



# Analytics of high average-utility patterns in the industrial internet of things

Jimmy Ming-Tai Wu<sup>1</sup> · Zhongcui Li<sup>1</sup> · Gautam Srivastava<sup>2,3</sup> · Unil Yun<sup>4</sup> · Jerry Chun-Wei Lin<sup>5</sup> 

Accepted: 5 August 2021  
© The Author(s) 2021

## Abstract

Recently, revealing more valuable information except for quantity value for a database is an essential research field. High utility itemset mining (HAUIM) was suggested to reveal useful patterns by average-utility measure for pattern analytics and evaluations. HAUIM provides a more fair assessment than generic high utility itemset mining and ignores the influence of the length of itemsets. There are several high-performance HAUIM algorithms proposed to gain knowledge from a disorganized database. However, most existing works do not concern the uncertainty factor, which is one of the characteristics of data gathered from IoT equipment. In this work, an efficient algorithm for HAUIM to handle the uncertainty databases in IoTs is presented. Two upper-bound values are estimated to early diminish the search space for discovering meaningful patterns that greatly solve the limitations of pattern mining in IoTs. Experimental results showed several evaluations of the proposed approach compared to the existing algorithms, and the results are acceptable to state that the designed approach efficiently reveals high average utility itemsets from an uncertain situation.

**Keywords** IoT · Uncertainty · Average-utility · Analytics · Sensor networks

## 1 Introduction

Sensor networks and the Internet of things (IoT) expand rapidly and several smart devices such as mobile phones, sensors, actuators, and RFID are widely used and applied in different industrial applications and domains [1, 2]. It is necessary to concern all the relevant factors such as weight, interestingness, utility, or frequency to discover meaningful and useful information for more specific applications,

tasks, and domains for decision making. The past works of frequent itemset mining (FIM) [3], sequential-pattern mining (SPM) [4], or high-utility itemset mining (or high-utility pattern mining, HUIM/HUPM) [5, 6] are the common tasks to reveal useful knowledge in the applications that can also be applied into real-world industrial applications and scenarios. The necessary information for decision-making needs, for example, to be found in the control signals provided to Internet-connected devices (e.g., light measurements, smart logs, or output data). For industrial areas [7], various models and approaches have been implemented and studied by applying data mining models [7] and those algorithms can be used for pattern evaluation in different industrial applications and domains.

Traditionally, FIM is used to find the relationships among the itemsets that can be widely applied to many tasks and applications, for example, supermarkets' basket-market analytics. It shows the correlations of the purchase of products or itemsets. SPM is also performed to find the relationships of the itemsets but focuses on the ordering sequences regarding timestamps, which can be widely applied to analyze the clickstream of the Webpage or the scheduling problem in industrial applications. Besides, HUIM can reveal the profitable products in supermarkets that consider both unit of profit and the quantity of the items

---

✉ Jerry Chun-Wei Lin  
jerrylin@ieee.org

<sup>1</sup> College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China

<sup>2</sup> Department of Math and CSC, Brandon University, Brandon, Canada

<sup>3</sup> Research Centre for Interneural Computing, Taichung, Taiwan

<sup>4</sup> Department of Computer Engineering, College of Electronics and Information Engineering, Sejong University, Seoul, South Korea

<sup>5</sup> Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway

that bring more effective knowledge for decision making. Several projects were carried out to quantify the necessary information efficiently regarding high-utility itemsets in a database by considering the improvements of mining efficiency and effectiveness of the discovered knowledge in different applications [8]. More than that, security issue [9, 10] regarding pattern analytics in IoT environment [11] has also arisen as an important and meaningful topic in recent years.

HUIM, on the other hand, has a significant weakness in pattern-mining tasks. With the length of a pattern, the utility value of an itemset increases as its size increases. Under this situation, if an item(set) is considered as a high-utility pattern, there is a very high possibility that any combination with this item(set) can also be considered as a high-utility pattern (i.e., any combination with the caviar such as apple or pear can also be considered as an itemset with very high utility in the basket-market analytics). This is inappropriate for pattern evaluation, particularly when the length (size) of a pattern (itemset) is very long. An alternative way to better reveal the utility of a pattern named high average-utility itemset mining (HAUIM) [12] was presented to figure out this limitation and provide better and fair results in pattern mining. Several works [13, 14] have been presented, and most of them only focus on handling the precise datasets. When the collected data include the uncertain factor, such as probability value in industrial applications in the WSN or IoTs, existing works could not handle this situation, and the discovered knowledge could be irrelevant to decision making.

Compare to traditional pattern-mining applications, IoT environments have more uncertainty. Even though the technique to enhance the reliability of IoT equipment has also growth recently, the value collected from the equipments of IoTs could not be considered as a precise value. For example, a temperature sensor is impossible to detect a precise value when it is placed at different locations in a room. Thus, the pattern-mining process should not ignore this limitation to establish knowledge for decision-making; the uncertainty should be involved in pattern analytics regarding IoT environments. PHAUI can reveal useful knowledge (patterns) from IoT environments. For instance, a sensor is designed to detect the various natural environment values. To understand the relationship with temperature, it is possible to apply the PHAUI for discovering the valuable patterns. For the traditional PHUI or HUI, they are used to generate a set of longer patterns, which may lead to produce the meaningless information; the designed PHAUI can be used to solve this limitation. Thus, we then presented APHAUIM, a new framework that considers the level-wise manner to efficiently disclose HAUIs from the collected uncertain databases of WSNs or IoTs in this work. This model is suitable and applicable

in the industrial WSNs or IoTs for further pattern-mining tasks. Main developments are then summarized below.

1. A Potential High Average-Utility Itemset Mining (APHAUIM) based on a level-wise model is introduced to discover the desired potential high average-utility itemsets efficiently from uncertain datasets used in industrial IoTs.
2. To further minimize the scale of the search space for discovering possible patterns of average-utility and uncertainty constraints, two effective upper-bounds are used. Those upper-bounds served as pruning strategies, reducing the search space size and allowing for more efficient discovery of the necessary information.
3. Experimental results indicated that the designed APHAUIM could reveal more significant and non-redundant patterns if the uncertain situation is considered more applicable and suitable in the industrial WSNs or IoTs.

## 2 Related work

As the rapid growth of computer techniques [15], analyzing data obtained from various devices, such as IoTs or WSNs, is not a simple process. Pattern mining is one of the techniques that reveals the relationships of the items in different domains and applications [16, 17]. The discovered information is presented as useful knowledge to show the implicit and potential information for making strategic decisions. The first generic algorithm to discover the correlation of the itemsets in the database is referred as association rule mining (or called ARM), which finds the set of frequent itemsets in the first stage by minimum support threshold then discovers the association rules in the second stage by minimum confidence threshold. As a result, it only deals with the precise dataset [18], however, data quantity and uncertainty exist in real-world domains and industries, which has been ignored and discarded in the generic ARM models. Quantitative factor [19] or interestingness [20] are also the alternative constraints that can be put into the FIM or ARM to reveal more information. The above algorithms handle the precise dataset that cannot be well-performed for handling the uncertain databases. In industrial applications such as manufacturing, the collected or received data may consist of the uncertainty factor; it is not possible to discover suitable knowledge for decision-making if the probability or uncertainty is involved in the pattern-mining tasks. Chui et al. [21] then presented an algorithm to discover the frequent expected patterns. Bernecker et al. [22] presented a model for handling the probabilistic frequent pattern mining from the uncertain databases. Several works regarding uncertain data mining can be studied [23].

