Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 08, August-2024

Development of Anomaly Detection System for Flight Data using AI

Takeshi Okazaki
School of Science and Engineering, Teikyo University
1-1 Toyosatodai, Utsunomiya, Tochigi, Japan

Yuichi Hasuda
School of Science and Engineering, Teikyo University
1-1 Toyosatodai, Utsunomiya, Tochigi, Japan

***Abstract*— It is essential for flight safety to support pilot in flight from the ground by air traffic controllers with high reliable IT systems. But such IT system needs huge budget, long developing period, and lacks some flexibility. We developed AI system which automatically configure normal flight pattern by learning the daily flight data, then automatically configure abnormal flight pattern. It can operate independently of existing IT systems. With this system, outlier pattern will be flexibly identified, then you can effectively detect some outlier data with less costs.**

***Keywords*— anomaly detection, flight data, CARATS, machine learning)**

I. INTRODUCTION

It is essential for flight safety to support pilot from the ground staffs and ground systems. Currently, these support from the ground is being carried out by the well-trained air traffic controllers 24 hours a day 365 days a year, using various air traffic control information processing systems [1].

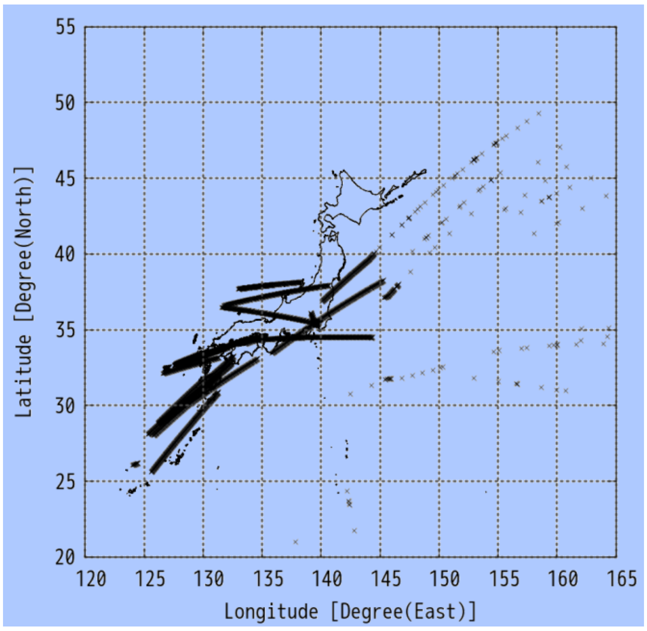
Although these IT systems have very high reliability and many functionalities, it is huge in scale and quite complicated so that it requires long period to develop. In addition, it cannot detect abnormal behavior unless it is pre-defined or pre-programmed it in advance. On the other hand, human can monitor with great flexibility and can detect anomalies which are not detected by such systems. But there are some differences in personal skills, and human errors such as oversight cannot be perfectly eliminated.

In this study, we will try to develop a system which utilizes AI to learn flight data and detect abnormal pattern. It automatically learns normal flight data and detects abnormal flight patterns. It will complement the functions and roles of both humans and IT systems with great flexibility, then improves surveillance capabilities.

