Received: November 24, 2022. Revised: December 30, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

A Novel Association Rule Mining Model for Generating Positive and Negative Association Rules with Hybridized Meta-Heuristic Development

Peddi Kishor[1]* **Sammulal Porika[2]**

[1]*R&D Cell, Jawaharlal Nehru Technological University Hyderabad, India*
[2]*Computer Science & Engineering Department, JNTUH University College of Engineering Jagtial, India*
* Corresponding author's Email: peddi.kishor@gmail.com

Abstract: Association rule mining (ARM) is to discover entire association rules with the confidence and support that are accordingly equivalent or above user-specific lower confidence and support. The existing suggested ARM methods can do the operations on the static datasets only, where the operation of the complete dataset must be accessible in the initial process of mining. Classical ARM methods are designed well on working with static data, and so, they cannot be used instantaneously for mining the association rules in data. To accomplish this, a novel framework is developed for association rule mining with the aid of an adaptive concept. Initially, the required data is collected from the benchmark datasets. The next step is to develop the novel association rule mining algorithm (NARMA), where the parameters are tuned by Modified Scavenger rate-based Dingo Billiards Optimization (MCR-DBO) by the hybridization of Dingo Optimizer (DOX) and Billiards-Inspired Optimization (BIO). This algorithm is used to identify the distinct itemsets of data. Consequently, the resultant dataset is given for finding the most eminent factors, where the itemsets are categorized into positive association rules as well as negative association rules. When the minimal threshold values are fixed with a minimum support count of 0.3, the proposed model achieved an average 81% confidence for generated top 1000 association rules from WebDocs dataset by minimizing the execution time and memory requirements. The outcomes disclose that the designed approach competently mines both the positive and negative rules and outperformed other recommended techniques.

Keywords: Association rule mining, Positive and negative rules, Modified scavenger rate-based dingo billiards optimization, Support, Confidence, Lift.

1. Introduction

In the recent world, the application of data mining is more vast [1]. Data mining generally stands for estimating a huge range of data by applying techniques along with the mining of hidden data. Data mining is eminently used in an extensive range of applications inclusive of trend and anomaly discovery, sequential pattern exploration, ARM, clustering, and classification [2]. Association analysis is a recent area of data mining research area. Generally, ARM is an efficient field and is also applicable in various applications for credit card business, fraud identification in the web, logistic regression, census data, protein sequences, bio-medical literature, medical detection field, market basket analysis, and so on [3]. The incorporation of events in data is revealed by the patterns. Consequently, data mining has various approaches, rules, and techniques for extracting specific data from the huge-scale dataset [4]. Thus, the association is generally utilized for making decisions with the factors like confidence, and support. Association is specifically helped in promoting the business with efficient decisions, which ensures the pre-described lower rate of two known metrics from a considered dataset [5].

Analyzing the association is an eminent research field in the field of data mining that is generally dependent on confidence and support factors for mining whether there is correlation existed among data [6]. ARM is one of the vastly preferred approaches for evaluating associations [7]. Here, the transactional analysis can be done via the association rules. There exist various association rules in a huge variety of user data. However, not entire association rules are helpful for consumers [8]. The exploration of association rules is generally separated into two processes. The initial process is to discover the whole group's concurrent itemsets in an iterative manner, where the consumers must hold the support rate as the lower as the minimum set [9]. The second step is to build the rule, where the users' settings must be higher than reliable. The exploration or prediction of entire recurrent itemsets is the interior part of discovering the association rules along with the ARM methods, which is also the primary and widest section of the computation [10]. An association rule is also named as a valid rule if it ensures a higher threshold of user-described minimum support and their confidence must be higher than user-specified lower threshold confidence. In general, a rule is initialized as $A \Rightarrow B$ that can specify the positive correlations among various items that is known as the Positive Association Rule (PAR) [11]. These rules can be presented at three diverse forms inclusive of $\neg A \Rightarrow \neg B$, $\neg A \Rightarrow B$, and $A \Rightarrow \neg B$. The negative correlations among items are specified as Negative Association Rules (NARs) [12].

On the other hand, enhancing the quality of ARM is more complicated work [13]. It is essential for adopting a huge range of approaches including post-estimation, process control, and prevention that utilize suitable strategies (like constraints), and thus, the customers participate in mining the data and also solved the issue of handling the association rules [14]. To be more specific, there are various recently suggested solutions for minimizing the count of association rules for the preservation of rules [15]. The most efficient algorithm is the FP-growth, which is also challenging to execute owing to the higher execution time, lower scalability, initialization of rules, etc [16]. Thus, heuristic strategies are assumed to the effective solutions for addressing several challenges in ARM methods. These methods aim to reduce the execution time and they do not require any information about the issue in a prior manner. Thus, integration of heuristic strategy with the ARM methods is helpful in mining of association rules. Some recent techniques suffer from the generation of incomplete and false rules [17]. Henceforth, this research contributes to the hybrid heuristic strategy along with the promotion of the ARM method. The innovative steps considered in this research work are ordered here.

- To design an intelligent ARM model with the contribution of mining the positive and negative association rules with help of hybridized meta-heuristic development along with the integration of ARM algorithm for helping the exploring the distributed, relational, and large-scale databases, etc for exploring the relationships among a set of items along with the promotion of scalability in the real-time scenarios.
- To reveal the NARMA strategy in designing the ARM model with the incorporation of the MCR-DBO algorithm for increasing the efficacy of mining the positive and negative association rules, where the factors of the apriori algorithm are optimized by the designed heuristic strategy with the aim of getting the optimal solutions with the higher convergence rate.
- To present the MCR-DBO algorithm for developing the NARMA strategy with the suitable tuning of factors like minimum support threshold, minimum confidence threshold, and minimum lift threshold to create the association rules in both positive and negative terms.

The residual parts of this research are illustrated here. Part 2 exhibit the related research works. Part 3 expresses the development of novel ARM of data using hybrid meta-heuristic algorithm. Part 4 designs the hybrid meta-heuristic with proposed MCR-DBO for implementing adaptive ARM. Part 5 reveals the ARM of data by NARMA. Part 6 explores the results and conclusion given in Part 7.

2. Literature survey

2.1 Traditional research works

In 2022, Chen et al. [18] have implemented a new data stream processing method with the adoption of a Type-2 Fuzzy Set with a mining technique known as Fuzzy Frequent Pattern. Initially, the sliding window approach was used for dynamically dividing the data streams, and then, from the numerical data stream, they have discovered the ambiguity a faster manner. The over-mining has been improved and the search dimension was reduced via designing a trees approach and also the framework of a dynamically compressed lists framework. Various experiments were explored under several sliding window sizes, and minimum support degrees to validate the effectiveness and efficacy of the proposed model regarding memory utilization and execution time. Thus, the outcomes have exhibited the contribution of exchanging preferences and information among the users for efficient cooperation in making better decisions.

In 2022, Koukaras et al. [19] have gathered Twitter data for mining the association rules and extract information regarding public attitudes around the world. This model was tried to explore the situation of use case scenario in handling the COVID-19 pandemic situations. They have also included topic gathering and visualizing approaches including wordclouds for forming themes or clusters of opinions. Then, this model has utilized ARM for discovering the recurrent wordsets and generated rules for inferring the client attitudes. The major aim of this approach was to use ARM as the post-processing method for enhancing the result of any topic-gathering approach. Thus, they have only saved the wordsets to discard the challenging ones. Finally, the outcomes have acquired the topics and lessened to a minimum number of topics while integrating the ARM with Latent Dirichlet Allocation. They have performed well and obtained generalizable and precise findings whereas the complications concerning user attitudes toward social media data were addressed.

In 2021, Antonello et al. [20] have offered an ARM model with a data-driven technique to identify the challenging functional dependencies between the elements of various complicated systems in infrastructures from processing the alarm messages of huge-range databases. This data was in a binary form, where the ARM technique was utilized to explore the complicated dependencies among elements of several systems. They have finally identified the sets of functionally derived elements. This model has exhibited superior outcomes in mining the ARM model. In 2020, Wang and Zheng [21] have implemented an improved Apriori technique for processing the time series of recurrent itemsets. From the time series data, the association rules were mined by estimating and analyzing the techniques. This process of mining was done based on time constraints. The final experimental outcomes have exhibited higher effectiveness regarding storage consumption.

In 2020, Dong et al. [22] proposed a new LOGIC approach for pruning and estimating the redundant association rules along with the logical reasoning. Additionally, several minimum confidences and correlation coefficients were integrated for guaranteeing a stronger correlation among the mined association rules, which must be controlled efficiently and flexibly. The findings revealed that this model has pruned to a higher level of mining.

In 2019, Bagui and Dhar [23] have designed a new Apriori algorithm along with a Hadoop framework using parallel and distributed MapReduce infrastructure. They have offered a framework for mining both negative and positive association rules in handling big data with the help of the Apriori technique and itemset mining. They have estimated and presented the outcomes of some optimization constraints in MapReduce Hadoop infrastructure. These experimental outcomes have offered superiority regarding the efficacy concerning runtime and also generated rules. Finally, a huge number of parallelization was found with the representation of higher block sizes, which has resulted in a higher effectiveness of algorithm.