In addition to traditional FIM or ARM, utility-oriented concept [24] is considered as an emerging and important task in recent years since it reveals more information for decision-making, especially the unit of profit and the quantity of the items are considered in the pattern-mining task. Compared to the support-based constraint, utility-oriented concept is more appropriate in real-world applications and domains, and the high-utility itemset mining (HUIM) [25] was then studied and implemented to better reveal HUIs for decision-making. Yao et al. [26] proposed the utility issue regarding the internal utility (quantity) and external utility (unit or profit) to mine HUIs. However, this work fails to handle the limitation so-called “combinational explosion”, thus the search space to discover the sufficient patterns is huge. To better solve this limitation, Liu et al. [27] then presented the TWU concept that takes the transaction-weighted utility as the measurement to estimate the value of upper-bound on the patterns. In addition, by investigating the high-transaction weighted utilization itemset (HTWUI) regarding the transaction-weighted downward closure (TWDC) property, the search space can be greatly cut down and the performance can be efficiently improved compared to the generic models. Ahmed et al. [28] then presented an incremental process called IHUP to revise the discovered HUIs in the dynamic database for transaction insertion. Tseng et al. then introduced the UP-growth+ method, which uses a condensed tree structure to shrink the original database and a variety of pruning techniques to shrink the search space for mining the required HUIs [29]. Then HUI-Miner approach proposed by Liu et al. [25] to use a compressed data structure for HUIM. Several extensions of HUIM regarding different topics and constraints are then discussed [30–32].

A key problem of HUIM is that the utility value gets higher along with the length of the itemset. The explanation for this is that even though an item in a high-utility pattern contributes very little utility, this pattern can still be called a HUI. To better reveal a good and an alternative solution for pattern mining regarding utility concept, a method called high average-utility itemset mining (HAUIM) [12, 13] was then proposed and presented that is used to find the average-utility of a pattern in the mining task. The first 2-phase TPAU algorithm [12] was presented to reveal the high average-utility itemsets (HAUIs) in a level-wise manner. It estimates the upper-bound value on the patterns using the average-utility upper-bound (*aub*), downsizing the search space to find the set of HAUIs. A projected algorithm called PAI [33] was presented to speedup mining performance by introducing the projection mechanism. A tree-based algorithm known as HAUP-growth [13] was also investigated to find the collection of HAUIs quickly using a compact tree structure. Besides, this structure can thus

avoid the multiple database scans compared to the level-wise manner. An attached array is utilized for each tree node; thus, the computational cost can also be reduced. An average-utility-list (AU-list) structure [14] was further studied to methodically reveal the set of HAUIs based on a condensed link-list structure. Two pruning strategies were further implemented to diminish the search space for pattern mining. Several extensions [34] were also presented and studied. However, the above works do not consider the uncertainty factor of the data that is collected in WSNs or IoTs. Thus, all the events and activities are treated as the same weight and importance, which is inappropriate for industrial and real-world applications.

In some industrial tasks and applications, the schedule for the manufacturing tasks can be arranged and organized by the knowledge discovered from the WSNs or IoTs [7]. Pattern analytics and mining play an important role to receive up-to-date information for the manufacturing industry [7]. For example, a record in a database can be considered a set of sensors receiving the values from WSNs or IoTs (e.g., temperature degree). Those values can be the abnormality or risk measure regarding the consideration of location or geographic issues. To better analyze the collected data in the industrial applications [16], the uncertainty constraint is one of the major factors that should be concerned for pattern analytics and knowledge evaluation. Several works put the uncertainty constraint into the pattern-mining tasks, for example, combining FIM and uncertainty as uncertain frequent pattern mining [35, 36]; taking the SPM and uncertainty as uncertain sequential pattern mining [37]; or integrating the HUIM and uncertainty as uncertain HUIM [38, 39].

The relevant utility-oriented pattern mining algorithms are organized in Table 1. Obviously, previous works seldom discussed the applications in IoT environments. In [38, 39], Lin et al. involved the uncertainty concept in the utility pattern mining field. As we know, uncertainty is one of the major characteristics in IoT data. Revealing valuable knowledge or information from IoT data is not a trivial task. Compared with the high-utility itemset (pattern) concept, high-average utility itemsets provide more precise and streamlined patterns than that of the high-utility concept. It is crucial to develop strategies or effective knowledge from IoT data. Up to now, fewer works considered the uncertainty factor especially in the high average-utility pattern mining. In addition, traditional utility pattern mining models do not consider the average concept, thus the generated patterns sometimes are too long and become meaningless. Thus, a high potential high average utility itemset algorithm (revealing high average utility itemsets in an uncertainty database) is proposed in this paper that can be adapted in IoT environments.

**Table 1** The mentioned utility mining works in this paper

Paper	Years	Authors	Patterns	Uncertainty
[28]	2009	Ahmed et al.	high-utility itemset	
[13]	2010	Lin et al.	high average-utility itemsets	
[12]	2011	Hong et al.	high average-utility itemsets	
[25]	2012	Liu et al.	high-utility itemsets	
[33]	2012	Lan et al.	high average-utility itemsets	
[29]	2013	Tseng et al.	high-utility itemsets	
[14]	2016	Lin et al.	high average-utility itemsets	
[31]	2016	Lin et al.	high-utility itemsets	
[32]	2016	Lin et al.	high-utility itemsets	
[34]	2017	Lin et al.	high average-utility itemsets	
[38]	2017	Lin et al.	high-utility itemsets	✓
[39]	2020	Ahmed et al.	high-utility itemsets	✓

### 3 Preliminary and problem statement

The first part of this section provides the foundational definitions of potential high average utility itemset mining (PHAUIM). The problem statement for PHAUIM is then presented in the Section 2. The last is PHAUIM's downward closure property.

Note that most of the previous utility research is based on transaction databases in trading market environments. In order to incorporate the earlier definitions in utility pattern mining, we followed the transaction and profit (utility) concepts from the earlier definitions. However, in an IoT environment, a transaction can be denoted as a record collected from sensors, and the value of utility can be indicated as the specific attribution value detected by sensors in a record. This situation also applies to an item in the sensor environments. Thus, an itemset is a combination of attributions in a record from a IoT data. Moreover, in order to simplify the content, the foundational definitions from the previous high-utility itemset mining would not be defined again here but refer to the previously average utility works [12, 13]. These foundational definitions include itemset average utility (denoted as  $au(X)$ , where  $X$  is an itemset), high average utility itemset (denoted as  $HAUI$ ), average utility upper bound (denoted as  $auub$ ), and high average utility upper bound itemsets (denoted as  $HAUUBI$ ). The other foundational definitions using in this work are provided in the following parts.

**Definition 1 (Itemset and Transaction Probability)** The corresponding probability  $p(X, T)$  of the itemset  $X$  in the transaction  $T$  is set as the same value of the corresponding

probability  $p(T)$  in the transaction  $T$ . The formula is given below:

$$p(X, T) = p(T) \quad (1)$$

**Definition 2 (Itemset Potential Probability in a Database)** The potential probability  $Pro(X)$  of the itemset  $X$  in the database  $D$  is defined below:

$$Pro(X) = \sum_{X \subseteq T \wedge T \in D} p(X, T) \quad (2)$$

**Definition 3 (High Probability Itemset, HPI)** If an itemset  $X$  in the database  $D$  is a high probability itemset (HPI) iff the following condition is satisfied:

$$Pro(X) \geq \mu \times |D|, \quad (3)$$

where  $\mu$  is the predefined threshold of the minimum potential probability.

HPIs are different from HAUIs, since HPIs do not need an upper-bound to keep the downward closure property. Thus, if the itemset  $X$  is not a HPI, all of the supersets of  $X$  is impossible as a HPI. Therefore, it can be applied in an Apriori-based algorithm directly.

**Definition 4 (Potential High Average Utility Itemset, PHAUI)** If an itemset  $X$  in the database  $D$  is a high potential average utility itemset (PHAUI) iff it is both a HAUI and a HPI simultaneously.

This work first proposed the concept of PHAUIs, and the proposed methods perform a Apriori-based process to reveal all of PHAUIs in a database. To keep the downward

closure property to reveal PHAUI, a superset of PHAUIs name as potential high average utility upper bound itemset (PHAUUBI) is introduced below.

