Applicable Metamorphic Testing for Erasable-Itemset Mining

Tzung-Pei Hong1, 2, Chen-Chia Chiu2, Ja-Hwung Su1* and Chun-Hao Chen3

1 Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung, Taiwan
2 Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan
3 Department of Information and Finance Management at National Taipei University of Technology, Taipei, Taiwan

Corresponding author: Ja-Hwung Su (e-mail: bb0820@ms22.hinet.net).

ABSTRACT Erasable-itemset mining is often used in production planning to identify itemsets that, if removed, would make little effect on the production profits. Consequently, much recent attention has been focused on increasing the mining efficiency. Yet, in addition to mining efficiency, guaranteeing mining correctness is also an important issue. In real applications, wrong mining results might lead a company to make inappropriate decisions. Therefore, in this paper, we use the metamorphic testing, a lightweight software-testing strategy, to check the mining results via discovering proper metamorphic relations. The core idea contains five metamorphic relations oriented from two aspects. In the first aspect, we alter maximum thresholds without changing the input data, and in the second aspect, we modify product databases and material items with fixing maximum thresholds. In the experiments, 11 datasets and 76 mutants generated by µJava were used to discuss the effects of database parameters, including the average number of materials in products, the number of different material items, the number of products, and the maximum threshold, respectively. The experimental results show that the metamorphic relations deliver good assessment performances in terms of effectiveness and efficiency, especially the fourth metamorphic relation.

INDEX TERMS data mining, erasable itemset, metamorphic testing, mining assessment, software testing.

I. INTRODUCTION

With the flourishing development of computer science, databases volumes are becoming larger and larger. Facing such a large volume of databases, people seek to obtain helpful information from these databases for decision supports. Thus, data mining techniques such as Apriori and FP-tree have been proposed to mine useful knowledge for decision-makers. According to the knowledge mined, decision-makers can effectively make the right decisions to meet the rapidly fluctuating commercial patterns. To this end, Agrawal et al. proposed data mining [2][3], which is widely used in fields like bioinformatics and predictive analysis. There are many applications —especially in industrial manufacturing — such as product production planning. The product production planning problem defines those itemsets (or the sets of raw materials of the products) that would not significantly impact product profit if removed. To alleviate this problem, Deng et al. proposed an erasable-itemset mining algorithm called META in 2009 [8].

Actually, for a manufacturing company owner, hiring a group of professionals for just mining the erasable itemsets will not be a good idea because the manual cost is so high. Additionally, with the development of cloud computing, more and more companies have begun outsourcing data mining tasks to third-party services. Due to such outsourcing service, during transfers, the returned itemsets or the mining program provided by third-party services could lose some important information. With this in mind, the correctness and integrity of the returned itemsets gains importance: if the returned itemsets are not accurate, it could result in an enormous loss for the company. In general cases, problems of the huge amount of the returned erasable itemsets and the diverse mining logic could incur high costs in both money and human resources. Therefore, verifying the mining correctness and maintaining the mining programs have been important research topics.

To aim at the issues above, a software-testing technique called metamorphic testing (MT) was proposed by Chen et al. in 1998 [5]. This approach defines some metamorphic relations (MRs), each of which includes an input relation and an output relation. A test case can be converted into another test case through the input relation. If the output results of two cases obtained from a program execution satisfy the output relation, then the program has high reliability. In contrast, if
the output results of two cases do not satisfy the output relation, the program apparently has logic errors.

By concluding above, the metamorphic relations are very helpful to cost reduction in validating the mining algorithms. In real applications, for a manufacturing company, how to reduce the manufacture cost by discovering redundant materials (erasable itemsets) is an important issue. If the erasable itemsets can be filtered out successfully by an erasable-itemsets mining algorithm, the manufacture cost will be decreased effectively. For this purpose, the metamorphic relation plays a critical role to assure the correctness of the program. If the metamorphic relations can be defined accurately, the manual cost for checking the mining algorithm can be saved. In the following, the main contributions are summarized into three-folds.

(1) From the technical viewpoint, although there have been several studies made on metamorphic testing and frequent-itemset-mining testing, no metamorphic relations are proposed for erasable-itemset mining.

(2) From the practical viewpoint, the proposed metamorphic relations can be applied in real applications of manufacturing. By the proposed methods, the manufacture cost can be reduced significantly.

(3) From the robustness viewpoint, the proposed metamorphic relations were evaluated by a number of experiments. The experimental results reveal that the proposed metamorphic relations were very promising in terms of effectiveness and efficiency.

The remainder of this paper is structured in the following. The literatures are briefly reviewed in Section 2. In Section 3, the proposed method for erasable-itemset mining assessments is presented in detail. The experimental analysis is interpreted in Section 4. Finally, the conclusions and future works are shown in Section 5.

II. Literature Review

The goal of this paper is to conduct a reliable assessment for erasable-itemset mining. Basically, it contains two major ideas, namely metamorphic testing and erasable-itemset mining. In this section, a brief review is made by the following categories.