In 2019, Su et al. [24] have incorporated the heuristic technique along with the ARM for mining and generating both negative and positive rules using FP-growth and particle swarm optimization (PSO). This optimizer has discovered the optimal solutions and discovered global support, where the association rules were mined using the FP-growth technique. Next, the information entropy was lastly measured for examining the efficacy of ARM, and then, the improved technique was used for analyzing the correlation among the social security events.

In 2022, Zheng et al. [25] have implemented an ARM technique to stream the temporal data for exploring the possible correlations to stream temporal data. The streaming temporal data-based fuzzy sets (STDFS) were recommended for creating the association rules from the data. Then, the experimental evaluation was done on a smaller dataset and compared the efficiency of this approach and showed the better mining of association rules in the temporal data with the aid of time stamps.

2.2 Problem definition

ARM is one kind of building the relationship among various itemsets of the data. Different algorithms are deployed for finding the association rules, but it stuck with certain limitations. Numerous studies are exploring the advantages and disadvantages of existing ARM in data that is given in Table 1. T2FS [18] resolves the drift issue and supports real-time data. But, the total transaction tends to cause a computational burden. latent dirichlet allocation [19] efficiently analyzed the word sets of data and uses them for real-time data. However, fluctuation occurs among the data since it handles large-scale dimensional data. Infrastructural ARM [20] mitigates the computational complexity and decreases the large set of generated rules. Yet, the reliability and availability of the model get affected. Improved Apriori algorithm [21] reduces the time and storage complexity. But, it does not suitable for practical implications. LOGIC [22] evades the weak correlation rules. On the other hand, it restricts large-scale dimensional datasets and features. Hadoop MapReduce [23] retains the minimum confidence and reduces time complexity. Due to large size

Received: November 24, 2022. Revised: December 30, 2022.

Table 1. Advantages and disadvantages of existing association rule mining models

Author [citation]	Methodology	Features	Challenges
Chen et al. [18]	T2FS	• Resolves the drift issue. • Supports for real-time data.	• The total transaction tends to cause a computational burden.
Koukaras et al. [19]	Latent Dirichlet Allocation	• Efficiently analyzed the word sets of data. • It uses for real-time data.	• The fluctuation occurs among the data since it handles large-scale dimensional data.
Antonello et al. [20]	CTI	• Mitigates the computational complexity. • Decreases the large set of generated rules.	• The reliability and availability of the model get affected.
Wang and Zheng [21]	Improved Apriori algorithm	• Reduces the time and storage complexity.	• It does not suitable for practical implications.
Dong et al. [22]	LOGIC	• Effectively evades the weak correlation rules.	• It restricts large-scale dimensional datasets and features.
Bagui and Dhar [23]	Hadoop MapReduce	• It retains the minimum confidence. • Reduces the time complexity.	• Due to large datasets and large block sizes, structural complexity occurs.
Suet al. [24]	PSO	• Enhances the efficiency of the model and evades the artificial blind setting.	• It is suggested to include the hybridized algorithm for further enhancement.
Zheng et al. [25]	STDFS	• Estimates the fuzzy rule to sustain better results.	• Owing to a large set of rules, overlap occurs that degrades the performance.

datasets and blocks, structural complexity occurs. PSO [24] enhances the efficiency of the model and evades the artificial blind setting. It is suggested to include the hybridized algorithm for further enhancement. STDFS [25] estimates the fuzzy rule to sustain better results. Owing to a large set of rules, overlap occurs that degrades the performance. To alleviate all the challenges, it provokes to design of a novel ARM model.

3. Development of novel association rule mining of data using hybrid meta-heuristic algorithm

3.1 Architecture of association rule mining

ARM is an essential process of declaring significant trends or patterns. It is specifically utilized for predicting the recurrent itemsets, trends, and patterns among those items. ARM is generally divided into two stages, where the initial step discovers the itemsets by determining the threshold of controlling the occurrences in the database, which are known as large or frequent itemsets. The second phase with the generation of association rules from the huge-scale itemsets with the factors of lower confidence. As the mining process is specifically helpful to extract the correlation among items from databases, it is more eminent for extracting the set of itemsets. Some challenges observed in the ARM process are listed here. ARM determines the functionality of association patterns intending to make suitable decisions that are integrated to be ideal within the process of association mining. The techniques must ensure the minimum I/O overhead and central processing unit (CPU) overhead. Moreover, while designing the algorithms for generating the association rules, various interesting factors must be incorporated that are chi-squared value, Gini, entropy gain, lift, conviction, Laplace value, gain, confidence, and support. These factors are also called indicators of the degree to which items in associations are correlated with each other. The major drawback is finding the associations by selecting the user-defined parameters. The most eminent and attractive parameters are confidence and support thresholds, which must be lower. It indicates that the amount of recurrent itemsets is increased that is directly proportional to the number of rules generated. However, most of these rules are redundant and thus, the choice of suitable values for these factors is highly challenging in the ARM model. Another problem is to design or develop techniques for several levels of association rules for achieving time efficiency and reducing the number of iterations. The time efficacy is attained by minimizing the database scans at every level. Last but not least issue is the generation of association rules in multiple levels or single levels, which must be formulated with

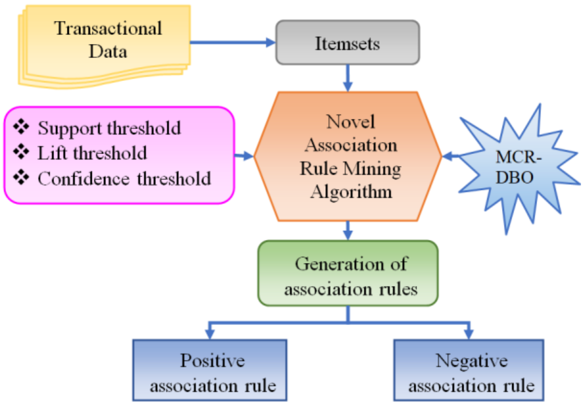


Figure. 1 ARM Architecture with the generation of positive and negative rules

precise data sources in suitable data format. Thus, this research explores a new ARM architecture with the adoption of intelligent methods, which is depicted in Fig. 1.

This research work explores a novel ARM model for generating positive and negative association rules with the design of a heuristic algorithm aided with the standard ARM method. It uses the MCR-DBO algorithm for the promotion of scalability in real-time scenarios.

Initially, the NARMA strategy is used in the ARM model with the incorporation of the MCR-DBO algorithm to create the itemsets. Thus, in the NARMA strategy, the factors like correlation coefficient, support, confidence, and lift factors are determined for estimating the rules, where the itemsets are categorized into positive association rules as well as negative association rules. Here, as the contribution, the factors of the apriori algorithm like minimum support threshold, minimum confidence threshold, and minimum lift threshold are optimized by the designed heuristic MCR-DBO strategy to get the optimal solutions with the higher convergence rate. Moreover, this model must ensure the minimum level of support, confidence, and lift levels for getting the efficient mined positive as well as negative association rules.

3.2 Datasets used for association rule mining

This ARM model gathers the WebDocs dataset from "http://fimi.uantw erpen .be/data/ : Access Date: 2022-10-20". It has a set of 1.7 million web HTML documents. These data were applied with a stemming technique to remove the HTML tags via filtering the documents. Next, every document is transformed into a diverse transaction comprising a group of several terms (called items) that are observed in the document. This database has 5,267,656 items with 1,692,082 transactions.

The gathered collection of the database is referred to as $\mathrm{DB}_{\mathrm{tr}}$, where $\mathrm{tr} = 1,2,3,\cdots,\mathrm{Tr}$, where the total number of acquired transactions is known as Tr.

A group of unique items is assumed as U, where $Y = \{y_1, y_2, \cdots, y_U\}$ and also assume $\mathrm{DB}_{\mathrm{tr}}$ – is the database of transactions as Tr, in which every transaction Tr can be a set of items or an item $\mathrm{IT}_{\mathrm{ik}}$ including a subset of Y. A unique identifier is used for associating every transaction. Thus, the association rule is formed as follows $E \Rightarrow F$, in which $E \subseteq Y$, $F \subseteq Y$ and $E \cap F = \varphi$.

4. Hybrid Meta-Heuristic with proposed MCR-DBOfor implementing novel association rule mining

4.1 Conventional DOX

DOX [26] is motivated by the dingo behavior in nature, which is a category of the dog in the canis family, which involves the community and collective habits of dingoes. These animals live in groups of 12-15 animals. As with other heuristic techniques, it follows some strategies for hunting inclusive exploring, encircling, and exploiting behaviors. This algorithm helps in handling practical engineering issues. These dogs follow social grading with the sorting of alpha dingoes, beta dingoes, and other remaining members. Here, the alpha plays a primary role in the group to assist the other members of the group with various decisions. The beta dingo is the second member to manage the decisions ordered by the alpha dingo whereas the residual members follow the decisions given by the alpha and beta members. The initial process of encircle behavior is modeled as follows.

$$\vec{J}_J = \left|\vec{A} \cdot \vec{B}_{\mathrm{prey}}(s) - \vec{B}(s)\right| \tag{1}$$

$$\vec{B}(t+1) = \vec{B}_{\mathrm{prey}}(t) - \vec{N}s \cdot \vec{J}(j) \tag{2}$$

$$\vec{A} = 2 \cdot \vec{a}_1 \tag{3}$$

$$\vec{N}s = 2\vec{n}s \cdot \vec{a}_2 - \vec{n}s \tag{4}$$

$$\vec{n}s = 3 - \left(T * \left(\frac{3}{T_{\max}}\right)\right) \tag{5}$$

In the aforementioned derivations, $\vec{A}$ and $\vec{N}s$ conveys the coefficient vectors, $\vec{B}_{\mathrm{prey}}$ denotes the location vector of prey, $\vec{a}_1$ and $\vec{a}_2$ expresses the random values in the range of [0, 1], $\vec{J}_J$ typifies the

Algorithm 1: depicts the pseudo code of DOX

Algorithm 1: DOX [26]

Derive the population of dingoes and corresponding constraints
Originate the initial solutions
For every solution
While the finishing criterion is not met ($t < T_{\max}$)
  Estimate the fitness values for every dingo and the intensity
  For every dingo
    Determine and update the recent solutions alpha, beta and other dingoes
    Re-update the solutions via Eq. (1)
   Execute the fitness of entire solutions and the corresponding intensity rate of all the dingoes
   Accumulate the finer solutions
  End for
$t = t + 1$
End while
End for
Acquire optimal solutions

distance between the dingo and prey, $\vec{n}s$is the linear value decreased value in the limit of [3, 0], $T_{\max}$ signifies the maximum number of iteration, t stands for the iteration, and $\vec{B}(s)$ refers to as the location vector of the dingo.

