INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM

A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.,

Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)],

Index Copernicus Publishers Panel, Poland with IC Value of 5.09 & number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 2840 Cities in 164 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

http://ijrcm.org.in/

CONTENTS

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography: Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript** ***anytime*** in ***M.S. Word format*** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. infoijrcm@gmail.com or online by clicking the link **online submission** as given on our website (***FOR ONLINE SUBMISSION, CLICK HERE***).

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____________

THE EDITOR
IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF** .

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '____' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:
a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mentionthe following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)
c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
d) The total size of the file containing the manuscript is required to be below **500 KB**.
e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE**: The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS**: The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT**: Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS**: Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.

6. **MANUSCRIPT**: Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS**: All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS**: All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT**: The main text should follow the following sequence:

   INTRODUCTION

   REVIEW OF LITERATURE

   NEED/IMPORTANCE OF THE STUDY

   STATEMENT OF THE PROBLEM

   OBJECTIVES

   HYPOTHESES

   RESEARCH METHODOLOGY

   RESULTS & DISCUSSION

   FINDINGS

   RECOMMENDATIONS/SUGGESTIONS

   CONCLUSIONS

   SCOPE FOR FURTHER RESEARCH

   ACKNOWLEDGMENTS

   REFERENCES

   APPENDIX/ANNEXURE

   It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES &TABLES**: These should be simple, crystal clear, centered, separately numbered &self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.

11. **EQUATIONS**:These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **REFERENCES**: The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.
- **ONLINE RESOURCES**
- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

STUDY ON IMPLEMENTING ASSOCIATION RULE MINING IN PARTICLE SWARM OPTIMIZATION

T. BHARATHI
RESEARCH SCHOLAR
RVS ARTS AND SCIENCE COLLEGE
COIMBATORE

DR. P. KRISHNAKUMARI
DIRECTOR OF MCA
RVS ARTS AND SCIENCE COLLEGE
COIMBATORE

ABSTRACT

There is a great need to discover association rules which are an important database mining problem. In the area of association rule mining is focused on improving computational efficiency in previous research. Determination of the threshold value of support and confidence affect the quality of association rule mining up to great extent. This paper gives attention to effectiveness of algorithm for association rule mining in order to improve computational efficiency as well as to automatically determine suitable threshold values. The Particle Swarm Optimization algorithm first search for the optimum fitness value of each particle and finds corresponding support and confidence as minimal threshold values after the data are transformed into binary values. In this paper the particle swarm optimization algorithm compared with a genetic algorithm.

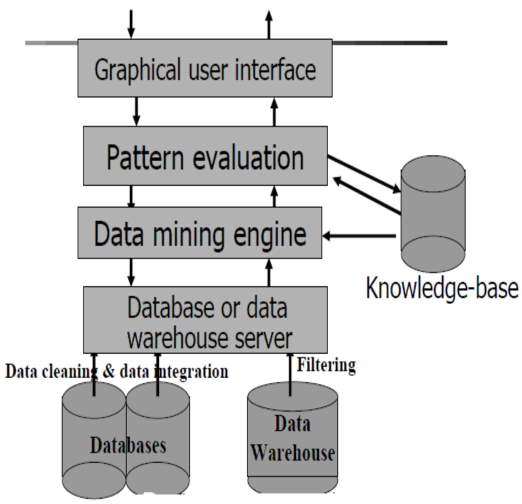
KEYWORDS

Association Rule Mining, Data Mining, Particle Swarm Optimization, Genetic Algorithm.

I. INTRODUCTION

Data Mining is process which finds useful patterns from large amount of data analysis is that the data mining is mine information and discover knowledge on the premise of no clear assumption.[1]. Generally data mining process is composed by preparation of data, data cleaning, data reduction and transformation, data mining, pattern evaluation and Knowledge presentation. Fig[1] shown the process of data mining.

FIGURE 1: PROCESS OF DATA MINING



Association rule mining is one of model among data mining several models. Association rule mining is used to searches for relationship between variables in database. Apriori algorithm is one of famous and oldest algorithm for discover the association rules. It consists of many modified algorithms to improving computational efficiency. The Support and Confidence is effect the quality of result and also it's used to decision making purpose. But thus previous algorithm doesn't consider that main two parameters. So this study mainly focused to improving algorithm can find minimum threshold value for support and confidence.

