Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Fuzzy-Assisted Mathematically Modified Cat Swarm Optimization Approach for Effective Sensitive Association Rule Hiding in Data Mining

Dr M Praneesh[1], R. Naveenkumar[2], Capt. Dr. M. Nalini[3], Dr. C. Karnan[4], Dr. R. Brindha[5], Shanthi R[6], Dr .P.Dharmendra Kumar[7]

[1]Assistant Professor, Department of Computer Science with Data Analytics, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamilnadu, India, Email: raja.praneesh@gmail.com Orcidid : 0000-0003-3691-1343.
[2]Dept of CSE, School of Engineering and Technology, CGC University Mohali-140307, Punjab India, Email: drnk1983@gmail.com, 0000-0001-9033-9400
[3]Principal & Associate Professor of Mathematics, J.K.K Nataraja College Of Arts & Science ,Kumarapalayam (TK), Namakkal (DT). Tamil Nadu. 638183, Email: naliniphd77@gmail.comhttps://orcid.org/0009-0000-9473-1549
[4]Assistant Professor, Department of Mathematics,K Ramakrishnan College of Engineering (Autonomous),Samayapuram, Trichy-621112, Tamilnadu, India, Email: karnankathir@gmail.com, Orchid: https://orcid.org/my-orcid?orcid=0009-0003-4097-4352
[5]Associate Professor, Department of Mathematics,Velalar College of Engineering and Technology, Erode 638 012, Email: brindhaaramasamy@gmail.com Orcid ID: 0009-0003-7748-8388
[6]Assistant Professor & HOD, Department of Mathematics, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, tamilnadu, India. Email: shanthir@maher.ac.in
[7]Assistant professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India, Email: pdharmendrakumar@kluniversity.in, dharmendra.phd.au@gmail.com, https://orcid.org/0009-0006-5087-9123

Abstract

Protecting and storing the confidential data imposes critical distress of privacy. Preserving the confidential data privacy is attained using Privacy Preserving Data Mining (PPDM). One of the significant problems in Privacy Preservation Data Mining is Association Rule Hiding (ARH) and it is utilised in hiding sensitive association rules. Every ARH algorithms intended to alter the original database such that no confidential association rules are derived from the transactional database. ARH approaches are extensively used in data mining to spot the association among the itemset. Most of the business organizations reveal certain information to the third party for the common benefit of identifying the needed knowledge for promoting the business schemes and decision making. Database may possess the private information where a business organization does not want to share that information to the third party. The problem of privacy plays a significant role when varied organizations share the data for the benefit by compromising the privacy of the individual person. Before revealing the information, confidential data in the database must be masked by PPDM approach, which is helpful in advance to security of the database. The proposed algorithmUnified Transaction Dimensionality Reduction Framework (UTDRF) uses Fuzzy Cat Swarm Optimization algorithm (FCSO) to hide the sensitive items in the transactional database. The proposed method is compared with the existing state-of-art algorithms. From the test it has been justified that the FCSO algorithm produced better results than existing algorithms.

Keywords: Sensitive rule, PPDM, bio-inspired, fuzzy, hiding, privacy, swarm intelligence, and security.

1. Introduction

Data mining has engrossed a grand deal of consideration in recent years. Data mining is employed to extort the potential information and unknown patterns from the massive data that is obtained from the repositories and data warehouses. It is an indispensable scheme for a massive data where intelligent approaches are implemented to uncover the potential data. The chief intent of data mining is to retrieve the needed knowledge from a data set and the knowledge is transformed into an intelligible and recognizable structure for later use

[1]. Apart from the raw data scrutinization, data mining incorporates the features of database management, pre-processing of data, data model consideration, data inference examination, metrics of interestingness, processing the identified pattern, visualization and renewal data. Business firms and industries use the data mining approaches such as neural networks, rule mining, usage mining and fuzzy logics to discover and perceive the necessary patterns [2].

