BASIC STUDY ON URGENCY CLASSIFICATION MODEL FOR SEWAGE PIPE USING MACHINE LEARNING

T. SUWA[1], M. FUJIU[2], Y. MORISAKI[3], T. FUKUOKA[3] and M. YOSHIKURA[4]

[1] Graduate Student, Division of Geosciences and Civil Engineering, Kanazawa University, Kanazawa, Japan,
[2] Associate Professor, Institute of Transdisciplinary Sciences, Kanazawa University, Kanazawa, Japan,
[3] Assistant Professor, Institute of Transdisciplinary Sciences, Kanazawa University, Kanazawa, Japan,
[4] Graduate Student，Division of Environmental Design, Kanazawa University, Kanazawa, Japan,
Correspond to Assoc. Prof. M. FUJIU (fujiu@se.kanazawa-u.ac.jp)

***Keywords**: Machine Learning, Sewage Pipe, Inspection Efficiency*

1. INSTRCUTIONS

In Japan, the sewerage system penetration rate will be about 80.1% by 2020, and the total length of sewage pipes will be approximately 490,000 km, which is a huge stock of sewage pipes. In addition, the percentage of sewage pipes that have passed the standard service life of 50 years will increase rapidly [1].

Currently, inspections of sewage pipes are conducted visually or by TV camera survey, and the degree of urgency is determined as shown in Table 1. However, in 2018, only 6686 km were surveyed, and this pace, it would take 72 years to survey all sewage pipes in a cycle. Therefore, it is difficult to equally survey, repair, and reconstruct the huge stock.

In order to conduct appropriate maintenance and management under limited personnel and budget constraints, it is necessary to estimate the urgency and prioritize sewage pipes that have not yet been inspected.

In this study, machine learning models for estimating urgency in sewage pipes are developed and evaluated. If it becomes possible to predict the deterioration of sewage pipes with high accuracy using information that can be obtained without conducting surveys, it will lead to the more efficient maintenance and management of sewage pipes.

2. DEVELOPMENT OF CLASSIFICATION MODEL

In this study, three typical ensemble learning models, Random Forest, XGboost and LightGBM are constructed and evaluated using Scikit-learn, which is a machine learning library.

In this study, pipe diameter, pipe length, number of pipes in the span, number of installed pipes in the span, and pipe age were used as input variables to the machine learning models, and default values for each model were used for the hyperparameters.

Table 1. Urgency Classification in sewage pipe

Degree of urgency	Classification of urgency
I	Immediate action is needed
II	Necessary measures can be extended to less than 5 years if simple measures are taken.
III	Necessary measures can be extended beyond 5 years if simple measures are taken.
IV	Soundness

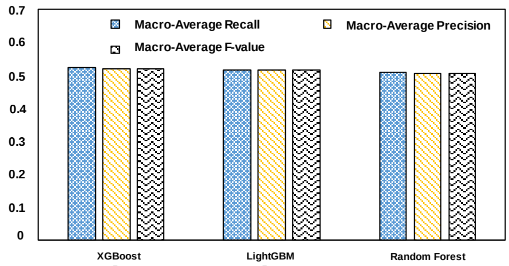

Figure 1. Results of the performance measures

The data used in this study are unbalanced. There is concern that if the number of data is biased, learning may not occur correctly. Therefore, in this study, random sampling was conducted from classes other than Urgency I, which has the least number of data, for number of Urgencies I data. Since there is concern that different sampling methods may result in different performances, 100 random samplings were used in this study. K-Fold Cross Validation (k=5) was applied to the 100 balanced data sets generated by the above procedure to verify the performance of the model. Three measures of performance were used: Macro-Average Precision, Macro-Average Recall and Macro-Average F-value. The results of the performance measures for each model are shown in Figure 1. Each value is the average of the results of the K-Fold Cross Validation on the data generated by 100 random samplings. As a result, the XGBoost classification model is the best representation.

3. CONCLUSIONS

In this study, an urgency classification model based on a machine learning method was developed and evaluated to improve the efficiency of maintenance management in sewage pipes. As a result, XGBoost performed best among the machine learning models. Since there is no significant difference in classification performance among the models, it is considered necessary to take into account feature values related to external forces such as traffic volume and rainfall to improve accuracy.

REFERENCES

[1] Ministry of Land, Infrastructure, Transport and Tourism. Available online:
https://www.mlit.go.jp/mizukokudo/sewerage/crd_sewerage_tk_000135.html,
(accessed on 15 September 2022).(In Japanese).