Review Article

A Model for Trend Analysis in the Online Shopping Scenario Using Multilevel Hesitation Pattern Mining

Abhishek Dixit ⓘ **, Akhilesh Tiwari** ⓘ **, and R. K. Gupta** ⓘ

Madhav Institute of Technology & Science, Gwalior, India

Correspondence should be addressed to Abhishek Dixit; abhishekdixit.cse@gmail.com

Academic Editor: Dilbag Singh

The present paper proposes a new model for the exploration of hesitated patterns from multiple levels of conceptual hierarchy in the transactional dataset. The usual practice of mining patterns has focused on identifying frequent patterns (i.e., which occur together) in the transactional dataset but uncovers the vital information about the patterns which are almost frequent (but not exactly frequent) called "hesitated patterns." The proposed model uses the reduced minimum support threshold (contains two values: attractiveness and hesitation) and constant minimum confidence threshold with the top-down progressive deepening approach for generating patterns and utilizing the apriori property. To validate the model, an online purchasing scenario of books through e-commerce-based online shopping platforms such as Amazon has been considered and shown that how the various factors contributed towards building hesitation to purchase a book at the time of purchasing. The present work suggests a novel way for deriving hesitated patterns from multiple levels in the conceptual hierarchy with respect to the target dataset. Moreover, it is observed that the concepts and theories available in the existing related work Lu and Ng (2007) are only focusing on the introductory aspect of vague set theory-based hesitation association rule mining, which is not useful for handling the patterns from multiple levels of granularity, while the proposed model is complete in nature and addresses the very significant and untouched problem of mining "multilevel hesitated patterns" and is certainly useful for exploring the hesitated patterns from multiple levels of granularity based on the considered hesitation status in a transactional dataset. These hesitated patterns can be further utilized by decision makers and business analysts to build the strategy on how to increase the attraction level of such hesitated items (appeared in a particular transaction/set of transactions in a given dataset) to convert their state from hesitated to preferred items.

1. Introduction

In this constantly changing technological scenario, exploring a nugget of information from the transactional dataset is very much essential for the discovery of new patterns and association rules. The business community and decision makers, taking crucial decisions on the basis of explored "information" or "knowledge," will have a better chance of survival in this competitive world. Moreover, in the recent past, e-commerce industry has emerged as one of the most preferred option of shopping in the online mode, and the example includes Amazon, Flipkart, and Snapdeal. This has extended ease and convenience to the customers and, at the same time, resulted in competition among the service providers. Due to this, it has become essential to know

something that nobody else knows in their business domain and make the difference. For this to happen, business houses and decision makers need to refer knowledge while doing crucial decision-making about products and promotional strategy planning for the growth of organization. This is where the present research work focuses.

Here, the concern is to analyze the transactional dataset, where each transaction is a record of items (purchased or almost purchased) placed in the cart (fully or partially executed). The objective of analysis is to know the buying patterns of customers on the basis of their liking and disliking. As evident from the literature, the analytics act has been exercised to reveal various types of patterns such as *Frequent Patterns* [1–5], *Profitable Patterns* [6], *Conditional Patterns* [7], *Calendar-Based Patterns* [8], and *Log Pattern Mining* [9]

using various techniques of pattern mining [10]. Moreover, after the success of mining knowledge from datasets, researchers deal with certain specific situations and perform various tasks such as mining on data streams [11, 12], recognition of handwritten expression [13], investigating customer buying behavior through Visual Market Basket Analysis (VMBA) [14], automated assessment of shopping behavior [15, 16], applying additional interestingness measures for association rule mining [17], and conditional discriminative pattern mining [18], and researchers also have to deal to improve the implementation of pattern mining algorithms using time stamp uncertainties and temporal constraints [19], privacy of frequent itemset mining using randomized response [20], and finding infrequent itemset to discover the negative association rule [21].

This work deals with *Hesitation Information Mining* [22] where the resultant will be in the form of patterns commonly termed **Hesitated Patterns**. Mining hesitated patterns are crucial for market basket analysis or online shopping scenario, where the retrieved patterns contain information about the items which are hesitated by the customers. Furthermore, the hesitated pattern is governed by some hesitation status [22], which works as a contributing state (or factor) for creating hesitation towards that item or itemsets (which constitutes hesitated pattern).

Related literature mentions vague set theory as an essential tool for generating vague association rules (VARs) [22–26] from the hesitated pattern set. Further, based on the currently available researchers and study in the field, it is concluded that mining of hesitated patterns at multiple levels of concept hierarchy (with different value of support threshold) is more sufficient and also helpful to expose the information from different levels of granularity. This is referred as Multilevel Association Rule Mining [27–30], and particularly, in the context of present research, it is known as Multilevel Hesitated Pattern Mining. In case of traditional association pattern mining [1], support and confidence measures are two important factors, which play a crucial role for generating frequent patterns and further for identification of association patterns or rules. Instead of usually applied support and confidence measures, here the proposed model utilizes two new measures, namely, attractiveness support value and hesitation support value, where the attractiveness and hesitation mean that the item or product is sold nicely, and item is hesitated by customers, respectively. As mentioned earlier, this paper proposes a new model for hesitation mining with following objectives:

(i) To mine hesitated patterns from multiple levels of concept hierarchy

(ii) To discover hesitated association patterns or rules

1.1. Background. Lu and Ng [22] introduced the concept of vague association rule (VAR) mining. They have handled vagueness and uncertainty using the concept of vague set theory. Lu et al. also coined some terminologies related to vague association rule (VAR) mining, as mentioned below:

Intent. This shows different states of an item or itemset such as support (liking), against (disliking), and hesitation (unclarity).

Hesitation Status. The stage or reason at which items are being hesitated or dropped out by the customer.

Attractiveness. This value indicates that how nicely and frequently the item is purchased by the customers i.e., item is currently purchased by the customer and will also be purchased by the customer in the near future.

Hesitation. This value indicates that how constantly customers are hesitated to purchase a particular item or set of items.