The process of identification of significant correlation among the items in a massive database is called as Association Rule Mining (ARM). By the consideration of interesting measures robust rules are identified from the database which is attained in ARM. Various application areas of ARH are Market Basket Illustration, Census Data, Customer Relationship Management and Medical Diagnosis [3].The main intent of ARM is spotting the regularities among the products in the database. Some of the association rules generating algorithms are Multidimensional association rule, Multilevel association rule and Quantitative association rule.The devices based on pervasive computing have increased data generation. The generated data is analysed for the growth of the business and many of the business firms consider data as the biggest asset [4]. Protecting and storing the confidential data imposes critical distress of privacy. Preserving the confidential data privacy is attained using Privacy Preserving Data Mining (PPDM). Some of the PPDM methods are Randomization method, Distributed privacy preservation, Downgrading application effectiveness, K-anonymity and l-diversity method [5].

Every business organization maintains the electronic data and generation of data is due to social networks as well as the business organizations. The maintenance and analysation of data is accomplished by numerous DM techniques. Data is considered as the biggest asset of every company. Confidentiality of the data is subjected to threat and the data need high privacy concern. Several PPDM techniques are projected for altering and changing the data to conserve privacy. The chief intent of PPDM is to assure the data protection without compromising the utility or nature of the data [6].

Advancement in the storage abilities of technology and the excess generation of data created from the social organization led the development of various data mining algorithms. Excessive collection and illustration of data is attained with various forms of DM algorithms based on the necessity [7]. ARM is one of the forms of data mining algorithm that retrieves the hidden patterns from the database. Retrieved patterns from the database may impose sensitive personal information and privacy of the data is not ensured with the traditional DM algorithms [8]. PPDM is developed to protect the confidential information of the user [9].

PPDM incorporates two chief factors that are used to ensure the guarantee of the data during processing of the application and the confidentiality of the data [10]. Confidential data about the individual is secured against the adverse receiving situations. Advancement in the data mining application is achieved with the PPDM approaches. Information with sensitive user data is masked with PPDM approaches which ensure the data confidentiality [11]. The significant purpose of PPDM is to reframe the original data in some way and establishing the relevant data mining applications. The key consideration of PPDM is anonymity of the data and data mining algorithm development [12].

One of the upcoming technologies in privacy-based data mining is ARH. The foremost feature of ARH is to conceal the sensitive association rules from the database. ARH secures the private data of the individual from exposing the data to the unauthorized persons. Privacy concern in data mining is amplified due to varied factors and the de-identification of the data assures the privacy of the data [13]. Certainly, illustration of data via intellectual scheme may mask the sensitive data from the unauthorized entities that is achieved by various ARH algorithms.ARH is attained by incorporating the external and visibly accessible sources of data in association with the publicly available knowledge to re-classify the hidden patterns of information [14]. ARH in data mining is primarily employed in hiding the sensitive data from exposing the data to the unauthorized users.

ARM approaches are extensively used in data mining to spot the association among the itemset. Most of the business organizations reveal certain information to the third party for the common benefit to identify the needed knowledge for promoting the business schemes and decision making. Database may possess the private information which the business organization does not want to share that information to the third party. The problem of privacy plays a significant role when varied organizations share the data for the benefit by

concealing the confidentiality of the individual person. Before revealing the information, confidential data in the database must be masked and to resolve the problem, PPDM approach is helpful to advance the security of the database.

The process of data illustration techniques facilitates organizations and governments to publish data containing particular information about individuals or organizations. The unconfined data sets not only provide important content to the beneficiaries, but also hold susceptible information about individuals whose privacy may be at risk. The disclosure risks of sensitive data also increases with the progress in data mining technology. The intent of database protection exploration society and the administration statistical organizations is to provide security to confidential data against unauthorized access. The sharing of knowledge for data processing brings lots of benefit for illustration and business collaboration.

The misuse of those techniques could cause the revealing of sensitive information. However, giant repositories of information that contain non-public data and sensitive rules need to be protected before being revealed. Motivated by the multiple conflicting necessities of knowledge sharing, privacy conserving data processing has become very popular in data processing and information security fields. This novel method can be applied directly to the data storage with other privacy preserving methods for better security.The primary objective of ARH is to identify the sensitive rules and modify the occurrences of confidential items to hide the confidential rules. In this phase two approaches are used, they are (i) heuristic approach and fuzzy approach. Fuzzy Cat Swarm Optimization algorithm (FCSO) is used to hide the sensitive items in the transactional database.