II. LITERATURE REVIEW

2.1 ASSOCIATION RULE MINING:

Association Rule analysis is useful for discovering interesting relationship hidden in large data sets. The uncovered relationship can be represented in the form of association rules. Agarwal et al was introducing association rules in 1993.Retailers are interested in analyzing the data to learn about the purchasing behavior of their customers. Such valuable information can be used to support a variety of business-related application such as marketing promotions, inventory management, and customer relationship management. An association rule is an implication expression of the form X Y, where X and Y are disjoint item sets. The strength of an association rule can be measured in terms of its Support and Confidence. Support determines how often a rule is applicable to a given data set, while confidence determines how frequently items in Y appear in transaction that contains X.

$$\text{Support, } s(X \rightarrow Y) = \frac{\partial\,(X \cup Y)}{N}$$

$$\text{Confidence, } c(X \rightarrow Y) = \frac{\partial\,(X \cup Y)}{N}$$

2.2 ASSOCIATION RULE ALGORITHM

2.2.1. Apriori Algorithms: The classical apriori algorithm as suggested by Agrawal et al in [1993] is one of the most important data mining algorithms. It is an influential algorithm for mining frequent item sets for Boolean association rules. Frequent Item sets: The sets of item which has minimum support. Apriori property: Any subset of frequent item set must be frequent. Joint operation: To find L_k, a set of candidate K-Item sets is generated by joining L_{k-1} with itself .

2.2.2. FP-Growth Algorithm: J.Han et al [2000], it's for complete set of frequent patterns, by pattern fragment growth. Efficiency of mining is achieved with three techniques: 1. A large database is compressed into a condensed smaller data structure , FP-tree which avoids the costly ,repeated database scans , 2. FP-tree based mining adopts a pattern – fragment growth method to avoid costly generation of large number of candidate sets, 3.A partitioning based ,divide-and-conquer method is used to decompose the mining task into a set of smaller tasks for mining confined patterns in conditional database , which dramatically reduces the search space.[2]

Join Step: C_k is generated by joining Lk-1with itself

Prune Step: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset

C_k: Candidate itemset of size k
L_k: frequent itemset of size k
L_1= {frequent items};
for(k= 1; L_k!=∅; k++) do begin
C_{k+1}= candidates generated from L_k;
for eachtransaction tin database do
increment the count of all candidates in C_{k+1}that are contained in t
L_{k+1}= candidates in $C_{k+1}$with min_support
end
return∪$_k$$L_k$;

2.2.3. Dynamic FP-Growth Algorithm: C.Gyorodi et al [2003] , It has focused in improving the FP-Tree algorithm construction based on two observed problems: 1. The resulting FP-Tree is not unique for the same "Logic "database. 2. The process needs two complete scans of the database. The above problems solved by G.Gyorodi et al. An important feature in this approach is that it's not necessary to rebuild the FP-Tree when the actual database is updated .It's only needed to execute the algorithm again taking into consideration the new transaction and stored FP-Tree.[3]

2.2.4. Partition Algorithm: A.Savasere et al [1995], Partion algorithm based on apriori algorithm, but it requires only two complete scans over the database .The Partition algorithm divided into two phases: 1. the database is divided into a number of non-overlapping partitions and frequent item sets local to partition are generated for each partition. The database is scanned completely for the first time. 2. Local frequent item sets from each partition are combined to generate global candidate item sets. Then the database is scanned second item to generate global frequent item sets [4].

2.2.5. DIC Algorithm: S.Brin et al [1997], DIS is a further variation of the Apriori-algorithm. DIC is a softens the strict separation between counting and generating candidates. Whenever a candidate reaches minsup that is even when this candidate has not yet "seen" all transactions, DIC starts generating additional candidates based on it. For that purpose a prefix-tree is employed. In contrast to the hash tree, each node-leaf node or inner node of the prefix-tree is assigned to exactly one candidate respectively frequent item set. In contrast to the usage of a hash tree that means whenever it reach a node it can be sure that the item set associated with this node is contained in the transactionfurthermore interlocking support determination and candidate generation decrease the number of database scans.[5]

2.2.6. Pincer Search Algorithm: This algorithm was proposed by Dao. I et al [1997].The algorithm was uses both the top-down and bottom –up approaches to Association rule mining. It is a slight modification to original Apriori algorithm. The main search direction is bottom -up expect that it conducts simultaneously a restricted top-down search. [6]

2.2.7. Pincer Search Algorithm: The concept of PSO was first suggested by Kennedy and Eberhart in 1995[5].Particle swarm optimization (PSO) is inspired by the social behavior observed in flocks of birds and schools of fish. In nature, there is a leader who leads the bird or fish groupto move, as illustrated in Fig. 2. Most members of the group follow the leader. In PSO, a potential solution to the considered problem is represented by a particle, similar to the individuals in the bird and fish groupEach particle travels in the solution space and attempts to move toward a better solution by changing its direction and speed based on its own past experience and the information from the current best particle of the swarm. [7]

2.2.8. The procedure of PSO is described as follows:

***A. Particle initialization*:**

An initial swarm of particles is generated in search space. Usually, the population size is decided by the dimension of problems.

