fig. 2, and the model is used to detect a disapproval or approval based on the input data. Recognition satisfying the criterion returns a value of “0,” and a disapproval that does not meet the criterion returns a value of “1.”

In this study, we compare the performance of three models, an RWN, an advanced DNN, and an SGAN, through a learning curve applied to select the model with the best performance. The selected model is verified through a performance comparison using the multi-layer perceptron model set as the standard.

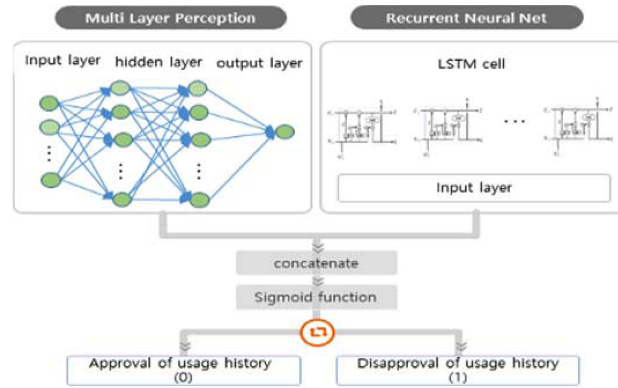

Fig. 2. Structure of disapproval detection model.

The first trained model is an RWN, which creates a neural network structure by topologically aligning the random graphs generated through the graph generation model.

The training of each model aims to reduce the loss value, and the training is set to 100 epochs. Fig. 3 shows a graph of the loss value in the learning curve of the RWN. It was confirmed that the loss value for the training data continued to decrease as the learning progressed.

Fig. 4 presents a graph of the recall and precision among the learning curves of the RWN, of which the recall is 0.84 and the accuracy is 0.75.

The second trained model was the advanced DNN, and the training was conducted for 100 epochs in the same way as with the RWN, with an aim at confirming the loss, recall, and precision, and reducing the loss value. Similar to the RWN, the loss values for the training and verification data continued to decrease as the learning progressed. In addition,

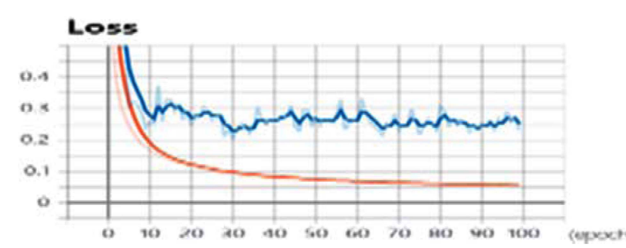

Fig. 3. Learning graph for loss values of an RWN.

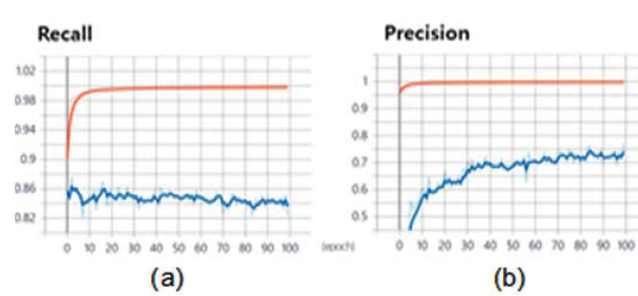

Fig. 4. Learning graphs for (a) recall and RWN and (b) precision and RWN.

Table 1. Performance of RWN with CE by threshold

Threshold	Accuracy	Precision	Recall	F1-Score
0.1	99.06%	52.24%	89.25%	65.91%
0.2	99.36%	63.17%	87.36%	73.32%
0.3	99.47%	68.86%	86.27%	76.59%
0.4	99.52%	72.24%	85.17%	78.17%
0.5	99.57%	75.85%	84.08%	79.75%
0.6	99.62%	81.92%	80.70%	81.30%
0.7	99.64%	86.08%	77.51%	81.57%
0.8	99.67%	91.42%	74.23%	81.93%
0.9	99.64%	93.58%	69.65%	79.86%

it was confirmed that the reproducibility was 0.86, and the precision was approximately 0.68.