2. Related Works

Privacy Preserving Data Mining (PPDM) has emerged as a critical paradigm to mitigate the risk of sensitive information disclosure during knowledge extraction processes. In the context of Association Rule Hiding (ARH), several optimization-driven and heuristic approaches have been proposed to balance privacy and data utility. Fuzzy-based ARH techniques enhance uncertainty handling and enable effective concealment of sensitive rules by leveraging membership functions and linguistic variables, thereby improving robustness in medical data applications (Krishnamoorthy & Murugesan, 2019). Metaheuristic algorithms such as Least Lion Optimization Algorithm (LLOA) introduce key-based sanitization mechanisms to minimize side effects while preserving non-sensitive patterns (Menaga & Revathi, 2018). Alternative strategies, including dummy item insertion, aim to distort transactional patterns without significantly altering database semantics (Kanekar& Dhanaraj, n.d.). Comprehensive analytical reviews emphasize that distortion-based sanitization outperforms blocking approaches in maintaining data utility (Telikani&Shahbahrami, 2018). Furthermore, PPDM frameworks incorporate multiple techniques such as anonymization, perturbation, and rule hiding to ensure confidentiality (Aldeen et al., 2015). Bio-inspired algorithms like cuckoo optimization demonstrate improved convergence in hiding sensitive rules with minimal information loss (Afshari et al., 2016). Additionally, swarm intelligence techniques such as Cat Swarm Optimization provide an efficient search mechanism for optimizing rule hiding processes (Chu et al., 2006).

3. Proposed Methodology

The exponential growth of high-dimensional transactional data in social networks and business systems necessitates intelligent and automated preprocessing mechanisms for efficient data mining. A critical challenge lies in handling noise and the curse of dimensionality, which degrade computational performance and pattern discovery accuracy. To address this, dimensionality reduction techniques are employed.

The Unified Transaction Dimensionality Reduction Framework (UTDRF) framework is introduced to eliminate redundant items and superfluous transactions by enforcing structural constraints on transactional data. Specifically, a minimum threshold (\tau = 10) is defined such that each transaction must contain exactly 10 items. Transactions exceeding this threshold are partitioned, retaining only the first 10 items while propagating the remaining items to subsequent transactions. Conversely, transactions with fewer than 10 items are merged with adjacent transactions until the threshold condition is satisfied.

This normalization mechanism ensures uniform transaction length, thereby reducing dimensional sparsity, improving computational tractability, and enhancing the effectiveness of downstream mining algorithms.