***B. Velocity and position update*:**

In each iteration, a new velocity value for each particle is calculated based on its current velocity, the distance from its previous best position, and the distancefrom the global best position. The new velocity value is then used to calculate the next position of the particle in the search space. The particle's velocity and position are dynamically updated as follows:

$$V_{id}^{new} = w \times V_{id}^{old} + c_1 \times rand \times (P_{id} - x_{id}^{old}) + c_2 \times rand \times (P_{gd} - x_{id}^{old}),$$

$$x_{id}^{new} = x_{id}^{old} + V_{id}^{new}.$$

The new velocity of a particle, Vnewid, is updated by taking into consideration of the particle's previous velocity, Voldid, and previous position, xoldid. w = [0.5+ rand/2] is an inertia weight and rand is a uniformlygenerated random number between 0 and 1. Thecognition parameter, c1, and social parameter, c2, areacceleration coefficients that are conventionally set to a fixed value 0–2.Pid is the previous individual bestposition of this particle and Pgd is the current global bestposition then calculates the new position of the particle, xnew id .[8]

a) Both particle's position and the global best positionare far from the optimum and the particle velocityis low compared to its distance to the optimum [9].

b) Global best position is close to the optimum and theparticle position is far from them resulting in asmall improvement region and a large next position region.

***C. Evaluation and update of best locations*:**

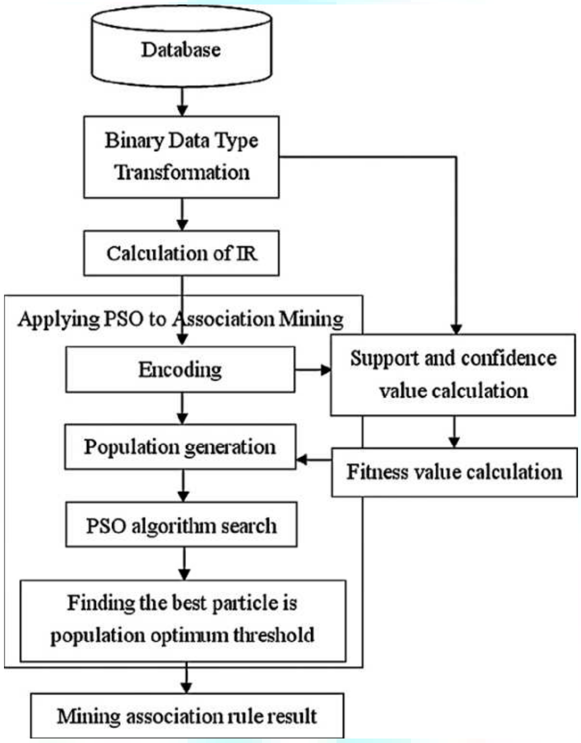
The fitness value of each particle is calculated by the objective function. The values of Pid and Pgd are then evaluated and replaced if better particle best position or global best position is obtained.

***D. Termination*:**

Steps (2) and (3) are repeated iteratively until the termination condition is met.

III METHODOLOGY

FIGURE:3THE PROPOSED ASSOCIATION RULE MINING ALGORITHM



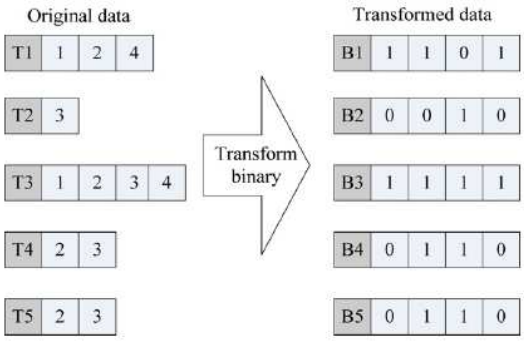
3.1. The proposed algorithm:

The proposed algorithm comprises two parts, preprocessing and mining. The first part providesprocedures related to calculating the fitness values ofthe particle swarm. Thus, the data are transformed andstored in a binary format. In the second part of thealgorithm, which is the main contribution of this study, the PSO algorithm is employed to mine the associationrules. First, it proceed with particle swarm encoding, this step is similar to chromosome encoding of geneticalgorithms. The next step is to generate a population ofparticle swarms according to the calculated fitnessvalue. Finally, the PSO searching procedure proceedsuntil the stop condition is reached, which means the best Particle is found. The support and confidence of the bestparticle can represent the minimal support and minimal confidence. [10]Thus, it can use this minimal support and minimalconfidence for further association rule mining. Fig.3 illustrates the algorithm structure.