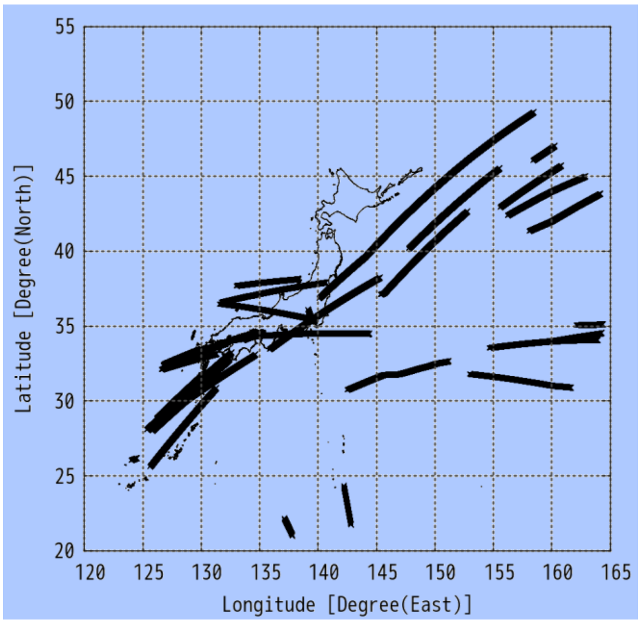
II. ANOMALY DETECTION BY MACHINE LEARNING

A. Data for Machine Learning

We have utilized CARATS (Collaborative Actions for Renovation of Air Traffic System) [2] Open Data as nominal flight data for our machine learning. It is the data sets published by the MLIT (Ministry of Land, Infrastructure, Transport and Tourism) Japan for R & D purpose and are used by many researchers.[3] This data consists of both regular airline flight data and weather data.



(a) Unprocessed data



(b) After linear interpolation
Fig.1 CARATS Open Data Example
(April 20,2020 00: 00: 00 ~ 00: 59: 59 JST)

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 08, August-2024

However, since the data interval can be extremely large in time over ocean area (Fig.1 (a)), we linearly interpolate the data and use them so that the data interval becomes 1 second or less (Fig.1 (b)).

B. Algorithm for Machine Learning

When performing abnormal detection in machine learning, there are mainly two ways in terms of the attributes of learning data you can use as described below.

1) both normal and abnormal data sets with labels.

2) normal data set only.

In most cases, including our research, you can obtain and utilize only normal data. So, it is necessary to learn only from the normal data set then determine abnormalities.
There are some methods for purpose including a method based on Support Vector Machine, a method based on a clustering, and a method based on the nearby distance.[4]
Here, we select four types of algorithms that seems to be applicable for flight anomaly detection and applied them to CARATS Open Data. The selected algorithm names and the application results are shown in table.1.

Table.1 Results of four Machine Learning Algorithm**.**

Name of algorithm	Outline	Judgment result
One Class SVM [5]	Abnormal judgment algorithm based on support vector machine.	Overall abnormal judgment is possible to some extent. But it is not suitable for the judgment of the flight data which includes the long and parallel shape.
Isolation Forest [6]	An algorithm that detects an isolated value using a determined tree.	Same as above. In addition, the area range determined to be normal tends to spread further than One Class SVM.
Local Outlier Factor [7]	An algorithm that determines outlier from the density of local data, not overall distribution.	It shows relatively good characteristics, but you need to be careful when applying, because the result is strongly affected by the density of the learning data.
K Nearest Neighbors [4]	It detects outliers from the number of nearby data, not from overall distribution.	A good judgment is obtained, and it doesn't affect too much against density changes in learning data. In addition, adjustment of the parameters is straightforward and easy.

Fig.2, Fig.3, Fig4, and Fig5 shows the normal and abnormal area judgement diagram based on the results of learning flight data by using each four algorithms. The green area shows the region which is judged to be normal, and the red area shows the region which is judged to be abnormal. The main parameters set for four machine learning algorithms are shown in table.2.
By adjusting the parameters on each algorithm, the characteristic and shape of the area shown in Fig.2, Fig.3, Fig4, and Fig5 changed to some degree, but there were no major changes in the tendency shown in Table.1.