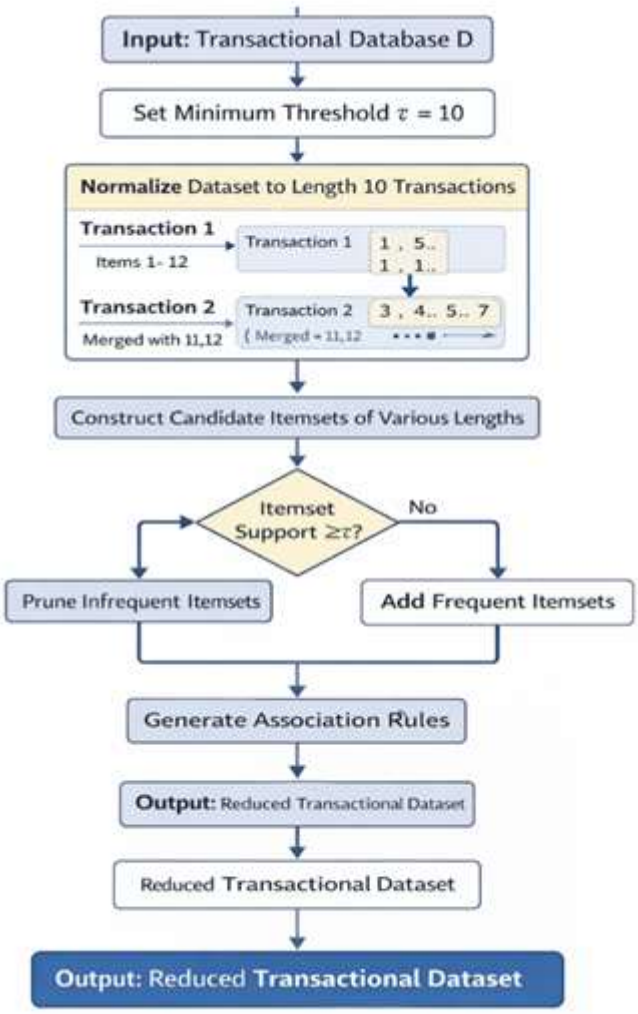

Figure 1: Dimensionality Reduction

Based on Individual Threshold concept, individual threshold value is given to every item in a transactional database. After assigning individual threshold value, a table is constructed with items, transaction, and the number of times each item occurred. In the first-item-set candidate generation, all the items will be evaluated based on the individual threshold value and the number of occurrences. If the occurrence is equal to or greater than the individual threshold value, the items are selected for the next candidate generation. In the next candidate generation, two item sets will be compared with the number of occurrence;from this minimum threshold values are selected. If the occurrence is equal to or higher than the minimum threshold value, the itemsets are selected for frequent itemset. This process is repeated until all candidate itemsets are generated. Frequent itemset generated association rules are identified by applying support and confidence values.

The fuzzy logic concept is based on computational model, where the information are manipulated and represented in mathematical way, moreover it resembles the human communication and reasoning processes. The Fuzzy Logic (FL) theory is intended to handle imprecision and improbability. In addition, fuzzy systems depend on fuzzy Logic and set theory which provide a meaningful and rich addition to traditional logic. According to FL, linguistic variables enhance the degree of knowledge representation. Conventional approach of knowledge representation uses bivalent logic which has major shortcomings like managing uncertainty and imprecision.

In ARM, support and confidence are the two most important measures used for evaluating the association rules.Support denotes how often the items displayed in the database, whereas Confidence quantifies the frequency with which an association rule $X \rightarrow Y$holds true, i.e., the proportion of transactions containing Xthat also contain Y. Association rules are considered valid only if they satisfy user-defined minimum thresholds for support and confidence. Here, $X$and Ydenote itemsets, and Trepresents the total number of transactions.

$$Support = \sigma \frac{(X,Y)}{|T|} \quad (1)$$

$$Confidence = \frac{\sigma(X,Y)}{\sigma(X)} \quad (2)$$

Fuzzy ARM is first originated in the form of knowledge discovery in fuzzy expert systems. Here in fuzzy expert system rather than using Boolean logic, fuzzy membership functions and rules are used to extract the information. In fuzzy expert system the rules are usually in the related form as follows: "If humidity is high then fan speed is high" Here if part is called as antecedent and then part is called as consequent.

The proposed algorithm uses Fuzzy Cat Swarm Optimization algorithm (FCSO) to hide the sensitive items in the reduced transactional database. The Cat Swarm Optimization algorithm (CSO) was introduced in [28], which simulates the common behavior of cats. The Cat Swarm Optimization CSO algorithm imitates the natural behavior of cats. The proposed method consists of five steps (i) Seeking Mode (ii) Seeking Range (iii) Count Dimension (iv) Self-Position and (v) Movement tracking. Initially the seeking mode and seeking range steps identify the sensitive items with respective ranges. Moreover in the count dimension step, number of modifications is calculated to modify the occurrences of sensitive items to change sensitive items as non-sensitive items. Finally sensitive items are modified until it reaches the goal and updated using self-positioning and movement tracking steps

Seeking Mode

In ARH it identifies the sensitive items by using user defined membership values. From the given formula it identifies the number of sensitive items from each transaction. Where F_{SI} represents the number of sensitive items in the transaction.I denotes the items present in the transaction, VT representsmembership value (0.75-1).

$F_{SI} = I \geq VT$ (3)

Seeking Range

Seeking range declares the mutative ratio for the selected dimensions. In seeking mode, if a dimension defined by Seeking range such as $F=x_i$ {VT, ET}with two membership values. Where μ represents the occurrences of items and its specified range, x_i denotes the support count of the items.

$$\mu(I_I) = \int_{0.50}^{1} (xi) \quad (4)$$

$$x_i = (VT, ET) (5)$$

Count Dimension/ Self-Position

In count Dimension step it identifies the number of modification to be done using below given formula. Select the transaction which has more number of sensitive items and reduce the support count of the item until all the modifications are done for all the fuzzy sensitive items. Where M_{SI} represents the number of modification, $O(F_{SI})$ denotes occurrence of fuzzy sensitive items.$IT(F_{SI})$ represents the individual support of fuzzy sensitive items.