Various researchers have suggested methodologies, where they have shown the computation mechanism based on attractiveness and hesitation values associated with each item corresponding to the database of the assumed scenario. Further, based on attractiveness ad hesitation values, the AH pair database was constructed, which has been utilized to establish four types of relationship between two or more items, namely, *A* (Attractiveness), *H* (Hesitation), and AH (Attractiveness-Hesitation): it gives an attractive and hesitation relation between pair of items which is used further for identification of hesitated patterns, and HA (Hesitation-Attractiveness) gives an hesitation and attractive relation between pair of items. For these relationships, four types of support (attractiveness support, hesitation support, attractiveness hesitation support, and hesitation attractiveness support) and four types of confidence (*A* confidence, *H* confidence, AH confidence, and HA confidence) were defined. It is observed that a few researchers addressed this domain considering different datasets, with varied constraints. The work conducted by Pandey et al. [23] mentions the computing mechanism for mining vague association rules (VARs) for class course information from the temporal database. Another dimension has been explored by Badhe et al. [6] in the form of new model for mining profitable patterns from the transactional dataset. In the sequence, the work mentioned in [24, 25] has presented genetic-based methodology for mining hesitated itemsets in the transactional dataset. In the recent research [26], authors have proposed elephant herding optimization-based vague association rule mining. This work also makes use of transactional data with focus on seasonal effect, for finding maximum profit. Dandotiya et al. [31] proposed a method to identify the optimized hesitation pattern from the transactional dataset using the weighted apriori and genetic algorithm. Dixit et al. [32] proposed a model for mining hesitated patterns from the transactional datasets using vague set theory with considering only one hesitation state.

Literature reveals that no direct competing methods are available, and the related existing work [18] is just introducing the concept of vague set theory and other formulations for vague association rule mining. However, the present paper proposes a novel method with the complete mechanism to handle the information pertaining to hesitated patterns at multiple levels of granularity, which can be readily used by the knowledge workers, organizations, and business analysts for

making strategies and planning for improving the attractiveness level of hesitated items or patterns.

The present paper is organized as follows: in Section 2, a new model for mining hesitated patterns is described; Section 3 illustrates the concept of model with the suitable example of online shopping, while the outcome of this model is discussed in Section 4 followed by conclusion drawn.

2. Proposed Model

2.1. Workflow Diagram of Proposed Model. The proposed model will include a premining phase that will process the data from the data source by cleaning and transforming it into the input ready dataset i.e., multilevel transactional dataset (for P^{th} Level) and then apply the multilevel hesitation mining algorithm; as a result of mining, hesitated patterns will be generated at each hesitation status, and this process will continue till the highest level of concept hierarchy. The generated pattern set will now be supplied to the hesitated association pattern module for generating interesting and potentially useful patterns, which will be used to facilitate the decision makers/knowledge workers for business strategy planning. The workflow diagram of the proposed model for multilevel hesitation pattern mining is shown in Figure 1.

2.2. Steps of Proposed Model. There can be n numbers of reasons, due to which the customer may hesitate to buy some products at the time of shopping, which may in turn probably result in the decrease of sale of those products or items. Therefore, it is required to identify such type of hesitated products so that promotional strategy can be framed. This section describes the step-by-step procedure of the proposed model, which helps in the exploration of hesitated patterns or items (based on some considered hesitation status) from the transactional dataset.

The steps involved to find hesitated patterns and generate rules are as follows:

Step 1: The given transactional dataset (TD) contains a set of transactions (T_j) (where $j = 1, 2, 3, \ldots, m$), and each transaction has a list of purchased and hesitated items. In this step, firstly, construct the concept hierarchy of the items present in the given dataset. Then, the given transactional dataset is transformed into the multilevel transactional dataset (MD). In this dataset, each item in a transaction is represented as

$$(\text{Item\_name, status of item}), \qquad (1)$$

where Item_name refers to the item, and it is written in multilevel taxonomy, while the status of the item gives the information about whether the item is purchased or hesitated. If the item is purchased, then the status of the item value is 1, but if the item is hesitated by the customer, then its value is one of the hesitation status ($h_1, h_2, \ldots, h_n$, where $h_i \subseteq h_{i+1}$) at which the customer is hesitated to purchase it.

Step 2: For finding hesitated frequent patterns at different levels in the hierarchy of the transactional dataset, a variable P is considered, where $P \in \{1, 2, 3, \ldots, M\}$, and this variable keeps track of the level number which is being processed.

This step encodes each item (either purchased or hesitated) which are present in the transaction of the multilevel transactional dataset. The encoding of each item is performed by using the sequence number of the item, which depends on the level in the hierarchy i.e., $\mathbf{L_i}$ (where $i = 1, 2, 3, \ldots, M$) and after considering the class replacing all the remaining numbers by the symbol " $*$ "

Step 3: Now, each item in an individual transaction is grouped according to its class, which depends on the level L_i (where $i = 1, 2, 3, \ldots, M$), and also, add their occurrences (due to which, the status value of the item, which is either purchased or hesitated, is changed) so that each grouped item in a transaction is in the form:

$$(\text{Item\_name, status of item purchase, status of item hesitated}), \qquad (2)$$

where the status of item purchase also means attractiveness of an item (s). This grouping is done in every transaction of the encoded multilevel transactional dataset individually.

Step 4: Consider another variable I, which is used to represent the length of the candidate pattern. It is represented as I-candidate pattern where $I \in \{1, 2, 3, \ldots, r\}$. For example, if the value of I is 1, then it is referred as 1-candidate pattern.

Step 5: For each level, it is necessary to define two value minimal threshold support (denoted as $\beta_{\mathbf{P}}$) represented as

$$\beta_{\mathbf{P}} = \left(\beta_{\mathbf{p}}, \beta_{\mathbf{h}}\right), \qquad (3)$$

where $\beta_{\mathbf{p}}$ value represents the minimal threshold support for purchased item(s) or attractiveness value, while another $\beta_{\mathbf{h}}$ value represents the minimal threshold support for hesitated item(s). This minimal threshold support will be taken as uniform or may be different for all the level in the hierarchy.

Step 6: Calculate the support of I-candidate pattern at each level P at different hesitation states $h_k = (1, 2, \ldots, n)$. This support also contains two support values, and it is represented as
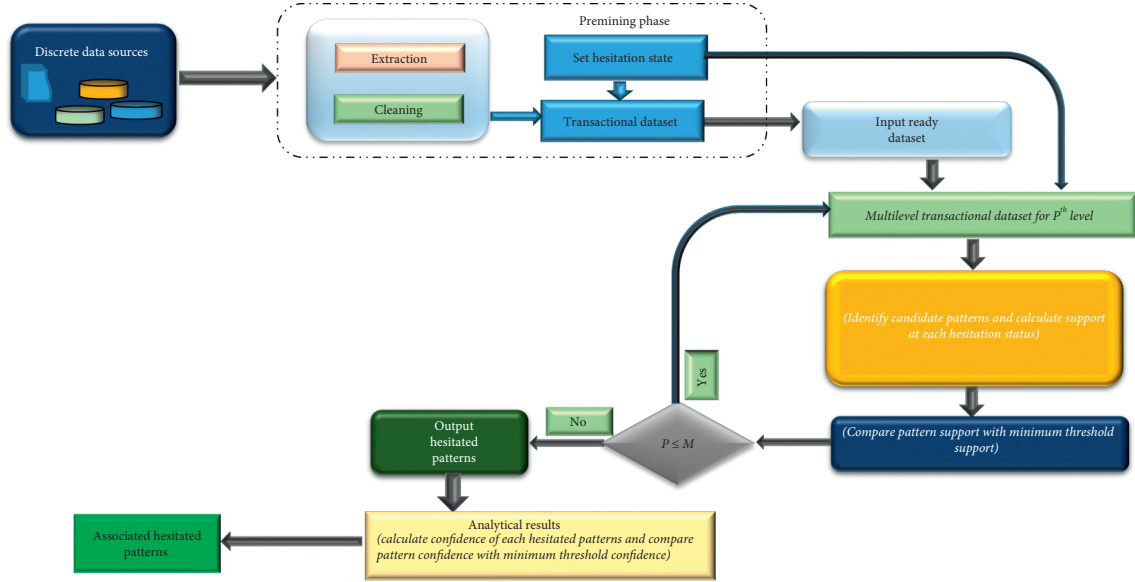
FIGURE 1: Workflow diagram of the proposed model.