A. Metamorphic Testing

Software testing is a way to assure and verify software quality, but it faces a fundamental problem: the oracle problem. The oracle problem can be identified that it is difficult to verify the results while performing a software test. In 1998, Chen et al. proposed metamorphic testing (MT), a new approach for generating the next test case [6]. This technique effectively alleviates the oracle problem and further triggers a set of related papers made gradually. Figure 1 shows the yearly numbers of MT-related publications from 2002 to 2019 in IEEE society. From Figure 1, we can know that, MT has become a popular topic recently due to the number of papers grew rapidly after the first International Workshop on Metamorphic Testing organized by Kanewala et al. in 2016.

Chen et al. provided a comprehensive review on MT in 2018 [7], which discussed the current challenges and present directions for the new research in the field of MT.

To explain MT in detail, a sine function is taken as an example:

$$\text{MR: } \{ \text{IR } (x_2 = \pi - x_1) \Rightarrow \text{OR } (\sin x_1 = \sin x_2) \}$$

(1)

In Eq. (1), MR, IR and OR indicate the metamorphic relation, the input relation and the output relation, respectively. According to the input relation in the related MR, MT uses the existing test case, known as the source input ($x_1$), to generate the subsequent test case, known as the follow-up input ($x_2$). Then, the program is examined with the source input ($x_1$) and the follow-up input ($x_2 = \pi - x_1$), and the results are obtained as the source output ($\sin x_1$) and the follow-up output ($\sin x_2$), respectively. Next, whether the source output and follow-up output satisfy the output relation in the same MR is determined. If the output results of two cases satisfy the output relation, the program is identified as the one with high correctness. However, implicit errors might still exist. In contrast, if the output results of two cases do not satisfy the output relation, the program is sure to have errors.

The most critical part of the MT procedure is designing reliable MRs. For that, Segura et al. proposed four aspects of a good MR design [19].

Figure 1. IEEE publications on metamorphic testing from 2002 to 2019.

Over the past few years, the core of MT has been used in various fields actually by designing specific MRs, such as domains of IoT (Internet Of Things) and data mining. In IoT, Azimian et al. adopted MT to search high-consuming applications for portable devices, such as smartphones [1]. In bioinformatics, Chen et al. used MT to test a range of bioinformatics programs [6]. In machine learning, Xie et al. proposed METTLE to evaluate and validate unsupervised learning systems [20]. In deep learning, Zhou and Sun used metamorphic testing to detect fatal errors in self-driving cars' onboard systems [23]. In data analytics software, Jarman et al. applied metamorphic testing to Adobe's data analytics software for time-series data [13]. In data mining, Zhang et al. used MT to verify the integrity of frequent itemsets and frequent-itemset mining programs [22]. Inspired by Zhang et al., we thus adopted this concept on the erasable-itemset mining. In [22], Zhang designed ten MRs, and in this paper, we refer to their idea to design five new MRs for erasable-itemset mining.
other one was proposed by Nam et al., which adopted the anti-
manner, the erasable patterns can be mined from the data
Through the proposed data structure and divide-and-conquer
pattern mining algorithm based on sliding windows [21].
Yun et al. presented an efficient erasable
pattern stream mining with a special data structure and a
approach to increase the efficiency [11]. Lee et al. then
then proposed a tree-based quasi-erasable-itemset mining
In 2017, Hong et al. proposed incremental erasable-
Algorithm called MERIT+ was presented by Le et al. [14],
rapidly. After that, in 2013, a revised version of the MERIT
algorithm called VME Algorithm [9], which used the structure
algorithms for erasable itemset mining have been proposed in
In 2009, erasable-itemset mining was introduced by Deng et
Apriori-based algorithm called META, which generates
candidate patterns level-by-level. Figure 2 shows the META
algorithm. In addition to the META algorithm, several
Example 1. To explain the above-mentioned erasable-
itemset mining, an illustrative example is shown before
presenting the details of the proposed approach. Table 1 shows
a product database consisting of five products and five
materials a-e. Assume the maximum threshold \( r \) is set at 40%.
The procedure of the algorithm is shown as follows.

**TABLE I**

| PID | Materials | Profit value |
|-----|-----------|--------------|
| P₁  | a, b, c   | 200          |
| P₂  | d         | 200          |
| P₃  | a, d      | 100          |
| P₄  | a, b      | 400          |
| P₅  | b, c, e   | 100          |

First, the total profit value of \( DB \) denoted as \( T \), is \( 200 + 200 + 100 + 400 + 100 = 1000 \). The gain values of candidate
erasable 1-itemsets, materials \( a \) to \( e \) appearing in the product
database, are calculated. Here, take material \( b \) as an example.
It appears in \( P₁, P₃, \) and \( P₅ \) in the product database. Thus, the
related gain value is \( 200 + 400 + 100 = 700 \). The gain values
of the other materials are calculated in the same way, as shown
in Table 2.

