Application of Machine Learning to Maritime Safety

Wu-Hsun Chung
Dept. of Transportation Science
National Taiwan Ocean University
Keelung, Taiwan
wxc218@email.ntou.edu.tw

Yi-Yang Hung
Dept. of Transportation Science
National Taiwan Ocean University
Keelung, Taiwan
11068016@mail.ntou.edu.tw

Chien-Chung Yuan
Dept. of Navigation
Taipei University of Marine Technology
Taipei, Taiwan
yuan.jj1031@gmail.com

***Abstract*—Machine learning in artificial intelligence (AI) has been widely applied to various areas. Port State Control (PSC) is an internationally agreed regime for maritime safety by inspecting foreign ships in national ports. Ship risk classification and on-board inspection for deficiencies are critical measures in PSC to verify whether the condition of ships selected complies with the requirements of international maritime regulations. This study proposes a hybrid machine-learning method combining a decision tree and association analysis to find the relationships between ship risk classification and deficiency association rules. The association rules of ship deficiencies with and without the intervention of a decision tree are compared to see if particular ship selection affects ship deficiency association. The results show that the intervention of a decision tree for particular ship classification at the pre-check stage can find more association rules than a traditional approach for ship selection does. In addition, the association rules identified have higher "Confidence" values, which means that they more likely occur when an antecedent deficiency item occurs. The findings can help PSC inspectors effectively and efficiently classify ship risks and identify ship deficiency items.**

***Keywords*—Machine Learning, Port State Control, Decision Tree, Association Analysis**

I. Introduction

The techniques of artificial intelligence (AI) and data mining have been widely applied to various fields, such as finance, manufacturing, healthcare, biology, etc.. Because of low shipping unit cost and high shipping volume compared with land and air transport, maritime transport has always been the primary transportation mode in world trades. In recent decades, the rapid growth of world trades significantly stimulated the demand for maritime transport. The report from UNCTAD (United Nations Conference on Trade and Development) shows that global shipping containers kept growing except the years for the financial crisis and COVID19 pandemic as shown in Fig. 1 [1]. This paper applies the techniques of AI and data mining to solve the problems of ship safety inspections in port management.

PSC (Port State Control) is a precaution measure against substandard shipping by allowing port administrations to perform inspections on foreign ships entering their ports. Therefore, PSC plays a critical role in the safety of vessels, personnel, cargo, and ports. The PSC problem involves two stages, the pre-check stage and the check stage, as shown in Fig 2. The pre-check stage is about ship selection that determines which ship to be inspected based on ship risks. After ships are selected for PSC inspections at the pre-check stage, the check stage is to identify potential deficiencies of the selected ships. Table I lists the primary items of deficiencies in PSC inspections [2]. If the association relationships among deficiencies are found, the PSC inspection activities become effective and efficient.

Many data mining techniques have been applied to PSC inspections and the shipping industry. For example, Yang et al. presented a data-driven Bayesian Network based model to analyze risk factors influencing PSC inspections and predicted the probability of ship detention in European countries [3]. Wang et al. also created a Bayesian Network based PSC risk probabilistic model to analyze the dependency and interdependency among the risk factors influencing PSC inspections based on the inspection database of Tokyo MoU [4]. In addition, Yang et al. also applied a Bayesian network. They incorporated a data-driven Bayesian network into the Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) method for ship detention risk control [5]. Yan et al. proposed a classification model using balanced random forest to predict ship detention by analyzing inspection records at the Hong Kong port [6].

On the other hand, some studies applied association analysis approaches to find the relationships between ship detention and inspection deficiencies. Gu et al. used FP-tree algorithm to find the correlations between inspection deficiencies and ship detention [7]. Zhang et al. used Apriori algorithm to analyze PSC inspection data for PSC decision support [8]. Chang et al. analyzed the PSC inspection records of Taiwan's major ports and identified the association relations of inspection deficiencies in terms of ship types, ship societies, and ship flags [9].

The major issues in past studies are that the problems occurring at the pre-check stage and check stage are highly related but usually dealt with separately. Few studies propose an integrated methodology to solve the separate problems occurring in two different stages. This paper presents a hybrid machine-learning method combining a decision tree and association analysis to find the relationships between ship selection and deficiency association rules. Understanding the connections can help PSC inspectors significantly improve the effectiveness of PSC inspection tasks.