The third trained model, SGAN, is a structure in which the classification and generation models complement each other during the training. A classification model that learns both labeled and unlabeled data determines whether the fake data generated by the generated model are authentic. The training was conducted in the same manner as in the previous two models, and it was confirmed that the loss value for the training data for SGAN continued to decrease as the training progressed. In the SGAN learning curve, a graph of the recall and precision confirmed that the recall rate was 0.4, and the precision was 1.

6) Model Evaluation

The evaluation data were used to evaluate the performance of the learning model. Evaluation indices such as the precision, recall, accuracy, and F1-Score were applied. An index for evaluating the learning model was set, and candidate models were compared using the corresponding index to select a model with a high performance.

As a result of comparing and evaluating the models, the performance of the RWN, which applied a data analysis and characteristic engineering, was determined to be the best. Table 1 shows the results of each performance index according to the threshold of the proposed model.

When the threshold value was 0.8, as listed in Table 1, the F1-Score was the highest at 81.93%, the precision (model hit rate) was 91.42%, and the recall (accident hit rate) was 74.23%.

D. Recommendation System for Expenditure Research Expenses

Fig. 5 shows a system configuration diagram for recommending suitable non-specific items for the usage history of each research fund. Similar to the research fund disapproval detection system, a system was built separately from the existing RCMS, and was used in the RCMS through the API. The recommendation module recommends non-specific items according to the user request, and when the user selects the recommended or other non-specific items, the selection value is learned again through the recommendation module.

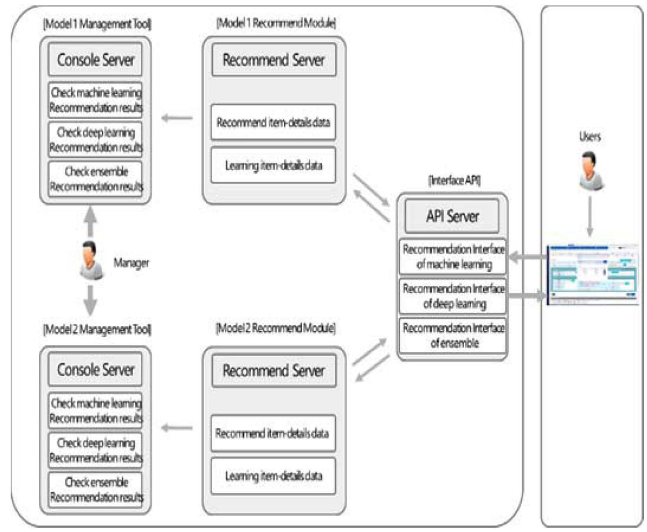

Fig. 5. Configuration diagram of recommendation system for research expenditure items.

The model flow used in the design of a non-specific recommendation model is the same as that applied for designing a disapproval research fund model. However, in the modeling of non-specific item recommendation models, characteristic engineering is not carried out because the structure and data type to be analyzed for non-specific item recommendation are simpler than those used in the detection of disapproved research expenditures.

1) Problem Definition

For implementation of a national R&D project, the researcher organizes a research budget and divides it based on the purpose of use, i.e., research expense item–detail–sub-detail. However, when trying to use research funds, researchers who are unfamiliar with accounting rules and regulations for managing such funds might become confused as to which non-tax items should be organized. Finding a method that can automatically inform researchers of the non-tax items of research funds that they intend to use is a problem with the system recommending non-tax items for the application of research funds.

2) Data Collection and Analysis

A total of 8,867,981 data points were collected and used for modeling. Errors were removed from the total data, and the number of data used for the actual modeling was 8,814,882. The error data include 53,099 cases (0.6%), and as the main type of error, no item information is available, which is 99.9% of the total error data.

3) Data Preprocessing

Normalization rules were applied to all data. Pre-processing of the syllable and morpheme analysis data was then conducted. Because the performance of the learning model depends on the preprocessing of the data, the consistency was verified by sampling the preprocessed data.