**TABLE II**

| Itemset | Gain value |
|---------|------------|
| a       | 700        |
| b       | 700        |
| c       | 300        |
| d       | 300        |
| e       | 100        |

**TABLE III**

| Itemset | Gain value |
|---------|------------|
| c       | 300        |
| d       | 300        |
| e       | 100        |
| c, e    | 300        |
| d, e    | 400        |

Accordingly, the maximum threshold value \( (T \times c) \) is
\( 1000 \times 0.4 = 400 \). Then the candidate erasable 1-itemsets
with gains lower than the threshold are filtered. From Table
2, we can know that \{c\}, \{d\} and \{e\} are erasable 1-itemsets
collected as the set of erasable 1-itemsets \( E₁ \). The set of candidate
erasable 2-itemsets \( C₂ \) is obtained from the
erasable 1-itemsets. This example yields three candidate
erasable 2-itemsets: \{c, d\}, \{d, e\}, and \{c, e\}. The gain
values of these candidate erasable 2-itemsets are calculated,
after which their gain values are compared with the
maximum threshold value.

Take a candidate erasable 2-itemset \{c, d\} as an example.
Because products \( P₁, P₃, \) and \( P₅ \) contain at least \( c \) or \( d \),
the sum of the profits of these products is set as the gain value
of the candidate erasable 2-itemset \{c, d\}, which is 600.
Because the gain value of the candidate erasable 2-itemset
{c, d} is larger than the maximum threshold value (400), the candidate erasable 2-itemset {c, d} is not an erasable 2-itemset. The other candidate erasable 2-itemsets {c, e} and {d, e} are processed in the same way. Then, no candidate erasable 3-itemsets are generated. Finally, all erasable itemsets obtained are collected as the output $E_I$, as shown in Table 3.

III. The Proposed Approach

A. Overview
The goal of this paper is to define MRs to validate the erasable-itemset mining programs. To reach this goal, in this paper, we propose five MRs from two aspects based on the input data of the mining problem. For the first aspect, we tune the maximum threshold without changing the product database and the set of materials. On the contrary, the threshold is fixed with changing the product database and the set of materials in the second aspect. This lightweight method based on metamorphic testing substantially reduces the company costs because it is automatic and easy to maintain. To evaluate the effectiveness of the proposed MRs, numerous experiments were conducted to simulate malicious attacks by mutating the erasable-itemset mining program (META) and using the mutation score to validate the performance of the proposed MRs.

B. Preliminary
A metamorphic relation (MR) consists of an input relation and an output relation. At the beginning of the metamorphic testing procedure, an existing test case called the source input is given. The input relation in a specific MR is used to generate a follow-up input by a source input, and the output relation is used to verify the relationship between a source output and the follow-up output. The source output indicates the results executed by the mining program with a source input, and the follow-up output indicates the results executed by the same mining program with a follow-up input. In the following, the used symbols in the proposed MRs are introduced. According to the concept of MT, these symbols can be divided into the source input, the follow-up input, the source output, and the follow-up output, individually.

1) SOURCE/FOLLOW-UP INPUTS
The source input refers to the existing test case, which consists of the product database, the set of materials, and the maximum threshold. The follow-up input refers to the results also consisting of the product database, the set of materials and the maximum threshold, where a source input is converted by a specific input relation. The symbols of the source and follow-up inputs to be used in the MRs are listed in Table 4.

2) SOURCE/FOLLOW-UP OUTPUTS
The source output refers to the results when the programmer executes the mining program with the specific source input. The follow-up output refers to the results when the programmer executes the same mining program with the follow-up input. The symbols of the source and follow-up outputs to be used in the MRs are listed in Table 5.

C. Metamorphic Relations
In this section, the proposed five MRs will be presented from two perspectives. In the first perspective, the maximum threshold will be tuned, keeping other input data constant. In the second perspective, the product database and the set of materials will be changed, keeping the maximum threshold constant. To better explain the design logic, for each MR, an explainable example is shown by following the examples taken in Section 2.

1) RELATIONS WITH VARYING MAXIMUM THRESHOLDS

(A) THE FIRST METAMORPHIC RELATION (MR1)
In MR1, we consider the impact on the number of erasable itemsets when decreasing the maximum threshold $r$, which is defined as follows:

Input relation: $I = I$, $D = D$ and $r' < r$.

Output relation: $|E| \leq |E|$ and $|E'| \leq |E|$.

In the above MR, the follow-up product database $D'$ is equivalent to the source product database $D$. The set of follow-up materials $I'$ is also equivalent to the set of source materials $I$. Then, the follow-up threshold $r'$ decreases as the follow-up maximum threshold $r'$ reduces, where the follow-up threshold is the result of the follow-up maximum threshold $r'$ multiplying the total profit of the follow-up product database. Thus, the gain values of source erasable itemsets may be larger than the follow-up threshold, and then remove them from the set of the source erasable itemsets $E$. Next, we glean the remaining source erasable itemsets as the set of the follow-up erasable itemsets $E'$. Besides, the gain values of the source nonerasable itemsets are still larger than the follow-up threshold, so they are still source nonerasable itemsets, and we glean them as the source nonerasable itemsets $NE'$. Thus, we conclude that the number of the follow-up erasable itemsets $|E'|$ is smaller than
or equal to that of the source erasable itemsets |E|. Thus, |E'| \leq |E| and |E'| \subseteq |E|.