$$\mu = \left(\mu_p, \mu_h\right), \tag{4}$$

where μ_p represents the support value of the purchased item or attractiveness value and μ_h represents the support value of the hesitated item. The support value of I-candidate pattern (x) at h_k is computed as

$$\mu_p(x) = \left(\sum_{j=1}^{m}\left(\frac{v_{T_j}(x)}{\eta_{T_j}}\right)\right), \tag{5}$$

$$v_{T_j} = \left(p_{T_j}(x) + \left((h_{k-1})_{T_j}(x)\right)\right), \tag{6}$$

where x is a pattern, m is the total number of transactions in the dataset, k is the number of the hesitation state, v is the total number of times an item is purchased (attractive) in the transaction T_i of level L_i dataset, and η is the total number of times an item is purchased and hesitated in the transaction T_i of level L_i dataset:

$$\mu_h(x) = \left(\sum_{j=1}^{m}\left(\frac{\iota_{T_j}(x)}{\eta_{T_j}}\right)\right), \tag{7}$$

$$\iota_{T_j}(x) = (h_k)_{T_j}(x), \tag{8}$$

where ι is the number of times an item is hesitated in the transaction T_i. So, the support value of I-candidate pattern (x) is

$$\mu = \left(\sum_{j=1}^{m}\left(\frac{v_{T_j}(x)}{\eta_{T_j}}\right), \sum_{j=1}^{m}\left(\frac{\iota_{T_j}(x)}{\eta_{T_j}}\right)\right). \tag{9}$$

In the normal form, by using equations (5)–(8),

$$\mu = \left(\sum_{j=1}^{m}\left(\frac{p_{T_j}(x) + \left((h_{k-1})_{T_j}(x)\right)}{\eta_{T_j}}\right), \right.$$
$$\left. \sum_{j=1}^{m}\left(\frac{(h_k)_{T_j}(x)}{\eta_{T_j}}\right)\right). \tag{10}$$

If the value of $\mu(x) \geq \beta_P$, i.e., support value of the pattern is greater than equal to the minimal threshold support, it means the pattern is referred as the hesitated frequent pattern.

Step 7: Now, using the hesitated frequent patterns generated in the Step 6, construct $(I+1)$-candidate patterns using the apriori candidate generation method [2–4], and their support is calculated as follows:

$$\mu_p(x, y) = \left(\sum_{j=1}^{m}\left(\min\left(\left(\frac{v_{T_j}(x)}{\eta_{T_j}}\right), \left(\frac{v_{T_j}(y)}{\eta_{T_j}}\right)\right)\right)\right), \tag{11}$$

where x and y are the two individual hesitated frequent patterns:

$$\mu_h(x, y) = \left(\sum_{j=1}^{m}\left(\min\left(\left(\frac{\iota_{T_j}(x)}{\eta_{T_j}}\right), \left(\frac{\iota_{T_j}(y)}{\eta_{T_j}}\right)\right)\right)\right). \tag{12}$$

So, by using equations (11) and (12), the support value is

$$\mu(x, y) = \left(\mu_p(x, y), \mu_h(x, y)\right). \tag{13}$$

Step 8: Now, repeat steps (2 to 7) at each level L_i to mine hesitated frequent patterns. The process is continued till each level in the hierarchy is traversed.

Step 9: Predefined minimal confidence is represented as $\alpha = (\alpha_p, \alpha_h)$. The confidence of each possible hesitated association pattern $(x = > y)$ is calculated as

$$\text{conf}(x \Longrightarrow y) = \frac{(\mu(x, y))}{\mu(x)}. \tag{14}$$

2.3. Multilevel Hesitated Pattern Algorithm.
The steps involved in the multilevel hesitated pattern algorithm are given in Algorithm 1.

2.4. Computational Complexity.
The computational complexity of Algorithm 1 is $O(M \times [h_k + N^2 \times 2^N \times h_k])$, where M is the highest level in the concept hierarchy, N is the maximum number of hesitated patterns in $(i-1)$ candidate pattern, and h_k is the number of hesitation states or status.

2.4.1. Computational Complexity of Multilevel Hesitated Pattern Algorithm.
In the pseudocode, outer while-loop repeats maximum number of levels i.e., M; thus, it takes $O(M)$ time. Now, first begin part inside while-loop calculate 1-candidate patterns using the mathematical formula which takes constant time but, at each level and at every hesitation state, we calculate 1-candidate patterns so it takes $O(M \times h_k)$ time. After calculating (i)-candidate patterns (i.e., $i = 1$), $(i + 1)$-candidate patterns at each hesitation state are calculated, and pruning is performed in the inner while-loop, if it is considered that the maximum number of the hesitated pattern is N in $(i-1)$-candidate patterns so this takes $O(N^2 \times h_k)$ time to generate $(i + 1)$-candidate patterns, and during pruning, it takes $O(2^N)$ time. Therefore, the time complexity of the algorithm is $O(M \times h_k + M \times N^2 \times 2^N \times h_k])$.

3. Illustration

3.1. Illustration 1.
It is well known that a number of courses are part of computer science discipline. The example includes Programming in C, Object-Oriented Programming, Data Structures, Theory of Computation, Operating Systems, and Database Management System. Further, to study and gain the knowledge about these courses, students have to refer some reference books. Therefore, they may purchase these reference books through the online mode or in the traditional mode.

In this illustration, the online purchasing scenario of reference books is considered, and courses relating to computer science discipline have been assumed, which includes Programming in C, Data Structures, and Analysis & Design of Algorithms. Moreover, it is also considered that, for a specific course, several reference books are available.

These books differ from one another in various aspects, such as content, publisher, and author; with this scenario, a concept hierarch is developed, as shown in Figure 2.

During the online purchasing process, the customer might hesitate to purchase books due to some reasons (hesitation status). These conditions may be

Firstly,

(a) Author of the book
(b) Price of the book
(c) Publication house

Secondly,

(a) Content of the book
(b) Delivery date

Third,

(a) Delivery not possible at the required place
(b) Extra delivery charges, etc.