$M_{SI} = [O(F_{SI}) - IT(F_{SI})-1)$ (6)

Movement Tracking

Movement tracking step used to check whether all the item are modified and the occurrences of the items are non-sensitive with specified range if it is satisfied update the position and modify the original database.

4. Result and Discussion

The performance metrics used for illustration are, Hiding Failure, Missing Cost, Execution Time and Memory Space. Finally, comparison is done between two approaches to identify the best approach. From the results, it is known that Fuzzy Approach has given better results.

4.1 Performances Measures

The hiding quality can be analyzed by computing the performances on the database. Therefore, estimation metrics for computing these performances are demonstrated below:

- Hiding failure (HF): denotes the failed rules from hiding and that can retrieved by the sanitization by:

$$HF=\frac{R_S(D')}{R_S(D)} \qquad (7)$$

- $|Rs(D)'|$ represents the number of confidential rules obtained from the sanitized database D', while $|Rs(D)|$ denotes the number of sensitive rules derived from the original database D. Table 1 presents a comparative illustration of the proposed and existing algorithms.

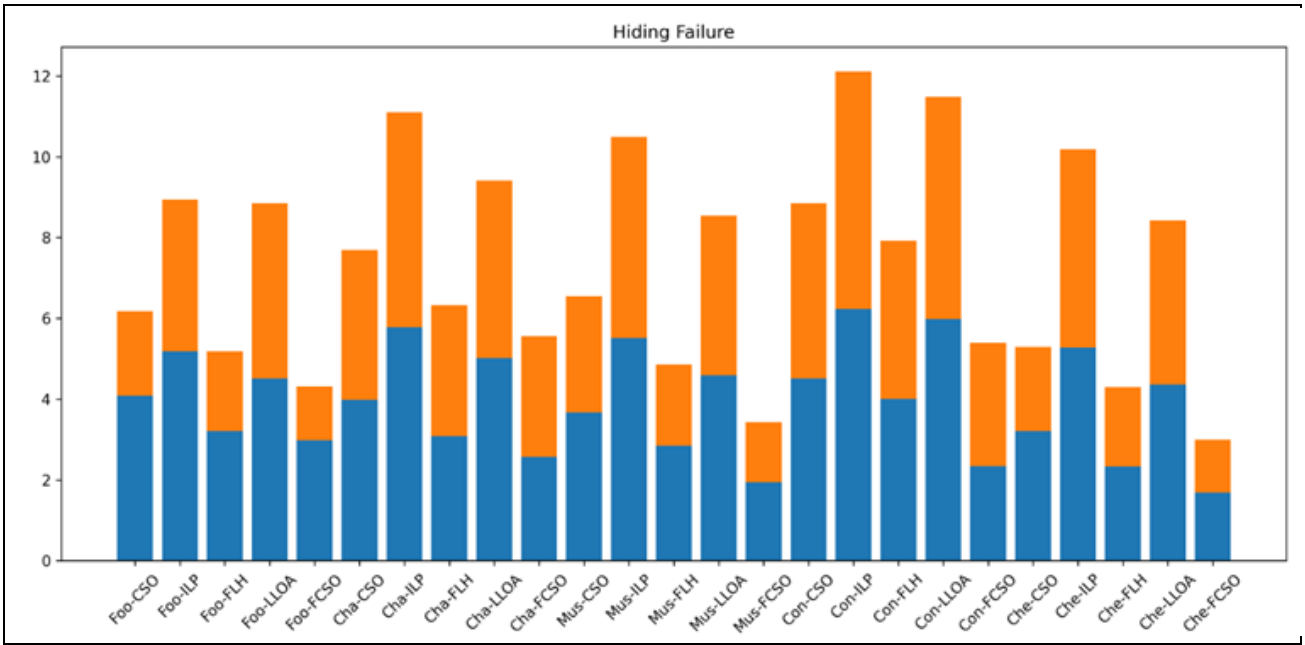

Figure 2: Performance of Hiding Failure

Figure 2, illustrate the illustration of hiding failure for all the datasets. The proposed algorithm is compared with the existing algorithm. The proposed algorithm FCSO gives better results than the existing algorithms.