4) Model Learning

A modeling technique suitable for problem solving was selected and applied to the data to create a learning model, which learns by dividing the data obtained into data for learning and evaluation. For the non-tax-specific recommendation system, two learning models were built, i.e., a support vector machine (SVM), which is machine learning approach, and word embedding (WE), which is a type of deep learning. In the non-tax item recommendation learning model, when an item comes in as input data, the non-detail–sub-detail should appear as the output data.

5) Model Evaluation

The evaluation data were used to evaluate the performance of the two learning models. Unlike the research fund disapproval detection system, the non-tax item recommendation system requires multiple classifications, and thus the performance evaluation index differs from that of the disapproval detection model. Because the evaluation index of the model indicates how many times the correct answer was recommended in the verification data, the performance evaluation index uses the *Accuracy*, a representative index that evaluates the predictive performance of a multi-classification model.

A 10-fold cross validation evaluation method was applied to verify the effectiveness of the recommendation model. The results show a uniform accuracy for all verification data fold in the machine learning (SVM) recommendation model shown in Table 2, and in the deep learning recommendation model provided in Table 3.

Table 2. Results of 10-fold cross-validation for SVM

Learn data	Number of learning	Verification data	Number of verification	Accuracy (1) (%)	Accuracy (2) (%)	Accuracy (3) (%)
1-9	7,933,360	0	881,522	90.48	94.93	96.43
0, 2-9	7,933,368	1	881,514	90.51	94.97	96.47
0, 1, 3-9	7,933,375	2	881,507	90.54	95.03	96.53
0-2, 4-9	7,933,381	3	881,501	90.41	94.89	96.41
0-3, 5-9	7,933,392	4	881,490	90.50	94.92	96.42
0-4, 6-9	7,933,399	5	881,483	90.49	94.86	96.38
0-5, 7-9	7,933,405	6	881,477	90.45	94.91	96.41
0-6, 8, 9	7,933,412	7	881,470	90.46	94.92	96.41
0-7, 9	7,933,418	8	881,464	90.50	94.94	96.47
0-8	7,933,428	9	881,454	90.48	94.93	96.43
Average	7,933,394	-	881,488	90.48	94.93	96.44

Table 3. Results of 10-fold cross-validation for WE

Learn data	Number of learning	Verification data	Number of verification	Accuracy (1) (%)	Accuracy (2) (%)	Accuracy (3) (%)
1-9	7,933,360	0	881,522	89.70	95.10	97.20
0, 2-9	7,933,368	1	881,514	89.70	95.20	97.20
0, 1, 3-9	7,933,375	2	881,507	89.70	95.20	97.20
0-2, 4-9	7,933,381	3	881,501	89.70	95.10	97.20
0-3, 5-9	7,933,392	4	881,490	89.70	95.20	97.20
0-4, 6-9	7,933,399	5	881,483	89.70	95.20	97.20
0-5, 7-9	7,933,405	6	881,477	89.70	95.20	97.30
0-6, 8, 9	7,933,412	7	881,470	89.80	95.20	97.20
0-7, 9	7,933,418	8	881,464	89.70	95.20	97.20
0-8	7,933,428	9	881,454	89.70	95.20	97.20
Average	7,933,394	-	881,488	89.71	95.18	97.21

IV. DISCUSSION AND CONCLUSIONS

In this paper, a new service based on artificial intelligence technology was proposed and implemented to improve the research fund management system. Such systems are used in the payment, usage, and calculation of research funds during the process of managing national government-funded R&D projects.

To implement the proposed service, we implemented a disapproval research fund detection model and a non-tax item recommendation model for research fund usage according to the design procedure of the artificial-intelligence-based model. Based on the data generated during the operation of the RCMS system, learning data were created through data collection, analysis, and pre-processing, and the best model was derived through learning and verification using various artificial intelligence models.