**Definition 5 (Potential High Average Utility Upper Bound Itemsets, PHAUBI)** If an itemset  $X$  in the database  $D$  is a high potential average utility upper bound itemset (HAUUBI) iff it is both a HAUUBI and HPI simultaneously.

Due to the above definitions, PHAUBIs is a superset of PHAUIs and also keep the downward closure property. Therefore, it is used to produce some candidate PHAUIs in order to narrow down the search space. However, the traditional  $auub$  is a very large and loose value thus the search space to find the promising HAUIs is very huge. This is reasonable since the generic  $auub$  does not include its own value. For example, suppose two itemsets  $A$  and  $B$  exist in a transaction, and based on  $auub$  model, the itemsets  $A$  and  $B$  have the same  $auub$  value in this transaction. Thus, a partial-upper bound called  $pub$  is then designed to solve this limitation and is explained as follows.

**Definition 6 (Active Item)** Let  $ai$  and  $ais$  respectively be the active item and active itemset in the designed algorithm for mining the HAUIs in a level-wise manner. We can have that the  $k$ -itemsets can be used for the next generation of  $(k+1)$ -itemsets, which is the generic progress of the Apriori-like approach. Thus, we can also assume an item or an itemset does not exist in the  $k$ -itemsets, it would not be involved in the next generation of  $(k+1)$ -itemsets. For the most Apriori-like algorithms, this model is used to generate the  $n$  itemsets from  $(n-1)$ -itemsets, and it is also called the downward closure property for pattern-mining progress.

Assume estimate the set of PHAUBIs with length  $n$ , an active item ( $ai$ ) shows that this specific item must appear in one of the PHAUBIs with the  $(n-1)$  length. We then can obtain the extended definition of active itemset of the candidate PHAUBIs with the length is  $n$  (called  $ais_n$ ) as follows:

$$ais_n = \left\{ ai \in x \mid x \in PHAUBIs^{n-1} \right\}, \tag{4}$$

where  $PHAUBIs^{n-1}$  can be considered as a PHAUBI with the  $(n-1)$  length.

For instance, suppose the itemsets  $\{A,B\}$ ,  $\{A,C\}$  and  $\{B,C\}$  are the discovered itemsets of the last iteration. In this situation, the active itemset for the next generation of PHAUBIs is considered as  $ais_3 = \{A,B,C\}$  with the length 3.

Thus, we can conclude that the active items can be used in the proposed model as the set of  $ais$  that holds the

downward closure property of the PHAUBIs. So that if an itemset is considered as a PHAUBI, of course, any subset of it is also considered as a PHAUBI.

**Definition 7 (Remaining Maximal Utility)** Suppose the remaining maximal utility of the active itemset in a transaction is then denoted as  $rmua$  in the proposed algorithm. Suppose that an itemset  $I = \{i_1, i_2, \dots, i_n\}$  in a transaction  $T = \{i_t, u_t\}$ ;  $I_t = \{i_{t1}, i_{t2}, \dots, i_{tm}\}$  can be considered as the purchase products in this record (or transaction in the basket-market dataset), and  $U_t = \{u_t(i_{t1}) = u_{t1}, u_{t2}, \dots, u_{tm}\}$  can be considered as the corresponding utility for all items existed in this transaction. Suppose we have the active itemsets called  $ais$ . The  $rmua$  of an itemset  $i$  in a transaction  $t$  is denoted as  $rmua(i, t)$  and defined below:

$$rmua(i, t) = \max \{u_t(i) \mid i \in (ais \cap i_t) \setminus i\} \tag{5}$$

**Definition 8 (Partial Maximal Utility Upper Bound)** Let partial maximal utility upper bound be denoted as  $(pub)$ . Suppose that an itemset is denoted as  $i$  and a record/transaction is denoted as  $D$ . Let  $pb_i$  be the  $pub$  value of  $i$  in  $D$ , that can also be defined below:

$$pb_i = \sum_{t \in D} pb_i^t \tag{6}$$

**Lemma 1** The anti-monotonicity property of  $pub$ .

Suppose that an itemset  $I$  is defined as  $I = \{i_1, i_2, \dots, i_n\}$  in a transaction  $T = \{i_t, u_t\}$ . In this case, the active itemsets are denoted as  $ais$ , and a superset of  $I$ ,  $I' = \{i_1, i_2, \dots, i_n, i_{n+1}, \dots, i_o\}$ , in which  $\forall x \in \{i_{n+1}, \dots, i_o\} \rightarrow x \in ais \setminus I$ , such that  $mu(I, T) \geq pb_i^t \geq au(I', T)$ .

**Lemma 2** The downward closure property of  $pub$  is stated as follows.

Assume an itemset  $I = \{i_1, i_2, \dots, i_n\}$  in a dataset  $D = \{t_1, t_2, \dots, t_2\}$ , active itemsets =  $ais$ , and a superset of  $i$ ,  $I' = \{i_1, i_2, \dots, i_n, i_{n+1}, \dots, i_o\}$ , in which  $\forall x \in \{i_{n+1}, \dots, i_o\} \rightarrow x \in ais \setminus i$ , such that  $auub(i) \geq pb_i \geq au(i')$ .

**Definition 9 (High Partial Upper Bound Itemset, HPUBI)** Let high partial upper bound itemset be denoted as  $pubi$ . Suppose an pre-defined threshold value is set as  $\varepsilon \times TU$ . An itemset  $i$  is considered as a  $pubi$  if it satisfies the condition as:

$$pb_i \geq \varepsilon \times TU \tag{7}$$

We can ensure that the constructed algorithm can retain the downward closure property for pattern generation based on the Apriori-like approach using the proposed Lemma

2. As a result, *pubi* collection can be used for the next generation of super-itemsets. As a result of the designed lemma, the search space can be greatly reduced, and the computational cost can be greatly reduced.

Since *pubi* can be used to reduce the size of the search space in the pattern-mining tasks of candidate generation, the subset of lead partial upper bound (named *lead-pub*) is also introduced here that can be further utilized in the designed algorithm and reduce the search space for candidate generation. In this situation, the orders of itemsets should be first pre-defined.

**Definition 10 (Lead Partial Upper Bound Itemset, *lead-pubi*)** Suppose that a predefined item order  $l$  is assumed initially and a candidate itemset is considered as  $I_c = \{i_1, i_2, \dots, i_n\}$  following the order of  $l$ . For the level-wise manner to for candidate generation progress, this candidate must be one of the subset of *pubi*, and the set of *pubi* is defined below:

$$\{\{i_1, i_2, \dots, i_{n-1}\}, \{i_1, i_2, \dots, i_{n-2}, i_n\}, \dots, \{i_2, i_3, \dots, i_n\}\} \quad (8)$$

In this case, the first element of this set is one of the *lead-pubis*.

Based on the above definition, we can have that a smaller upper-bound than that of the *pub* of a candidate can be defined as follows. This upper-bound of remaining maximal utility of active itemset can be used to further minimize the size of search space.