Missing cost (MS):

Missing Cost denotes the items that don't pose any sensitive values that are concealed by the sanitization process and cannot be retrieved from the sanitization database D'. The value of LR is calculated by the following formula:

$$MC=\frac{|\sim R_S(D)|-\sim R_S(D')||}{|R_S(D)|} \qquad (8)$$

In the equation, | ~ Rs (D) | signifies the count of the items that doesn't poses any sensitive values that are retrieved from the determined database D and | ~ Rs (D') | signifies the count of the items that doesn't poses any sensitive values that are retrieved from the sanitization database D

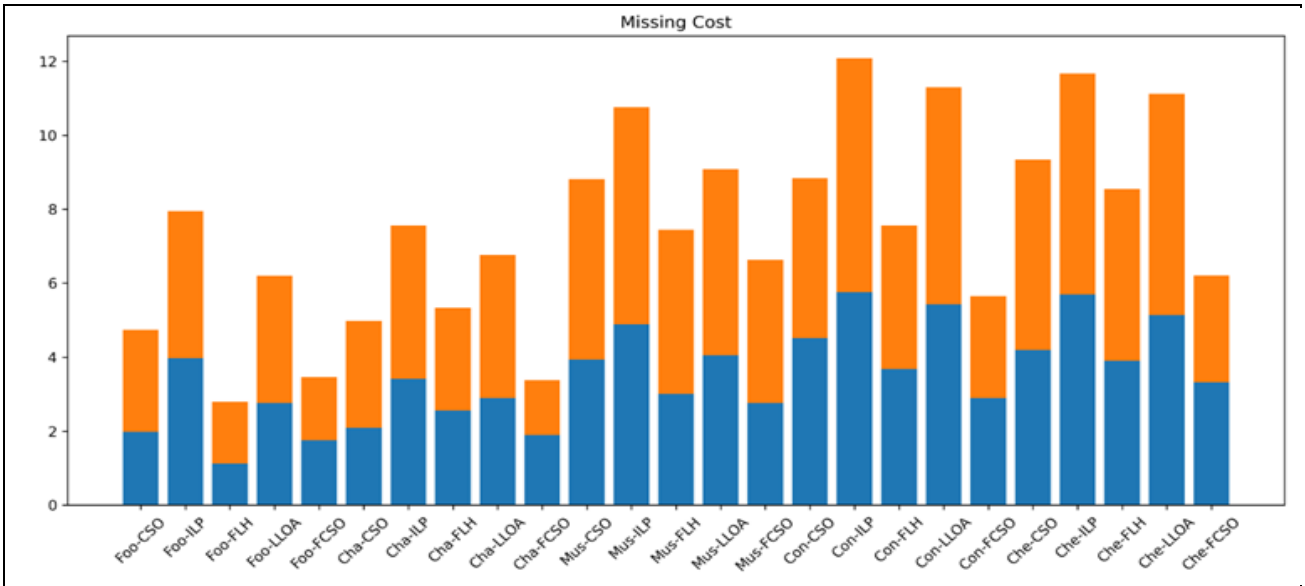

Figure 3: Performance of Missing Cost

Figure 3 determine the illustration of missing cost for all the datasets. The proposed algorithm is compared with the existing algorithms. From the illustration the proposed algorithm FCSO gives the best results than the existing algorithm

➢ Runtime: The time spent to attain the optimized. The results are evaluated with the proposed and existing algorithms. Execution time is calculated in milliseconds