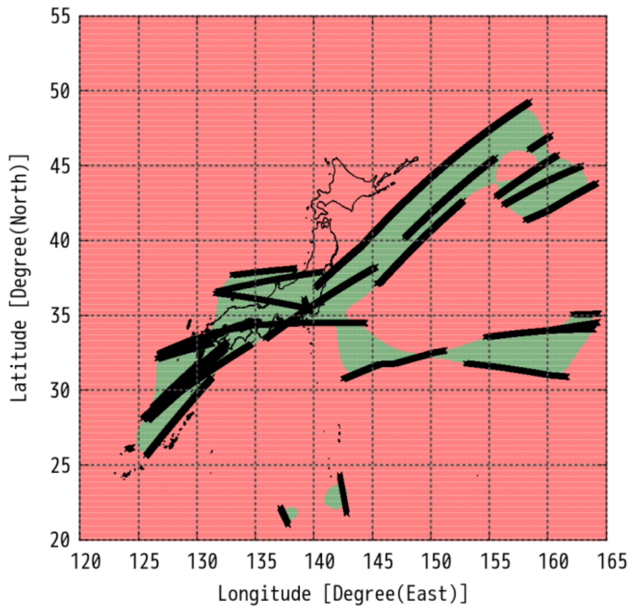


Fig.2 Area Judgment by One Class SVM

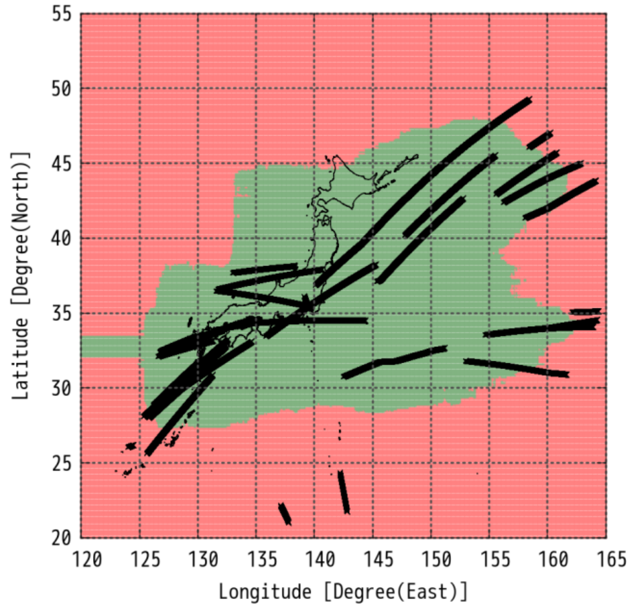


Fig.3 Area Judgment by Isolation Forest

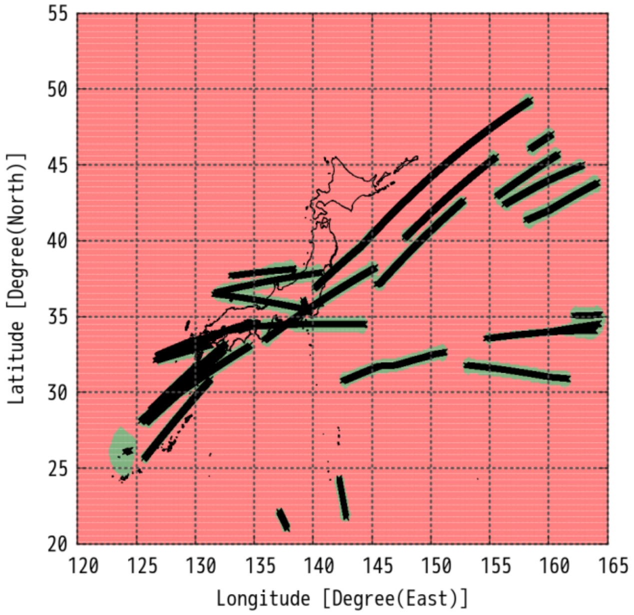


Fig.4 Area Judgment by Local Outlier Factor

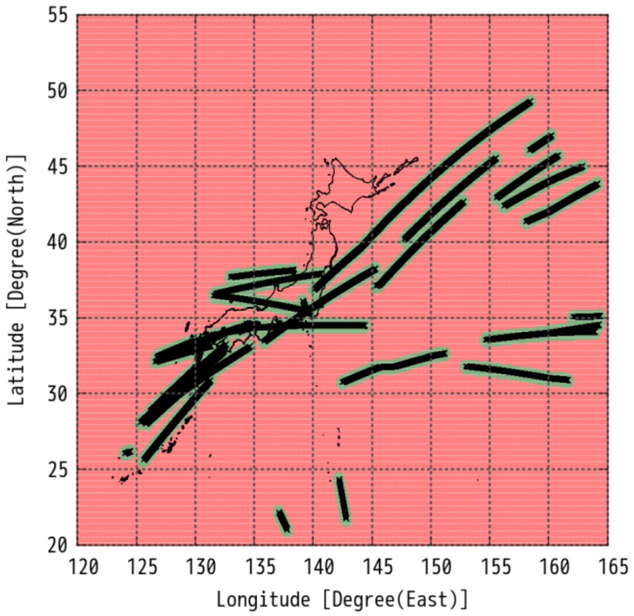


Fig.5 Area Judgment by K Nearest Neighbors

C. Evaluation of each algorithm for regional judgment

Both One Class SVM and Isolation Forest are not suitable for this decision because the shape of the boundary between the normal area and abnormal areas are not desirable. Further it is difficult to predict and/or control the border shape from the main parameters in advance. Local Outlier Factor changes the shape of normal area depending on the density of the learning data, so care must be taken when using it for this judgment.

On the other hand, K Nearest Neighbors has a desirable characteristic because the shape of the normal area is like the flight pattern. Furthermore, it is ideal for this judgment because it is not greatly affected by the density of the learning data, and you can directly control the width of the normal area by adjusting the main parameters.

Table.2 Main parameters for four Machine Learning

Name of algorithm	Main parameters
One Class SVM	kernel = RBF nu = 0.1 / gamma = 0.1
Isolation Forest	n_estimators = 100 contamination = 0.1
Local Outlier Factor	n_neighbors = 200 contamination = 0.0027(3σ)
K Nearest Neighbors	n_neighbors = 5 radius = 0.5

III. COST PERFORMANCE S

The main ground IT-Systems are TAPS for around airport, TEPS for enroute, and TOPS for ocean area.[1] The estimated cost for these IT-Systems exceeds about 23.8 billion yen for 10 years from 2015. [8][9][10] On the other hand, this system can be configured using only one standard PC and open-source software, The cost of this functionality is extremely small.

IV. CONCLUSION S

