**This is the Pre-Published Version.**
Electronic version of an article published as Asia-Pacific Journal of Operational Research, v. 38, no. 03, 2140013, https://doi.org/10.1142/S0217595920400138, ©World Scientific Publishing Co. & Operational Research Society of Singapore. The journal url is https://www.worldscientific.com/worldscinet/apjor

# Development of two highly-efficient and innovative inspection schemes for PSC inspection

Ran Yan, Dan Zhuge*, Hans Wang
*Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*

## Abstract

Port state control (PSC) inspection contributes a lot to improving maritime safety and protecting marine environment. After selecting the ships coming to a port for inspection, one critical challenge faced by the PSC authorities is deciding what deficiency items should be inspected. To address this problem, two innovative and high-efficient PSC inspection schemes describing specific PSC inspection sequences are proposed for the inspectors' reference when time and resources are limited, especially when there are difficulties in estimating the possible deficiencies in advance. Both schemes take the occurrence probability, inspection cost, and ignoring loss of each deficiency item into account. More specifically, the first inspection scheme is based on the occurrence probabilities of the deficiency items in the whole data set, while the second scheme further considers the correlations among the deficiency items extracted by association rules. The results of numerical experiments show that the efficiency of the two proposed inspection schemes is 1.5 times higher than that of the currently used inspection scheme. In addition, the second inspection scheme performs better than the first inspection scheme, especially when inspecting ships with no less than 5 deficiency

* Corresponding author. Email: dan.zhuge@connect.polyu.hk

items and limited inspection resources.

*Keywords*: port state control (PSC) inspection, deficiency item, inspection scheme, association rule

## 1. Introduction

Marine casualties and incidents can bring about considerable losses to the shipping industry and society (Liu et al., 2016). It is reported by European Maritime Safety Agency (EMSA) that from 2011 to 2017, there were a total of 20,616 maritime casualties and incidents with 23,264 ships involved. Due to the accidents, 6,812 people were injured and 683 died (EMSA, 2018). As the consequences of maritime accidents are unbearable to ships, human beings, and cargos, marine safety is gaining increasing attention in recent years. Meanwhile, reducing environment pollutions related to international shipping is receiving wide notice in recent decades (IMO, 2011). Various international conventions and documents aiming to improve marine safety and protect marine environment are introduced by the International Maritime Organization (IMO), such as the International Convention for the Safety of Life at Sea (SOLAS), the International Convention for the Prevention of Pollution from Ships (MARPOL), the International Convention on Standards of Training, Certification and Watchkeeping for Seafarers (STCW), and the Maritime Labour Convention (MLC).

Memorandum of Understanding (MoU) on port state control (PSC), which was first signed by 14 European countries in 1982, is an international inspection regime aiming to guarantee the foreign ships coming to the port state to comply with the various international conventions. The main procedure of a typical PSC inspection is as follows. When a ship comes to a port, the port state authority needs to decide whether or not to inspect the ship based on some criteria. For example, Tokyo and Paris MoUs adopt the New Inspection Regime (NIR) to select the inspected ships while other MoUs may adopt different criteria. If the ship is decided to be inspected, one or more PSC officers (PSCOs) will get on board and conduct an initial inspection, which mainly aims to check the certificates and documents of the ship and crew as required by the relevant conventions. Usually, if the required certificates are valid and the PSCOs are satisfied with the general impression and overall observations of the ship, its equipment, and its crew, the inspection could be terminated. However, if there are clear grounds for believing that the condition of the ship or its equipment does not correspond substantially with the related conventions, or that the master and crew are not familiar with essential shipboard procedures, a more detailed inspection will be carried out (IMO, 2017; Tokyo MoU, 2017). During an inspection, the conditions that are not in compliance with the requirements of the relevant conventions are recorded as ship deficiencies. According to the documents of Tokyo MoU, there are 18 types of deficiency items related to the international maritime conventions, including but not limited to SOLAS, MARPOL, the International Convention on Tonnage Measurement of Ships, and the International Convention on Civil Liability for Oil Pollution Damage (CLC), as listed in Table 1(IMO, 2019; Tokyo MoU, 2018b). If serious deficiencies are found, the ship is likely to be detained until it can proceed to sea without presenting a danger to the ship or persons on board. After each inspection, an inspection report containing information on the inspected ship (e.g., ship IMO number, ship flag, ship operating company, ship type, ship recognized organization, etc.) and its inspection results (ship deficiencies and detention) will be generated and kept in the relevant PSC database.

Table 1: List of deficiency codes and items【加上参考文献，例如东京 MoU】

| Code | Deficiency item | Code | Deficiency item | Code | Deficiency item |
|---|---|---|---|---|---|
| D1 | Certificates and documentation | D7 | Fire safety | D13 | Propulsion and auxiliary machinery |
| D2 | Structural condition | D8 | Alarms | D14 | Pollution prevention |
| D3 | Water/Weathertight condition | D9 | Working and living conditions | D15 | ISM |
| D4 | Emergency system | D10 | Safety of navigation | D16 | ISPS |
| D5 | Radio communication | D11 | Life saving appliances | D18 | MLC |
| D6 | Cargo operations including equipment | D12 | Dangerous goods | D99 | Other |

To guarantee inspection efficiency, it is clearly stated that the main purpose of PSC is to prevent a ship proceeding to the sea if it is unsafe to the marine environment and to avoid unnecessary ship detention or delay (IMO, 2017). Thus, not every deficiency item of all the coming ships will be inspected. Instead, only some deficiency items of the high-risk ships will be inspected due to limited time and human resources. However, in practice, since there are rare instructions on the inspection sequence for the PSCOs, the inspected areas of a ship and to what extent they will be inspected are highly dependent on PSCOs' expert judgments. Nevertheless, personal judgments might be biased and inaccurate. First, if the possible deficiencies can be estimated in advance by the PSCOs, since some of them may lack experience, it is likely that the limited resources are allocated to inspect those less important or less frequently occurring deficiencies so that the relative serious deficiencies are ignored. As a result, inequality and inefficiency may be caused and the detected deficiencies of a single ship can be quite different when inspected by different PSCOs. Second, some ship deficiency items might be too veiled to be easily judged in advance even if the PSCOs are professional enough and familiar with the ship conditions. Therefore, if the inspection decisions are purely dependent on expert estimation, fatal deficiencies may be missed.

One possible way to improve the effectiveness of the inspection sequence is to develop inspection schemes that could identify as many deficiency items as possible after inspecting a certain number of deficiencies. In this study, we develop two instructive inspection schemes based on historical PSC inspection data and association rule learning method to draw a balance between the limited inspection resources and ship safety. First, we develop a new inspection scheme which takes the value of each deficiency item into account. The value of a deficiency item comprises the possibility of occurrence of the deficiency item, the cost of inspecting the deficiency item, and the loss of ignoring the deficiency item. To better illustrate the relationship between the deficiency items, we then develop another inspection scheme by considering the correlations among the deficiency items, which means that the probability of the occurrence of a deficiency when its related deficiencies are detected is higher than that when no related deficiencies are detected. The relevance between the deficiencies is identified by the association rules that are derived from the frequent itemset using Apriori algorithm (Agrawal and Srikant, 1994). Thus, the inspection decisions are dynamic since the possibility of detecting a certain deficiency item depends on the previously detected deficiencies. By selecting the deficiency item with the highest value in the remaining deficiencies, the PSCOs can make the subsequent inspection decisions more accurately and efficiently. The results of the numerical experiments show that both of the newly proposed inspection schemes can identify the deficiency items 1.5 times more efficiently than the currently used inspection scheme. Moreover, the second inspection scheme, which takes the relevance among the deficiency items into consideration, is better than the first inspection scheme when inspecting ships containing no less than 5 deficiency items while the inspection time and resources only allow 5 or 6 deficiency items to be inspected.