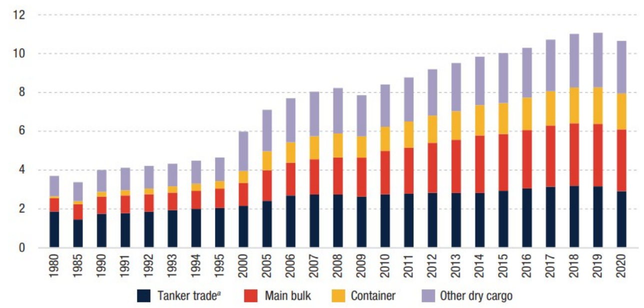

Fig. 1. Statistics of world shipping container volume between 1980 and 2020 (Source: UNCTAD 2021) [1]

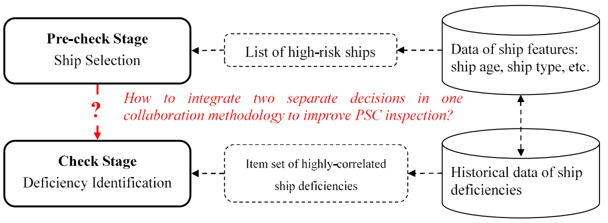

Fig. 2. Two stages of PSC inspections

TABLE I. PRIMARY ITEMS OF DEFICIENCY CODES IN TOKYO MOU [2]

Code	Description	Code	Description
01000	Certificates & Documentation	10000	Safety Of Navigation
02000	Structural Conditions	11000	Life Saving Appliances
03000	Water/Weathertight Conditions	12000	Dangerous Goods
04000	Emergency Systems	13000	Propulsion and Auxiliary Machinery
05000	Radio Communications	14000	Pollution Prevention
06000	Cargo Operations Including Equipment	15000	ISM (International Safety Management)
07000	Fire Safety	16000	ISPS (International Safety for Port and Ships)
08000	Alarms	99000	Other
09000	Working and Living Conditions	18000	Labor Conditions

II. METHOD

Fig. 3 shows the analysis framework in this study. First, the data associated with ship features, such as ship type, ship age, ship tonnage, etc., were extracted from the historical PSC inspection records. A classification approach, a decision tree, uses the feature data to classify ships into several groups with different risk levels. Second, association analysis was performed to find the association relationships among deficiency items in these groups.

This study introduces decision trees at the pre-check stage for ship selection and performs deficiency association analysis. In addition, to observe the influence of ship selection on the deficiency association at the following check stage, we compare the association analysis of ship deficiencies with and without the intervention of a decision tree. The detainment status of ships is used as the target value for decision tree classification.

A. *Decision Tree in Machine Learning*

A decision tree is a tree-like decision support tool, which leads to many possible consequences, including chance outcomes, associated costs, and utility. It is also a supervised learning technique that can be used for both classification and regression problems. Generally, a decision tree consists of root node, internal node, branch, and leaf node, as shown in the example in Fig. 4. This paper adopts the CART algorithm (Classification and Regression Tree algorithm) to form the decision tree for PSC ship selection.

The decision tree for PSC ship selection is illustrated in Fig. 5. The ship features, including ship type, ship age, gross tonnage, ship flag, and classification society, are used as the nodes in the decision tree to grow branches. After going down to several levels of nodes, the decision tree can classify ships into groups of ships based on ship risks or detainment status. The decision tree can also predict the number of deficiencies found in PSC inspections with regression mode.

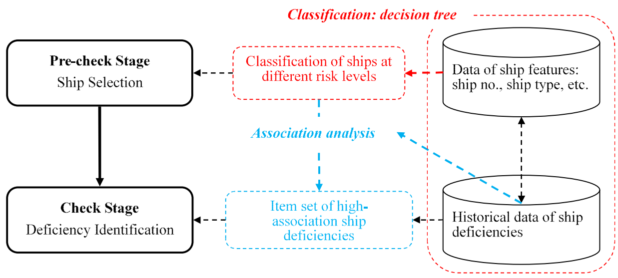

Fig. 3. Analysis framework

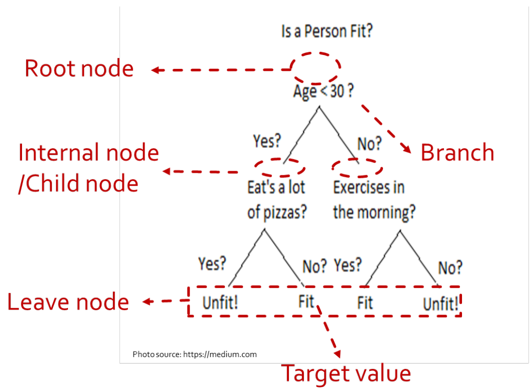

Fig. 4. A example of a decision tree [10]