3.2. Preprocessing of PSO association rule mining

Binary transformation: In this transform the transaction data into binary type data, each recorded andstored as either 0 or 1[8]. This approach can acceleratethe database scanning operation, and it calculatessupport and confidence more easily and quickly. Thetransformation approach is explained by an example inFig. 4. In Fig. 4, there are five records, say T1 to T5, in theoriginal data. Each of these records is transformed andstored as a binary type. For instance, there are a total ofonly four different products in the database, so four cellsexist for each transaction. Take B4 as an example, thistransaction only purchased products 2 and 3, so thevalues of cells 2 and 3 are both "1s," whereas cells 1and 4 are both "0s."Conversion shown in figure 6.Fig. 4: Data type transformation

FIGURE 4: DATA TYPE TRANSFORMATION



3.3. Application of PSO to association rule mining

Applying PSO to association mining is the mainpart of this study. We use PSO as a module to mine bestfitness value. The algorithmic process is quite similar tothat of genetic algorithms, but the proposed proceduresinclude only encoding, fitness value calculation, population generation, best particle search, andtermination condition. Each of the steps in the PSOalgorithm and the process of generating associationrules are explained as follows:

Encoding: According to the definition of associationrule mining, the intersection of the association rule ofitem set X to item set Y (X→Y) must be empty. Itemswhich appear in item set X do not appear in item set Y, and vice versa. Hence, both the front and back partitionpoints must be given for the purpose of chromosomeencoding. The item set before the front partition point iscalled "item set X," while that between the front

Partition and back partition points is called "item set Y."

3.4. Fitness value calculation: The fitness value in thisstudy is utilized to evaluate the importance of eachparticle. The fitness value of each particle comes from the fitness function. Here, employ the target function [11] to determine the fitness function value as shown inEq.

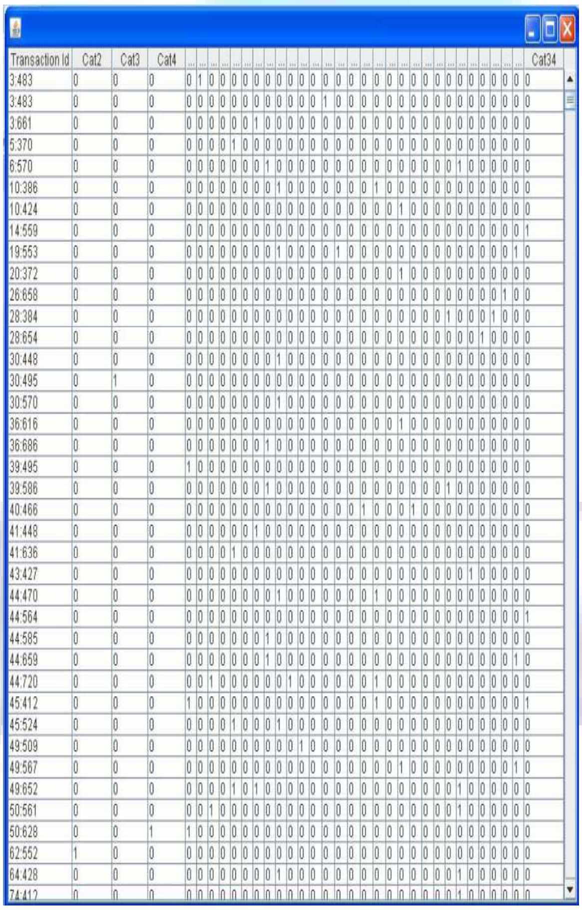
Fitness (k) = confidence (k) X log (support (k)

X length (k) + 1)

Fitness (k) is the fitness value of association rule type k.Confidence (k) is the confidence of association rule typek.Support (k) is the actual support of association rule typek.Length (k) is the length of association rule type k.The objective of this fitness function ismaximization. The larger the particle support andconfidence, the greater the strength of the association, meaning that it is an important association rule.