## 2. Literature review

### 2.1 Studies of PSC inspection

Research on maritime transportation attracts wide attention in recent years (Sun et al., 2015; Hu and Liu, 2016; Hu et al., 2017). Especially, there has been an increasing amount of literature on PSC inspection, some of which pays particular attention to developing ship selection criteria for PSC authorities and identifying the effects of PSC inspection. Before conducting a PSC inspection, the decision of what ships should be selected for inspection among all the coming ships is one of the critical issues faced by the port state, since limited time and resources need to be allocated to inspect the ships with worse conditions. One of the most popular methods used in developing the ship selection scheme is machine learning method. Xu et al. (2007a) proposed a Support Vector Machine (SVM) model to distinguish the ships between high risk and low risk. They then improved the performance of the risk assessment system by combining web mining technologies (Xu et al., 2007b). Later, Gao et al. (2008) proposed an ensemble model of K-nearest neighbor and SVM (KNN-SVM) that can identify the high-risk ships more accurately. Zhou and Sun (2010) proposed a ship target system which could be automatically optimized and was self-evolutional by using Generalized Additive Modelling (GAM) approach. Recently, Yang et al. (2018) developed a Bayesian network-based model to predict the detention probabilities of the coming ships, which could be used to help the port states to allocate inspection resources.

As outcomes of PSC inspection, the effects on maritime safety as well as on the inspected ships also attract much attention. Regarding the influence on maritime safety, Li and Zheng (2008) concluded that the PSC program was effective in raising maritime safety level after analyzing marine casualty statistics and PSC inspection database. Knapp et al. (2011) suggested that the monetary savings due to reducing maritime accidents brought by PSC inspections was from 70 to 190 thousand dollars. Based on Bayesian network, Hänninen and Kujala (2014) pointed out that the most influential indicators of ship accident involvement were the knowledge on ship type, PSC inspection type, and the number of structural conditions related deficiencies. Heij and Knapp (2018) also indicated that the PSC inspection outcomes had the predictive power in predicting the vessel accident involvement in the next year. In respect of the effects on the inspected ships, Cariou et al. (2007) pointed out after a PSC inspection, the length of inspection interval of the following two successive PSC inspections was reduced for some types of vessels, and the reported deficiencies during next inspection was reduced by 63% on average.

Apart from the aforementioned research areas, previous studies also reported that both ship factors and non-ship factors could have an impact on PSC inspection results. Ship factors mainly include ship age, ship type, and ship operating company, etc.

(Cariou et al., 2007; Cariou et al., 2009; Cariou et al., 2015; Yang et al., 2018; Tsou, 2018), while non-ship factors comprise the number and backgrounds of the PSCOs, the professional profile of the inspectors, and the area where the inspection is conducted (Knapp and Franses, 2007; Ravira and Piniella, 2016; Graziano et al., 2018).

**2.2 Applying association rule learning method to transport research**

Association rule learning algorithm is a rule-based learning method to discover the inherent and interesting rules between variables in large database. The concept of association rule was proposed by Agrawal et al. (1993). Popular algorithms used to mine association rules include but are not limited to Apriori algorithm, Eclat algorithm, and FP-growth algorithm (Zhang and Zhang, 2002). In the past decade, there has been an increasing number of studies that apply association rule learning method to road transport research. Among them, various studies applied association rule mining methods to analyze road transport casualties, such as Weng et al. (2016), Ait-Mlouk et al. (2017), Besharati and Tavakoli Kashani (2018), Yu et al. (2019), Kumar and Toshniwal (2016), and Zhang et al. (2018). Association rule mining methods are also employed to extract the transition patters in public transport, such research includes Zhao et al. (2018) and Zhao et al. (2019). The concept of association rule is also used in the field of rail transport, and the representative studies are Mirabadi and Sharifian (2010), Tang and Qin (2015), and Ghomi et al. (2016).

With regard to the field of air transport and maritime transport, there are much fewer studies. In air transport field, Sternberg et al. (2016) applied data indexing techniques together with association rules to identify the hidden patterns of flight delays in Brazil. In maritime transport field, contributory factors to both nonserious and serious shipping accidents were listed respectively by using association rules (Weng and Li, 2019). Correlations among the detention deficiencies and external factors were examined by applying association rule mining algorithms to the ship detention records in Tokyo MoU database (Tsou, 2018).

From the above-mentioned literature, it can be seen that on the one hand, despite a large number of studies on PSC inspection, to the best of our knowledge, the inspection sequence of the deficiency items has seldom been studied in the existing literature. On the other hand, although association rule learning method performs well in the field of road transport, there is rare attempt in applying this method to maritime transport research. Thus, in this study, two new PSC deficiency item inspection schemes are developed based on historical inspection records and association rule mining method. The hidden correlations among the deficiency items are extracted by the association rules and the new schemes can give instructions on ship inspection to the PSCOs.

## 3. Development of Inspection Scheme I for PSC inspection

### 3.1 Data set, indexes and definitions

In this study, we use the initial inspection records at the Port of Hong Kong from January 1, 2018 to June 30, 2018 with at least one deficiency item detected as the whole data set. Totally, there are $M = 297$ records and $N = 18$ types of deficiencies. The types and detected times of the deficiency items are shown in Table 2.

Table 2. Types and detected times of ship deficiency items

| Deficiency item in $I$ | Deficiency code | Deficiency type | Total detected times |
|---|---|---|---|
| $it_1$ | D1 | Certificates and documentation | 87 |
| $it_2$ | D2 | Structural condition | 17 |
| $it_3$ | D3 | Water/Weathertight condition | 97 |
| $it_4$ | D4 | Emergency system | 42 |
| $it_5$ | D5 | Radio communication | 46 |
| $it_6$ | D6 | Cargo operations including equipment | 8 |
| $it_7$ | D7 | Fire safety | 164 |
| $it_8$ | D8 | Alarms | 33 |
| $it_9$ | D9 | Working and living conditions | 115 |
| $it_{10}$ | D10 | Safety of navigation | 133 |
| $it_{11}$ | D11 | Life saving appliances | 120 |
| $it_{12}$ | D12 | Dangerous goods | 2 |
| $it_{13}$ | D13 | Propulsion and auxiliary machinery | 30 |
| $it_{14}$ | D14 | Pollution prevention | 89 |
| $it_{15}$ | D15 | ISM | 23 |
| $it_{16}$ | D16 | ISPS | 0 |
| $it_{17}$ | D18 | MLC | 0 |
| $it_{18}$ | D99 | Other | 12 |

The set of inspection records is denoted by $R = \{R_{1,\ldots,} R_M\}$. A certain inspection, which can also be called an experiment, is denoted by $R_m \in R$. The set of deficiency items is denoted by $I = \{it_1, \ldots, it_N\}$, which contains the total 18 types of deficiency items as required by Tokyo MoU. Regarding each record, we denote the deficiency set of a record $R_m$ with $N_m$ detected deficiency items as $D_{R_m} = \{D_{R_m,1}, \ldots, D_{R_m,N_m}\}$. Note that

$D_{R_m} \subseteq I$ and $D_{R_m} \neq \varnothing$, as we only take the inspections with deficiencies detected into consideration.

To develop Inspection Scheme I, we first introduce the concept of an itemset. An itemset is a specific collection of deficiencies. An itemset containing $i \in [1, N]$ deficiency items is called an $i$-itemset and is denoted by $I_i$. We then define the event of observing a particular itemset $I_i$ as $E(I_i)$, which means after inspecting a ship, it is found that the ship has all the deficiency items in the itemset $I_i$. We define $P(E(I_i))$ as the proportion of the $M$ records that have all the deficiencies in the itemset $I_i$, i.e., the probability of the occurrence of $E(I_i)$. Note that a record that has all the deficiency items in the itemset $I_i$ may also include deficiency items not in $I_i$.