In this illustration, all these abovementioned reasons are considered for the formulation of hesitation status h_1, h_2, and h_3, respectively.

Hence, the objective is to explore or find frequently hesitated books (due to any of the described hesitation status). The proposed model is applied on the considered dataset. The step-by-step procedure of the model is as follows:

Step 1: Let us consider a transactional dataset (TD), which contains ten numbers of transactions (T_j), namely, $(T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9,$ and $T_{10})$ and three hesitation status (h_1, h_2, h_3). The transactional dataset is shown in Table 1.

In multilevel taxonomy, items

I1: Let Us C (BPB Publications)
I2: The Complete Reference (TATA McGraw-Hill)
I3: Data Structures (Technical Publication Pune)
I4: Data Structures, Algorithm and Application in C++ (Universities Press)
I5: Fundamentals of Computer Algorithms (Universities Press)
I6: Design and Analysis of Computer Algorithms (Pearson)

will be denoted as {11}, {12}, {21}, {22}, {31}, and {32}, respectively. Now, according to the model, this transactional dataset is converted into the multilevel transactional dataset (MD), as shown in Table 2.

Step 2: The items present in the hierarchy are encoded in this step. The concept hierarchy, as shown in Figure 1, contains reference books for computer science discipline; as a root node, it is referred as level 0. However, Programming in C, Data

Input: transactional dataset, minimum threshold support, minimum threshold confidence, number of hesitation status.
Output: hesitated patterns, hesitated association patterns
TD: Initial Transactional Dataset
MD: Multilevel Transactional Dataset//after transforming TD into multilevel taxonomy
M: highest level in the concept hierarchy//input
P: store the currently processing level
CP_i: candidate pattern of size i//$i = 1, 2, \ldots, t$
HP_i: hesitated Patterns of size i//$i = 1, 2, \ldots, t$
β_P: minimal threshold support as (β_p, β_h)
//different for each level in the hierarchy
//β_p is the attractiveness support and β_h is the hesitation support of an itemset.
$\alpha =$ minimal threshold confidence as (α_p, α_h)
//α_p is the attractiveness confidence and α_h is the hesitation confidence of an itemset.
h_k: hesitation status//$k = 1, 2, \ldots, n$
Initialize: $P = 1$
While $(P! = M)$ **do**
  begin
    //for each class at each hesitation status
        Initialize: $i = 1$
        Support_calculation for i-candidate patterns
        $\mu_p(x, y) = (\sum_{j=1}^{m} (\min((v_{T_j}(x)/\eta_{T_j}), (v_{T_j}(y)/\eta_{T_j}))))$
        $\mu = (\sum_{j=1}^{m} (p_{T_j}(x) + ((h_{k-1})_{T_j}(x))/\eta_{T_j}), \sum_{j=1}^{m} ((h_k)_{T_j}(x)/\eta_{T_j}))$
        $CP_1 = \{$candidate patterns$\}$
        $HP_1 = \{$hesitated patterns$\}$//after comparing Support with minimum threshold support (β_P)
        $i = i + 1$
          While $(HPi - 1! = \varnothing)$ **do**
          Begin
         //gen_candidate_patterns from HP_{i-1};//according to hesitation status
            $CP_i = \varnothing$
            for all pattern hp_1 belongs to HP_{i-1} **do**
            for all pattern hp_2 belongs to HP_{i-1} **do**
            if $hp_1[1] = hp_2[1] \wedge hp_1[2] = hp_2[2] \wedge \cdots \wedge hp_1[i - 1] < hp_2[i - 1]$
            then $CP = hp_1[1], hp_1[2], \ldots hp_1[i - 1], hp_2[i - 1]$
            $CP_i: = CP_i U\{CP\}$
              Prune (CP_i);
               for all CP belongs to CP_i
              for all $(i - 1)$ subsets b of CP **do**
                if b **does not belong to** HP_{i-1}
                then $CP_i = CP_i/\{CP\}$
            $CP_i = \{$candidate patterns$\}$
            Calculate the support of each prune candidate patterns at each Hs.
            $\mu_p(x, y) = (\sum_{j=1}^{m} (\min((v_{T_j}(x)/\eta_{T_j}), (v_{T_j}(y)/\eta_{T_j}))))$
            Where x and y are the two individual hesitated frequent patterns.
            $\mu_h(x, y) = (\sum_{j=1}^{m} (\min((\iota_{T_j}(x)/\eta_{T_j}), (\iota_{T_j}(y)/\eta_{T_j}))))$
            $\mu(x, y) = (\mu_p(x, y), \mu_h(x, y))$
            $HP_i = \{$hesitated patterns$\}$
        end
          $i = i + 1$;
        end
  end
  $P = P + 1$;
end
  Association Pattern Generation
    for all item in HP **do**
      Construct association $\{HP - \{$item$\} \Longrightarrow \{$item$\}\}$
      Calculate confidence $\{HP - \{$item$\} \Longrightarrow \{$item$\}\}$
        $\text{conf}(x \Longrightarrow y) = ((\mu(x, y))/\mu(x))$
      conf $\{HP - \{$item$\} \Longrightarrow \{$item$\}\} = S\{HP - \{$item$\}U\{$item$\}\}/S\{HP - \{$item$\}\}$
      if confidence $\{HP - \{$item$\} \Longrightarrow \{$item$\}\} \geq \min\_\text{conf}(\alpha)$ **then**
      Output $\{HP - \{$item$\} \Longrightarrow \{$item$\}\}$
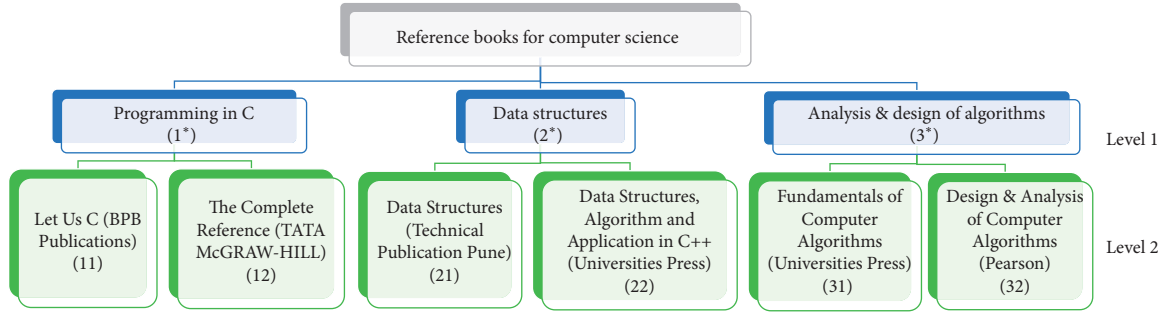
ALGORITHM 1: Mining multilevel hesitated pattern algorithm.