The position of prey is assisted in formulating the position of the dingo. The hunting is done by formulating orders from the leader of the pack known as the alpha dingo. The location of dingoes in the pack is equated in evaluating the place of optimal search solutions as computed here.

$$\vec{J}_\alpha = |\vec{A}_1 \cdot \vec{B}_\alpha - \vec{B}| \tag{6}$$

$$\vec{J}_\beta = |\vec{A}_2 \cdot \vec{B}_\beta - \vec{B}| \tag{7}$$

$$\vec{J}_{\text{od}} = |\vec{A}_3 \cdot \vec{B}_{\text{od}} - \vec{B}| \tag{8}$$

$$\vec{J}_1 = |\vec{B}_\alpha - \vec{N}s \cdot \vec{B}_\alpha| \tag{9}$$

$$\vec{J}_2 = |\vec{B}_\beta - \vec{N}s \cdot \vec{B}_\beta| \tag{10}$$

$$\vec{J}_3 = |\vec{B}_{\text{od}} - \vec{N}s \cdot \vec{B}_{\text{od}}| \tag{11}$$

Further, the intensity of every dingo is formulated with the estimation of the fitness among the alpha, beta, and other solutions, which are correspondingly specified as H_α, H_β and H_{od}.

$$\vec{I}T_\alpha = \log\left(\frac{1}{H_\alpha - 1Z - 100} + 1\right) \tag{12}$$

$$\vec{I}T_\beta = \log\left(\frac{1}{H_\beta - 1Z - 100} + 1\right) \tag{13}$$

$$\vec{I}T_{\text{od}} = \log\left(\frac{1}{H_{\text{od}} - 1Z - 100} + 1\right) \tag{14}$$

Finally, the position of prey is determined with the help of fitness among the solutions. These solution updating is equated for giving the precise optimal solutions.

4.2 Conventional BIO

BIO [27] is a recent meta-heuristic strategy motivated by the billiards game. BIO is designed by formulating the characteristics of a multi-dimensional billiards ball by assuming a number of decision parameters. BIO is initiated by generating the balls in random allocation; the pockets are specified as the optimal solutions. Here, the considered solutions in terms of balls are categorized into normal balls and cue balls. The target ball is hit by every cue ball, which goes to a pocket. While encountering other balls by the cue balls, collision and kinematic laws are governed.

Initialization process is the primary task of any heuristic scheme, where BIO is designed in Eq. (15).

$$J^0_{w,v} = \text{rn}^0_v + d_{[0,1]}(\text{rn}^{\max}_v - \text{rn}^{\min}_v) \tag{15}$$

In Eq. (15), the number of variables and search individuals are correspondingly indicated by v and w, where $v = 1,2,3, \cdots 2V$, and $w = 1,2,3, \cdots W$, the preliminary values of the v^{th} variable with the w^{th} ball is referred to as $J^0_{w,v}$, a random number with the standardized allocation of $[0, 1]$ is particularized as $d_{[0,1]}$, and the maximum and minimum acceptable constraints of the v^{th}variable is respectively framed as $\text{rn}^{\max}_w$ and $\text{rn}^{\min}_w$.

The next process is to estimate the locations of pockets and balls depending on the objective formulation. Latter, the pockets are estimated by offering the exploitation capability of this technique and also have a memory of saving the top optimal solutions attained so far. The presence of this memory enhances the efficiency which leads to lower computational expensiveness. It finally updates the memory and replaces it with better balls obtained locations in every iteration process. Then, the balls are grouped and ordered by taking the fitness values and categorized into two similar sets of cue balls and usual balls. It is generally followed by the colliding bodies optimization technique. Further, the allotment of pockets to the balls is done by using the roulette-wheel-choosing strategy by choosing a destination pocket for every common ball. The lower fitness of pockets is taken more advantage of, where the probability of selecting a pocket is derived in Eq. (16).

$$\Re_l = \frac{e^{-\xi H_l}}{\sum e^{-\xi H_l}} \tag{16}$$

Here, the choice pressure is represented as ξ, and the value of the objective function in the l^{th} pocket is known as H_l. Finally, the target balls can be reached by the most current cue ball and forwarded to the pockets. The location updating for the balls is carried out after occurring colliding with the balls due to the shorter precision. Thus, it increases the exploitation capability by decreasing the error. It is carried out for the common balls using Eq. (17).

$$J_{w.v}^{\text{new}} = d_{[-\text{Err},\text{Err}]}(1-\ell)\left(J_{w,v}^{\text{old}} - \text{SA}_{l,v}^{w}\right) + \text{SA}_{l,v}^{w} \tag{17}$$

In Eq. (17), the older and newer values of the v^{th} variable from the w^{th} common ball are accordingly specified as $J_{w.v}^{\text{old}}$ and $J_{w.v}^{\text{new}}$, $d_{[-\text{Err},\text{Err}]}$ illustrate the arbitrary number in the range of error rate with$[-\text{Err},\text{Err}]$is as $\text{rn}_{[-\text{er},\text{er}]}$, and the l^{th} pocket of the v^{th} variable to w^{th} common ball pair is specified as $\text{SA}_{l,v}^{w}$.

The precision rate PC is formulated with the iteration and a maximum number of iterations.

$$\text{PC} = \frac{t}{t_{\max}} \tag{18}$$

The next step is to discover the velocities of common balls after the collision occurs.

$$\vec{v}_v' = \sqrt{2g\overrightarrow{J_w^{\text{old}} J_w^{\text{new}}}}\ X\ \widehat{J_w^{\text{old}} J_w^{\text{new}}} \tag{19}$$

In Eq. (19), g specifies the acceleration rate, $\widehat{J_w^{\text{old}} J_w^{\text{new}}}$ stands for the vector of unit movement regarding the w^{th} common ball after the collision, $\overrightarrow{J_w^{\text{old}} J_w^{\text{new}}}$ refers to the forwarding of the w^{th}common ball, and $\vec{v}_v'$ specifies the velocity of the w^{th}common ball. Then, the velocity for the cue balls collision occurring after and before the process is consequently illustrated as $\vec{v}_{v+V}'$, and $\vec{v}_{v+V}$, and equated here, where the user definite variable in the range of [0, 1) is indicated asϖ.

$$\vec{v}_{v+V} = \frac{\vec{v}_v'}{\widehat{J_w^{\text{old}} J_w^{\text{new}}} \cdot J_w^{\text{old}} J_w^{\text{new}}} J_{v+V}^{\text{old}} J_{v+V}^{\text{old}} \tag{20}$$

$$\vec{v}_{v+V}' = \varpi\left(1 - \frac{t}{t_{\max}}\right)\left(\vec{v}_{v+V} - \vec{v}_v'\right) \tag{21}$$

The kinematic correlation is used for determining the locations of cue balls along with their velocity determination as formulated in Eq. (22).

$$\overrightarrow{J_{v+V}^{\text{new}}} = \frac{\vec{v}_{v+V}'}{2g}\vec{v}_{v+V}' + \overrightarrow{J_w^{\text{old}}} \tag{22}$$

The next stage is to avoid the exploration ability via utilizing the threshold for escaping, where the dimension of the updated ball varies or not. The threshold *Thr* is derived for estimating the uniformly allocated arbitrary number for every updated ball. If $d <$ *Thr* is verified, then the regeneration of the modernized ball is done with arbitrary dimensions as formulated in Eq. (23).

$$J_{w,v} = \text{rn}_v^{\min} + d_{[0,1]}\left(\text{rn}_v^{\max} - \text{rn}_v^{\min}\right) \tag{23}$$

Finally, the location of the balls is updated whereas some balls may be positioned outer side of the specified boundary. Thus, there is a need of examining the boundary often. The search procedure is stopped after verifying a particular condition of reaching the final number of iterations. If it is verified, the process is stopped and obtained the optimal solutions and thus, it delivers the optimal pockets, or else the process will be reinitiated.