We then define the probability of observing the event $E(I_i)$ as the Support of the itemset $I_i$, i.e., $Sup(I_i) = P(E(I_i))$, and thus $Sup(I_i) \in [0,1]$. It is obvious that the larger the Support value is, the more frequently this itemset occurs in the inspection records. In order to find out the itemsets that frequently appear in the $M$ records, we define the minimum threshold of Support as $\min Sup$. The itemsets with their Support values no less than $\min Sup$ are called large itemsets, i.e., if and only if $I_i^* \subseteq I$ is a large itemset, $Sup(I_i^*) \geq \min Sup$ (Tan et al., 2015).

### 3.2 Generation of large itemsets

Given the value of $\min Sup$, an algorithm called Apriori is adopted to generate the large itemsets (Agrawal and Srikant, 1994). This algorithm is used to discover useful and hidden relationships between data. We assume that the items in each deficiency set $D_{R_m}$ and by all itemsets are ordered in the alphabet. The Apriori algorithm is based on the following two properties of large itemsets (Agrawal and Srikant, 1994).

*Property I.* Any non-empty and strict subset of a large itemset is large.

*Property II*. Any superset of a non-large itemset cannot be large.

Now we describe the Apriori algorithm for generating the large itemsets (Agrawal and Srikant, 1994; Tan et al., 2015). We denote a large itemset containing $k$ items as a large $k-$itemset. Denote $L_k$ as the set of all large $k-$itemsets. Denote $C_k$ as the set of candidate large $k-$itemsets. Denote $Num(I_i)$ as the occurrence times of itemset $I_i$ in the record set $R$.

***Algorithm 1.*** Generate large itemsets $L_k$, $K=1,2,...,N$.

**Step 1:** $k=1$; //generate all large 1-itemsets
$L_k = \varnothing$;
for all $it_n \in I$
  $Sup(it_n)=0$;
  $Num(it_n)=0$;
    for all $R_m \in R$
      if $it_n$ is contained in $R_m$
        $Num(it_n)=Num(it_n)+1$;
      end if;
    end for;
    $Sup(it_n)=\dfrac{Num(it_n)}{M}$;
    If $Sup(it_n) \geq \min Sup$
      $L_1 = L_1 \cup \{it_n\}$;
    end if;
  end for.

**Step 2:** for ( $k=2$; $L_{k-1} \neq \varnothing$ and $k \leq N$; $k++$ ) //generate all large $k-$itemsets,

$C_k =$ generate_candidate ( $L_{k-1}$ ) //generate candidate large $k-$itemsets from the existing large $(k-1)-$itemsets by using Algorithm 2.

$L_k = \varnothing$ ;

for each $c \subset C_k$

$Num(c) = 0$ ;

$Sup(c) = 0$ ;

for all $R_m \in R$

if $c$ is contained in $R_m$

$Num(c) = Num(c) + 1$ ;

end if;

end for;

$Sup(c) = \frac{Num(c)}{M}$ ;

if $Sup(c) \geq \min Sup$

$L_k = L_k \cup \{c\}$

end if;

end for;

end for.

Denote a pair of large itemsets in $L_{k-1}$ by $I^{*\prime}_{k-1} = \{it'_1, it'_2, ..., it'_{k-2}, it'_{k-1}\}$ and $I^{*\prime\prime}_{k-1} = \{it''_1, it''_2, ..., it''_{k-2}, it''_{k-1}\}$. We use "<" to denote that the left-hand side item precedes the right-hand side item in the alphabet.

***Algorithm 2.*** *generate_candidate (* $L_{k-1}$ *).*

**Step 1:** Joining Step

$C_k = \varnothing$ ; //Based on Property I

for all pairs of itemsets in $L_{k-1}$, $k \geq 2$

if ( $k = 2$ or $it'_1 = it''_1$, $it'_2 = it''_2$, …, $it'_{k-2} = it''_{k-2}$ )

if $it'_{k-1} < it''_{k-1}$

$C_k = C_k \cup \{it'_1, it'_2, ..., it'_{k-2}, it'_{k-1}, it''_{k-1}\}$ ;

else

$C_k = C_k \cup \{it'_1, it'_2, ..., it'_{k-2}, it''_{k-1}, it'_{k-1}\}$ ;

end if;

end if;

end for.

**Step 2:** Pruning Step

for all itemsets $c \subset C_k$ //Based on Property II

for all subsets $s$ containing $(k-1)$ items of $c$

if $s \not\subset L_{k-1}$

```
delete c from C_k;
end if;
end for;
end for;
return C_k.
```

In Algorithm 1, the first step is to find all large 1-itemsets by scanning the whole record set $R$. By iteration, the set of all large $k-$itemsets ($k \geq 2$) $L_k$ is found based on the candidate large $k-$itemsets $C_k$ generated by the set of large $(k-1)-$itemsets $L_{k\text{-}1}$. The algorithm terminates until all the large itemsets are found. Algorithm 2 describes a way to find the set of candidate large $k-$itemsets $C_k$ based on $L_{k-1}$. A candidate large $k-$itemset is a combination of a pair of large itemsets which have the same first $(k-2)$ items and a different $(k-1)\text{th}$ item. After the combinations are formulated in Step 1, the subsets containing $(k-1)$ items of each combination are checked in Step 2. If any subset of a candidate itemset is not a large itemset, then this candidate itemset is deleted from the set of candidate large itemsets. After the Joining Step and Pruning Step, all the candidate large $k-$itemsets $\overline{I}_k$ can be found.

### 3.3 Description of Inspection Scheme I

Inspection Scheme I (short for Scheme I) is based on the large 1-itemsets. We set $\min Sup = 0.1$. After applying the Apriori algorithm in the input data set, large 1-itemsets can be generated as shown in Table 3.

Table 3. Large 1-intemsets

| Large 1-intemset | Support |
|---|---|
| {D7 - Fire safety} | 0.55 |
| {D10 - Safety of navigation} | 0.45 |
| {D11 - Life saving appliances} | 0.40 |
| {D9 - Working and living conditions} | 0.39 |
| {D3 - Water/Weathertight condition} | 0.33 |
| {D14 - Pollution prevention} | 0.30 |
| {D1 - Certificates and documentation} | 0.29 |
| {D5 - Radio communication} | 0.15 |
| {D4 - Emergency system} | 0.14 |
| {D8 – Alarms} | 0.11 |
| {D13 - Propulsion and auxiliary machinery} | 0.10 |

To develop the inspection scheme, we take the probability of a deficiency item occurs, the cost of inspecting the deficiency item and the loss of ignoring the deficiency item into consideration. The possibility of the occurrence of $it_i$ is denoted by $P_{it_i}$. Denote the inspection cost of deficiency item $it_i$ by $C_{it_i}$, $C_{it_i} > 0$. If an existing deficiency item is not identified, the loss is huge and denoted by $L_{it_i}$, $L_{it_i} > C_{it_i}$. Note that ideally, the cost and loss values of a deficiency item are not only at financial level, but also reflect the effects on marine safety and environment, time delay, allocated inspection resources, etc. Denote the value of inspecting deficiency item $it_i$ as $V_{it_i}$, and we have $V_{it_i} = P_{it_i} \times L_{it_i} - C_{it_i}$. The larger the value of a deficiency item, the more worthy of being inspected.