Among the artificial intelligence models implemented in this study, the accuracy and recall of the research fund rejection detection model were evaluated as 91.42% and 74.23%, respectively. If this model is applied in practice, it will be possible to detect 130,000 cases, i.e., 74.23% of the 170,000 cases receiving a disapproval of use during the last 5 years. Researchers will be able to contribute to the efficient use of research funds for national R&D projects by re-examining the detected usage details and resolving insufficient parts to be recognized, or by converting the corresponding research funds into other uses.

For the artificial intelligence model applied to the non-tax item recommendation service for research funding implemented in this study, when the user inputs an item, three items in order of their similarity rate are recommended, i.e., the most suitable item–detail–sub-detail. The accuracy of the recommendation model was 97.21%. Because inquiries on the selection of item–detail–sub-detail account for the largest share of the total number, researchers can eliminate the time wasted on such matters, and a considerable reduction in the number of phone inquiries at RCMS operating institutions will be achievable.

REFERENCES

[1] S. O. Yoon, "A study on the main issues of artificial intelligence-based public services," *Korea Public Management Review*, vol. 32, no. 2, pp. 83-104, 2018. DOI: 10.24210/kapm.2018.32.2.004.

[2] J. Y. Lee and I. S. Kim, "Detecting abnormalities in fraud detection system through the analysis of insider security threats," *The Journal of Society for e-Business Studies*, vol. 23, no. 4, pp. 153-169, 2018. DOI: 10.7838/jsebs.2018.23.4.153.

[3] M. A. Oh, H. S. Choi, S. H. Kim, J. H. Jang, J. H. Jin, and M. K. Cheon, "A study on social security big data analysis and prediction model based on machine learning," *Research Report of Korea Institute for Health and Social Affairs 2017-46*, Dec. 2017.

[4] J. W. Kim, H. A. Pyo, J. W. Ha, C. K. Lee, and J. H. Kim, "Various deep learning algorithms and applications," *Communications of the Korean Institute of Information Scientists and Engineers*, vol. 33, no. 8, pp. 25-31, 2015.

[5] S. H. Park and D. S. Choi, "Experiments on performance of loan screening model using multi-layer perceptron," *Proceedings of the Korean Institute of Information Scientists and Engineers*, pp. 1899-1900, 2017.

[6] T. H. Hong, S. H. Kim, and E. M. Kim, "An intelligent personal credit rating model based on deep learning using GAN and DNN," *The Journal of Internet Electronic Commerce Research*, vol. 19, no. 1, pp. 1-16, 2019. DOI: 10.37272/JIECR.2019.02.19.1.1.

[7] J. A. Jeong, K. H. Lee, and H. K. Jung, "Prediction model for unpaid customers using big data," *Journal of the Korea Institute of Information and Communication Engineering*, vol 24. no. 7, pp. 827-833, 2020. DOI: 10.6109/jkiice.2020.24.
7.827.

[8] C. H. Hwang, H. S. Kim, and H. K. Jung, "Detection and correction method of erroneous data using quantile pattern and LSTM," *Journal of Information and Communication Convergence Engineering*, vol 16, no. 4, pp. 242-247, Dec. 2018. DOI: 10.6109/jicce.2018.16.4.242.

Yongkuk Kim
received his PhD in 2021 from the Department of Computer Engineering at PaiChai University, Republic of Korea. He has been working in the Department of Digital Operations Team at the Korea Evaluation Institute of Industrial Technology since 2000, working on the planning, development, and operation of industrial technology R&D information services. His current research interests include big data, artificial intelligence, and robotic process automation.

Tan Juan
received her MS degree in computer technology from Ocean University of China in 2010. Since 2005, Tan Juan has been working in the Department of Software Technology at Weifang University of Science and Technology in China, where she teaches courses in computer science. Her current research interests include artificial intelligence, big data, deep learning, and blockchain.

Hoekyung Jung
received his MS degree and PhD in 1987 and 1993, respectively, from the Department of Computer Engineering, Kwangwoon University, Korea. From 1994 to 1995, he worked for the Electronics and Telecommunications Research Institute as a researcher. Since 1994, he has worked in the Department of Computer Engineering at PaiChai University, where he now works as a professor. His current research interests include machine learning, IoT, big data, and machine learning.