### Example 2.
Here, an example is given to illustrate MR1. The follow-up product database D' and the set of the follow-up materials I' keep constant, as shown in Table 1. The source maximum threshold r is 40%, and we reduce it to 30% as the follow-up maximum threshold r'. Thus, the follow-up maximum threshold value is 1000 \times 0.3 = 300. Then, we execute the mining program with the follow-up input data. The set of the follow-up candidate erasable 1-itemsets CE1 is the same as the set of the source candidate erasable 1-itemsets CE1 shown in Table 2. Then the follow-up candidate erasable 1-itemsets not exceeding the follow-up maximum threshold are filtered. From Table 2, we know that \{c\}, \{d\} and \{e\} are the follow-up erasable 1-itemsets. Next, the set of the follow-up candidate erasable 2-itemsets CE2 is obtained from the set of the follow-up candidate erasable 1-itemsets E1'. In this example, we have three follow-up candidate erasable 2-itemsets: \{c, d\}, \{d, e\}, and \{c, e\}. The gain values of three candidate erasable 2-itemsets are calculated and compared with the maximum threshold of 300. Then, no follow-up candidate erasable 3-itemsets are generated. Finally, Table 6 shows the follow-up erasable itemsets. By comparing Tables 3 and 6, we can know that the number of the follow-up erasable itemsets is smaller than that of the source erasable itemsets.

### Table VI

| Itemset | Gain value |
|---------|------------|
| c       | 300        |
| d       | 300        |
| e       | 100        |

### Example 3.
Based on the above examples, given 60% as the follow-up maximum threshold r', the follow-up maximum threshold value is 1000 \times 0.6 = 600. Then, as shown in Table 7, \{c\}, \{d\} and \{e\} are the follow-up erasable 1-itemsets, and \{c, d\}, \{c, e\} and \{d, e\} are the follow-up erasable 2-itemsets because the referred gain values \{300, 300, 100, 600, 300, 400\}, respectively are all larger than the follow-up threshold in this example.

### Table VII

| Itemset | Gain value |
|---------|------------|
| c       | 300        |
| d       | 300        |
| e       | 100        |

2) **RELATIONS WITH VARYING MATERIALS AND THE DATABASE D**

(A) **THE THIRD METAMORPHIC RELATION (MR3)**

In MR3, we consider the impact on the number of erasable itemsets when removing an erasable 1-itemset from the product database and the set of materials. The third MR is defined as follows:

**Input relation:** I' is constructed by removing an erasable 1-itemset from I. D' is constructed by removing the same erasable 1-itemset from D. Then, r' = r.

**Output relation:** |E'| \leq |E| and |E'| \subseteq |E'|

In the above MR, the set of the follow-up materials I' are equivalent to the source product database D and the set of source materials I, respectively. Then, the follow-up threshold increases as the follow-up maximum threshold r' increases, where the follow-up threshold is derived as the same as that in MR1. Thus, the gain values of source nonerasable itemsets may be smaller than or equal to the follow-up threshold and remove them from the source nonerasable itemsets NE. Then, we glean the remaining source nonerasable itemsets as the set of the follow-up nonerasable itemsets NE'. Because the gain values of source erasable itemsets are still smaller than the follow-up threshold value, they are still erasable itemsets. Thus, the set of the follow-up erasable itemsets E' is obtained as the follow-up erasable itemsets plus those itemsets removed from the set of the source nonerasable itemsets NE. According to the above, we conclude that the number of the follow-up erasable itemsets |E'| is larger than or equal to the number of the source erasable itemsets |E|. Thus, |E| \leq |E'| and |E| \subseteq |E'|.

### Example 4.
The follow-up product database D' is constructed by removing a source erasable 1-itemset e from the source product database D, but the profit values of each product are not changed. From Table 3, we have all source erasable 1-itemsets. Assume that we remove item e from the source product database to form the follow-up product...
database $D'$ as shown in Table 8. Then, the set of the follow-up product items $I'$ also removes item $e$ from the set of the source product items. The maximum threshold value, which is 400, stays constant. Finally, \{c\} and \{d\} are the follow-up erasable 1-itemsets and no follow-up erasable 2-itemset is generated in this case. By comparing this result with Table 3, we can conclude that the number of the follow-up erasable itemsets is smaller than the number of the source erasable itemsets.

(B) THE FOURTH METAMORPHIC RELATION (MR4)

In MR4, we consider the impact on the number of erasable itemsets while adding new material. The fourth MR is defined as follows:

**Input relation:** $I'$ is constructed by adding a new material to $I$. $D'$ is constructed by adding the same material in each product materials. Then, $r' = r$.