FIGURE 2: Concept hierarchy of reference books for computer science.

TABLE 1: Transactional dataset.

T_{ID}	I1	I2	I3	I4	I5	I6
T_1	1	h_1	h_1	0	h_1	0
T_2	1	1	0	h_2	h_2	0
T_3	0	h_2	0	h_2	1	h_2
T_4	1	1	0	0	1	0
T_5	0	1	1	0	1	1
T_6	0	0	h_3	h_2	h_2	h_3
T_7	0	0	0	h_2	1	h_2
T_8	1	0	0	0	0	1
T_9	0	1	0	1	1	0
T_{10}	h_2	0	h_2	0	1	h_2

Structures, and Analysis and Design of algorithms are the internal nodes (all are at level 1) and are encoded as {1*}, {2*}, and {3*}, respectively. Moreover, Let Us C (BPB Publications), The Complete Reference (TATA McGraw-Hill), Data Structures (Technical Publication Pune), Data Structures, Algorithm and Application in C++ (Universities Press), Fundamentals of Computer Algorithms (Universities Press), and Design and Analysis of Computer Algorithms (Pearson) (all are at level 2) are encoded as {11}, {12}, {21}, {22}, {31}, and {32}, respectively.

Step 3: The model will traverse all the level one by one. In the considered example, the hierarchy has two levels for traversing (level 0 is not considered) i.e., $P = 1$ and $P = 2$. For level 1, group the items present in the individual transaction of MD. After grouping, the modified multilevel transactional dataset for level 1 is transformed into new layout, which is shown in Table 3.

Step 4: The next task after grouping the items is to find hesitated patterns, and these patterns have some length which is denoted by I. Now, the procedure for mining hesitated frequent patterns at various levels over all hesitation status is described in the step 5, 6, and 8.

Step 5: Consider two different minimal threshold support for level 1 and level 2, which are $\beta_1 = (0.80, 0.50)$ and $\beta_2 = (0.60, 0.40)$, respectively.

Step 6: Calculate hesitated frequent patterns, for **Level P = 1:**

For I = 1,

1-candidate patterns:
In the concept hierarchy, there are three numbers of 1-candidate patterns that are present i.e., {1*}, {2*}, and {3*}, and the dataset has three hesitation status h_1, h_2, and h_3. So, the support of each pattern at every hesitation status is calculated (using equation (10)) as follows.

{1*} (h_1) = {(0.25, 0.25), (0.50, 0.0), (0.66, 0.0), (0.25, 0.0), (0.50, 0.0), (0.33, 0.0)} = (2.49, 0.25)

{1*} (h_2) = {(0.50, 0.0), (.50, 0.0), (0.0, 0.25), (0.66, 0.0), (0.25, 0.0), (0.50, 0.0), (0.33, 0.0), (0.0, 0.25)} = (2.74, 0.50)

{1*} (h_3) = {(0.50, 0.0), (0.50, 0.0), (0.25, 0.0), (0.66, 0.0), (0.25, 0.0), (0.50, 0.0), (0.33, 0.0), (0.25, 0.0)} = (3.25, 0.0)

{2*} (h_1) = {(0.0, 0.25), (0.25, 0.0), (0.33, 0.0)} = (0.58, 0.25)

{2*} (h_2) = {(0.25, 0.0), (0.0, 0.25), (0.0, 0.25), (0.25, 0.0), (0.0, 0.25), (0.0, 0.33), (0.33, 0.0), (0.0, 0.25)} = (0.83, 1.33)

{2*} (h_3) = {(0.25, 0.0), (0.25, 0.0), (0.25, 0.0), (0.25, 0.0), (0.25, 0.25), (0.33, 0.0), (0.33, 0.0), (0.25, 0.0)} = (2.16, 0.25)

{3*} (h_1) = {(0.0, 0.25), (0.25, 0.0), (0.33, 0.0), (0.50, 0.0), (0.33, 0.0), (0.50, 0.0), (0.33, 0.0), (0.25, 0.0)} = (2.49, 0.50)

TABLE 2: Multilevel transactional dataset.

T_{ID}	Items
T_1	(11, 1), (12, h_1), (21, h_1), (31, h_1)
T_2	(11, 1), (12, 1), (22, h_2), (31, h_2)
T_3	(12, h_2), (22, h_2), (31, 1), (32, h_2)
T_4	(11, 1), (12, 1), (31, 1)
T_5	(12, 1), (21, 1), (31, 1), (32, 1)
T_6	(21, h_3), (22, h_2), (31, h_2), (32, h_3)
T_7	(22, h_2), (31, 1), (32, h_2)
T_8	(11, 1), (32, 1)
T_9	(12, 1), (22, 1), (31, 1)
T_{10}	(11, h_2), (21, h_2), (31, 1), (32, h_2)

TABLE 3: Modified multilevel transactional dataset (for level 1).

T_{ID}	Items
T_1	(1^*, 1, h_1), (2^*, 0, h_1), (3^*, 0, h_1)
T_2	(1^*, 2, 0), (2^*, 0, h_2), (3^*, 0, h_2)
T_3	(1^*, 0, h_2), (2^*, 0, h_2), (3^*, 1, h_2)
T_4	(1^*, 2, 0), (3^*, 1, 0)
T_5	(1^*, 1, 0), (2^*, 1, 0), (3^*, 2, 0)
T_6	(2^*, 0, $h_2 + h_3$), (3^*, 0, $h_2 + h_3$)
T_7	(2^*, 0, h_2), (3^*, 1, h_2)
T_8	(1^*, 1, 0), (3^*, 1, 0)
T_9	(1^*, 1, 0), (2^*, 1, 0), (3^*, 1, 0)
T_{10}	(1^*, 0, h_2), (2^*, 0, h_2), (3^*, 1, h_2)

$\{3^*\}$ (h_2) = {(0.25, 0.0), (0.0, 0.25), (0.25, 0.25), (0.33, 0.0), (0.50, 0.0), (0.0, 0.25), (0.33, 0.33), (0.50, 0.0), (0.33, 0.0), (0.25, 0.25)} = (2.74, 1.33)

$\{3^*\}$ (h_3) = {(0.25, 0.0, (0.25, 0.0), (0.50, 0.0), (0.33, 0.0), (0.50, 0.0), (0.25, 0.25), (0.66, 0.0), (0.50, 0.0), (0.33, 0.0), (0.50, 0.0)} = (4.07, 0.25)
The support of each 1-candidate patterns with their support is shown in Table 4.

Now, compare the support of every candidate patterns with the minimal threshold support (β). Those patterns whose support is greater than or equal to minimal threshold support (i.e., for this level is (0.80, 0.50)) are referred as hesitated frequent patterns and are shown in Table 5.