**Definition 11 (Remaining Maximal Utility of Active itemset, *lrmua*)** Suppose that an itemset is considered as  $I = \{i_1, i_2, \dots, i_n\}$  and a transaction in the database is defined as  $T = \{i_t, u_t\}$ .  $I_t = \{i_{t1}, i_{t2}, \dots, i_{tm}\}$ . Each transaction consists of the purchase items in the database. Suppose that  $U_t = \{u_t(i_{t1}) = u_{t1}, u_{t2}, \dots, u_{tm}\}$  is considered as the corresponding utility for all items existed in this transaction. Based on the above definitions and properties, the active itemsets is denoted as *ais* and a predefined item order  $L = \{i_1^l, i_2^l, \dots, i_p^l\}$ . Suppose that  $i_n = i_w^l$ , and set an itemset  $s = \{i_{w+1}^l, i_{w+2}^l, \dots, i_p^l\}$ . Thus, the *lrmua* of  $i$  in  $T$  is denoted as  $lrmua(i, T)$ , which is defined below:

$$lrmua(i, T) = \max \{u_t(i) | i \in (ais \cap i_t \cap s) \setminus i\}. \quad (9)$$

**Definition 12 (Lead Partial Maximal Utility Upper Bound for an Itemset in a Transaction)** Suppose that an itemset  $I$  is considered as  $I = \{i_1, i_2, \dots, i_n\}$  in a transaction  $T = \{i_t, u_t\}$ . Again, the  $i_t = \{i_{t1}, i_{t2}, \dots, i_{tm}\}$  is the purchase items in this transaction,

and  $U_t = \{u_t(i_{t1}) = u_{t1}, u_{t2}, \dots, u_{tm}\}$  is considered as the corresponding utility of all items existed in this transaction. In this case, we can have that the active itemsets is denoted as *ais*. Also, we can suppose that the predefined item order is set as:  $L = \{i_1^l, i_2^l, \dots, i_p^l\}$ . The *lead-pub* of  $i$  in  $T$  is denoted as  $lpb_i^T$ , and defined below:

$$lpb_i^l = \begin{cases} \frac{u(i,T)+m \times lrmua(i,T)}{|i|+m} & , \text{ if } lrmua(i, T) > au(i, T) \\ \frac{u(i,T)+lrmua(i,T)}{|i|+1} & , \text{ if } 0 < lrmua(i, T) \leq au(i, T) \\ 0 & , \text{ if } lrmua(i, T) = 0, \end{cases} \quad (10)$$

where  $m$  is considered as the number of  $(ais \cap i_t \cap s) \setminus i$ .

**Definition 13 (Lead Partial Maximal Utility Upper Bound for an Itemset)** Suppose that an itemset is denoted as  $i$  and a database is denoted as  $D$ . The *lead-pub* of  $i$  in  $D$  can thus be denoted as  $lpb_i$ , which is defined below:

$$lpb_i = \sum_{t \in D} lpb_i^T \quad (11)$$

**Definition 14 (Lead High Partial Upper Bound Itemset, *LHPUBI*)** Suppose that a pre-defined threshold is set as  $\varepsilon \times TU$ , an itemset  $i$  is considered as a *lead-pubi* if it holds the condition as:

$$lpb_i \geq \varepsilon \times TU \quad (12)$$

In conclusion, the *lead-pubi* holds the similar definition as the *pubi*. That is, the *lead-pubi* can also hold the same downward closure property as *pubi*. However, the number of the candidates in the set of *lead-pubi* is much smaller than that of the *pubi* since  $lead-pubi \subset pubi$ . Thus, the *lead-pubi* can be used to further reduce the search space of the potential candidates. Based on the above definitions of two upper-bound values, two candidate itemsets are considered in the designed model that can be utilized by the level-wise manner (or so-called Apriori-like approach). Those two candidates for finding the required patterns are then defined and stated below.

**Definition 15 (Potential High Partial Upper Bound Itemset, *PPUBI*)** The itemset  $X$  in the database  $D$  is a potential high partial upper-bound itemset (PPUBI) iff it is both a HPUBI and HPI simultaneously.

**Definition 16 (Potential LEAD High Partial Upper Bound Itemset, *PLPUBI*)** The itemset  $X$  in the database

$D$  is a potential lead high partial upper bound itemset (PLPUBI) iff it is both a LHPUBI and HPI simultaneously.

**Problem statement** The intention of PHAUM is revealing all of the PHAUMs in an uncertain dataset. Suppose that an input dataset is set as  $D$ , the unit profit table of the itemsets is set as  $p$ , the pre-defined utility threshold is set as  $\varepsilon$ , a pre-defined potential probability threshold is set as  $\mu$ . The output of the designed algorithm is to discover the set of potential high average-utility (PHAUMs) from the uncertain database  $D$ .

---

#### Algorithm 1 APHAUM.

---

**Input:**  $D$ , a dataset;  $\varepsilon$ , a predefined minimum utility threshold and  $\mu$ , a pre-defined minimum potential probability threshold.

**Output:** PHAUMs, a set of potential high utility itemset.

```

1: set  $PH = \emptyset$  for PHAUMs;
2: for each transaction (record)  $t$  in the database  $D$  do
3:   calculate the maximal utility  $m$  in  $t$ ;
4:   for each item (attribution)  $i$  in  $t$  do
5:     set  $A[i] = A[i] + m$ ;
6:     calculate  $au(i) = au(i) + u(i, t)$ ;
7:     calculate  $Pro(i) = Pro(i) + p(t)$ ;
8:   end for
9: end for
10: for each item (attribution)  $i$  do
11:   if  $Pro(i) \geq \mu \times |D|$  then
12:     if  $au(i) \geq \varepsilon \times TU$  then
13:       PHAUMs  $\leftarrow i$ ;
14:     end if
15:     set  $PPUBIs \leftarrow \emptyset$ ;
16:     set  $PLPUBIs \leftarrow \emptyset$ ;
17:     if  $A[i] \geq \varepsilon \times TU$  then
18:        $PPUBIs \leftarrow i$ ;
19:        $PLPUBIs \leftarrow i$ ;
20:     end if
21:   end if
22: end for
23: while  $PPUBIs \neq \emptyset$  or  $PLPUBIs \neq \emptyset$  do
24:   for each itemset (pattern)  $e$  in  $PLPUBIs$  do
25:     scan (combine)  $pubs$  by  $e$  to generate candidate
itemsets  $\rightarrow C$  ▷ By  $PUPUBIs$  and  $(PPUBIs)$ 
26:   end for
27:   set  $PPUBIs \leftarrow \emptyset$ ;
28:   set  $PLPUBIs \leftarrow \emptyset$ ;
29:   for each transaction (record)  $t$  in  $D$  do
30:     scan the information of utility and probability for
each candidate itemset (pattern) in  $t$ ;
31:   end for
32:   for each candidate itemset (pattern)  $c$  in  $C$  do
33:     if  $Pro(c) \geq \mu \times |D|$  then
34:       if  $pub_c \geq p$  then
35:          $PPUBIs \leftarrow c$ ;
36:       end if
37:       if  $lpub_c \geq \varepsilon \times TU$  then
38:          $PLPUBIs \leftarrow c$ ;
39:       end if
40:       if  $au(c) \geq \varepsilon \times TU$  then
41:         PHAUMs  $\leftarrow c$ ;
42:       end if
43:     end if
44:   end for
45: end while
46: return PHAUMs;

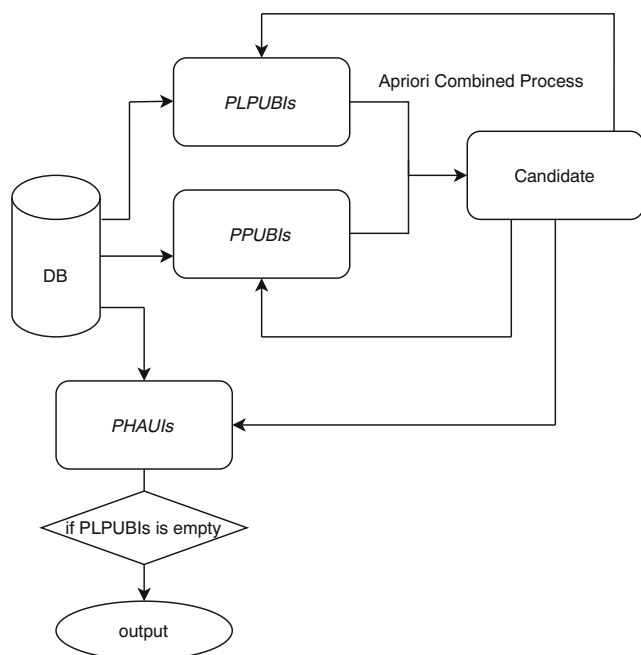
```

---

## 4 Designed Apriori-based PHAUM, APHAUM

The detailed algorithm of the developed APHAUM is described as follows. The proposed APHAUM utilized two tighter upper-bounds such as  $pub$  and  $lead-pub$  to hold the downward closure property. Based on those two upper-bounds, the designed algorithm maintains two itemsets respectively called  $phps$  and  $plhps$  to maintain the downward closure property. The developed algorithm can thus efficiently reveal all the satisfied PHAUMs by holding the completeness and correctness based on the maintained two itemsets. The pseudo-codes of the developed algorithm are respectively described in Algorithm 1. Note that all the acronyms used in the developed algorithm are then shown in the Appendix of Table 3. Due to the same reason mentioned in the definition section, we also made use of the concepts from the applications in trading markets. The pseudo-code uses the definitions from transaction databases like the word “transaction” and “itemset”. For instance, a transaction can also be transferred into a record in a IoT database.