**Output relation:** $|E| = |E'|$ and $|NE_k'| = |NE_k| + |NE_{k-1}| + |E_{k-1}|$, where $|NE_0| = 1$ and $|E_0| = 0$.

In the above MR, the set of the follow-up materials $I'$ is constructed by adding a new material, which we assume is $g$, to the set of the source materials $I$. The follow-up product database $D'$ is constructed by adding material $g$ to each product in the source product database $D$. The follow-up maximum threshold $r'$ is equivalent to the source maximum threshold $r$. Suppose the set of the source erasable 1-itemsets $E_1$ is $\{i_1, i_2, i_3, ..., i_n\}$. Because material $g$ is added to each product, its gain value is equal to the total profit of the follow-up product database $D'$. Thus, its gain value must be larger than or equal to the follow-up threshold value. Material $g$ should belong to the set of the follow-up nonerasable 1-itemsets $NE_1'$. Thus, the set of the follow-up erasable 1-itemsets $E_1'$ should be equivalent to the set of the source erasable 1-itemsets $E_1$. Because $E_1' = E_1$, we infer the set of the follow-up erasable 2-itemsets $E_2'$ is also equivalent to the set of the source erasable 2-itemsets $E_2$. Finally, we conclude that the number of the source erasable itemsets $|E|$ is equal to the number of the follow-up erasable itemsets $|E'|$.

Next, the number of the source nonerasable 1-itemsets $|NE_1|$ is obtained while the number of all source 1-itemsets minus the number of the source erasable 1-itemsets $|E_1|$. In addition, the number of the follow-up nonerasable 1-itemsets $|NE_1'|$ is obtained while the number of all follow-up 1-itemsets minus the number of the follow-up erasable 1-itemset $|E_1'|$. Thus, $|NE_1'| = |NE_1| + |NE_0| + |E_0|$, where $|NE_0| = 1$ and $|E_0| = 0$. Then, the number of the source nonerasable 2-itemsets $|NE_2|$ is obtained while the number of all source 2-itemsets minus the number of the source erasable 2-itemsets $|E_2|$, where the source 2-itemsets are generated by combinations of all source 1-itemsets. Then, the number of the follow-up nonerasable 2-itemsets $|NE_2'|$ is obtained while the number of all follow-up 2-itemsets minus the number of the follow-up erasable 2-itemsets $|E_2'|$. Thus, $|NE_2'| = |NE_2| + |NE_1| + |E_1|$, where $|NE_1| = 1$ and $|E_1| = 0$.

Example 5. The follow-up product database $D'$ is constructed by adding a new item $g$ to each product in the source product database $D$, but the profit values of each product are not changed. Table 9 shows the follow-up product database $D'$. Then, material $g$ is added to the set of materials, and the threshold is 400. The maximum threshold value, which is 400, stays constant. Next, we remove the erasable 1-itemsets \{c\}, \{d\} and \{e\} to obtain the set of the follow-up erasable 1-itemsets $NE_1'$, \{a, b, g\} with the gain set \{700, 700, 1000\}, respectively. By the same method, we have the set of the source nonerasable 1-itemsets $NE_1'$, \{a, b\} with the gain set \{700, 700\}, respectively. In this case, the number of the follow-up nonerasable 1-itemsets $|NE_1'|$ is 2. Thus, we can infer that $|NE_1'| = |NE_1| + |NE_0| + |E_0|$, where we set $|NE_0| = 1$ and $|E_0| = 0$. Next, \{c, e\} and \{d, e\} are the follow-up erasable 2-itemsets.

(C) THE FIFTH METAMORPHIC RELATION (MR5)

In MR5, we consider the impact on the number of erasable itemsets when adding a new product with a new material to the product database. The fifth MR is defined as follows:

**Input relation:** $I'$ is constructed by adding a new material to $I$. $D'$ is constructed by adding a new product $P_{n+1}$ to $D$, where the material of $P_{n+1}$ is the new material, and its profit is zero. Then, $r' = r$.

**Output relation:** $|E| \leq |E'|$.

In the above MR, the set of the follow-up product items $I'$ is constructed by adding a new material $h$ to the set of the source materials $I$. The follow-up product database $D'$ is constructed by adding a new product $P_{n+1}$ to the source product database $D$, where its material is $h$, and the profit of this product is zero. Due to the profit of $P_{n+1}$ is zero, the gain value of $g$ is also zero. Thus, $h$ must belong to the set of the

---

**TABLE VIII**

| PID | Materials | Profit value |
|-----|-----------|--------------|
| $P_1$ | a, b, c | 200 |
| $P_2$ | d | 200 |
| $P_3$ | a, d | 100 |
| $P_4$ | a, b | 400 |
| $P_5$ | b, c | 100 |

**TABLE IX**

| PID | Materials | Profit value |
|-----|-----------|--------------|
| $P_1$ | a, b, c, g | 200 |
| $P_2$ | d, g | 200 |
| $P_3$ | a, d, g | 100 |
| $P_4$ | a, b, g | 400 |
| $P_5$ | b, c, g | 100 |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
follow-up erasable 1-itemsets, and the number of follow-up erasable 1-itemsets $|E'_1|$ increases thereby. Then, the set of the follow-up erasable 2-itemsets $E'_2$ is generated by the set of the follow-up erasable 1-itemsets $E'_1$. Further, the number of the follow-up erasable 1-itemsets $|E'_1|$ increases, and the number of the follow-up erasable 2-itemsets $|E'_2|$ might increase, too. Finally, we conclude that the number of the source erasable itemsets $|E|$ is smaller than the number of the follow-up erasable itemsets $|E'|$, i.e., $|E| \leq |E'|$.