Since regular airline flights goes almost the same pattern, it is considered possible to automatically judge the abnormal flight data, by comparing data at the same time on the previous day or on the previous week and detecting the deviation from there. In this study, it was up to the two-dimensional area decision between normal region and abnormal areas by using machine learning. But in the future, we will try to perform outlier detection by using ADS-B broadcast message receiving from aircraft by using K Nearest Neighbors, which were optimal for our machine learning, then try to three-dimensional machine learning including altitude.

REFERENCES

[1] Outline of air traffic control information processing system, https: //www.mlit.go.jp/koku/content/001493947.pdf
[2] Long -term vision for future air traffic systems, https: //www.mlit.go.jp/koku/carats/, 2010 view
[3] Providing air traffic data (carats Open Data) and utilization for R & D. Japan Aeronautics Society Magazine Volume 68, No. 4 (April 2020) Copyright (C) 2020 Japan Aeronautics Society
[4] Tsuyoshi Ide (2015). Practical guide by abnormal detection -R by introductory machine learning-Corona
[5] SkLearn.svm.oneclassvm https://scikit-learn.org/stable/modules/generated/sklearn.svm.oneclassvm.html, viewed on November 25, 2024
[6] Fei Tony Liu, Kai Ming Ting, Zhi-Hua Zhou, Isolation Forest. 2008 Eighty IEEE International Conference on Data Mining (ICDM).
[7] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. NG, Jörg Sander, Lof: Identify Density-Based Local Outliers, ACM SIGMOD 2000.
[8] Japanese government procurement of Mitsubishi Electric (MLIT) https://irbank.net/chotatu/4010001008772?n=2000012100001, viewed on May 06, 2024
[9] Japanese government procurement of NTT Data (MLIT) https://irbank.net/chotatu/9010601021385?n=2000012100001, viewed on May 06, 2024
[10] Japanese government procurement of NEC (MLIT) https://irbank.net/chotatu/7010401022916?n=2000012100001, viewed on May 06, 2024