4.3 Proposed MCR-DBO

In the suggested ARM model, a new heuristic strategy known as MCR-DBO is designed for promoting the performance of the ARM process. Here, the negative and positive association rules are mined with the help of a newly recommended NARMA technique, where the MCR-DBO is implemented for tuning the parameters like minimum support threshold, minimum confidence threshold, and minimum lift threshold in the Apriori Algorithm to create the association rules in both positive and negative terms. In this paper, the DOX is selected for enhancing the performance, where the features of DOX are efficiently protect the solutions from the local optima, are utilized in several optimization problems, require lower computational efforts, and also reach the global optimal solutions optimally. However, DOX suffers from converging ability and thus, the BIO is introduced in this research work for integrating the features of BIO. BIO holds better convergence capability, has higher sensitivity

Received: November 24, 2022. Revised: December 30, 2022.

Algorithm 2: specifies the pseudo-code of the BIO

Algorithm 2: BIO [27]

Derive the population and formulate the necessary parameters
Examination of the fitness among entire search individuals
Determine the random population by Eq. (15).
while ($t < T_{\max}$)
  Formulate the objective functional value of the locations of balls and pocket
  Evaluate the population and pocket memory
  Produce a set of ordinary balls and cue balls
  For every pair of ball
    Generate population with roulette-wheel strategy
  End for
  Use Eq. (17) for determining the location of an ordinary ball
  After the collision, compute the velocity of the common ball and cue ball by Eq. (19) and Eq. (21)
  Formulate the velocity before collision for cue ball by Eq. (20)
  Generate the location of the cue ball by Eq. (22)
  if $d <$ Thr
    Regenerate the locations by Eq. (23)
  End if
  Examine the specified boundary and verify it
Verify the termination criterion
$t = t + 1$
End while
End
Acquire the optimal solutions

behaviors, uses limited parameters, along with effective exploitation properties, along with stable in getting the final results.

Thus, the incorporation of BIO into DOX gives precise outcomes in the ARM model. It also helps in generating the association rules for both negative and positive aspects. Finally, MCR-DBO is designed by modifying the parameter in DOX. Here, the hunting or Scavenger rate in DOX is known as $\vec{B}$, which is mathematically modeled in Eq. (24).

$$\vec{B} = \frac{(\mathrm{Thr} \times \mathrm{CT})}{g*1000} \tag{24}$$

In Eq. (24), g is the acceleration rate, *Thr* is the escaping threshold in BIO, and *CT* is determined by discovering how many several fitness solutions are greater than mean fitness, and $\vec{a}_1$ stands for the random parameter lie among [0, 1], which is used for updating the solutions.

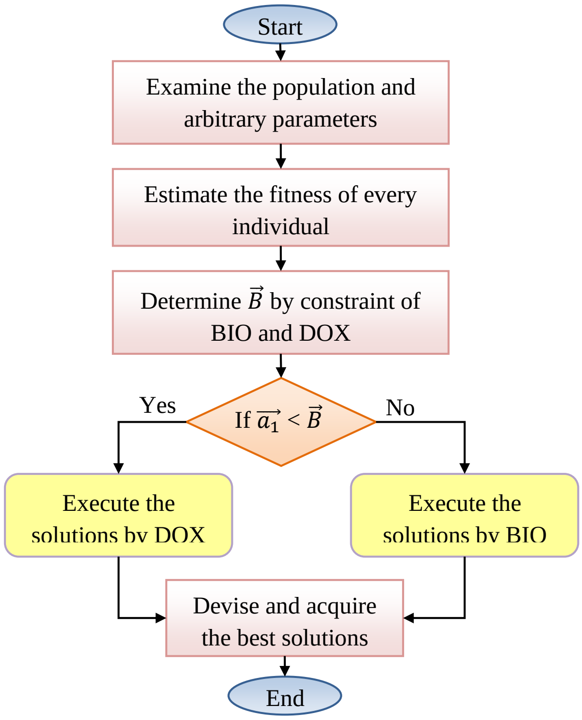


Figure. 2Flowchart of the MCR-DBO for ARM model

Algorithm 3: The pseudo-code of the MCR-DBO

Algorithm 3: MCR-DBO

Derive the population & formulate the necessary parameters
Examination of the fitness among entire search individuals
while ($\mathrm{t} < \mathrm{T}_{\max}$)
  Determine hunting or Scavenger rate in DOX $\vec{B}$ by Eq. (24)
  If $\vec{a}_1 < \vec{B}$
    Execute solution updating by DOX method using Algorithm 1
  Else
    Execute solution updating by BIO method using Algorithm 2
  End if
  Verify the termination criterion
  $\mathrm{t} = \mathrm{t} + 1$
End while
End
Acquire the optimal solutions

If $\vec{a}_1 < \vec{B}$, then the solution updating is done by the DOX method, or else the solutions are executed by the BIO method. The pseudo-code of the recommended MCR-DBO is given in Algorithm 3. The flowchart of the recommended MCR-DBO for the ARM model is given in Fig. 2.

5. Association rule mining of data using NARMA

5.1 Concept of Apriori algorithm

The eminent ARM algorithms are Apriori techniques generally used for mining positive associational rules. Here, the positive association rules are mined for discovering the items, which are positively correlated with one another. The negative association rules are described as items, which are correlated negatively, which is one item goes down and another goes up. The Apriori technique is one of the eminent utilized techniques for ARM. The recurrent patterns are discovered using the Apriori technique, in which these patterns occurred regularly in the data. It uses an iterative technique, in which *ik*-itemsets are utilized for exploring $(ik + 1)$ itemsets. The recurrent itemsets are discovered by finding the initial of recurrent 1-itemsets via scrutinizing the database and accumulating their counts. The itemsets are derived by verifying the minimum support threshold. It results in finding the recurrent 2-itemsets. This procedure is continued until reaching an empty set of newly produced itemsets, which has to meet the minimum support threshold until emptying the itemsets. Further, there is a need of verifying the itemsets over a lower confidence level for estimating the association dataset. It is very simpler for executionusually helped in mining the entire recurring itemsets.

For example, the count of candidate items in the set is given as $\mathrm{IT_{ik}}$, which is estimated by the Apriori technique. It is estimated with support level after it is discovered. Next, the pruning is carried out for the itemsets, which is less than the support level. Then,$\mathrm{JL_{ik-1}}$will be used for connecting themselves and leading to$\mathrm{IT_{ik}}$. Then, the pruning of$\mathrm{JL_{ik-1}}$is done, where the number of entire items occurred in$\mathrm{JL_{ik-1}}$ and so, these itemsets with lower than $(ik - 1)$ is deleted in $\mathrm{JL_{ik-1}}$. Likewise, the count of related itemsets will minimize for declining the number of candidate itemsets.

The representation of the Apriori algorithm is given in Fig. 3.

5.2 Novel association rule mining algorithm

The major limitation in executing the Apriori-based technique is identifying the functional dependencies for alarming the databases. It is caused due to the lower-level support assigned threshold. It has resulted in unfavourable computational expensiveness by discovering the amount of recurring patterns. It does not permit the exploration of the rules with integrations of frequent and rare items. Though, the applications of this technique try to estimate the huge-scale databases and fail. Although, it is simpler and clear, it wastes more time for holding a huge range of candidate sets with more recurring itemsets with large-scale itemsets and lower support levels. It requires higher memory capacity and suffers from inefficiency in processing huge-scale transactions. Thus, this research work explores a new MCR-DBO algorithm for mining the association rules in an effective manner. Thus, tuning the parameters like minimum support threshold, minimum confidence threshold, and minimum lift threshold in Apriori Algorithm are done by the same MCR-DBO technique to create the association rules in both positive and negative terms. The generation of positive and negative rules is created by NARMA with the formulation of the following metrics. The major aim of this model is mathematically derived in Eq. (25).

$$H = \underset{\{\mathrm{Lft,Cft,Sut}\}}{\mathrm{argmax}} \left(\mathrm{LF} + \mathrm{CF} + \mathrm{SU} + \mathrm{CC}\right) \tag{25}$$

Here, the threshold values of support, confidence, and lift are mathematically represented by Sut, Cft, and Lft metrics, where the range of the minimum support threshold is given among [0.01 - 0.99], the minimum confidence threshold is specified

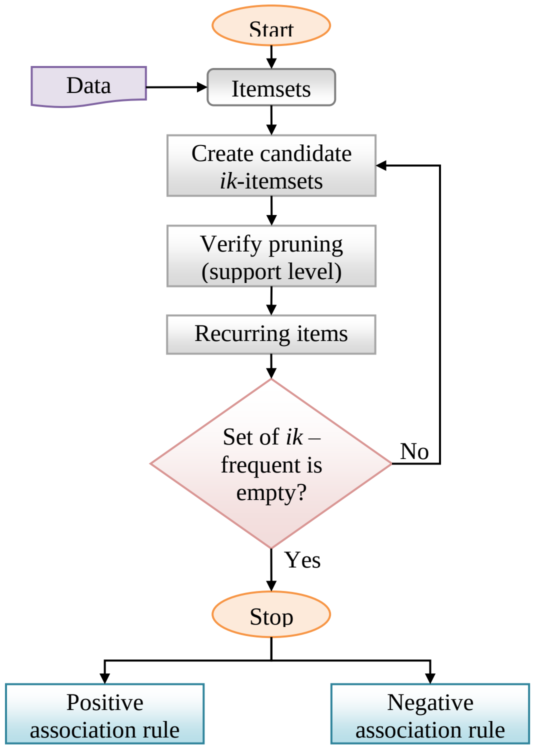


Figure. 3 Working process of Apriori algorithm for mining the association rules in the ARM model

among [0.01 - 0.99] and the minimum lift threshold lies among [0.01 - 0.99]. The major factors in constructing the association rules by NARMA are organized here to get the maximum rate of lift, confidence, correlation coefficient and support.