Example 6. The follow-up product database $D'$ is constructed by adding a new product $h$ to the source product database and the related profit is zero. Table 10 shows the follow-up erasable 1-itemsets, and the number of follow-up erasable 2-itemsets.

Example 7. The mutation score is the rate of detected mutants over the total mutants for a specific MR. If a mutant program violates an MR, this mutant program can be identified as killed. For example, if there are totally $t$ mutants and $k$ detected mutants for a MR, then the performance of this MR is $k/t$. The erasable-itemset mining programs in this paper were implemented in Java, running on a server with Intel Core i7-9700 3GHz and 8 GB RAM.

TABLE XI

| Parameters | Description |
|------------|-------------|
| $T$        | Average number of materials per product |
| $N$        | Number of different materials in the dataset |
| $D$        | Total number of products in the dataset |

B. Experimental Results

The major intent of evaluations is to know the performances of proposed MRs. All evaluations were conducted in 3 aspects, including the impact of the average number of items in each product, the number of different items in the datasets and the number of products in the datasets, and the maximum threshold.

1) Influences of average numbers of materials per product holds

The experiments include two major sources, namely mutated programs and transaction data. The goal of mutated programs is to simulate the implementation errors in a mining program. For this purpose, we used the $\mu$Java tool, a mutation system for Java programs [17]. In the experiments, 76 invalid programs were yielded for the META algorithm, each of which had only one fault, called mutant. An example is shown in Figure 3. The correct program is to calculate the sum accumulated from 0 to 9, referring to Line3 of $j = j + 1$. After using $\mu$Java, Line 3 is altered to $j = j - 1$.

For experimental transaction data, the IBM Synthetic Data Generator was used to generate synthetic datasets [24]. The parameters for creating synthetic datasets are shown in Table 11. To evaluate the performance of the proposed metamorphic relations (MRs), we adopted the mutation score metric:

\[
m\text{utation score} = \frac{\text{the number of killed programs}}{\text{the total number of mutant programs}}. \tag{2}\]

FIGURE 3. Arithmetic implementation error.

FIGURE 4. Mutation scores under different average numbers of materials per product.

In overall, it delivers some points. First, MR4 derives the highest performance in detecting implementation errors. Second, the average mutation scores of MR1 and MR2 are pretty close and low in contrast. That is, MR1 and MR2 are not easy to detect. Third, MR3, MR4, and MR5 are sensitive for the average number of materials being 80. Especially, MR4 and MR5 decrease while MR3 increases. The potential
reason is that the addition operation is harder to detect than the removal operation. Figure 5 shows the time to execute all mutants in every MR. As the average number of materials increases, the execution time increases accordingly. In summary, the threshold-varying MRs are more difficult to detect than the data-varying MRs for the transaction length, and all execution time is very close.

In addition to the above parameters, the goal of this evaluation is to clarify the impacts on different numbers of products. By referring to Table 11, $D$ in this evaluation was set to 1000(1K), 5500(5.5K) and 10000(10K), where $N=1000$, $T=10$ and the maximum threshold was also 0.04%. Figure 8 shows that the performances, which include a set of aspects. First, MR1 and MR2 perform the same and much worse than the others. Second, MR1, MR2 and MR4 are not influenced by the number of products in our datasets. On the contrary, performances of MR5 and MR3 slightly change. Figure 9 shows the execution time. It can be observed that the execution time of all MRs grows significantly as the number of products rises.

2) INFLUENCES OF NUMBERS OF UNIQUE MATERIALS

To observe the influence of the number of different items in the datasets on our MRs, we altered parameter $N$ (the number of different materials in the datasets) to generate datasets. In this experiment, $N$ is set to 3000(3K), 5000(5K), 7000(7K) and 9000(9K), where $T=5$, $D=10000$ and the maximum threshold is set to be 0.04%. Figure 6 shows the mutation score of each MR, which delivers some observations. First, MR1 and MR2 are with the same performances. Second, for MR3, MR4 and MR5, the performances are stable and higher than those of MR1 and MR2. Figure 7 shows the time taken to execute all mutants in every MR. The execution time of all MRs linearly increases as the number of unique materials in the datasets increases. Yet, MR2 needs much more time than the others on a large number of unique items. This is because MR2 raises up the maximum threshold and therefore costs explosively to search items.