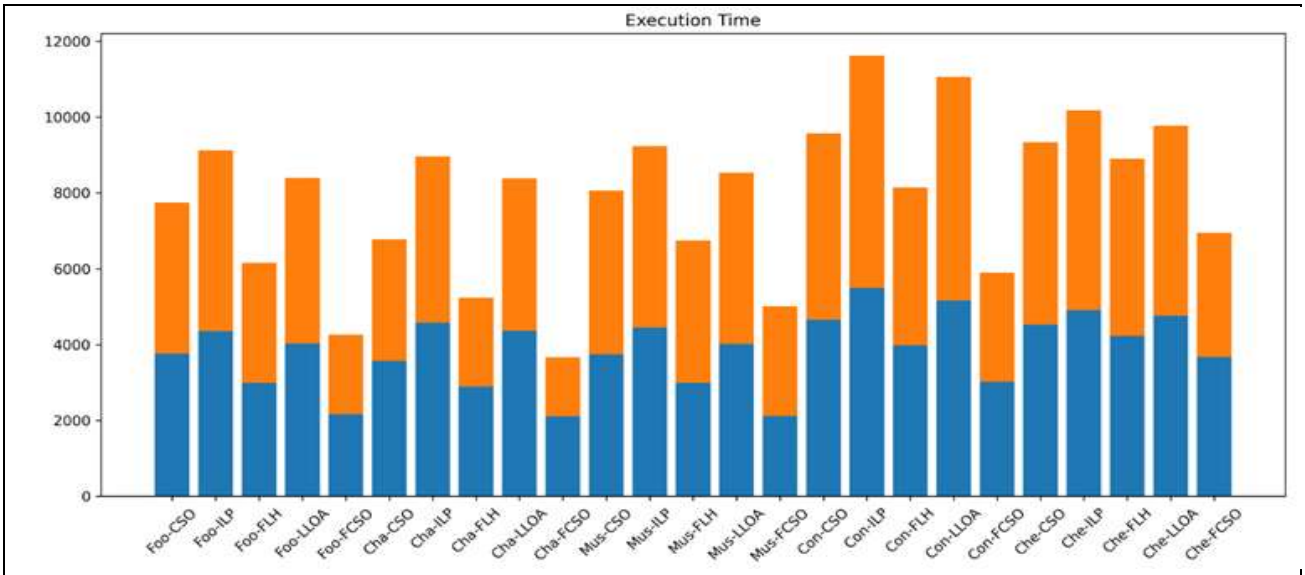

Figure 4: Performance of Execution Time

Figure 4 gives the execution time performance of proposed and existing algorithms for all the datasets. From the analyses it has been observed that the proposed FCSO algorithm gives better results than the existing algorithms.

➢ Memory: The usage of the storage space occupied by the algorithms. Table 4 gives the utilization of memory space for the proposed and existing algorithms.

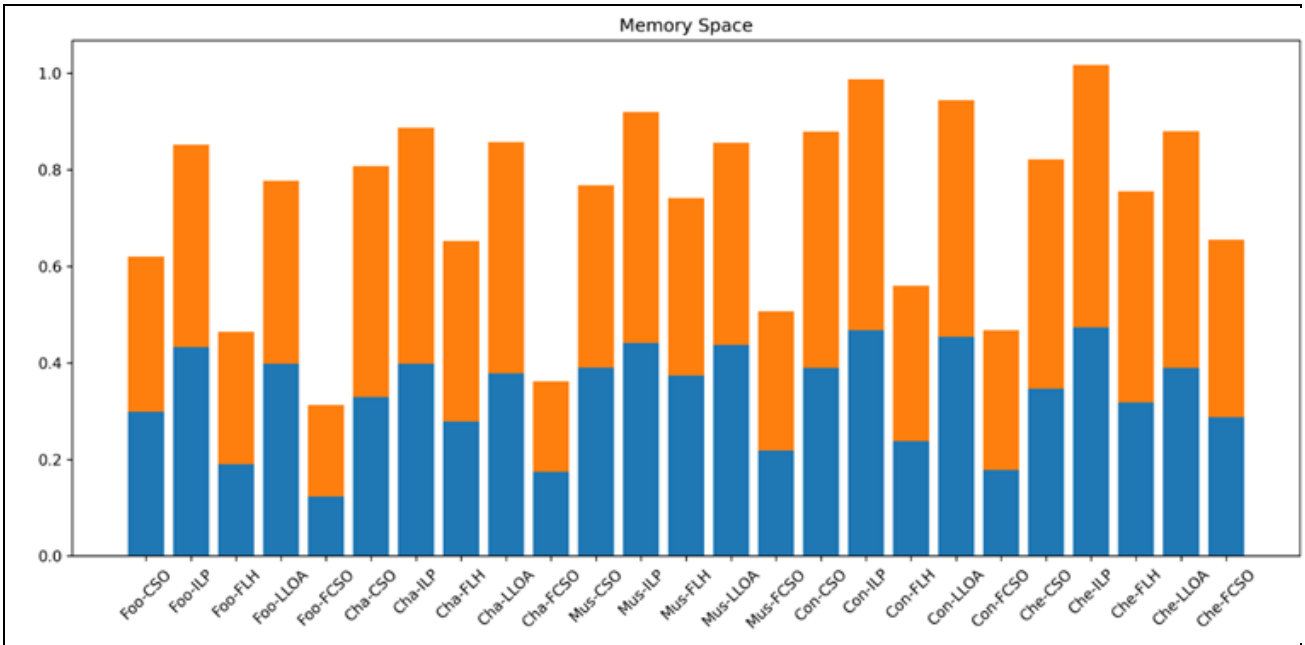

Figure 5: Illustration of Memory Space

Figure 5 give the performances of the proposed and existing algorithms for all the datasets. The proposed algorithm FCSO gives better results than the existing algorithms.