Step 7: Using these hesitated frequent patterns, 2-candidate patterns are generated by using the concept of the apriori candidate generation method [2, 3] (this method is applicable only on h_2 hesitation status because, for pairing, a sufficient number of patterns are available only in this hesitation status). After applying this method, the result will generate in the form as follows: (1^*, 2^*), (1^*, 3^*), and (2^*, 3^*).

Now, **For I = 2,**

2-candidate pattern:

TABLE 4: 1-candidate pattern with their support.

Pattern	Hesitation status		
	h_1	h_2	h_3
$\{1^*\}$	(2.49, 0.25)	(2.74, 0.50)	(3.25, 0.0)
$\{2^*\}$	(0.58, 0.25)	(0.83, 1.33)	(2.16, 0.25)
$\{3^*\}$	(2.49, 0.25)	(2.74, 1.33)	(4.07, 0.25)

Calculate the support of these generated 2-candidate patterns at h_2 hesitation status by using equation (13).

$\{1^*, 2^*\}$ (h_2) = {(0.25, 0.0), (0.0, 0.25), (0.25, 0.0), (0.33, 0.0), (0.0, 0.25)} = (0.83, 0.50)

$\{1^*, 3^*\}$ (h_2) = {((0.25, 0.0), (0.0, 0.25), (0.33, 0.0), (0.25, 0.0), (0.50, 0.0), (0.33, 0.0), (0.0, 0.25)} = (1.66, .50)

$\{2^*, 3^*\}$ (h_2) = {(0.25, 0.0), (0.0, 0.25), (0.0, 0.25), (0.25, 0.0), (0.0, 0.25), (0.0, 0.33), (0.33, 0.0), (0.0, 0.25)} = (0.83, 1.33)

Thus, the 2-candidate patterns support is compared with minimal threshold support. The 2-candidate patterns, which are hesitated frequent patterns, are shown in Table 6.

Using these generated hesitated frequent patterns, generate 3-candidate patterns. So, the 3-candidate pattern generated is $\{1^*, 2^*, 3^*\}$.

For I = 3,

3-candidate pattern:
Now, calculate the support (by using equation (13)) for this generated pattern and compare it with the predefined support.

$\{1^*, 2^*, 3^*\}$ (h_2) = {(0.25, 0.0), (0.0, 0.25), (0.25, 0.0), (0.33, 0.0), (0.0, 0.25)} = (0.83, 0.50).
The support value is greater or equal to the minimal threshold support. So, it is a hesitated frequent pattern. No further candidate pattern is generated. The process will stop at this level and move to the next level in the hierarchy.

Now, repeat steps (2–7) to calculate hesitated frequent patterns, **for Level P = 2**; after encoding of items according to level 2, the dataset is updated, as shown in Table 7.

For **I = 1,**

1-candidate pattern:
$\{21\}$ (h_2) = {(0.25, 0.0), (0.25, 0.0), (0.0, 0.25)} = (0.50, 0.25).

$\{22\}$ (h_2) = {(0.0, 0.25), (0.0, 0.25), (0.0, 0.25), (0.0, 0.33), (0.33, 0.0)} = (0.33, 1.08).

$\{31\}$ (h_2) = {(0.25, 0.0), (0.0, 0.25), (0.25, 0.0), (0.33, 0.0), (0.25, 0.0), (0.0, 0.25), (0.33, 0.0), (.33, 0.0), (0.25, 0.0)} = (1.99, 0.50).

$\{32\}$ (h_2) = {(0.0, 0.25), (0.25, 0.0), (0.0, 0.33), (0.50, 0.0), (0.0, 0.25)} = (0.75, 0.83).

Table 5: Hesitated frequent 1-candidate pattern (at level 1).

Pattern	Hesitation status		
	h_1	h_2	h_3
{1*}	—	(2.74, 0.50)	—
{2*}	—	(0.83, 1.33)	—
{3*}	—	(2.74, 1.33)	—

Table 6: Hesitated frequent 2-candidate pattern (at level 1).

Pattern	h_2
{1*, 2*}	(0.83, 0.50)
{1*, 3*}	(1.66, 0.50)
{2*, 3*}	(0.83, 1.33)

Compare the support of every candidate item with the minimal threshold support β_2 i.e., (0.33, 0.30). After comparing, the hesitated frequent patterns are shown in Table 8.

Now, 2-candidate patterns are generated using the hesitated frequent patterns.

For **I = 2,**

2-candidate pattern:
The 2-candidate patterns generated from 1-candidate hesitated frequent patterns are {22, 31}, {22, 32}, and {31, 32}. Here, calculate the support of these patterns to identify hesitated frequent patterns.

{**22, 31**} **(h_2)** = {(0.0, 0.25), (0.0, 0.25), (0.33, 0.0)} = (0.33, 0.50)

{**22, 32**} **(h_2)** = {(0.0, 0.25), (0.0, 0.33)} = (0.0, 0.58)

{**31, 32**} **(h_2)** = {(0.25, 0.0)} = (0.25, 0.0)

After comparing the support with the minimal threshold support for this level, it is identified that only {22, 31} is a hesitated frequent patterns with support (0.33, 0.50), while {22, 32} and {31, 32} patterns are not qualified to be hesitated frequent patterns (because their support is less than minimal threshold support).

Step 8: Now, no further candidate pattern is generated, and all the levels of the hierarchy have traversed. So, stop the process.

Step 9: Calculate the confidence of all hesitated frequent patterns, which are mined at each level by using equation (14), Consider that the minimal threshold confidence is (0.60, 0.45).

conf (1^* => 2^*) = (0.83, 0.50)/(2.74, 0.50) = (0.30, 1)

conf (1^* => 3^*) = (1.66, 0.50)/(2.74, 0.50) = (0.60, 1)

conf (2^* => 3^*) = (0.83, 1.33)/(0.83, 1.33) = (1, 1)

conf (1^*, 2^* => 3^*) = (0.83, 0.50)/(0.83, 0.50) = (1, 1)

Table 7: Updated multilevel transactional dataset (for level 2).

T_{ID}	Items
T_1	(11, 1, 0), (12, 0, h_1) (21, 0, h_1), (31, 0, h_1)
T_2	(11, 1, 0), (12, 1, 0), (22, 0, h_2), (31, 0, h_2)
T_3	(12, 0, h_2), (22, 0, h_2), (31, 1, 0), (32, 0, h_2)
T_4	(11, 1, 0), (12, 1, 0), (31, 1, 0)
T_5	(12, 1, 0), (21, 1, 0), (31, 1, 0), (32, 1, 0)
T_6	(21, 0, h_3), (22, 0, h_2), (31, 0, h_2), (32, 0, h_3)
T_7	(22, 0, h_2), (31, 1, 0), (32, 0, h_2)
T_8	(11, 1, 0), (32, 1, 0)
T_9	(12, 1, 0), (22, 1, 0), (31, 1, 0)
T_{10}	(11, 0, h_2), (21, 0, h_2), (31, 1, 0) (32, 0, h_2)

Table 8: Hesitated frequent 1-candidate pattern (at second level).