3) INFLUENCES OF NUMBERS OF PRODUCTS

To observe the influence of the number of different items in the datasets on our MRs, we altered parameter $N$ (the number of different materials in the datasets) to generate datasets. In this experiment, $N$ is set to 3000(3K), 5000(5K), 7000(7K) and 9000(9K), where $T=5$, $D=10000$ and the maximum threshold is set to be 0.04%

FIGURE 5. Execution time on different average numbers of materials per product.

FIGURE 6. Mutation scores under different numbers of unique materials.

FIGURE 7. Execution time on different numbers of unique materials.

FIGURE 8. Mutation scores under different numbers of products.

FIGURE 9. Execution time on different numbers of products.
4) EXPERIMENTAL SUMMARY

Based on the above experimental evaluations, the results are summarized in the following.

1) To ensure the effectiveness of the developed models, in the experiments, a mutation score is proposed to evaluate the effectiveness. This measure shows the rate of successful detections for biased programs. The experimental results show MR4 performed better than the other MRs. In the future, we will keep testing the proposed 5 MRs by real data.

2) By comparing the mutation scores above, the threshold-varying MRs are not easy to detect. In contrast, data-varying MRs are easy to detect, especially for MR4. The potential explanation is that a new material is added in each product. The change is clearer than that in the others. In detail, in the mining programs, the correct results rely heavily on the coding logics and thresholds. Therefore, the threshold changes probably recover the loss results by biased coding. This is the potential reason that the threshold-varying MRs are not easy to detect in the experimental cases.

3) From the execution-time viewpoint, the heavy data changes will result in high execution time. Especially for product length, it needs much time to compose and decompose the products. This problem often happens in Apriori-based algorithms. Moreover, by comparing MR1 and MR2 in Figure 7, the threshold is sensitive to the numbers of unique materials. This is because the mutated threshold is lower than the source threshold for MR1. Further, more candidates are pruned to speed up the mining if facing a larger number of unique materials.

4) In overall, MR1, MR2 and MR4 perform more stably than MR 3 and MR5. However, MR1 and MR2 are much harder to detect than MR4. This is because the programming logic and the threshold are insensitive to the results. Besides, lower thresholds in erasable-itemset mining generate fewer results. However, the biased logics might recover the lost results, which makes the biased programs uneasy to detect. For example, if the itemsets \{A, D\} are eliminated by a threshold, the biased logic still generates them. As a result, the biased program is not detected successfully in this case. This problem will be researched in our future intent.

V. Conclusions

Recently, cost-down has been a hot topic for a manufacturing company. To this end, mining of erasable materials is proposed as a solution. However, it needs much manual effort to verify the correctness of mining programs. To remedy this, in this paper, we use metamorphic testing, a lightweight method to verify the correctness of the erasable itemsets mining. In the proposed method, five metamorphic relations (MRs) are proposed according to the properties of the erasable itemset mining. Each MR is analytically evaluated by the proposed mutation score. These MRs provide manufacturing companies with a solution to effectively verify the correctness of erasable itemsets. Although this paper provides metamorphic testing for erasable data mining, there remain a number of works to do in the future. First, the threshold-varying problem will be researched for a robust performance. Second, not only checking the mining correctness, but also logical errors will be recognized by a level-wise metamorphic testing in the future. Third, correlations with real companies will be conducted using real datasets.

REFERENCES

[1] F. Azimian, F. Faghii, M. Kargahi and S. M. Mahdi Mirdehghan, “Energy Metamorphic Testing for Android Applications,” The 30th International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1-6, 2019.

[2] R. Agrawal, T. Imielinski and A. Swami, “Mining association rules between sets of items in large database,” The ACM SIGMOD International Conference on Management of Data, pp. 207-216, 1993.

[3] R. Agrawal and R. Srikant, “Fast algorithm for mining association rules,” The 20th International Conference on Very Large Data Bases, pp. 487-499, 1994.

[4] Y. Baek, U. Yun, H. Kim, H. Nam, G. Lee, E. Yoon, B. Vo and J. C.-W. Lin, “Erasable pattern mining based on tree structures with damped window over data streams,” Engineering Applications of Artificial Intelligence, Vol. 94, 103735, 2020.

[5] T. Y. Chen, S. C. Cheung and S. M. Yiu, “Metamorphic testing: a new approach for generating next test cases,” Technical Report HKUST-CS98-01, Department of Computer Science, Hong Kong University of Science and Technology, 1998.

[6] T. Y. Chen, J. W. Ho, H. Liu and X. Xie, “An innovative approach for testing bioinformatics programs using metamorphic testing,” BMC Bioinformatics, Vol. 10, Issue 1, pp. 24, 2009.

[7] T. Y. Chen, F. C. Kuo, H. Liu, P. I. Poon, D. Towey, T. H. Tse and Z. Q. Zhou, “Metamorphic Testing: A review of Challenges and Opportunities,” ACM Computing Survey, Vol. 51, No. 51, Article 4, 2019.