Lift LF**:** It is the relation of confidence to estimated confidence. It is a value that permits taking information about the augment in the possibility of the "then" (consequent) specified in the "if" (antecedent) section. It is formulated in Eq (26).

$$LF(E \Rightarrow F) = \frac{su(E \Rightarrow F)}{su(E)su(F)} \quad (26)$$

Confidence CF**:** It is the fraction of the amount of transactions, which comprise entire itemsets in the antecedent and also consequent (known as the support) to the count of transactions, which contain entire items in the ancestor. It is formulated in Eq (27).

$$\mathrm{CF}_{\mathrm{EF}} = \frac{\sigma_{\mathrm{EF}}}{\sigma_E} \quad (27)$$

In Eq. (27), the number of transactions with deviations in F^{th} index is known as σ_{EF} and for E^{th} the index is specified as σ_E.

Support SU**:** It is generally known as the amount of transactions which contain entire items in the consequent and ancestor modules of the rule. It is formulated in Eq (28).

$$SU(E \Rightarrow F) = \frac{\sigma_{EF}}{nj} \quad (28)$$

Here, nj is the distribute network.

Correlation coefficient CC**:** It is a measure of verifying the mining of rules, where the search space is reduced via a pruning scheme. It is formulated in Eq (29).

$$\mathrm{CC}_{\mathrm{EF}} = \frac{\mathrm{su}(E \cup F)}{\mathrm{su}(E)\mathrm{su}(F)} \quad (29)$$

The support of itemset F is defined as $SU(F)$ and the support of itemset E is known as $SU(E)$.

The threshold value must be maintained in the designed ARM model, which is derived as follows.

$$\mathrm{SU}(E \Rightarrow F) \geq \mathrm{Sut} \quad (30)$$

$$\mathrm{LF}(E \Rightarrow F) \geq \mathrm{Lft} \quad (31)$$

$$\mathrm{SU}(E \Rightarrow F) \geq \mathrm{Cft} \quad (32)$$

These constraints must be satisfied for getting the final association rules. The design of NARMA technique using the MCR-DBO algorithm is given in Fig. 4.

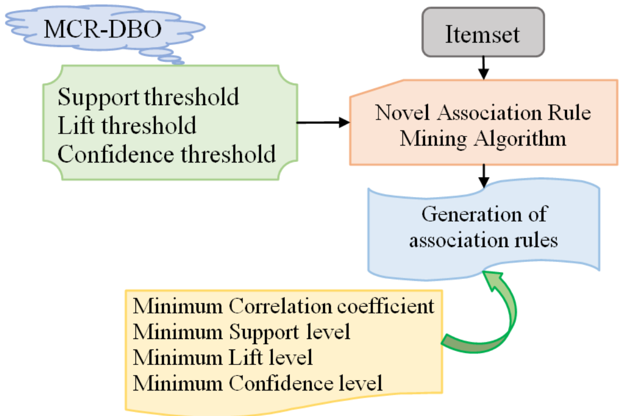


Figure. 4 NARMA using MCR-DBO algorithm for generating the association rules

5.3 Creating positive and negative association rules

An ARM can be done by following the derived terminologies.

- ❖ Association rule is derived as specified here $E \Rightarrow F$, in which $E \subseteq Y$, $F \subseteq Y$ and $E \cap F = \varphi$.
- ❖ The support for a rule is formulated in a transaction set DB that is the possibility of E existing in DB.
- ❖ The conditional probability is derived as the confidence of a rule, in which the succeeding F is true to the specified antecedent E.
- ❖ A positive item y_{ik} is presented in a Tr.
- ❖ A negative item $\neg y_{ik}$ is included in a Tr.
- ❖ Minimum support, confidence, lift, and correlation coefficient threshold must be satisfied for the generated positive association rule, which is generally represented by $E \Rightarrow F$.
- ❖ A negative association rule is derived as $E \Rightarrow F$, in which E or F either has one negative item.

Finally, if the confidence, correlation coefficient, and support of either (E and not F) or (not E and F) or (not E and not F) is higher than the lower confidence, support threshold and the lift of this similar rule is also larger than 1, then it is known as a negative association rule. If the confidence, support, and correlation coefficient of ($E \Rightarrow F$) is higher than the minimum threshold confidence, support threshold, and the lift ($E \Rightarrow F$) is higher than 1, then this is called as a positive association rule.

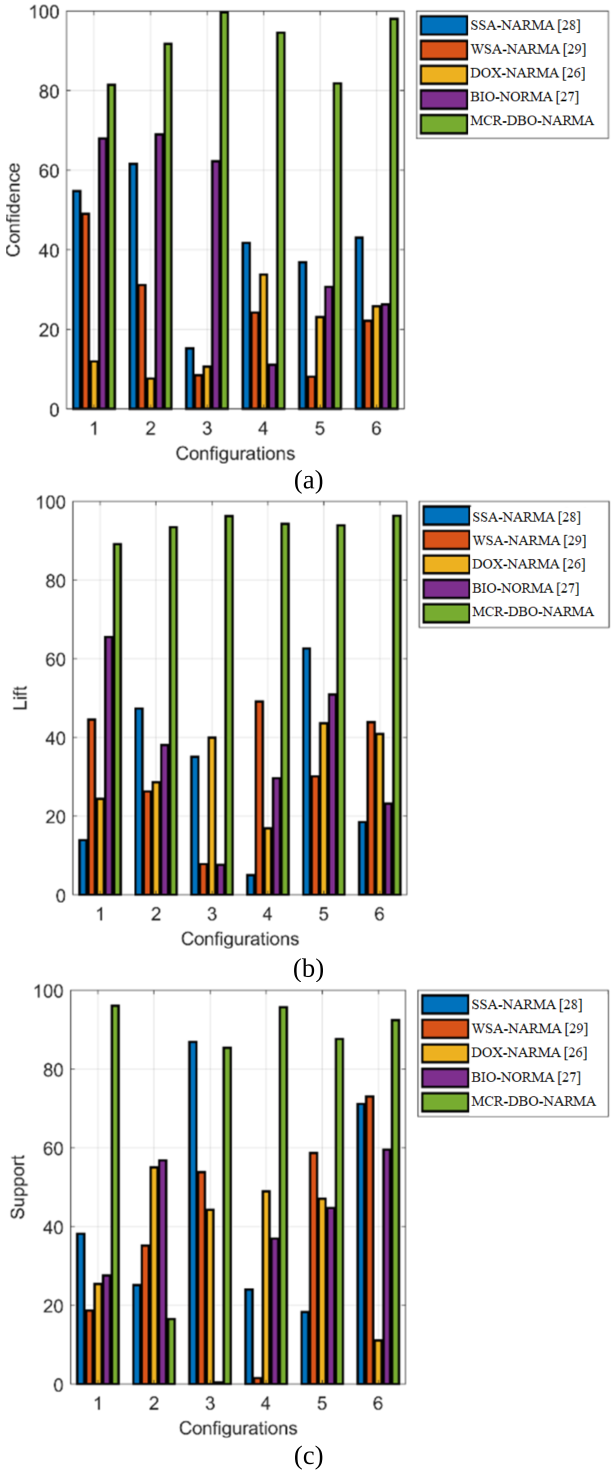


Figure. 5 Performance investigation on the ARM model with six different configurations over heuristic strategies regarding: (a) Confidence, (b) Lift, and (c) Support