Pattern	h_2
{22}	(0.33, 1.08)
{31}	(1.99, 0.50)
{32}	(0.75, 0.83)

conf (1^* => 2^*, 3^*) = (0.83, 0.50)/2.74, 0.50) = (0.30, 1)

conf (22 => 31) = (0.33, 0.50)/(0.33, 1.08) = (1, 0.46)

3.2. Illustration 2. Similar to illustration 1, another online purchasing scenario (with relatively large concept hierarchy) of grocery items (along with the items such as rice, flours, masala, and oil) can be considered. Moreover, the type and brand for a specific grocery item can also be considered. These items may differ from one another in their price, quantity, quality, etc. Considering this scenario, the concept hierarchy that will be developed is shown in Figure 3. In order to explore the hesitated items, the procedure depicted in illustration 1 is to be applied.

4. Discussion and Results

The proposed model is competent enough to explore hesitated patterns from multiple levels in the concept hierarchy related to the target dataset. It is observed that the proposed model is generating the hesitated patterns as per the expectation, but the results can also be analyzed in terms of quantitative and qualitative dimensions. If the generated results are considered along with the quantitative dimension, then the proposed model is effectively generating all the patterns (that is, completeness). The model is complete in nature because it generates all hesitated patterns (of all sizes) at each level of granularity. Hence, the produced results are covering all the pattern set which shows the sufficiency of the model from the quantitative point of view.

In the present work, certain hesitation status has been considered for validating the proposed model. It is observed that the model is producing quality results (in terms of accuracy) which are dependent on the considered hesitation status. It is realized that the inclusion of more hesitation status (by means of various ways such as surveying, buying behavior analysis of customers, experience,
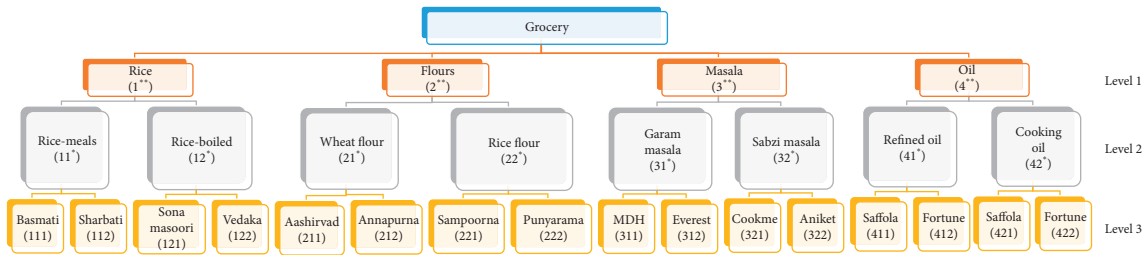
FIGURE 3: Concept hierarchy of grocery items.

and common intelligence of knowledge workers) may further help in improving the quality of generated hesitated patterns.

The results show that the model is revealing the hesitated pattern set from multiple levels of granularity with the desired level of quality. Further, the quality is very concerned with the considered hesitation status, and the same may be improved by taking into account more hesitation status during the computation process of the proposed model. Moreover, the quality can also be improved by associating appropriate choose optimization mechanism at the post-mining stage to refine the generated hesitated pattern set on the basis of various interestingness factors.

Along with the qualitative and quantitative aspects, the parallel implementation of the proposed model (for larger size catalogs) can be achieved by exploring the hesitated patterns with respect to every hesitation status at each level of the concept hierarchy in a distributed manner i.e., level-wise hesitated patterns can be calculated on separate machines, that is, the exploration of level 1 hesitated patterns on machine 1, level 2 hesitated patterns on machine 2, and so on, and aggregate the result in the form of the global hesitated pattern set.

The proposed model is applied on the considered transactional dataset of the online book purchasing scenario through e-commerce platforms such as Amazon and Flipkart. As a consequence, hesitated frequent patterns are generated as follows: $\{1^*\}$, $\{2^*\}$, $\{3^*\}$, $\{1^*, 2^*\}$, $\{1^*, 3^*\}$, $\{2^*, 3^*\}$, $\{1^*, 2^*, 3^*\}$, $\{22\}$, $\{31\}$, $\{32\}$, and $\{22, 31\}$ at h_2 (hesitation status). Subsequently, the hesitated association patterns or rules discovered from the hesitated patterns include $(1^* => 3^*)$, $(2^* => 3^*)$, $(1^*, 2^* => 3^*)$, and $(22 => 31)$. Theses hesitated patterns and association patterns can be interpreted in the following manner. The association pattern $\{22 => 31\}$, implies that {Data Structures, Algorithm and Application in C++ (Universities Press) => Fundamentals of Computer Algorithms (Universities Press)} books are associated with each other i.e., hesitated by most of the customer. Particularly, this association pattern shows the certainty of hesitation of other book titles (right side of the rule or pattern set), when the book title (left side of the rule or pattern set) is hesitated. This is because of the hesitation status; content of the books is not much more different. Based on this hesitation information, the attractiveness of hesitated patterns or hesitated association pattern sets can be increased. In this way, organizations and business houses may plan their promotional strategies.

5. Conclusions

This work presents a new model for exploration and discovery of the hesitated pattern set from the transactional data relating to the online shopping scenario. The model is effective and useful for generating hesitated patterns, which can be further utilized for crucial decision-making purposes within an organization. This will enable the organization and business houses to survive in this competitive age. Using the proposed model, hesitated patterns can be identified and considered for turning hesitated items into the preferred ones (by improving the attractiveness value). Moreover, the proposed model is capable in handling hesitated information from different levels of granularity, which reveals the effectiveness in the generated hesitated patterns. However, when the dataset increases, then the large number of hesitated patterns will be generated. This will consume lots of processing time and may result in the degradation of the efficiency of algorithms. Further, to handle this situation, one possible way is to make use of appropriately chosen optimization mechanism [33–35].

Data Availability

No data are available. However, for the purpose of modeling, the data were assumed on the basis of the current online shopping scenario.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Record*, vol. 22, no. 2, pp. 207–216, 1993.

[2] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings of the International Conference on Very Large Data Bases*, pp. 487–499, Santiago, Chile, September 1994.

[3] J. Han, M. Kamber, and J. Pei, *Data Mining–Concepts & Techniques*, Morgan Kaufmann, Burlington, MA, USA, 3rd edition, 2011.