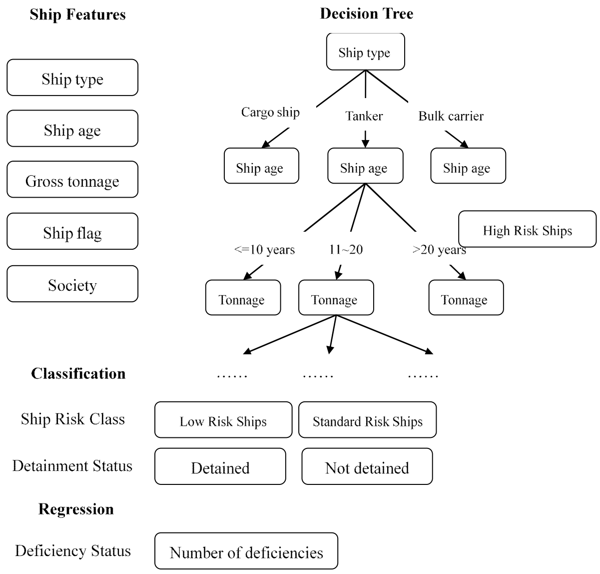

Fig. 5. Decision trees for ship inspections

B. *Associaiton Analysis in Machine Learning*

Association rule analysis is a rule-based machine learning method for identifying the relationships between variables in a database. This study adopts the Apriori algorithm for the association analysis of PSC deficiency items in this study. The Apriori algorithm proposed by Agrawal and Srikant [11] is a popular algorithm that has been adopted widely in different areas. This algorithm uses a bottom-up approach, where frequent subsets are extended one item at a time, and groups of candidates are tested by a given threshold value. This algorithm contains three primary parameters: "support", "confidence", and "lift".

The three parameters is based on conditional probability, as defined in Equation (1)~(3). X and Y represent two types of itemsets. $X \rightarrow Y$ means an association rule that Y (consequent) occurs when X (antecedent) occurs. In this analysis, X and Y represent two different deficiency itemsets. $X \rightarrow Y$ is the association rule that deficiency itemset Y occurs once deficiency itemset X occurs.

Support indicates how frequently the itemset of which the antecedent X and the consequent Y both occur in the whole data set. The support is the probability that deficiency itemset X and deficiency itemset Y occur simultaneously as defined in Equation (1). Confidence means the conditional probability that deficiency itemset Y occurs under the condition that deficiency itemset X occurs. Confidence can be interpreted as how frequent the rule has been identified as defined in Equation (2). A higher confidence value indicates the rule more likely occurs. Lift is defined as the ratio of the conditional probability of occurrence of the antecedent X and that of the consequent Y to the occurrence probability of the antecedent X as shown in (3). It determines if deficiency itemset X and deficiency itemset Y were dependent. If the lift value is greater than one, it means that one itemset is dependent on one another, and vice versa.

$$\text{Support } (X \rightarrow Y): P(X \cap Y) \tag{1}$$

$$\text{Confidence } (X \rightarrow Y): P(Y|X) \tag{2}$$

$$\text{Lift } (X \rightarrow Y): P(Y|X) / P(X) \tag{3}$$

Table II lists the parameter setting in the Apriori algorithm. This study sets "Minimal support" at 0.1 and 0.2, respectively, to compare two analysis scenarios. Meanwhile, the two scenarios have the same "Minimal confidence" setting at 0.9. In addition, "Minimal itemset" and "Maximal itemset" are used to determine the size of the itemsets in the association analysis and set at 1 and 3, respectively.

TABLE II. PARAMETER SETTING OF THE APRIORI ALGORITHM

Parameter	Description	Threshold setting
Minimal support	The threshold value of support for the frequent itemsets.	10%, 20%
Minimal confidence	The threshold value of confidence for the frequent itemsets.	90%
Lift	The measure to determine the dependence of itemsets	>1
Minimal itemset	The integer value for the minimal number of items in an itemset.	1
Maximal itemset	The integer value for the maximal number of items in an itemset.	3

III. CASE

The case of PSC inspections occurs in four main ports in Taiwan, namely, Keelung, Taichung, Kaohsiung and Hualien. This study analyzes the historical PSC inspection records in the ports collected from 2014 to 2021. The data source comes from the database of the Maritime Transport Network Portal (MTNet) of the Ministry of Transportation and Communications of Taiwan.

IV. RESULTS AND DISCUSSIONS

The analysis results based on the proposed method and the case are shown below. We discuss them in two aspects: the number of rules identified and the quality of identified rules, to see the relationship between ship selection and ship deficiency association and the influence of ship selection on ship deficiency association.