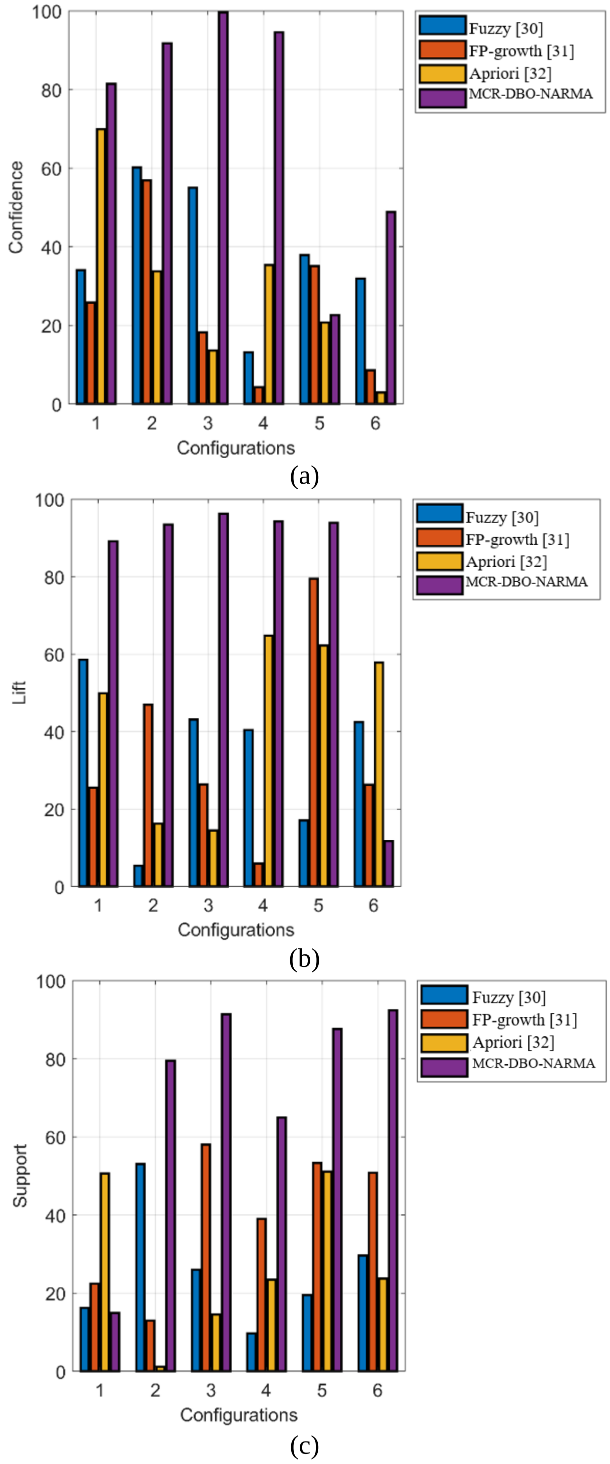
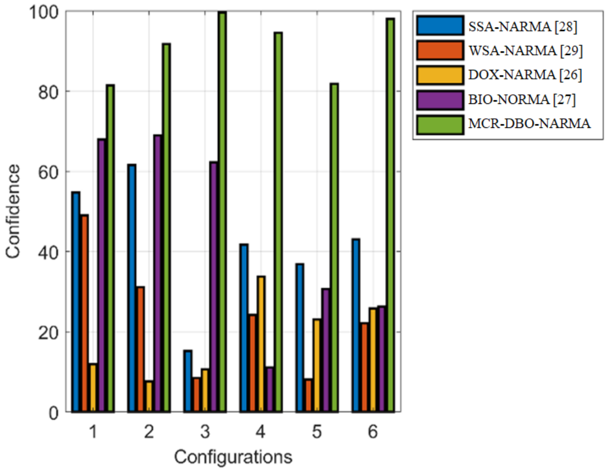


Figure. 6 Performance investigation on the ARM model with six different configurations over traditional ARM algorithms regarding: (a) Confidence, (b) Lift, and (c) Support

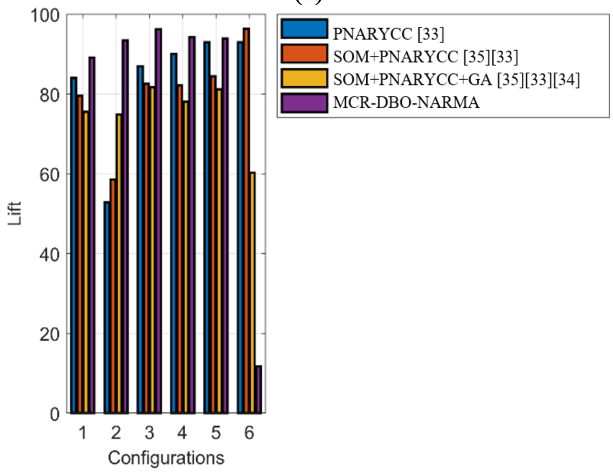
6. Results and evaluation

6.1 Proposed model and description

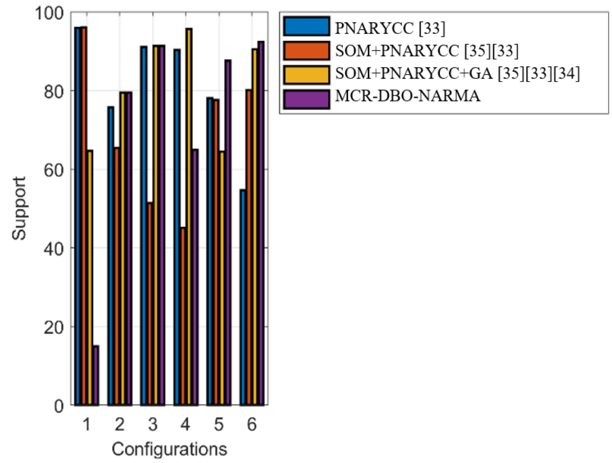
The recommended model was evaluated in MATLAB 2020a and the estimation was carried out by different measures. The effectiveness of the model was also done with comparative analysis over six configurations with the constraint of varying the transaction from [10000, 20000, 30000, 40000, 50000, and 60000], respectively along with the 100 number of iterations. The designed strategy was



(a)



(b)



(c)

Figure. 7 Performance investigation on the ARM model with six different configurations over traditional ARM algorithms regarding: (a) Confidence, (b) Lift, and (c) Support

efficiently analyzed over conventional methods like Salp Swam algorithm (SSA) [28], water strider algorithm (WSA) [29], DOX [26], BIO [27], classical ARM methods inclusive Fuzzy [30], FP-growth [31], Apriori [32] along with the standard known approaches were "Positive and negative ARM using Yules correlation coefficient (PNARYCC) [33]", genetic algorithm (GA) [34] and self-organizing map (SOM) [35].

6.2 Investigation over heuristic methods

In the designed ARM model, the MCR-DBO-NARMA scheme is estimated in terms of six configuration schemes as depicted in Fig. 5. The recommended MCR-DBO-NARMA increases the higher level of confidence, lift and support values by fixing the minimum similar threshold values. Thus, the MCR-DBO-NARMA efficiently creates both the negative and positive association rules in the recommended ARM model.

6.3 Investigation over traditional ARM methods

In the proposed ARM model, the estimation on the conventional approaches are evaluated over six different configurations as depicted in Fig. 6 and Fig. 7. For example, MCR-DBO-NARMA reaches higher level of confidence, support and lift levels to gain the superiority over conventional approaches. It finally ensures the higher effectiveness of the designed model to demonstrate the superior level of generated association rules.

6.4 Convergence estimation

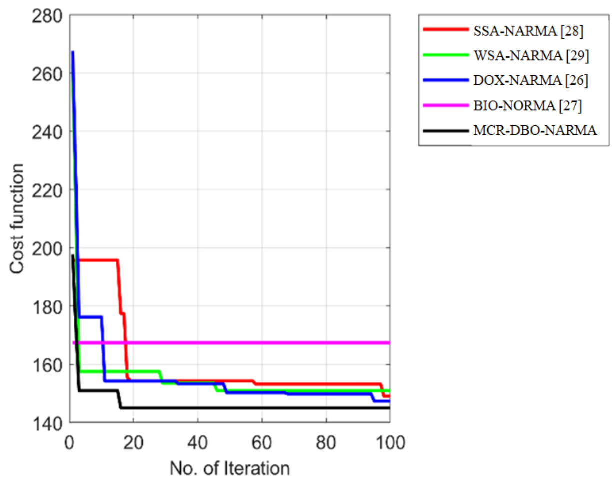
The convergence estimation is analyzed regarding proposed and traditional methods as given in Fig. 8. The MCR-DBO-NARMA ensures the highest convergence rate for all the six configurations with the minimum cost function rate. The convergence rate of MCR-DBO-NARMA is 7%, 6.4%, 3.3%, and 13.6% correspondingly superior to SSA-NARMA, WSA-NARMA, DOX-NARMA and BIO-NARMA for 1st configuration in 100th iteration. Thus, similarly, the recommended model finally ensures the superior ARM model over traditional methods.

6.5 Statistical evaluation

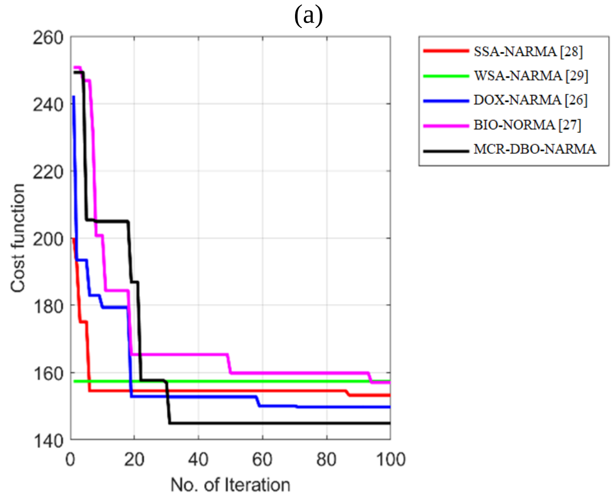
In the designed model, the statistical examination is conducted over six configurations as depicted in Table 2. From the analysis, it is stated that the MCR-DBO-NARMA strategy helps in promoting the efficiency over traditional approaches. As the objective value is execution time assumed as the minimization function, the best of the MCR-DBO-NARMA value is lower while determining with the conventional techniques. Thus, the proposed scheme results in better performance via recommending the implemented approach and thus, it ensures the better ARM efficiency in creating the rules.

Received: November 24, 2022. Revised: December 30, 2022.