[4] A. K Pujari, *Data Mining Techniques*, University Press, Hyderabad, India, 3rd edition, 2013.

[5] A. Tiwari, R. K. Gupta, and D. P. Agrawal, "A survey on frequent pattern mining: current status and challenging

issues," *Information Technology Journal*, vol. 9, no. 7, pp. 1278–1293, 2010.

[6] V. Badhe, R. S. Thakur, and G. S. Thakur, "Vague set theory for profit pattern and decision making in uncertain data," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 6, pp. 182–186, 2015.

[7] A. Adhikari and P. R. Rao, "Mining conditional patterns in a database," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1515–1523, 2008.

[8] A. K. Mahanta, F. A. Mazarbhuiya, and H. K. Baruah, "Finding calendar-based periodic patterns," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1274–1284, 2008.

[9] J. Chen, P. Wang, S. Du, and W. Wang, "Log pattern mining for distributed system maintenance," *Complexity*, vol. 2020, Article ID 6628165, 2020.

[10] B. N. Asmat, M. Ullah, and D. Rukhshanda, "Basic pattern mining: a review on techniques and applications," *International Journal of Scientific & Engineering Research*, vol. 9, pp. 31–44, 2018.

[11] A. Farhat, M. S. Gouider, and L. B. Said, "New algorithm for frequent itemsets mining from evidential data streams," *Procedia Computer Science*, vol. 96, pp. 645–653, 2016.

[12] L. Bustio-Martínez, A. Muñoz-Briseño, R. Cumplido, R. Hernández-León, and C. Feregrino-Uribe, "A novel multi-core algorithm for frequent itemsets mining in data streams," *Pattern Recognition Letters*, vol. 125, pp. 241–248, 2019.

[13] A. D. Le, B. Indurkhya, and M. Nakagawa, "Pattern generation strategies for improving recognition of handwritten mathematical expressions," *Pattern Recognition Letters*, vol. 128, pp. 255–262, 2019.

[14] V. Santarcangelo, G. M. Farinella, A. Furnari, and S. Battiato, "Market basket analysis from egocentric videos," *Pattern Recognition Letters*, vol. 112, pp. 83–90, 2018.

[15] M. C. Popa, L. J. M. Rothkrantz, C. Shan, T. Gritti, and P. Wiggers, "Semantic assessment of shopping behavior using trajectories, shopping related actions, and context information," *Pattern Recognition Letters*, vol. 34, no. 7, pp. 809–819, 2013.

[16] M. C. Popa, L. J. M. Rothkrantz, P. Wiggers, and C. Shan, "Shopping behavior recognition using a language modeling analogy," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 1879–1889, 2013.

[17] F. Benites and E. Sapozhnikova, "Hierarchical interestingness measures for association rules with generalization on both antecedent and consequent sides," *Pattern Recognition Letters*, vol. 65, pp. 197–203, 2015.

[18] Z. He, F. Gu, C. Zhao, X. Liu, J. Wu, and J. Wang, "Conditional discriminative pattern mining: concepts and algorithms," *Information Sciences*, vol. 375, pp. 1–15, 2017.

[19] S. S. Titarenko, V. N. Titarenko, G. Aivaliotis, and J. Palczewski, "Fast implementation of pattern mining algorithms with time stamp uncertainties and temporal constraints," *Journal of Big Data*, vol. 6, pp. 1–34, 2019.

[20] C. Sun, Y. Fu, J. Zhou, and H. Gao, "Personalized privacy-preserving frequent itemset mining using randomized response," *Scientific World Journal*, vol. 2014, Article ID 686151, 2014.

[21] S. Mahmood, M. Shahbaz, and A. Guergachi, "Negative and positive association rules mining from text using frequent and infrequent itemsets," *Scientific World Journal*, vol. 2014, Article ID 973750, 2014.

[22] A. Lu and W. Ng, "Mining hesitation information by vague association rules," in *Proceedings of the Conceptual Modeling-ER 2007, 26th International Conference on Conceptual Modeling*, pp. 39–55, Auckland, New Zealand, November 2007.

[23] A. Pandey and K. R. Pardasani, "A model for mining course information using vague association rule," *International Journal of Computer Applications*, vol. 58, no. 20, pp. 1–5, 2012.

[24] A. K. Singh and A. Tiwari, "Vague set based association rule mining for profitable patterns," *International Journal for Science and Advance Research in Technology*, vol. 2, no. 2, pp. 1–6, 2016.

[25] P. Shrivastava and A. Tiwari, "On the use of vague set theory and genetic algorithm for hesitation information mining," *International Journal for Science and Advance Research in Technology*, vol. 2, no. 2, pp. 7–12, 2016.

[26] B. Sowkarthika, A. Tiwari, and U. Pratap Singh, "Elephant herding optimization based vague association rule mining algorithm," *International Journal of Computer Applications*, vol. 164, no. 5, pp. 15–23, 2017.

[27] J. Han and Y. Fu, "Mining multiple-level association rules in large databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, no. 5, pp. 798–805, 1999.

[28] P. Gautam and K. R. Pardasani, "A novel approach for discovery multi level fuzzy association rule mining," *Journal of Computing*, vol. 2, no. 3, pp. 56–64, 2010.

[29] F. A. El-Mouadib and A. O. El-Majressi, "A study of multilevel association rule mining," in *Proceedings of the International Arab Conference on Information Technology (ACIT)*, Benghazi, Libya, December 2010.

[30] M. Vanarse and S. Kasar, "Multilevel association rule mining for large datasets: a review," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 8, pp. 583–586, 2017.

[31] M. Dandotiya and M. Parmar, "Optimal hesitation rule mining using weighted apriori with genetic algorithm," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 3321–3328, 2019.

[32] A. Dixit, A. Tiwari, and R. K. Gupta, "Investigating multilevel hesitated patterns using vague set theory," in *Soft Computing: Theories and Applications*, H. Zolfagharinia, M. Pant, T. Kumar Sharma, R. Arya, and B. C. Sahana, Eds., Springer, Singapore, Asia, pp. 335–345, 2020.

[33] X.-S. Yang and X. He, "Nature-inspired optimization algorithms in engineering: overview and applications," *Nature-Inspired Computation in Engineering*, vol. 637, pp. 1–20, 2016.

[34] H. S. Pannu, D. Singh, and A. K. Malhi, "Improved particle swarm optimization based adaptive neuro-fuzzy inference system for benzene detection," *CLEAN-Soil Air Water*, vol. 46, no. 5, 2018.

[35] H. S. Pannu, D. Singh, and A. K. Malhi, "Multi-objective particle swarm optimization-based adaptive neuro-fuzzy inference system for benzene monitoring," *Neural Computing and Applications*, vol. 31, no. 7, pp. 2195–2205, 2019.