[8] Z. H. Deng, G. D. Fang, Z. H. Wang and X. R. Xu, “Mining erasable itemsets,” The International Conference on Machine Learning and Cybernetics, pp. 12-15, 2009.

[9] Z. H. Deng and X. R. Xu, “An efficient algorithm for mining erasable itemsets,” The 9th International Conference on Advanced Data Mining and Applications, pp. 214-225, 2010.

[10] Z. H. Deng and X. R. Xu, “Fast mining erasable items using NC_sets,” Expert Systems with Applications, Vol. 39, pp. 4453–4463, 2012.

[11] T. P. Hong, L. H. Chen, S. L. Wang, C. W. Lin and B. Vo, “Quasi-erasable itemset mining,” The 2017 IEEE International Conference on Big Data, pp. 1816–1820, 2017, USA.

[12] T. P. Hong, K. Y. Lin, C. W. Lin and B. Vo, “An incremental mining algorithm for erasable itemsets,” The 2017 IEEE International Conference on Innovations in Intelligent Systems and applications, pp. 286-289, 2017, Poland.

[13] D. C. Jarman, Z. Q. Zhou and T. Y. Chen, “Metamorphic testing for Adobe data analytics software,” The 2nd International Workshop on Metamorphic Testing, pp. 21-27, 2017.

[14] T. Le, F. Coenen and B. Vo, “An efficient algorithm for mining erasable itemsets using the difference of NC Sets,” The IEEE International Conference on Systems, Man, and Cybernetics Manchester, pp. 2270-2274, 2013.

[15] T. Le and B. Vo, “MEI: An efficient algorithm for mining erasable itemsets,” Engineering Applications of Artificial Intelligence, Vol. 27, pp. 155–166, 2014.

[16] G. Lee and U. Yun, “Single-pass based efficient erasable pattern mining using list data structure on dynamic incremental databases,” Future Generation Computer Systems, Vol. 80, pp. 12-28, 2018.

[17] Y. S. Ma, J. Offutt and Y. R. Kwon, “MuJava: an automated class mutation system,” Software Testing, Verification and Reliability, Vol. 15, Issue 2, pp. 97-133, 2015.
H. Nam, U. Yun, E. Yoon and J. C.-W. Lin, "Efficient Approach for Incremental Weighted Erasable Pattern Mining with List Structure," *Expert Systems with Applications*, Vol. 143. 113087, 2020.

S. Segura, G. Fraser, A. B. Sanchez and A. Ruiz-Cortes, "A survey on metamorphic testing," *IEEE Transactions on Software Engineering*, Vol. 42, Issue. 9, 2016.

X. Xie, Z. Y. Zhang, T. Y. Chen, Y. Liu, P. L. Poon and B. Xu, "METTLE: A METamorphic Testing Approach to Assessing and Validating Unsupervised Machine Learning Systems," *IEEE Transactions on Reliability*, 2020, doi: 10.1109/TR.2020.2972266.

U. Yun, G. Lee and E. Yoon, "Advanced approach of sliding window based erasable pattern mining with list structure of industrial fields," *Information Sciences*, Vol. 494, pp. 37–59, 2019.

J. W. Zhang, X. Y. Xie and Z. Y. Zhang, "How reliable is your outsourcing service for data mining? a metamorphic method for verifying the result integrity," *The 8th International Conference on Software Analysis, Testing, and Evolution*, pp. 120-136, 2018.

Z. Q. Zhou and L. Sun, "Metamorphic testing of driverless cars," *Communications of the ACM*, Vol. 62, No. 3, 2019.

IBM, "Quest synthetic data generation code," *IBM Quest Data Mining Project*, 1996.

**Chun-Hao Chen** is an associate professor at Department of Information and Finance Management at National Taipei University of Technology, Taipei, Taiwan. Dr. Chen received his Ph.D. degree with major in computer science and information engineering from National Cheng Kung University, Taiwan, in 2008. He has a wide variety of research interests covering data mining, time series, machine learning, evolutionary algorithms, and fuzzy theory. Research topics cover portfolio selection, trading strategy, business data analysis, time series pattern discovery, etc. He serves as the associate editor of the International Journal of Data Science and Pattern Recognition, and IEEE Access. He is also a member of IEEE.

**Chun-Chia Chiu** received his B.S. degree Department of Computer Science and Information Engineering at National University of Kaohsiung in 2020. His research interests include data mining and machine learning.

**Ja-Hwang Su** received his Ph.D. degree in the Department of Computer Science and Information Engineering at National Cheng Kung University in 2010. Currently, he is an assistant professor in the Department of Computer Science and Information Engineering at National University of Kaohsiung. His research interests include Machine Learning and Multimedia Information Retrieval. He has held 16 patents in USA and R.O.C. and published more than 70 research papers in some premier journals and International conferences, like IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Multimedia, IEEE Intelligent Systems, ICME, etc. Also he served as program committee members and reviewers in these journals and international conferences, such as SIGKDD, ICDM, CIKM, DASFAA, TKDD, KNOYS, CIM, JBIH and so on.