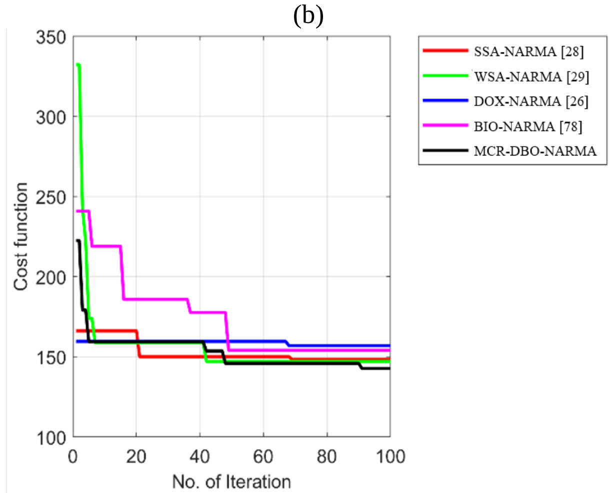
The execution time and memory usage of classical Apriori algorithm grow exponentially if the number of items and transactions in the databases increases. When the number of transactions and distinct items in them is in millions, it is very difficult or highly impossible for Apriori to generate high confidence rules. On other side, Apriori cannot be



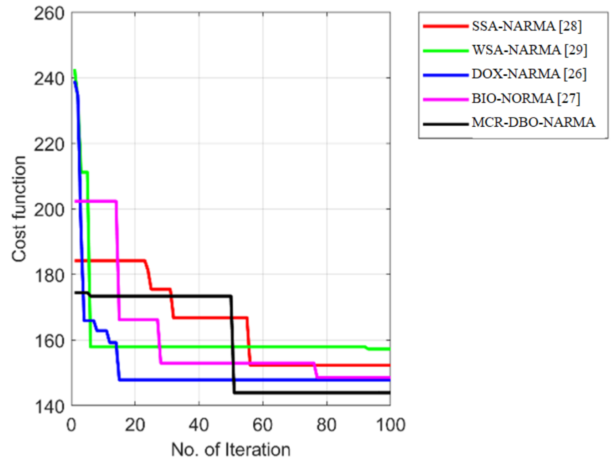
(a)



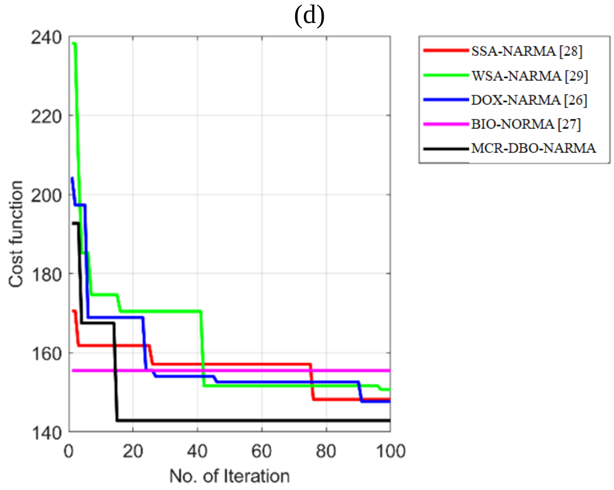
(b)



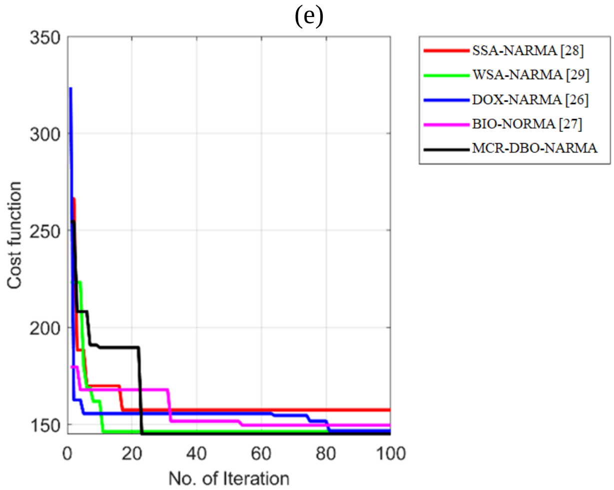
(c)



(d)



(e)



(f)

Figure. 8 Convergence estimation on the ARM model with over traditional algorithms regarding: (a) configuration 1, (b) configuration 2, (c) configuration 3, (d) configuration 4, (e) configuration 5, and (f) configuration 6

applicable to dynamic databases. These issues are addressed with the use of optimization algorithms in association rule mining. Fig. 5 shows that the proposed optimization model outperforms the DOX [26], BIO [27], SSA [28], and WSA [29] optimization models across six different configurations from the WebDocs dataset in terms of confidence, lift, and support. Fig. 8 show that the proposed model convergences more quickly than the other four optimization models. This enhancement has occurred due to the use of hybridization of Dingo Optimizer (DOX) [26] and Billiards-inspired Optimization (BIO) [27] in the proposed model. The DOX effectively protects solutions from local optima. However, it lacks converging ability. BIO has greater convergence capability, higher sensitivity behaviors, efficient exploitation properties, and uses fewer parameters to achieve the desired results. The hybridization of DOX and BIO combines their strengths to converge more quickly and produces rules of high quality. The findings from Fig. 6 show that the proposed model outperforms Apriori [32], Fuzzy-ARM [30], and FP-Growth [31]. Since Apriori and FP-growth purely depends on static data, they may not produce high quality association rules if any new transactions were added after these algorithms established the initial support count of each distinct item. When the minimal threshold values are fixed with a minimum support count of 0.3, the suggested model generates many associations rule when minimum confidence is not provided. Before convergence, the proposed model generates a large number of association rules in every round. However, after completion of each round, only the top 1000 association rules generated from the first to the current iteration are considered and the rest discarded. Finally, the proposed model, when applied to the WebDocs dataset, achieved an average confidence level of 81% for the top 1000 association rules generated.

Table 2. Statistical examination on the recommended association rule mining over various optimization techniques for six different configurations

Measures	**SSA-NARMA [28]**	**WSA-NARMA [29]**	**DOX-NARMA [26]**	**BIO-NARMA [27]**	**MCR-DBO-NARMA**
Transaction as 10000 (configuration 1)					
Best	149.11	151	147.38	167.42	145.09
Worst	195.76	263.02	267.52	167.42	197.73
Mean	160.43	154.85	155.34	167.42	146.65
Median	154.38	151	150.26	167.42	145.09
SD	15.313	12.647	15.064	2.86×10^{-14}	6.1359
Transaction as 20000 (configuration 2)					
Best	153.24	157.35	149.78	157.03	144.85
Worst	200.06	157.35	242.36	250.79	249.24
Mean	155.78	157.35	157.67	170.58	159.86
Median	154.51	157.35	152.75	159.82	144.85
SD	6.8353	1.71×10^{-13}	15.114	22.914	28.159
Transaction as 30000 (configuration 3)					
Best	148.34	147.11	156.92	154.12	142.71
Worst	166.15	332.11	159.6	240.82	222.41
Mean	152.74	157.13	158.71	174.39	153.25
Median	150.09	147.11	159.6	154.12	145.91
SD	6.7819	28.316	1.2647	25.829	12.612
Transaction as 40000 (configuration 4)					
Best	152.35	157.26	147.86	148.53	143.9
Worst	184.2	242.61	239.05	202.31	174.43
Mean	165.05	161.05	151.75	160.52	158.7
Median	166.8	157.9	147.86	152.95	158.65
SD	13.023	14.366	13.854	17.731	14.873
Transaction as 50000 (configuration 5)					
Best	148.19	150.67	147.69	155.5	142.82
Worst	170.51	238.14	204.38	155.5	192.68
Mean	156.24	161.85	157.7	155.5	147.03
Median	157.14	151.61	152.59	155.5	142.82
SD	5.2947	15.909	11.531	3.43×10^{-13}	11.174
Transaction as 60000 (configuration 6)					
Best	157.38	146.22	146.7	149.59	144.93
Worst	266.37	223.3	323.69	179.55	254.46
Mean	161.85	150.56	155.31	156.01	156.83
Median	157.38	146.22	155.53	151.57	144.93
SD	16.309	15.817	17.433	8.927	24.201

7. Conclusion

A novel framework was designed for ARM with the aid of an adaptive mining technique. Initially, the essential data from the standard database was collected from the benchmark datasets. The first step was to develop the NARMA with the incorporation of MCR-DBO. It was helped to identify the distinct itemsets of data. Consequently, the itemsets were categorized into positive association rules as well as negative association rules with the adoption of factors like support, confidence, lift, and correlation coefficient. The efficacy was estimated and evaluated over other techniques. The convergence rate of proposed model is 7%, 6.4%, 3.3%, and 13.6% correspondingly superior to SSA-NARMA, WSA-NARMA, DOX-NARMA and BIO-NARMA respectively. When the minimal threshold values are fixed with a support count of 0.3, the proposed model achieved an average 81% confidence for generated top 1000 association rules from WebDocs dataset. At last, the outcomes have disclosed that the designed approach using MCR-DBO competently mined both the positive and negative rules.