A. *Number of Rules Identified*

Table III shows the number of rules identified under the influence of decision trees at two parameter settings. "Yes" means that a decision tree is implemented for ship selection. "No" means that ship selection is performed in a traditional manner while a decision tree is not implemented. At the first parameter setting (Support=0.1, Conference=0.9), the intervention of a decision tree can help us identify more association rules (48 rules) among fewer inspection records (389) after ship selection. When we go to the second parameter setting (Support=0.2, Conference=0.9), the result goes in a similar fashion. However, the number of rules identified is reduced significantly because "Support" increases from 0.1 to 0.2. The results show that the intervention of a decision tree at pre-check stage identifies more association rules than a traditional approach for ship selection.

TABLE III. NUMBER OF RULES IDENTIFIED UNDER THE INFLUENCE OF DECISION TREES

Use of decision trees	Number of classified records	Number of rules identified	Number of consequent types	Rule duplicates
Parameter: Support=0.1, Conference=0.9				
Yes	389	48	4	15
No	786	21	2	15
Parameter: Support=0.2, Conference=0.9				
Yes	389	8	2	1
No	786	2	2	1

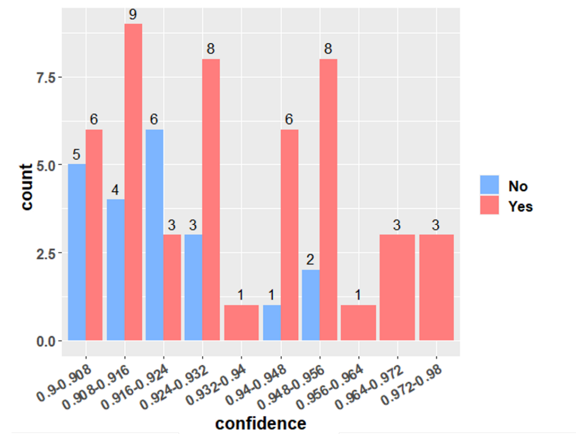

Fig. 7. "Confidence" dsitribution of the association rules with and without the intervention of a decision tree (Support=0.1, Conference=0.9).

B. *Quality of Identified Rules*

Fig. 6 and Fig. 7 illustrate "Support" and "Confidence" distribution of the association rules with and without the intervention of a decision tree. Fig. 8, and Fig. 9 show the same distribution but at the second parameter setting (Support=0.2, Conference=0.9). In most of "Support" intervals, either in Fig. 6 or in Fig. 8, the number of rules identified with the intervention of a decision tree is more than the one without the intervention of a decision tree. In most of "Confidence" intervals, either in Fig. 7 or in Fig. 9, the number of rules identified with the intervention of a decision tree is also more than the one without the intervention of a decision tree. Especially in the range of high values (>0.932), the number of rules identified with a decision tree is far more than those without a decision tree. The results show that the rules more likely occur when an antecedent occurs.

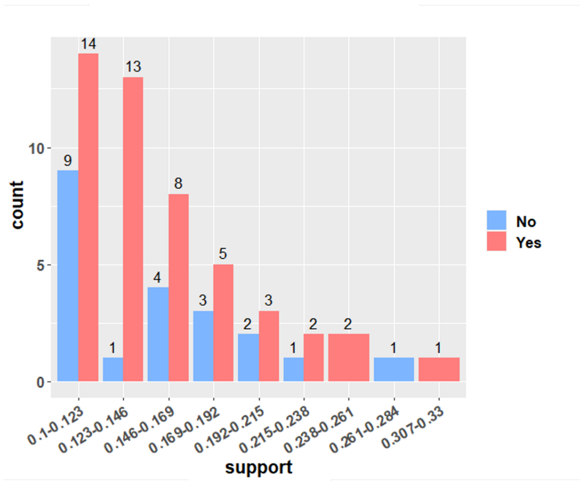

Fig. 8. "Support" dsitribution of the association rules with and without the intervention of a decision tree (Support=0.2, Conference=0.9).

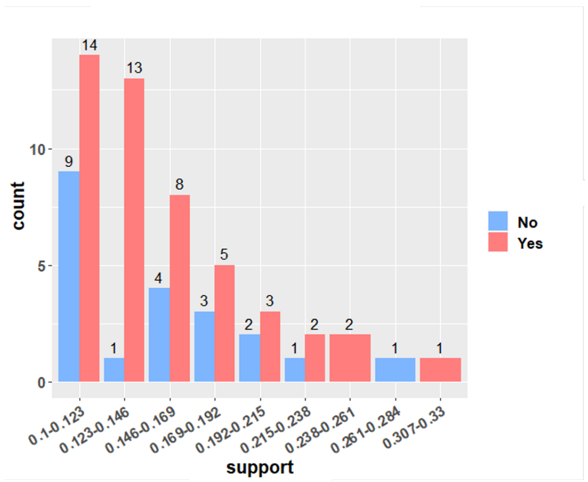

Fig. 6. "Support" dsitribution of the association rules with and without the intervention of a decision tree (Support=0.1, Conference=0.9).