5. Conclusion

ARH approaches are extensively used in data mining to spot the association among the itemset. Most of the business organizations reveal certain information to the third party for the common benefit of identifying the needed knowledge for promoting the business schemes and decision making. Database may possess the private information where a business organization does not want to share that information to the third party. The problem of privacy plays a significant role when varied organizations share the data for the benefit by compromising the privacy of the individual person. Before revealing the information, confidential data in the database must be masked by PPDM approach, which is helpful in advance to security of the database. In Fuzzy approach the proposed algorithm uses Fuzzy Cat Swarm Optimization algorithm (FCSO) to hide the susceptible items in the reduced transactional database. From the illustration states that the proposed algorithm has produced improved results than existing algorithms.

Reference

1. Gunawan, D. (2020). Classification of privacy preserving data mining algorithms: a review. *JurnalElektronika dan Telekomunikasi*, *20*(2), 36-46.
2. Afshari, M. H., Dehkordi, M. N., & Akbari, M. (2016). Association rule hiding using cuckoo optimization algorithm. *Expert Systems with Applications*, *64*, 340-351.
3. Suma, B., & Shobha, G. (2021). Association rule hiding using integer linear programming. *International Journal of Electrical and Computer Engineering (IJECE)*, *11*(4), 3451-3458.
4. Krishnamoorthy, S., & Murugesan, K. (2019). Protecting the privacy of cancer patients using fuzzy association rule hiding. *Asian Pacific Journal of Cancer Prevention: APJCP*, *20*(5), 1437.
5. Menaga, D., & Revathi, S. (2018). Least lion optimisation algorithm (LLOA) based secret key generation for privacy preserving association rule hiding. *IET Information Security*, *12*(4), 332-340.
6. Kanekar, R. P., & Dhanaraj, R. Adding Dummy Items To Hide Sensitive Association Rules.
7. Telikani, A., &Shahbahrami, A. (2018). Data sanitization in association rule mining: An analytical review. *Expert Systems with Applications*, *96*, 406-426.

8. Aldeen, Y. A. A. S., Salleh, M., & Razzaque, M. A. (2015). A comprehensive review on privacy preserving data mining. *SpringerPlus*, *4*(1), 1-36.
9. Afshari, M. H., Dehkordi, M. N., & Akbari, M. (2016). Association rule hiding using cuckoo optimization algorithm. *Expert Systems with Applications*, *64*, 340-351.
10. Chu, S. C., Tsai, P. W., & Pan, J. S. (2006, August). Cat swarm optimization. In *Pacific Rim international conference on artificial intelligence* (pp. 854-858). Springer, Berlin, Heidelberg.
11. A.Velliangiri. (2026). Design and Experimental Validation of a Bidirectional Converter for Battery–Grid Interface in Renewable Systems. *Transactions on Power Electronics and Renewable Energy Systems*, 8-18.
12. Vishnu Vardhan Reddy Kavuluri. (2025). Software Dependency Analysis Using Graph Learning for Large Codebases. *Journal of Wireless Intelligence and Spectrum Engineering*, 28–32.
13. F Rahman. (2025). An Adaptive Computational Acoustics Model for Noise Control and Sound Radiation Analysis in Engineering Structures. *Advanced Computational Acoustics Engineering*, *3*(2), 31-36.
14. Rajan.C, &N .Saranya. (2025). Nonlinear Dynamical Modeling and Control Strategies for Vibration Mitigation in Smart Structural Systems. *Journal of Applied Mathematical Models in Engineering*, *1*(2), 32–39.
15. Petra Novak, & Marko Jurić. (2025). AR in Tourism Creates Authentic Guest Experiences: Real Cases from Top Hotels. *Journal of Tourism, Culture, and Management Studies*, *2*(1), 38–46.