Due to the lack of data and the sake of simplicity, we assume that the value of $L_{it_i}$ and the value of $C_{it_i}$ are identical to each deficiency item, respectively. It should be noted that it is reasonable to assume the ignoring loss and inspection cost are the same for each deficiency item respectively for two reasons. First, 【通常不要提具体人，

以免给该人带来不必要的麻烦（特别是政府官员）；如果需要提具体人，需要征得他同意】as suggested by a senior PSCO in a port under the Tokyo MoU, all the deficiency items are related to important international maritime regulations and conventions, and thus they can be viewed of the same level of importance and their loss values can be viewed as identical. Second, as suggested by the PSCOs we interviewed, they usually walk around the ship to observe its conditions as well as to inspect the deficiency items, and thus the cost of inspecting a deficiency item can also be roughly treated as the same. As positive $V_{it_i}$ indicates that the deficiency item is worthy of being inspected, we need to compare $P_{it_i} \times L_{it_i} - C_{it_i}$ and 0, i.e., we need to know the value of $C_{it_i} / L_{it_i}$. To determine the inspection sequence of the deficiency items, we also need to compare the values of $P_{it_i}$ of all the deficiency items with positive $V_{it_i}$. As estimating the value of $C_{it_i} / L_{it_i}$ is quite complicated and there are few references, for the sake of simplicity, we set $C_{it_i} / L_{it_i}$ equal to the PSC inspection rate at the Port of Hong Kong during the time period from 2015 to 2017. According to the annual reports of Tokyo MoU in 2015, 2016, and 2017, there were a total of 10,239 ships visiting the Port of Hong Kong and 1,324 of them were inspected during this period (Tokyo MoU, 2016; Tokyo MoU, 2017; Tokyo MoU, 2018a). Therefore, we set $C_{it_i} / L_{it_i} = 0.1293$. By converting $V_{it_i} = P_{it_i} \times L_{it_i} - C_{it_i}$ to $\frac{V_{it_i}}{L_{it_i}} = P_{it_i} - \frac{C_{it_i}}{L_{it_i}}$, we can view $L_{it_i}$ as the unit of $V_{it_i}$, and the value of a deficiency item equals the difference between $P_{it_i}$ and $\frac{C_{it_i}}{L_{it_i}}$. The value of each deficiency item is listed in Table 4.

Table 4. Values of all deficiency items in large1-itemsets

| Large 1-intemset | Value (unit: $/L_{it_i}$ ) |
|---|---|
| {D7 - Fire safety} | 0.4207 |
| {D10 - Safety of navigation} | 0.3207 |
| {D11 - Life saving appliances} | 0.2707 |
| {D9 - Working and living conditions} | 0.2607 |
| {D3 - Water/Weathertight condition} | 0.2007 |
| {D14 - Pollution prevention} | 0.1707 |
| {D1 - Certificates and documentation} | 0.1607 |
| {D5 - Radio communication} | 0.0207 |
| {D4 - Emergency system} | 0.0107 |
| {D8 – Alarms} | −0.0193 |
| {D13 - Propulsion and auxiliary machinery} | −0.0293 |

Based on $V_{it_i}$ of each deficiency item, we first propose Inspection Scheme I for the PSCO's reference when conducting PSC inspection. The inspection scheme lies on two basic assumptions:

(a) The cost of inspecting a deficiency is identical no matter if this deficiency item exists.

(b) If an existing deficiency item exists and it is inspected, it can be detected.

***Inspection Scheme I***

The values of the deficiency items in Inspection Scheme I are based on the Support values of the items in large 1-itemsets, and the value of each deficiency item $V_{it_i} = P_{it_i} \times L_{it_i} - C_{it_i}$ is fixed. The general inspection sequence is as follows: staring from inspecting D1- certificates and documentation as suggested by Tokyo MoU (2018b), all the remaining deficiency items with positive $V_{it_i}$ will be inspected from larger $V_{it_i}$ to smaller $V_{it_i}$. Totally, 9 deficiencies, namely D1, D7, D10, D11, D9, D3, D14, D5, and D4, are worthy of being inspected. The inspection sequence is shown in Figure 1.

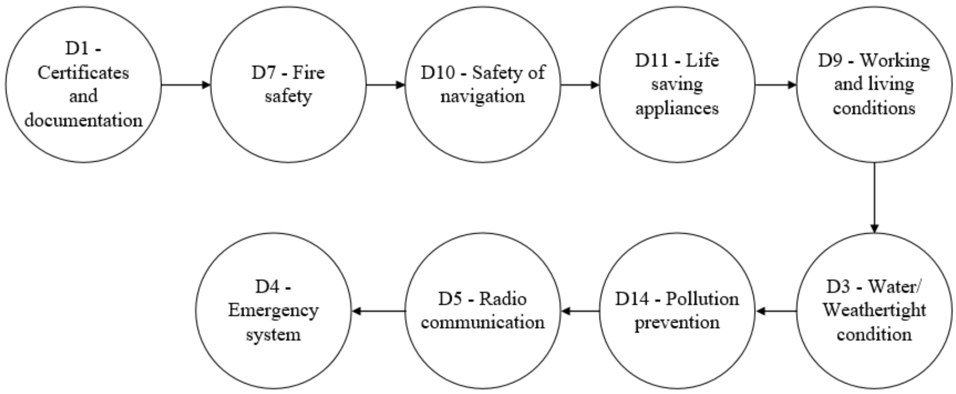


Figure 1. Inspection sequence of Inspection Scheme I

## 4. Development of Inspection Scheme II for PSC inspection

### 4.1 Indexes and definitions

The data set used to develop Inspection Scheme II (short for scheme II) is the same as that is used for developing scheme I. In scheme II, we further consider the relevance among the deficiency items to better illustrate their relationships, i.e., the deficiency items are dependent and the probability of the occurrence of each deficiency item is influenced by other deficiency items. The dependency is presented by the association rules generated from the large 2-itemsets and large 3-itemsets, which are shown in Table 5 and Table 6. Note that no Support of the itemsets containing 4 items is greater than or equal to $\min Sup$, and hence the biggest large itemsets only contain 3 items. One main reason for generating the association rules from the large itemsets is that only the rules with occurrence beyond the minimum support threshold are statistically significant and worth being considered (Agrawal, 1993).

Table 5. Large 2-intemsets

| Large 2-intemset | Support | Large 2-intemset | Support | Large 2-intemset | Support |
|---|---|---|---|---|---|
| {D7, D10} | 0.28 | {D1, D7} | 0.17 | {D3, D9} | 0.15 |
| {D7, D11} | 0.24 | {D3, D10} | 0.17 | {D3, D14} | 0.13 |
| {D7, D9} | 0.23 | {D9, D11} | 0.17 | {D1, D14} | 0.11 |
| {D10, D11} | 0.21 | {D1, D10} | 0.17 | {D4, D11} | 0.10 |
| {D7, D14} | 0.19 | {D1, D11} | 0.17 | {D9, D14} | 0.10 |
| {D9, D10} | 0.19 | {D3, D11} | 0.16 | {D1, D9} | 0.10 |
| {D3, D7} | 0.18 | {D11, D14} | 0.16 | {D1, D3} | 0.10 |
| {D10, D14} | 0.18 | | | | |

Table 6. Large 3-intemsets

| Large 3-intemset | Support | Large 3-intemset | Support | Large 3-intemset | Support |
|---|---|---|---|---|---|
| {D7, D10, D11} | 0.14 | {D10, D11, D14} | 0.12 | {D1, D7, D11} | 0.10 |
| {D7, D9, D10} | 0.13 | {D3, D10, D11} | 0.11 | {D1, D10, D14} | 0.10 |
| {D7, D10, D14} | 0.13 | {D7, D9, D11} | 0.11 | {D3, D7, D11} | 0.10 |
| {D1, D7, D10} | 0.12 | {D1, D10, D11} | 0.11 | {D3, D7, D10} | 0.10 |
| {D7, D11, D14} | 0.12 | | | | |