The proposed method (Algorithms 1) is a standard Apriori-based process to find all PHAUMs in a database. In Algorithm 1, the proposed APHAUM first calculates the maximal utility for each transaction, utility information and probability information for each item by a loop (lines 2 to 9). Then the proposed method finds the PHAUMs with one item and the initial  $PPUBIs$  and  $PLPUBIs$  in lines 15 and 22. Note that, all of itemsets are estimated if its probability is larger than  $\mu \times |D|$  (line 11). The loop function in line 23 reveals PHAUMs with different items. Algorithms 1 applies the upper-bounds  $pub$  and  $lead-pub$  to reduce the searching space. In fact, we can only use  $pub$  in the algorithm. When the proposed method applies  $lead-pub$  concept in the process, the first  $n-1$  items (a sub itemset) in the combined itemset (with  $n$  items) need to be existed in  $PLPUBIs$ . For example, if a combined itemset (candidate itemset) is  $\{a, b, c\}$ , the itemset  $\{a, b\}$  must exist in  $PLPUBIs$ ,  $\{b, c\}$  and  $\{a, c\}$  are existed in  $PPUBIs$ . In the loop function started at line 23, the proposed APHAUM first generates candidate itemsets from the previous round  $PPUBIs$  and  $PLPUBIs$  (lines 24 to 26). Then the new  $PPUBIs$  and  $PLPUBIs$  for the next round in line 27 and line 28 are performed. In lines 29 to 31, the progress rescans the database to check the utility information and probability information for each candidate itemset. The following loop addresses the candidate itemsets one by one to the set of  $PPUBIs$ ,  $PLPUBIs$  and  $PHAUMs$ . In order to keep the downward closure property of high-potential itemsets, all of itemsets are estimated if its probability is larger than  $\mu \times |D|$  (line 33). In the case of only using  $pub$  upper-bound in the process, the algorithm does not need to maintain the set of  $PLPUBI$  and use  $PPUBIs$  instead of  $PLPUBIs$  in line 24. In the experiments, the performance



**Fig. 1** The simple flowchart of the proposed method

of these kinds of situations (one upper-bound and two upper-bounds) will be compared and evaluated. Figure 1 shows a simple flowchart of the proposed method. The proposed method generates Potential LEAD High Partial Upper Bound Itemset (*PLPUBIs*) and Potential High Partial Upper Bound Itemset (*PPUBIs*) to obtain candidate itemsets based on the Apriori-like method. Then the process will check the candidate itemsets to see whether it is Potential High Average Utility Itemset (*PHAUIs*) (also generates *PLPUBIs* and *PPUBIs* for the next round). If *PLPUBIs* is an empty set, then the progress is terminated and the output is produced *PHAUIs*.

## 5 Experimental evaluation

In this section, the proposed APHAUIMs showed the performance to reveal PHAUIs efficiently. APHAUIM(I) is the proposed method with *lead-pub* and APHAUIM is the proposed method without *lead-pub*. The experiments also showed the results of the previous AHAUIM (Apriori-based HAUIM) [12] in revealing HAUIs. Results also showed the performance of APHAUIM and AHAUIM in four standard datasets that have been widely used in utility-oriented pattern-mining tasks [40]. Table 2 states the properties of the four used datasets datasets in the experimental evaluation. All algorithms were performed by Mac PC, running on Apple M1 processor with 8GB main memory by macOS Big Sur OS. In additions, the algorithms are then written by Java language and will be released to public for the further usage.

**Table 2** Properties of used datasets

Dataset	# D	# I	AvgLen	Density
Mushroom	8,124	119	23	19.3277%
Foodmart	21,557	1,559	4	0.2838%
Chess	3,196	75	37	49.33333%
Accidents	34,018	468	34	7.2239%

### 5.1 Runtime

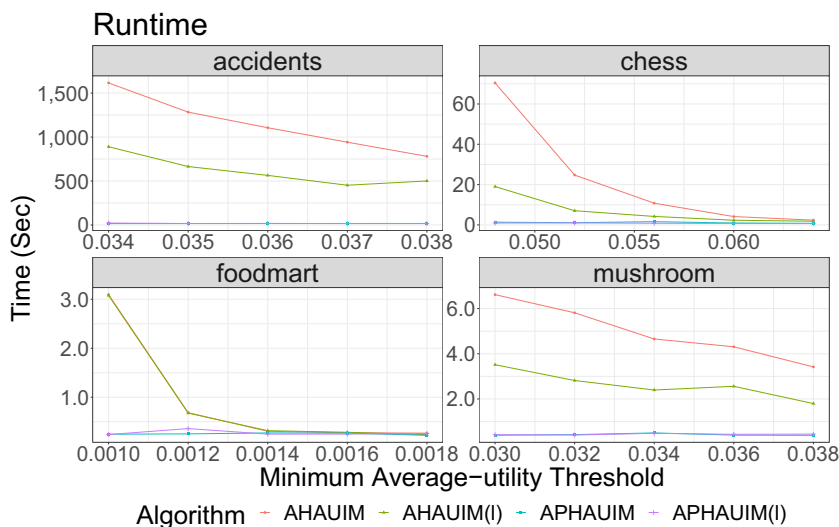
In this section, the runtimes of the proposed APHAUIM and the previous AHAUIM with different minimum average-utility thresholds in the four datasets from SPMF [40] are then evaluated and compared. The results of the experiments are indicated in Fig. 2. Obviously, the algorithms applied *lead-pub* usually have better performance than the algorithms without *lead-pub*. However, the influence of *lead-pub* is more conspicuous in revealing HAUIs than revealing PHAUIs. Applying *lead-pub* is also suffered the computation cost in calculating one more upper-bound. Due to fewer candidate itemsets in PHAUIs, the benefit of *lead-pub* is not very obvious, especially in a small and loose database. The runtime of APHAUIM(I) is even more than APHAUIM in foodmart database with minimum average-utility threshold 0.0012. However, in most situations, *lead-pub* is a useful technique to reduce the computational cost in the whole process. Thus, the designed model is acceptable and efficient to mine the required information (e.g., potential high average-utility itemsets) from the uncertain databases.

### 5.2 Number of candidates

In this section, the amount of candidates (PHAUIs and HAUIs) for the proposed APHAUIM and the previous AHAUIM with different minimum average-utility thresholds in the four datasets are compared and evaluated. The results of the experiments are indicated in Fig. 3. The first point of these experiments is the different numbers of the candidates of PHAUIs and HAUIs. In PHAUI mining algorithm, there is one more upper-bound (limitation) than the traditional HAUIM mining algorithms. Therefore, the number of candidates in PHAUI mining algorithms is significantly less than the candidates of HAUIs. Note that the proposed method utilized and applied the downward closure property of probability value to generate the candidate itemsets in the Apriori process. The second point of this section is showing the power of *lead-pub*. Whether in the field of PHAUI or HAUIM, *lead-pub* can always reduce the numerous candidates in the mining process. It is always useful to reduce the computation time in the process. Moreover, the effect of *lead-pub* is more effective in



**Fig. 2** Runtimes for different minimum average-utility thresholds



a loosen minimum average-utility threshold. The Apriori-based algorithms without using *lead-pub* will produce a massive number of candidate itemsets when the value of the threshold is small. The *lead-pub* can control and suppress the increase of candidate itemsets effectively when the value threshold is declining down. Thus, it is necessary to apply *lead-pub* when a mining process is set a small value of the threshold.