Notation list

Symbol	Description
$\vec{J}_J$	Distance between dingo and prey
$\vec{B}_{\text{prey}}$	Location vector of prey
$\vec{B}(s)$	Location vector of the dingo
$T_{\max}$	Maximum number of iterations
H_α	Fitness of α dingo

$\vec{IT}_{\alpha}$	Intensity of α dingo
$J_{w,v}^{0}$	Initial value of v^{th} variable of w^{th} ball
$\text{rn}_{v}^{\max}$	Maximum value of v^{th} variable of w^{th} ball
H_l	Objective function of l^{th} pocket
$\Re_l$	Probability of selecting l^{th} pocket
ξ	Choice pressure
$\text{SA}_{l,v}^{w}$	The l^{th} pocket of the v^{th} variable to w^{th} common ball pair
PC	Precision rate
g	Acceleration rate
$\overgroup{J_w^{\text{old}} J_w^{\text{new}}}$	Unit movement of w^{th} common ball after collision
$\overrightarrow{J_w^{\text{old}} J_w^{\text{new}}}$	Displacement of the w^{th} common ball
$\vec{v}_v$	Velocity of cue ball
$\text{rn}_{v}^{\max}$	Maximum acceptable constraint of the v^{th} variable
Thr	Threshold value
$\vec{B}$	Scavenger rate in DOX
CT	No. of fitness solutions > mean fitness
$\vec{a}_1$	random parameter lie among [0, 1]
CF	Confidence of an association rule
SU	Support count of item(s)
LF	Lift of an association rule
CC	Correlation coefficient
$E \Rightarrow F$	Association rule
Sut	threshold support value
Lft	Threshold lift value
Cft	Threshold confidence value

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data collection, writing original draft preparation, writing review and editing, visualization, have been done by 1st author. The supervision and project administration have been done by 2nd author.

References

[1] M. Thamer, S. E. Sappagh, and T. E. Shishtawy, "A Semantic Approach for Extracting Medical Association Rules", *International Journal of Intelligent Engineering Systems*, Vol. 13, No. 3, pp. 280-293, 2020, doi: 10.22266/ijies2020.0630.26.

[2] J. Liu, D. Shi, G. Li, Y. Xie, K. Li, B. Liu, and Z. Ru, "Data-driven and association rule miningbased fault diagnosis and action mechanism analysis for building chillers", *Energy and Buildings*, Vol. 216, p. 109957, 2020.

[3] P. D. S. Smart, K. K. Thanammal, and S. S. Sujatha, "A novel linear assorted classification method based association rule mining with spatial data", *Sādhanā*, Vol. 46, No. 7, pp. 1-12, 2021.

[4] Y. Rochd and I. Hafidi, "An Efficient Distributed Frequent Itemset Mining Algorithm Based on Spark for Big Data", *International Journal of Intelligent Engineering Systems*, Vol. 12, No. 4, pp. 367-377, 2019, doi: 10.22266/ijies2019.0831.34.

[5] L. I. Chengyan, S. Feng, and G. Sun, "DCE - miner: an association rule mining algorithm for multimedia based on the MapReduce framework", *Multimedia Tools and Applications*, Vol. 79, pp. 16771–16793, 2020.

[6] L. Qiu, Y. Li, and X. Wu, "Preserving privacy in association rule mining with bloom filters", *Journal of Intelligent Information Systems*, Vol. 29, pp. 253-278, 2007.

[7] Z. Abedjan and F. Naumann, "Improving RDF Data through Association Rule Mining", *Datenbank-Spektrum*, Vol. 13, pp. 111-120, 2013.

[8] H. Zhang, J. Chen, Y. Qiang, J. Zhao, J. Xu, X. Fan, Y. Yang, and X. Zhang, "DART: a visual analytics system for understanding dynamic association rule mining", *The Visual Computer*, Vol. 37, pp. 341-357, 2021.

[9] D. Menaga and S. Saravanan, "GA-PPARM: constraint-based objective function and genetic algorithm for privacy preserved association rule mining", *Evolutionary Intelligence*, Vol. 15, pp. 1487-1498, 2022.

[10] B. N. Keshavamurthy, A. M. Khan, and D. T. Niwal, "Privacy preserving association rule mining over distributed databases using genetic algorithm", *Neural Computing and Applications*, Vol. 22, No. 1, pp. 351-364, 2013.

[11] T. Wang, B. Xiao, and W. Ma, "Student Behavior Data Analysis Based on Association Rule Mining", *International Journal of Computational Intelligence Systems*, Vol. 15, No. 32, pp. 1-9, 2022.

[12] S. Nagaraj and E. Mohanraj, "A novel fuzzy association rule for efficient data mining of ubiquitous real-time data", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 11, No. 11, pp. 4753–4763, 2020.

[13] A. Borah and B. Nath, "Rare association rule mining from incremental databases", *Pattern Analysis and Applications*, Vol. 23, No. 1, pp. 113-134, 2020.

[14] G. Kalyani, M. V. P. C. S. Rao, and B. J. Kiramaiah, "Particle Swarm Intelligence and Impact Factor-Based Privacy Preserving Association Rule Mining for Balancing Data Utility and Knowledge Privacy", *Arabian Journal for Science and Engineering*, Vol. 43, No. 8, pp. 4161-4178, 2018.

[15] A. Das, M. M. Ahmed, and A. Ghasemzadeh, "Using trajectory-level SHRP2 naturalistic driving data for investigating driver lane-keeping ability in fog: An association rules mining approach", *Accident Analysis & Prevention*, Vol. 129, pp. 250-262, 2019.

[16] N. A. Funde, M. M. Dhabu, A. Paramasivam, and P. S. Deshpande, "Motif-based association rule mining and clustering technique for determining energy usage patterns for smart meter data", *Sustainable Cities and Society*, Vol. 46, p. 101415, 2019.

[17] S. Çokpınar and T. İ. Gündem, "Positive and negative association rule mining on XML data streams in database as a service concept", *Expert Systems with Applications*, Vol. 39, Issue. 8, pp. 7503-7511, 2012.

[18] J. Chen, P. Li, W. Fang, N. Zhou, Y. Yin, H. Xua, and H. Zheng, "Fuzzy Association Rules Mining based on Type-2 Fuzzy Sets over Data Stream", *Procedia Computer Science*, Vol. 199, pp. 456-462, 2022.

[19] P. Koukaras, C. Tjortjis, and D. Rousidis, "Mining association rules from COVID-19 related twitter data to discover word patterns, topics and inferences", *Information Systems*, Vol. 109, p. 102054, 2022.

[20] F. Antonello, P. Baraldi, A. Shokry, E. Zio, U. Gentile, and L. Serio, "A novel association rule mining method for the identification of rare functional dependencies in Complex Technical Infrastructures from alarm data", *Expert Systems with Applications*, Vol. 170, p. 114560, 2021.

[21] C. Wang and X. Zheng, "Application of improved time series Apriori algorithm by frequent itemsets in association rule data mining based on temporal constraint", *Evolutionary Intelligence*, Vol. 13, pp. 39-49, 2020.

[22] X. Dong, F. Hao, L. Zhao, and T. Xu, "An efficient method for pruning redundant negative and positive association rules", *Neurocomputing*, Vol. 393, pp. 245-258, 2020.

[23] S. Bagui and P. C. Dhar, "Positive and negative association rule mining in Hadoop's MapReduce environment", *Journal of Big Data*, Vol. 6, No. 75, pp. 1-16, 2019.

[24] T. Su, H. Xu, and X. Zhou, "Particle Swarm Optimization-Based Association Rule Mining in Big Data Environment", *IEEE Access*, Vol. 7, pp. 161008-161016, 2019.

[25] H. Zheng, P. Li, and J. He, "A Novel Association Rule Mining Method for Streaming Temporal Data", *Annals of Data Science*, Vol. 9, No.4, pp. 863-883, 2022.

[26] A. K. Bairwa, S. Joshi, and D. Singh, "Dingo Optimizer: A Nature-Inspired Metaheuristic Approach for Engineering Problems", *Mathematical Problems in Engineering*, Vol. 2021, 2021.

[27] A. Kaveh, M. Khanzadi, and M. R. Moghaddam, "Billiards-inspired optimization algorithm; a new meta-heuristic method", *Structures*, Vol. 27, pp. 1722-1739, October 2020.

[28] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems", *Advances in Engineering Software*, Vol.114, pp. 163-191, Dcember 2017.

[29] A. Kaveh and A. D. Eslamlou, "Water strider algorithm: A new metaheuristic and applications", *Structures*, Vol. 25, pp. 520-541, June 2020.

[30] O. Dogan, F. C. Kem, and B. Oztaysi, "Fuzzy association rule mining approach to identify e-commerce product association considering sales amount", *Complex & Intelligent Systems*, Vol. 8, pp. 1551–1560, 2022.

[31] Y. Zeng, S. Yin, J. Liu, and M. Zhang, "Research of Improved FP-Growth Algorithm in Association Rules Mining", *Scientific Prograing*, Vol. 2015, pp. 1-6, 2015.

[32] M. A. Mottalib, K. S. Arefin, M. M. Islam, M. A. Rahman, and S. A. Abeer, "Performance Analysis of Distributed Association Rule Mining with Apriori Algorithm", *International Journal of Computer Theory and Engineering*, Vol. 3, No. 4, p. 484, August 2011.

[33] P. Kishor and S. Porika, "An efficient approach for mining positive and negative association rules from large transactional databases", *In: Proc. of International Conference on Inventive Computation Technologies (ICICT)*, Coimbtore, India, pp. 1-5, 2016.

[34] K. Indira and S. Kanmani, "Association Rule Mining Using Genetic Algorithm: The role of Estimation Parameters", *In: Proc. of International Conference on Advances in*

Computing and Communications, Vol. 190, pp 639–648, 2011.

[35] W. C. Rojas, C. Vargas, and C. M. Villegas, "Interactive Visualization of Association Rules Model Using SOM", *In: Proc. of the XV International Conference on Human Computer Interaction*, Tenerife, Spain, pp. 1–4, September 2014.