A rule is generated by dividing a large $i$- itemset $I_i^*$ ($i \geq 2$) into two mutually exclusive and non-empty deficiency itemsets, $I_j$ and $I_k$, with $I_j \cup I_k = I_i^*$. Note that both $I_j$ and $I_k$ are large itemsets. To determine whether the rule from $I_j$ to $I_k$ (denoted by $I_j \rightarrow I_k$) is an association rule, we further introduce two indexes: Confidence and Lift (McNicholas et al., 2008). The Confidence of $I_j \rightarrow I_k$ (denoted by $Conf(I_j \rightarrow I_k)$) can be interpreted as the conditional probability of the event of $E(I_k)$ under the condition that the event of $E(I_j)$ has occurred, i.e., $Conf(I_j \rightarrow I_k) = \frac{P(E(I_k) \cap E(I_j))}{P(E(I_j))} = P(E(I_k) \mid E(I_j))$. $Conf(I_j \rightarrow I_k) \in [0,1]$. The larger value the

Confidence is, the more likely the deficiency items in $I_k$ will be detected after the deficiency items in $I_j$ are detected. Lift of $I_j \rightarrow I_k$ (denoted by $Lift(I_j \rightarrow I_k)$) is the measure of the influence of the occurrence of event $E(I_j)$ on the occurrence of event $E(I_k)$. $Lift(I_j \rightarrow I_k) = \frac{P(E(I_k) \cap E(I_j))}{P(E(I_j)) \times P(E(I_k))} = \frac{P(E(I_k) \mid E(I_j))}{P(E(I_k))}$ and $Lift(I_j \rightarrow I_k) \in [0, +\infty)$. It represents the ratio of the probability of the occurrence of event $E(I_k)$ under the condition that event $E(I_j)$ occurs and the probability that event $E(I_k)$ occurs unconditionally in the record set. If $Lift(I_j \rightarrow I_k) = 1$, i.e., $P(E(I_j), E(I_k)) = P(E(I_j)) \times P(E(I_k))$, $E(I_j)$ and $E(I_k)$ are independent. If $Lift(I_j \rightarrow I_k) \in [0,1)$, the occurrence of $E(I_j)$ reduces the probability that $E(I_k)$ occurs. If $Lift(I_j \rightarrow I_k) \in (1, +\infty)$, the occurrence of $E(I_j)$ increases the probability of the occurrence of $E(I_k)$. After introducing the indexes, we can now define an association rule:

**Definition 1**: Suppose that there is a large $i$- itemset $I_i^*$ ($i \geq 2$) and its two mutually exclusive and non-empty deficiency itemsets $I_j$ and $I_k$ such that $I_j \cup I_k = I_i^*$. Given the minimum threshold of Confidence, $\min Conf$, and the minimum threshold of Lift, $\min Lift$, the rule $I_j \rightarrow I_k$ is an association rule if and only if $Conf(I_j \rightarrow I_k) \geq \min Conf$ and $Lift(I_j \rightarrow I_k) \geq \min Lift$.

The implication of this association rule is that during the PSC inspection if the deficiency items in $I_j$ are detected, there is a high probability that this ship also has deficiency items in $I_k$. The left-hand side of an association rule is called antecedent and the right-hand side is called consequent (Agrawal et al., 1993).

### 4.2 Generation of association rules

After all the large $k-$itemsets are obtained and the values of $\min Conf$ and $\min Lift$ are given, we can then generate the corresponding association rules. Similar to Property I and II, we can have the following Property III (Agrawal et al., 1993):

*Property III*. Partition a large $i-$itemset $I_i^*$ ($i \geq 2$) into two itemsets $I_j$ and $I_k$. The rule from $I_j$ to $I_k$ is denoted by $I_j \rightarrow I_k$. $Conf(I_j \rightarrow I_k) < \min Conf$. For any non-empty and strict subset of $I_j$, denoted by $\underline{I}_j$, and the superset of $I_k$, denoted by $\overline{I}_k = I_i - \underline{I}_j$, the rule from $\underline{I}_j$ to $\overline{I}_k$ is called a sub-rule of $I_j \rightarrow I_k$, and $Conf(\underline{I}_j \rightarrow \overline{I}_k) < \min Conf$ (Agrawal and Srikant, 1994).

*Proof:*

We first denote the events of observing $I_j$ and $I_k$ as $E(I_j)$ and $E(I_k)$, respectively, and the events of observing $\underline{I}_j$ and $\overline{I}_k$ as $E(\underline{I}_j)$ and $E(\overline{I}_k)$, respectively. The Confidence of $I_j \rightarrow I_k$ can be presented as $Conf(I_j \rightarrow I_k) = \frac{P(E(I_j) \cap E(I_k))}{P(E(I_j))} = \frac{P(E(I_i))}{P(E(I_j))}$, and the Confidence of $\underline{I}_j \rightarrow \overline{I}_k$ can be presented as $Conf(\underline{I}_j \rightarrow \overline{I}_k) = \frac{P(E(\underline{I}_j) \cap E(\overline{I}_k))}{P(E(\underline{I}_j))} = \frac{P(E(I_i))}{P(E(\underline{I}_j))}$. As $\underline{I}_j \subset I_j$, we have $P(E(\underline{I}_j)) \geq P(E(I_j))$ and $Conf(\underline{I}_j \rightarrow \overline{I}_k) \leq Conf(I_j \rightarrow I_k) < \min Conf$. Therefore, we can conclude that $Conf(\underline{I}_j \rightarrow \overline{I}_k) < \min Conf$. ■

It can be seen from the above property that the sub-rules of a rule with its

Confidence less than $\min Conf$ cannot be association rules. We can use this property to simplify the process by ignoring the sub-rules of the rules with Confidence less than $\min Conf$.

We now describe the process of generating association rules of all large $k-$ itemsets in $L_k$ (Agrawal and Srikant, 1994). A consequent containing $m$ ($1 \le m < k$) items is denoted by $h_m$ and the set of all $h_m$ is denoted by $H_m$. we use a recursive algorithm called Association rules generation involving Ap-AssRule, which can first generate the rules with their Confidence larger than or equal to $\min Conf$ and then generate the set of association rules by deleting rules with Lift less than $\min Lift$.

***Algorithm 3.*** *Association rules generation.*

| | |
|---|---|
| **Step 1:**<br>Generating_Rules | $Rules = \varnothing$;<br>for each large $k-$itemset $I_k^*$, $k \ge 2$<br>    Ap-AssRule ( $I_k^*$); //recursively call the function<br>end for; |
| **Step 2:**<br>Pruning_Rules<br>( $Rules$ ) | for each $rule$ in $Rules$<br>  Calculate $Lift(rule)$;<br>  if $Lift(rule) < \min Lift$<br>    Delete $rule$ from $Rules$; //Filter rules by Lift<br>  end if;<br>end for;<br>Return $Rules$. |

***Algorithm 4.*** *Ap-AssRule* ( $I_k^*$).

$m = 1$;
$H_m = \{h_m \mid h_m \subset I_k^*\}$; //generate all consequents containing one item
while ( $k \ge m+1$ )
  for each $h_m \subset H_m$
    //Divide $I_k^*$ into two parts with $h_m$ as the consequent

$rule = I_k^* - h_m \to h_m$ ;

Calculate $Conf(rule)$ ;

if ( $Conf(rule) \geq \min Conf$ )

$Rules \cup rule$ ; //Filter rules by Confidence

else

Delete $h_m$ from $H_m$ ; //Based on Property III

end if;

end for;

$m = m + 1$ ;

$H_m$ = *generate_candidate* ( $H_{m-1}$ ); //generate $H_m$ from $H_{m-1}$ by calling Algorithm 2.

end while; //loop until $k < m + 1$

### 4.3 Description of Inspection Scheme II

Inspection Scheme II is based on the association rules of the deficiency items. We set $\min Conf = 0.6$ and $\min Lift = 1.2$ as the thresholds and the generated association rules are presented in Table 7. Except for Rule NO. 4, which is generated by a large 2-itemset, all the other association rules are generated by the large 3-itemsets. As the Confidence value is used to determine the strongness of an association rule, and the Lift value is used to verify if it is meaningful, the association rules with higher Confidence values are of higher priority to be adopted.