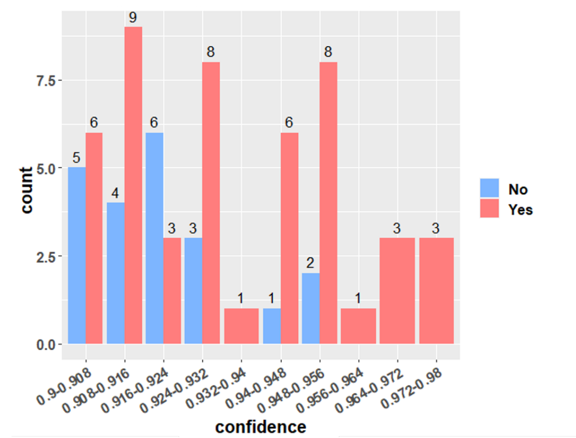

Fig. 9. "Confidence" dsitribution of the association rules with and without the intervention of a decision tree (Support=0.2, Conference=0.9).

V. Conclusions and Future Work

This paper applies machine learning techniques, decision trees and association analysis, to explore the relationship between ship selection at the pre-check stage and ship deficiency association at the check stage. The association rules of ship deficiencies with and without the intervention of a decision tree are compared to see if particular ship selection affects ship deficiency association at the following check stage.

The results show that the intervention of a decision tree for particular ship classification at the pre-check stage can find more association rules than a traditional approach for ship selection does. Also, the association rules identified have higher "Confidence", meaning they more likely occur when an antecedent occurs. The findings can provide PSC inspectors helpful guidelines for making PSC inspections effective and efficient.

There are still some insufficiencies existing in this study for future work. For example, this study uses only a decision tree and the Apriori algorithm for ship classification and association analysis. Different machine learning techniques can be considered for further investigation to find a better combination of ship classification and association analysis.

Acknowledgment

The study is supported by the Center of Excellence for Ocean Engineering (CEOE) of National Taiwan Ocean University (NTOU) and the Ministry of Science and Technology (MOST), Taiwan, R.O.C. under Grant MOST 110-2221-E-019-071.

References

[1] UNCTAD, "Review of Maritime Transport 2021," 2021.

[2] Tokyo MOU. (2019), "Tokyo MOU Deficiency Codes." Retrieved August 22, 2022, from http://www.tokyo-mou.org/publications/tokyo_mou_deficiency_codes.php

[3] Yang, Z., Z. Yang, and J. Yin, (2018). "Realising advanced risk-based port state control inspection using data-driven Bayesian networks." Transportation Research Part A: Policy and Practice, 110: 38-56.

[4] Wang, Y., F. Zhang, Z. Yang, and Z. Yang, (2021). "Incorporation of deficiency data into the analysis of the dependency and interdependency among the risk factors influencing port state control inspection." Reliability Engineering & System Safety, 206, 107277.

[5] Yang, Z., C. Wan, Z. Yang, and Q. Yu, (2021). "Using Bayesian network-based TOPSIS to aid dynamic port state control detention risk control decision." Reliability Engineering & System Safety, 213, 107784.

[6] Yan, R., S. Wang, and C. Peng, (2021). "An artificial intelligence model considering data imbalance for ship selection in port state control based on detention probabilities." Journal of Computational Science, 48, 101257.

[7] Gu, X., S. Hu, J. Wu and X. Chen (2015). "Correlation analysis of ship detention reasons based on FP-tree algorithm." Journal of Shanghai Maritime University 36(2): 60-64.

[8] Zhang, X.-n., N. Li and F.-e. Zhu (2016). "Research and analysis of ship defect data using Apriori algorithm." Ship Science and Technology 38(7A): 139-141.

[9] Chang, C.M., W.H. Chung, S.L. Kao, and C.C. Yuan, "Association Analysis on Inspection Deficiencies of Ships in Port State Control," International Forum on Shipping, Ports and Airports (IFSPA). Hong Kong, May 2017.

[10] Simple Decision Tree. (2018), " Decision Trees Explained Easily." Retrieved August 22, 2022, from https://chirag-sehra.medium.com/decision-trees-explained-easily-28f23241248

[11] Agrawal, R. and R. Srikant (1994). Fast algorithms for mining association rules. Proc. 20th int. conf. very large data bases, VLDB.