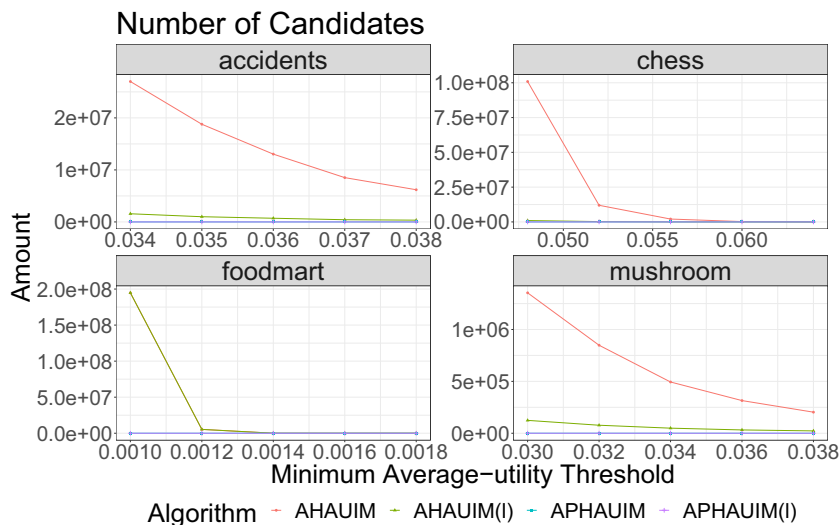
### 5.3 Discovered PHAUIs or HAUIs

In this section, the number of PHAUIs or HAUIs with different minimum average-utility thresholds in the four datasets are compared and shown in Fig. 4. The experimental results showed that if we apply a traditional HAUIs mining algorithm in an uncertainty database, it will generate many overestimated itemsets in the results. If the average probability of the whole uncertainty database

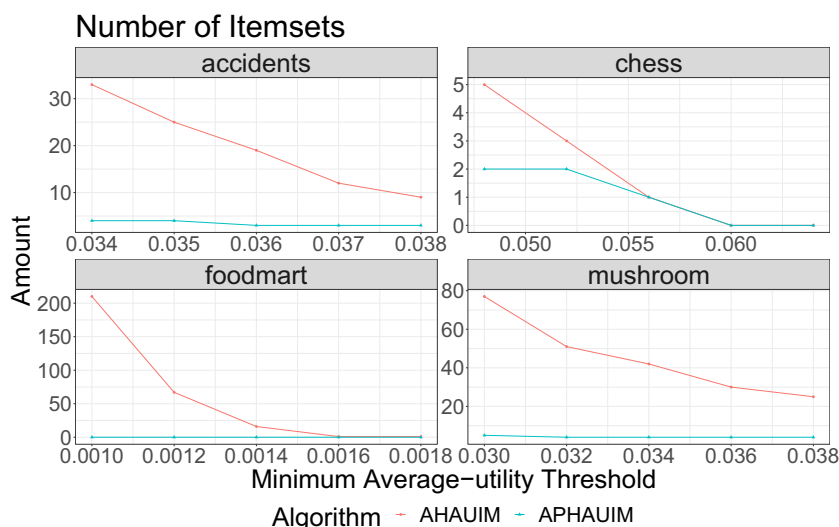
is low, the meaning of the revealed itemsets is unusable, and the real usable knowledge will be hidden in the huge number of itemsets. That is to say, if the level of uncertainty is very high in an uncertainty database, the previous HAUI algorithms could not be applied to discover effective and valuable knowledge. After involving in the concept of the probability, the proposed method can precisely estimate interesting itemsets with an uncertain probability. Especially for a large minimum average-utility threshold, the different numbers of the candidates of PHAUIs between HAUIs are larger than ten million, and this big gap shows the effectiveness of the designed model.

Generally, the designed model is capable and effective to handle the mining task regarding the utility and uncertainty constraint, which is very applicable in the WSNs or IoTs scenario. Also, the discovered patterns are meaningful in the WSNs or IoTs since the uncertainty factor should be considered and implemented while evaluating the

**Fig. 3** Numbers of candidates for different minimum average-utility thresholds



**Fig. 4** Numbers of revealing itemsets for different minimum average-utility thresholds



effectiveness of the patterns especially the average-utility model provides an alternative and fair measurement to establish the value of the pattern regardless the length of it. In addition, the designed upper-bound models are also effective to accelerate the mining progress compared to the generic models in the task of HAUIM.

### 5.4 Memory usage

This experimental section shows the memory usage of the compared algorithms. Due to the Java environment, it is very hard to evaluate the precise memory usage of running processes. Here we monitored the maximal memory usage for the Java machine when a specific algorithm is performed. Results are then shown in Fig. 5.

From the results, we can observe that due to the small searching space, the memory usages of PHAUIM-based algorithms are less than HAUIM-based approaches. However,

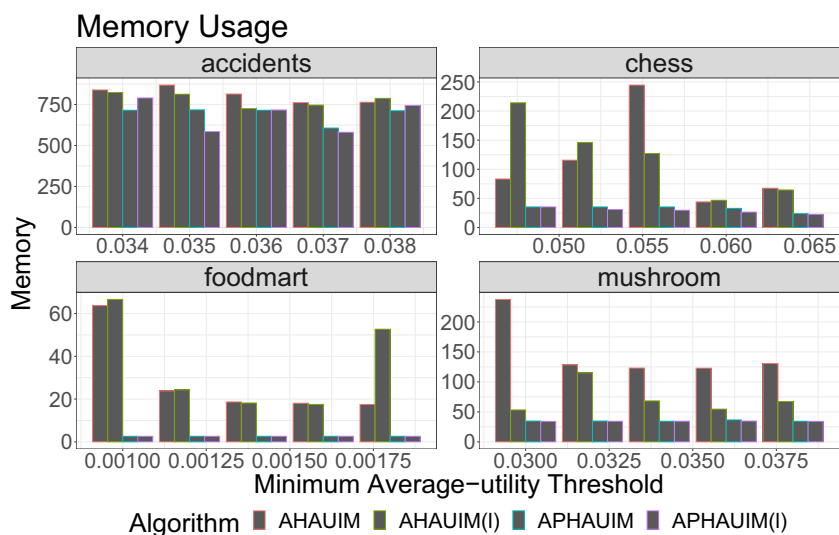
the influence of the proposed *lead-pub* is not clear in the memory usage. Even though *lead-pub* can effectively reduce the number of candidates, it however, still suffers the extra computation process and a more memory requirement. Normally, if a mining process needs more runtime, it also needs more memory to handle the entire progress for obtaining the final results.

### 5.5 Scalability

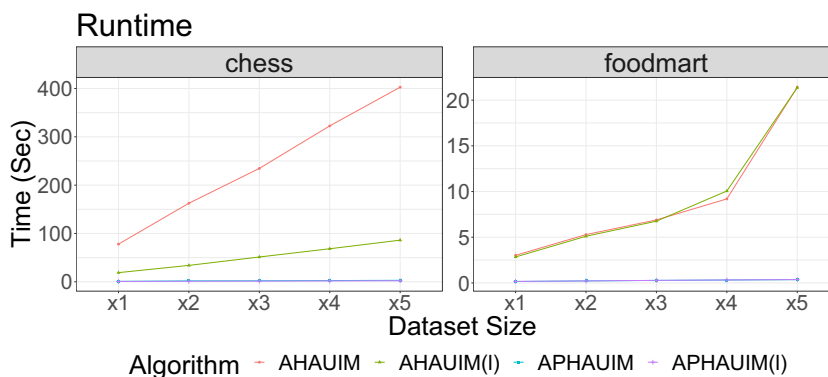
Following the results of the previous section, the scalability of the proposed PHAUIM is discussed in this section. Figures 6 and 7 showed runtimes and the number of candidates of the compared algorithms in terms of the different scales of database sizes.