Table 7. Association rules of the deficiency items

| Rule NO. | Left-hand side | Right-hand side | Confidence | Lift | Rule NO. | Left-hand side | Right-hand side | Confidence | Lift |
|---|---|---|---|---|---|---|---|---|---|
| 1 | D1, D14 | D10 | 0.91 | 2.03 | 12 | D7, D14 | D10 | 0.66 | 1.49 |
| 2 | D11, D14 | D10 | 0.77 | 1.72 | 13 | D7, D14 | D11 | 0.65 | 1.61 |
| 3 | D11, D14 | D7 | 0.77 | 1.40 | 14 | D1, D10 | D11 | 0.64 | 1.58 |
| 4 | D4 | D11 | 0.74 | 1.83 | 15 | D1, D11 | D10 | 0.64 | 1.43 |
| 5 | D1, D10 | D7 | 0.74 | 1.34 | 16 | D3, D10 | D11 | 0.63 | 1.55 |
| 6 | D1, D7 | D10 | 0.73 | 1.62 | 17 | D10, D11 | D14 | 0.61 | 2.02 |
| 7 | D10, D11 | D7 | 0.72 | 1.30 | 18 | D1, D7 | D11 | 0.61 | 1.50 |
| 8 | D10, D14 | D7 | 0.72 | 1.28 | 19 | D7, D11 | D10 | 0.61 | 1.35 |
| 9 | D10, D14 | D11 | 0.70 | 1.73 | 20 | D1, D10 | D14 | 0.60 | 2.00 |
| 10 | D9, D10 | D7 | 0.70 | 1.26 | 21 | D3, D7 | D10 | 0.60 | 1.35 |
| 11 | D3, D11 | D10 | 0.68 | 1.52 | | | | | |

***Inspection scheme II***

In Inspection Scheme II, we also consider the value of each deficiency item $V_{it_i}$, which contains its occurrence probability, the cost of inspecting it and the loss of ignoring it the same as that of Inspection Scheme I. Regarding $V_{it_i}$ of all the deficiency items, 9 deficiency items, namely D1, D7, D10, D11, D9, D3, D14, D5, and D4, are worthy of being inspected in total. The differences between Inspection Schemes I and II are that in Inspection Scheme II, the values for some deficiency items are dynamic after some certain deficiencies are detected according to the association rules as indicated in Table 5, while the values for all deficiency items in Inspection Scheme I are static. There are four types of deficiency items on the right-hand side among all the correlation rules: D7, D10, D11, and D14, which means that only $P_{D_7}$, $P_{D10}$, $P_{D_{11}}$, and $P_{D_{14}}$ are dynamic and related to the detected deficiencies, while the probabilities of other deficiency items are static. Starting from inspecting D1 in the first inspection round, the probabilities of the uninspected deficiency items are updated based on the association rules. Then, the deficiency item with the highest probability to occur is selected to be inspected in the next round. Note that as the minimum Confidence value of the association rules is 0.6, if items on the left-hand side of an association rule are detected, the value of the deficiency item on the right-hand side is the largest (larger than 0.4207) among all the uninspected items and should be inspected in the next round. The first 6 rounds of inspection are presented in Figure 2. In this figure, “Y” means the deficiency is detected while “N” means the deficiency does not exist. The red nodes represent that the probabilities of these deficiency items are updated according to the association rules. The brackets contain the left-hand side of the used association rules. It should be noted that the first 5 rounds inspection have already updated all the updatable deficiency probabilities among the uninspected deficiencies (recall that only $P_{D_7}$, $P_{D10}$, $P_{D_{11}}$, and $P_{D_{14}}$ can be updated). From the 6th round of inspection, the probability of each deficiency item is equal to its Support value.

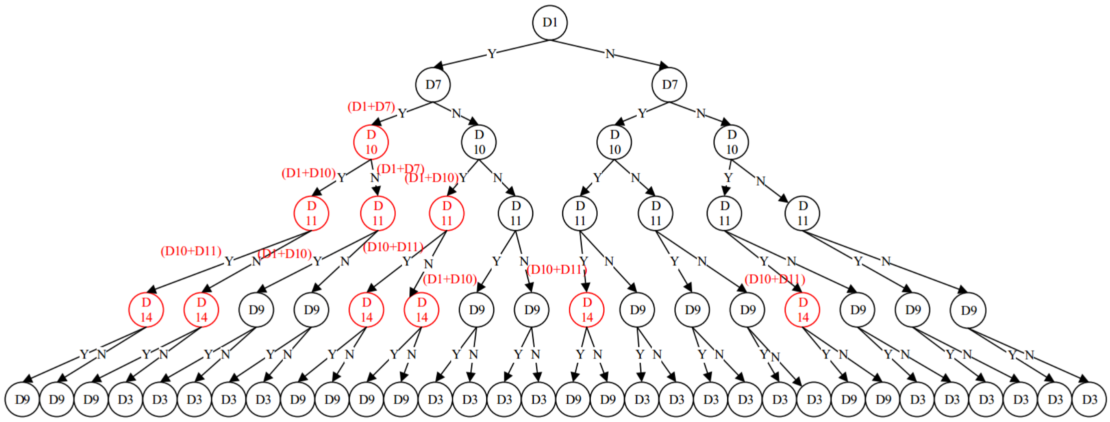


Figure 2. Inspection sequence of Inspection Scheme II

## 5. Numerical experiments

### 5.1 Comparison of the current and new inspection schemes

To the best of our knowledge, the inspection sequence of the deficiency items is not clearly stated in any documents, and thus we assume that the current inspection scheme requires all the deficiency items to be inspected in theory. The ratio of the total detected deficiency items and the number of inspection rounds can be regarded as the inspection efficiency, as it indicates the probability of detecting a deficiency item after inspecting each one of them. By using the testing data set $\Psi_1$ which contains the detected deficiency items in initial PSC inspection with at least one deficiency item detected from 1 July 2018 to 31 August 2018 at the Port of Hong Kong, we compare the inspection efficiency of the current inspection scheme and the two new inspection schemes. Totally, there are 138 inspection records and 519 detected deficiency items in the testing data set $\Psi_1$. The inspection efficiency of the three schemes is shown in Table 8. The current inspection scheme requires all the deficiency items listed in Table

1 to be inspected, and hence there are totally $18\times138=2484$ inspection rounds. Schemes I and II both require 9 out of the total 18 deficiency items to be inspected, so there are $9\times138=1242$ rounds of inspection.

Table 8. Inspection efficiency of the current and new inspection schemes

| Inspection scheme | Inspection rounds | Total detected deficiency items | Inspection efficiency |
|---|---|---|---|
| Current | 2,484 | 519 | 20.89% |
| Scheme I | 1,242 | 435 | 35.02% |
| Scheme II | 1,242 | 435 | 35.02% |

It can be seen from Table 8 that each of the two new inspection schemes can identify 83.82% of the real detected deficiency items by inspecting only 50% of all the possible deficiency items. It also shows that the inspection efficiency of the two new inspection schemes is the same and is almost 1.5 times higher than that of the currently used inspection scheme. Therefore, we can conclude that the new schemes have higher efficiency and perform better than the currently used one, which means that if one of the new inspection schemes is adopted, more deficiency items can be detected after conducting a certain number of inspection rounds.

**5.2 Comparison of the two new inspection schemes**

As both of the new inspection schemes contain 9 rounds of inspection, we also compare their performance. Data set $\Psi_1$ is used in the first comparison. The total number of identified deficiencies after each inspection round of the two schemes is shown the Table 9.