From the experimental results, if *lead-pub* is not considered, a large-scale database size and high-density database should logically requires more time to reveal

**Fig. 5** Memory usages for different minimum average-utility thresholds



**Fig. 6** Runtimes for different data scales



HAUIs or PHAIUs. However, these two factors did not show a high relationship with runtimes in the experimental results. Even though a database has more candidates, but it is possible to spend less time than a database that has fewer candidates. In Fig. 6, the relationship between the average length and the runtime for a database is obviously higher than the other factors. That is because the scanning process of Apriori should take more time. There is no doubt that HAUIs and PHAIUs methods have this same characteristic in the above conclusion.

In addition, the proposed *lead-pub* should have the outstanding performance in most situations. It should be noted that if a database has a low average length and a low density, a well-defined pruning strategy is unfruitful (for example, in foodmart dataset). The difference between a number of normal patterns and a number of HAUIs (PHAUIs) is huge. Thus, a precise cutting method usually cannot diminish the searching space effectively. Fortunately, this database type does not require higher computational sources in most cases.

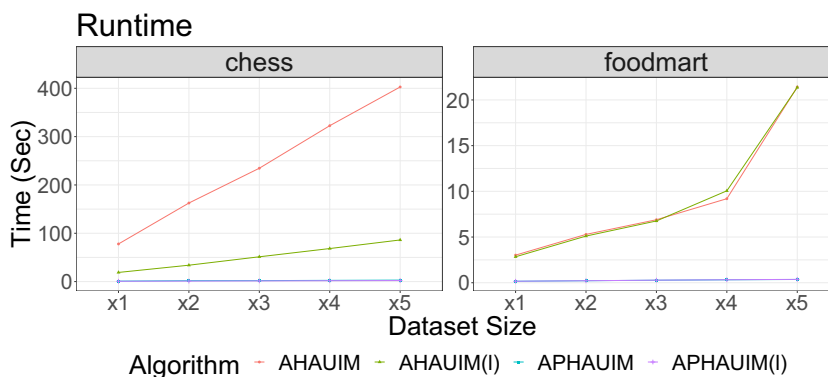
Finally, due to the simple duplication of the original database, the numbers of candidates are absolutely fixed. Figure 7 shows this characteristic and makes sure that the proposed method can work well in large-scale databases.

### 6 Conclusion

In recent decades, the collected data from the wireless sensor networks (WSNs) and Internet of Things (IoTs) have dramatically increased; thus, the uncertainty in the database is necessary to be considered for the pattern-mining task, especially for some tasks in the industrial applications and domains. However, most existing algorithms in pattern mining especially for the HAUIM cannot be performed to handle the collected data with uncertainty, which is inappropriate in real situations. In this work, a new model called Apriori-based potential high average-utility itemset mining (APHAUIM) is presented to reveal the potential high average-utility patterns from an uncertain database. Experimental evaluation is then performed to present the effectiveness and efficiency of the developed algorithms compared to the generic approach of HAUIM, and the designed APHAUIM with two upper-bounds (pruning strategies) is acceptable for performance evaluation used in the uncertain IoT databases.

In the future, a compressed data structure will be explored to solve the limitation of the level-wise progress for pattern-mining tasks especially the tree or list structures could be involved and concerned to improve the mining performance. More online or stream tasks in the real

**Fig. 7** Numbers of candidates for different data scales



industry can also be extended and considered for further studies. Besides, it is also an interesting topic to consider the dynamic environments to develop the efficient algorithms regarding transaction insertion, deletion, and modification in the databases.

## Appendix

**Table 3** Acronym/Notation Table

Acronym	Notation
ARM	association-rule mining
AU-list	average-utility-list
auub	average utility upper bound
FIM	frequent itemset mining
HAUIM	high-average-utility itemset mining
HAUUBI	high average utility upper bound itemset
HAUI	high average-utility itemset
HUI	high-utility itemset
HUIM	high utility itemset mining
HUPM	high-utility pattern mining
HPI	high probability itemset
HPUBI	high partial upper bound itemset
HTWUI	high-transaction weighted utilization itemset
IoT	Internet of Thing
LHPUBI	lead high partial upper bound itemset
PHAUIM	potential high average utility itemset mining
PHAU	potential high average utility itemset
PHAUUBI	potential high average utility upper bound itemset
PPUBI	potential high partial upper bound itemset
PLPUBI	potential lead high partial upper bound itemset
SPM	sequential-pattern mining
TWDC	transaction-weighted downward closure
TWU	transaction weighted utilization
WSN	Wireless Sensor Network

**Acknowledgment** This research is supported by Shandong Provincial Natural Science Foundation (ZR201911150391).

**Funding** Open access funding provided by Western Norway University Of Applied Sciences.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended

use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Maddikunta PKR, Gadekallu TR, Kaluri R, Srivastava G, Parizi RM, Khan MS (2020) Green communication in iot networks using a hybrid optimization algorithm. *Comput Commun*
- Maddikunta PKR, Srivastava G, Gadekallu TR, Deepa ND, Boopathy P (2020) Predictive model for battery life in iot networks. *IET Intell Transp Syst* 14(11):1388–1395
- Agrawal R, Srikant R (1994) Fast algorithms for mining association rules in large databases. In: *International conference on very large data bases*, pp 619–624
- Agrawal R, Srikant R (1995) Mining sequential patterns. In: *IEEE international conference on data engineering*, pp 3–14
- Liu Y, Liao W, Choudhary A (2005) A two-phase algorithm for fast discovery of high utility itemsets. In: *The pacific-asia conference on knowledge discovery and data mining*, pp 689–695
- Yao H, Hamilton HJ, Butz CJ (2004) A foundational approach to mining itemset utilities from databases. In: *SIAM international conference on data mining*, pp 211–225
- Chen F, Deng P, Wan J, Zhang D, Vasilakos AV, Rong X (2015) Data mining for the internet of things: literature review and challenges. *Int J Distrib Sensor Netw* 11(8):431047
- Gan W, Lin JCWL, Fournier-Viger P, Chao HCC, Tseng V, Yu P (2019) A survey of utility-oriented pattern mining. *IEEE Trans Knowl Data Eng*
- Lin CW, Hong TP, Hsu HC (2014) Reducing side effects of hiding sensitive itemsets in privacy preserving data mining. *Sci World J* 2014:1–12
- Lin CW, Hong TP, Yang KT, Wang SL (2015) The ga-based algorithms for optimizing hiding sensitive itemsets through transaction deletion. *Appl Intell* 42:210–230
- Lin JCW, Srivastava G, Zhang Y, Djenouri Y, Aloqaily M (2020) Privacy preserving multi-objective sanitization model in 6g iot environments. *IEEE Internet of Things J*
- Hong TP, Lee CH, Wang SL (2011) Effective utility mining with the measure of average utility. *Expert Syst Appl* 38(7):8259–8265
- Lin CW, Hong TP, Lu WH (2010) Efficiently mining high average utility itemsets with a tree structure. *Lect Notes Comput Sci* 5990:131–139
- Lin JCW, Li T, Fournier-Viger P, Hong TP, Zhan J, Voznak M (2016) An efficient algorithm to mine high average-utility itemsets. *Adv Eng Inform* 30(2):233–243
- Lin JCW, Shao Y, Djenouri Y, Yun U (2020) Asrnn: A recurrent neural network with an attention model for sequence labeling. *Knowl-Based Syst*
- Wan J, Tang S, Li D, Wang S, Liu C, Abbas H, Vasilakos AV (2017) A manufacturing big data solution for active preventive maintenance. *IEEE Trans Ind Inform* 13(4):2039–2047
- Liu Y, Weng X, Wan J, Yue X, Song H, Vasilakos AV (2017) Exploring data validity in transportation systems for smart cities. *IEEE Commun Mag* 55(5):26–33
- Han J, Pei J, Yin Y, Mao R (2004) Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Min Knowl Disc* 8(1):53–87
- Hong TP, Kuo CS, Chi SC (1999) Mining association rules from quantitative data. *Intell Data Anal* 3(5):363–376