Table 9. Identified deficiencies of the two new inspection schemes in the 1st comparison

| Inspection Scheme | 1st round | 2nd round | 3rd round | 4th round | 5th round | 6th round | 7th round | 8th round | 9th round |
|---|---|---|---|---|---|---|---|---|---|
| I | 42 | 124 | 186 | 242 | 298 | 333 | 378 | 401 | 435 |
| II | 42 | 124 | 186 | 242 | 299 | 337 | 378 | 401 | 435 |

Table 9 indicates that the first 4 rounds of inspection are the same in both schemes, and after the 5th and 6th inspection rounds, the performance of Inspection Scheme II is a little better than that of Inspection Scheme I, with 1 and 4 more deficiencies detected. The differences between the two schemes depend on the inspected item in the 5th inspection round: if the detected deficiency items in the first 4 rounds contain D1 and D10, or D10 and D11, D14 will be inspected in Scheme II, while D9 will be inspected in Scheme I no matter what deficiencies are detected; Otherwise the inspection sequences are the same in the two schemes.

To further compare their performance, we do the second comparison by selecting the deficiency items in the inspection records with larger than or equal to 5 deficiency items detected from 1 September 2018 to 31 December 2018 at the Port of Hong Kong as testing data set $\Psi_2$. There are 53 records in total, with 380 deficiency items detected. Then, we use the two proposed inspection schemes to conduct PSC inspection and the total number of identified deficiency items after each inspection round is shown in Table 10.

Table 10. Identified deficiencies of the two new inspection schemes in the 2nd comparison

| Inspection Scheme | 1st round | 2nd round | 3rd round | 4th round | 5th round | 6th round | 7th round | 8th round | 9th round |
|---|---|---|---|---|---|---|---|---|---|
| I | 40 | 89 | 132 | 163 | 192 | 205 | 246 | 258 | 278 |
| II | 40 | 89 | 132 | 163 | 204 | 228 | 246 | 258 | 278 |

The above table shows that when inspecting the ships with deficiency items no less than 5, these two inspection schemes can identify 73.16% deficiencies after inspecting 50% of all the possible deficiency items. Besides, it is indicated that Scheme II outperforms Scheme I, with 12 (9.41%) and 23 (11.22%) more deficiency items detected after finishing the 5th and 6th rounds of inspection. Thus, we can conclude that although Inspection Scheme I is intuitive and easy to understand, Inspection Scheme II works better than Inspection Scheme I, especially when inspecting ships with no less than 5 deficiency items.

## 6. Discussion

In this study, we used 297 PSC inspection records with no less than 2 deficiency items detected at the Port of Hong Kong as the training set to calibrate the inspection scheme models. Although some interesting insights are generated, such as the large itemsets and valid association rules, it is worth mentioning that if more inspection data can be incorporated, for example, inspection records of 12 to 24 months, we can find more comprehensive and accurate correlations among the deficiency items. Meanwhile, these two innovative inspection schemes may also cause some possible consequences. First, the ship operators may take some measures before the inspection to prevent their ships from being identified the related deficiencies and even detained if they are aware of the inspection schemes. Second, only some of the deficiency items are included in these two inspection schemes while other deficiency items are omitted. Regarding the first consequence, it is believed that the goal of PSC inspection is to guarantee the ships to comply with the various international conventions by conducting inspection as well as its deterrence. Thus, if the ship operators are willing to spare their efforts to keep the ships in satisfactory condition and conforming to the regulations, we can say that the goal of PSC inspection has been achieved. Regarding the second consequence, both the relevant documents on PSC and the PSCOs we interviewed suggest that in practice, after checking the documentation, the PSCOs will walk around the ship to observe its overall condition. If deficiencies are detected, they will pay more attention to the corresponding deficiency categories and conduct a more comprehensive inspection. If there are no clear findings, they may let the ship go without further inspection. Under this situation, both of the inspection schemes are designed to give some instructions and reference to the PSCOs when the time and inspection resources are limited and the deficiencies are not that obvious instead of interfering with their own expert judgment.

## 7. Conclusion

PSC inspection is viewed as an effective way to contribute to the enhancement of maritime safety and security, and the prevention of marine pollution. Due to the limited time and human resources, not every deficiency item listed by the MoUs can be inspected. Thus, it is worthy of developing inspection schemes that can give instructions to the PSCOs in order to improve inspection efficiency. The goal of the inspection schemes is to identify as many deficiency items as possible after conducting certain rounds of inspections. In this paper, two inspection schemes based on the inspection value of each deficiency item are proposed. The inspection value of a deficiency item comprises its probability of occurrence, the cost of inspecting it and the loss of ignoring it. To be more specific, the probabilities in Inspection Scheme I are the occurrence probabilities of the deficiency items in the whole data set and are static, while the probabilities in Inspection Scheme II also depend on the interdependencies among the deficiency items and are dynamic. As the data and references are limited, we approximate the values of the cost and loss by setting the ratio of cost and loss equal to the PSC inspection rate at the Port of the Hong Kong from 2015 to 2017.

Both of the inspection schemes suggest that 9 deficiency items with positive values, i.e. D1, D7, D10, D11, D9, D3, D14, D5, and D4, should be inspected. The inspection sequence of Scheme I is fixed, while in Scheme II, 4 types of deficiency items occur on the right-hand side of the generated association rules: D7, D10, D11, and D14, which means that their probabilities (i.e., values) and inspection sequence are dynamic. Thus, the inspection sequence of Scheme II is dynamic and is related to the detected deficiencies. The detailed inspection sequences of the schemes are also provided.

To the best of our knowledge, this is the first research on developing inspection schemes containing detailed inspection sequence for PSC inspection. Numerical experiments show that both the newly proposed inspection schemes are about 1.5 times more efficient when used to identify the deficiency items compared with the currently used inspection scheme. Further, the performance of Inspection Scheme II is better than Inspection Scheme I, with 9.41% and 11.22% more deficiency items detected after finishing 5th and 6th rounds of inspection when inspecting ships with no less than 5 deficiency items. In the future research, the value of the inspection cost and ignoring loss of each deficiency item can be estimated more accurately to further improve the performance of the two schemes.

**Acknowledgements**

This study is supported by the National Natural Science Foundation of China (Grant Numbers 71701178, 71771050, 71831008).

**Reference**

Agrawal, R and R Srikant (1994). Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference of Very Large Data Bases*, 1215, 487-499.

Agrawal, R, T Imieliński and A Swami (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2), 207-216.

Ait-Mlouk, A, F Gharnati and T Agouti (2017). An improved approach for association rule mining using a multi-criteria decision support system: a case study in road safety. *European Transport Research Review*, 9(3), 40-53.

Besharati, M M and A Tavakoli Kashani (2018). Which set of factors contribute to increase the likelihood of pedestrian fatality in road crashes? *International Journal of Injury Control and Safety Promotion*, 25(3), 247-256.

Cariou, P, M Q Mejia Jr and F C Wolff (2007). An econometric analysis of deficiencies noted in port state control inspections. *Maritime Policy & Management,* 34(3), 243-258.

Cariou, P, M Q Mejia and F C Wolff (2009). Evidence on target factors used for port state control inspections. *Marine Policy,* 33(5), 847-859.

Cariou, P and F C Wolff (2015). Identifying substandard vessels through port state control inspections: a new methodology for concentrated inspection campaigns. *Marine Policy,* 60, 27-39.

EMSA (2018). Annual overview of marine casualties and incidents 2018. < http://www.emsa.europa.eu/news-a-press-centre/external-news/item/3406-annual-overview-of-marine-casualties-and-incidents-2018.html> (accessed 31.05.19).