20. Geng L, Hamilton HJ (2006) Interestingness measures for data mining: A survey. *ACM Comput Surv* 38(3):9
21. Chui CK, Kao B, Hung E (2007) Mining frequent itemsets from uncertain data. In: Pacific-asia conference on knowledge discovery and data mining. Springer, pp 47–58
22. Bernecker T, Kriegel HP, Renz M, Verhein F, Zuefle A (2009) Probabilistic frequent itemset mining in uncertain databases. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 119–128
23. Aggarwal CC, Yu PS (2009) A survey of uncertain data algorithms and applications. *IEEE Trans Knowl Data Eng* 21(5):609–623
24. Gan W, Lin JCW, Fournier-Viger P, Chao HC, Tseng VS, Yu PS (2019) A survey of utility-oriented pattern mining. *IEEE Trans Knowl Data Eng* :1–22. <https://doi.org/10.1109/TKDE.2019.2942594>
25. Liu M, Qu J (2012) Mining high utility itemsets without candidate generation. In: Proceedings of the 21st ACM international conference on information and knowledge management. ACM, pp 55–64
26. Yao H, Hamilton HJ, Butz CJ (2004) A foundational approach to mining itemset utilities from databases. In: Proceedings of the SIAM international conference on data mining. SIAM, pp 482–486
27. Liu Y, Liao WK, Choudhary A (2005) A two-phase algorithm for fast discovery of high utility itemsets. In: Pacific-asia conference on knowledge discovery and data mining. Springer, pp 689–695
28. Ahmed CF, Tanbeer SK, Jeong BS, Lee YK (2009) Efficient tree structures for high utility pattern mining in incremental databases. *IEEE Trans Knowl Data Eng* 21(12):1708–1721
29. Tseng VS, Shie BE, Wu CW, Yu PS (2013) Efficient algorithms for mining high utility itemsets from transactional databases. *IEEE Trans Knowl Data Eng* 25(8):1772–1786
30. Gan W, Lin JCW, Fournier-Viger P, Chao HC, Hong TP, Fujita H (2018) A survey of incremental high-utility itemset mining. *Wiley Interdiscipl Rev Data Min Knowl Discov* 8(2)
31. Lin JCW, Gan W, Fournier-Viger P, Hong TP, Tseng VS (2016) Fast algorithms for mining high-utility itemsets with various discount strategies. *Adv Eng Inform* 30(2):109–126
32. Lin JCW, Fournier-Viger P, Gan W (2016) FHN: An efficient algorithm for mining high-utility itemsets with negative unit profits. *Knowl-Based Syst* 111:283–298
33. Lan GC, Hong TP, Tseng VS (2012) Efficiently mining high average-utility itemsets with an improved upper-bound strategy. *Int J Inf Technol Decis Making* 11:1009–1030
34. Lin JCW, Ren S, Fournier-Viger P, Hong TP, Su JH, Vo B (2017) A fast algorithm for mining high average-utility itemsets. *Appl Intell* 47(2):331–346
35. Leung CKS, Mateo MAF, Brajczuk DA (2008) A tree-based approach for frequent pattern mining from uncertain data. In: The pacific-asia conference on knowledge discovery and data mining, pp 653–661
36. Wang L, Cheng R, Lee SD, Cheung D (2010) Accelerating probabilistic frequent itemset mining: A model-based approach. In: ACM international conference on information and knowledge management, pp 429–438
37. Ge J, Xia Y, Wang J, Nadungodage CH, Prabhakar S (2017) Sequential pattern mining in databases with temporal uncertainty. *Knowl Inf Syst* 53(3):821–850
38. Lin JCW, Gan W, Fournier-Viger P, Hong TP (2017) Efficiently mining uncertain high-utility itemsets. *Soft Comput* 21(11):2801–282
39. Ahmed U, Lin JCW, Srivastava G, Yasin R, Djenouri Y (2020) An evolutionary model to mine high expected utility patterns from uncertain databases. *IEEE Trans Emerg Top Comput Intell*
40. Fournier-Viger P, Lin JCW, Gomariz A, Gueniche T, Soltani A, Deng Z, Lam HT (2016) The spmf open-source data mining library version 2. In: Joint european conference on machine learning and knowledge discovery in databases, pp 36–40

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Jimmy Ming-Tai Wu** is currently an Assistant Professor at College of Computer Science and Engineering, Shandong University of Science and Technology. He was also an Assistant Professor at School of Computer Science and Technology, Harbin Institute of Technology-Shenzhen, China. His current research work is related to Data Mining, Big Data, Cloud Computing, Artificial Intelligence, Evolutionary Computation, Machine Learning and Deep Learning. He was graduated and earned his Ph.D. degree with major in Computer Science and Engineering from National Sun Yat-sen University, Kaohsiung, Taiwan. He worked in an IC design company in Taiwan as a firmware developer and information technology manager in two years. He also was a research scholar in the Department of Computer Science and Information Engineering in National University of Kaohsiung, Kaohsiung, Taiwan and in the Department of Computer Science, College of Engineering in University of Nevada, Las Vegas.



**Zhongcui Li** is now a postgraduate in School of computer science and engineering at shandong University of Science and Technology. Her research interests include Deep Learning and Data Mining.



**Gautam Srivastava** was awarded his B.Sc. degree from Briar Cliff University in USA in the year 2004, followed by his M.Sc. and Ph.D. degrees from the University of Victoria in Victoria, British Columbia, Canada in the years 2006 and 2012, respectively. In 2014, he joined a tenure-track position at Brandon University in Brandon, Manitoba, Canada, and was promoted to the rank Associate Professor in January 2018. He has published a total

of 200 papers in high-impact conferences and journals.



**Unil Yun** received the M.S. degree in computer science and engineering from Korea University, Seoul, South Korea, in 1997, and the Ph.D. degree in computer science from Texas A&M University, College Station, TX, USA, in 2005. He was a Postdoctoral Associate for almost one year with the Computer Science Department, Texas A&M University. He was with the Multimedia Laboratory, Korea Telecom, from 1997 to 2002.

He was a Senior Researcher with the Electronics and Telecommunications Research Institute. In March 2007, he joined the School of Electrical and Computer Engineering, Chungbuk National University, South Korea. Since August 2013, he has been with the Department of Computer Engineering, Sejong University, Seoul. His research interests include data mining, information retrieval, database systems, artificial intelligence, and digital libraries.



**Jerry Chun-Wei Lin** received his Ph.D. from the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan in 2010. He is currently a full Professor with the Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway. He has published more than 400 research articles in refereed journals (IEEE TKDE, IEEE

TFS, IEEE TNNLS, IEEE TCYB, IEEE TII, IEEE TITS, IEEE TNSE, IEEE TETCI, IEEE SysJ, IEEE SensJ, IEEE IOTJ, ACM TKDD, ACM TDS, ACM TMIS, ACM TOIT, ACM TIST) and international conferences (IEEE ICDE, IEEE ICDM, PKDD, PAKDD), 15 edited books, as well as 33 patents (held and filed, 3 US patents). His research interests include data mining, soft computing, artificial intelligence/machine learning, and privacy preserving and security technologies. He is the Editor-in-Chief of the International Journal of Data Science and Pattern Recognition, the Guest Editor/Associate Editor for several IEEE/ACM journals such as IEEE TFS, IEEE TII, ACM TMIS, ACM TOIT, and IEEE Access. He has recognized as the most cited Chinese Researcher respectively in 2018, 2019, and 2020 by Scopus/Elsevier. He is the Fellow of IET (FIET), senior member for both IEEE and ACM.