Gao, Z, G M Lu, M J Liu and M Cui (2008). A novel risk assessment system for port state control inspection. In: *IEEE International Conference on Intelligence and Security Informatics (ISI) Proceedings*, 242-244.

Ghomi, H, M Bagheri, L P Fu and L F Miranda-Moreno (2016). Analyzing injury severity factors at highway railway grade crossing accidents involving vulnerable road users: a comparative study. *Traffic Injury Prevention,* 17(8), 833-841.

Graziano, A, P Cariou, F C Wolff, M Q Mejia and J U Schröder-Hinrichs (2018). Port state control inspections in the European Union: do inspector's number and background matter? *Marine Policy,* 88, 230-241.

Hänninen, M and P Kujala (2014). Bayesian network modeling of port state control inspection findings and ship accident involvement. *Expert Systems with Applications,* 41(4), 1632-1646.

Heij, C and S Knapp (2018). Predictive power of inspection outcomes for future shipping accidents–an empirical appraisal with special attention for human factor aspects. *Maritime Policy & Management,* 45(5), 1-18.

Hu, L and Y Liu (2016). Joint design of parking capacities and fleet size for one-way station-based carsharing systems with road congestion constraints. *Transportation Research Part B: Methodological*, *93*, 268-299.

Hu, H T, X Z Chen and Z Sun (2017). Effect of water flows on ship traffic in narrow water channels based on cellular automata. *Polish Maritime Research*, *24*(s3), 130-135.

IMO (2011). IMO and the environment. < http://www.imo.org/en/OurWork/Environment/Documents/IMO%20and%20the%20Environment%202011.pdf> (accessed 4.6.19).

IMO (2017). Resolution A. 1119(30): Procedures for port state control. < http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Assembly/Documents/A.1119%2830%29.pdf> (accessed 22.5.19).

IMO (2019). Implementation, control and coordination. < http://www.imo.org/en/OurWork/MSAS/Pages/ImplementationOfIMOInstruments.aspx> (accessed 13.03.19).

Knapp, S, G Bijwaard and C Heij (2011). Estimated incident cost savings in shipping due to inspections. *Accident Analysis & Prevention,* 43(4), 1532-1539.

Knapp, S and P H Franses (2007). Econometric analysis on the effect of port state control inspections on the probability of casualty: can targeting of substandard ships for inspections be improved? *Marine Policy,* 31(4), 550-563.

Kumar, S and D Toshniwal (2016). A data mining approach to characterize road accident locations. *Journal of Modern Transportation,* 24(1), 62-72.

Li, K X and H S Zheng (2008). Enforcement of law by the port state control (PSC). *Maritime Policy & Management,* 35(1), 61-71.

Liu C C, X Xiang, C R Zhang and L Zheng (2016). A decision model for berth allocation under uncertainty considering service level using an adaptive differential evolution algorithm. *Asia-Pacific Journal of Operational Research*, 33(06), 1-28.

McNicholas, P D, T B Murphy and M O'Regan (2008). Standardising the lift of an association rule. *Computational Statistics & Data Analysis,* 52(10), 4712-4721.

Mirabadi, A and S Sharifian (2010). Application of association rules in Iranian Railways (RAI) accident data analysis. *Safety Science,* 48(10), 1427-1435.

Ravira, F J and F Piniella (2016). Evaluating the impact of PSC inspectors' professional profile: a case study of the Spanish Maritime Administration. *WMU Journal of Maritime Affairs,* 15(2), 221-236.

Sternberg, A, D Carvalho, L Murta, J Soares and E Ogasawara (2016). An analysis of Brazilian flight delays based on frequent patterns. *Transportation Research Part E: Logistics and Transportation Review,* 95, 282-298.

Sun, Z, Z L Chen, H T Hu and J F Zheng (2015). Ship interaction in narrow water channels: a two-lane cellular automata approach. *Physica A: Statistical Mechanics* *and Its Applications*, *431*, 46-51.

Tan, P N, M Steinbach and V Kumar (2005). Association Analysis: Basic Concepts and Algorithms. *Introduction to Data mining*. Addison-Wesley, Boston.

Tang, Y Y and Y Qin (2015). Railway freight demand analysis based on multidimensional association rules mining. In: *Fifth International Conference on Transportation Engineering Proceedings*, 2075-2081.

Tokyo MoU (2012). Tokyo MoU deficiency codes. < http://www.tokyo-mou.org/publications/tokyo_mou_deficiency_codes.php> (accessed 13.03.19).

Tokyo MoU (2016). Annual report 2015 on port state control in the Asia-Pacific region. < http://www.tokyo-mou.org/doc/ANN15.pdf> (accessed 28.05.19).

Tokyo MoU (2017). Annual report 2016 on port state control in the Asia-Pacific region. <http://www.tokyo-mou.org/doc/ANN16.pdf> (accessed 2 7.12.18).

Tokyo MoU (2018a). Annual report 2017 on port state control in the Asia-Pacific region. <http://www.tokyo-mou.org/doc/ANN17.pdf> (accessed 2 9.05.19).

Tokyo MoU (2018b). Memorandum of Understanding on port state control in the Asia-Pacific Region. < http://www.tokyo-mou.org/doc/Memorandum%20rev18.pdf > (accessed 13.01.19).

Tokyo MoU (2019). Annual report 2018 on port state control in the Asia-Pacific region. <http://www.tokyo-mou.org/doc/ANN18.pdf> (accessed 27.12.18).

Tsou, M C (2018). Big data analysis of port state control ship detention database. *Journal of Marine Engineering & Technology,* 17, 1-9.

UNCTAD (2018). Review of maritime transport 2018. < https://unctad.org/en/PublicationsLibrary/rmt2018_en.pdf> (accessed 29.05.31).

Weng, J X and G R Li (2019). Exploring shipping accident contributory factors using association rules. *Journal of Transportation Safety & Security,* 11(1), 36-57.

Weng, J X, J Z Zhu, X D Yan and Z Y Liu (2016). Investigation of work zone crash casualty patterns using association rules. *Accident Analysis & Prevention,* 92, 43-52.

Xu, R F., Q Lu, W J Li, K X Li and H S Zheng (2007a). A risk assessment system for improving port state control inspection. In*: International Conference on Machine Learning and Cybernetics Proceedings in 2007*, 818-823.

Xu, R F, Q Lu, K X Li and W J Li (2007b). Web mining for improving risk assessment in port state control inspection. In: *International Conference on Natural Language Processing and Knowledge Engineering Proceedings in 2007*, 427-434.

Yang, Z S, Z L Yang and J B Yin (2018). Realising advanced risk-based port state control inspection using data-driven Bayesian networks. *Transportation Research Part A*, 110, 38-56.

Yu, S, Y H Jia and D Y Sun (2019). Identifying factors that influence the patterns of road crashes using association rules: a case study from Wisconsin, United States. *Sustainability*, 11(7), 1925-1939.

Zhang, C Q and S C Zhang (2002). Association rule mining: models and algorithms. Germany, Springer-Verlag Publishing Company.

Zhang, Z H, Q H, J Gao and M Ni (2018). A deep learning approach for detecting traffic accidents from social media data. *Transportation Research Part C: Emerging Technologies*, 86, 580-596.

Zhao, D, W Wang, G P Ong and Y J Ji (2018). An association rule based method to integrate metro-public bicycle smart card data for trip chain analysis. *Journal of Advanced Transportation,* 2018, 1-11.

Zhao, D, W Wang, C Y Li, Y J Ji, X J Hu and W F Wang (2019). Recognizing metro-bus transfers from smart card data. *Transportation Planning and Technology,* 42(1), 70-83.

Zhou, C and J Sun (2010). Automatically optimized and self-evolutional ship targeting system for port state control. In: *IEEE International Conference on Systems Man and Cybernetics (SMC) Proceedings 2010*, 791-795.