3.5.Population generation: *In* order to apply the evolution process of the PSO algorithm, it is necessary to first generate the initial population. In this study, it has select particles which have larger fitness values as the population. The particles in this population are called initial particles.

FIGURE: 5 BINARY TRANSFORMATIONS



Fitness value calculation is shown in figure 5

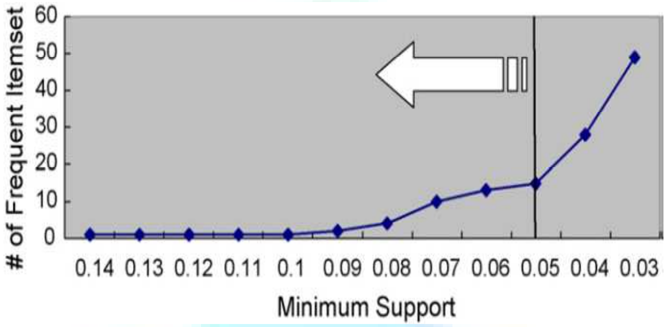
3.6. Search the best particle: First, the particle with themaximum fitness value in the population is selected as the "gbest." It designed a method to constrain thesearch. The constrained method is to calculate thedistance between the particle's new position and all thepossible particles inside the constrained range before theparticle's position is updated. Definitely, the particlewith the smallest distance will be selected and treated asthe particle's new position. In the distance measuring function it use traditional "Euclidean distance"as shown in Eq.

$$dist(x^n, y^m) = \sqrt{\sum_{1}^{d}(x_i^n - y_i^m)^2}$$

Where x_n is the position of the particle at nth update and ym is the possible particle number min the constrainedrange. In addition, d is the dimension of the search space. The nearest possible particle is selected to be thetarget particle's new position. This method can preventa particle from falling beyond the search space when its position is updated.

3.7. Termination condition: To complete particle evolution, the design of a termination condition is necessary. Inthis study, the evolution terminates when the fitnessvalues of all particles are the same. In other words, thepositions of all particles are fixed. Another terminationcondition occurs after 100 iterations and the evolutionof the particle swarm is completed.Finally, after the best particle is found, its supportand confidence are recommended as the value ofminimal support and minimal confidence as shown inFig[6]. These parameters are employed forassociation rule mining to extract valuable information. [10]

FIGURE 6: NUMBER OF HIGH-FREQUENCY ITEM SETS UNDER DIFFERENT MINIMAL SUPPORTS



IV.Model evaluation results and discussion

This section will use the database provided by Microsoft SQLServer 2000 to verify the feasibility of the proposed algorithm. Adetailed discussion is provided as follows.

4.1. Experimental platform and database

This study's experiment was conducted in the environment ofMicrosoft Windows XP using an IBM compatible computer withInter Pentium IV 1.60GHz and 512MB RAM. The algorithm wascoded by Borland C++ Builder 6.In regard to the experimental testing database, its source was aFoodMart2000 retail transaction database embedded in a MicrosoftSQL Server 2000, as illustrated in Fig. 6.Since there are different kinds of transaction databases inFoodMart2000, it only select sales fact 1997 data table for assessment. The number of product items in this data table is 1560.

In order to effectively mine meaningful association rules, thisexperiment categorizes the products into groups according to theproduct category provided by the data table. Thus, products are classified into 34 categories, each with a corresponding productcategory id. In regard to data selection, 6000 customers' arerandomly selected along with their corresponding transaction dataat different times. After arrangement, there are a total of 12,100transaction records for these 6000 customers.

FIGURE:7 THE DATA TABLE OF THE FOODMART2000 DATABASE

customer_id	time_id	product_category_id	product_category
5845	727	31	Snack Foods
5845	727	23	Meat
5845	727	23	Meat
5845	727	3	Beer and Wine
5845	727	29	Seafood
5846	386	14	Electrical
5846	386	15	Frozen Desserts
5846	386	19	Hygiene
5846	386	17	Fruit
5846	386	20	Jams and Jellies
5846	536	13	Eggs
5846	536	17	Fruit
5846	536	7	Canned Soup
5846	536	27	Pizza
5846	585	8	Hot Beverages
5846	585	26	Paper Products
5846	585	4	Bread
5846	585	6	Candy
5846	585	31	Snack Foods
5846	600	31	Snack Foods
5846	600	21	Kitchen Products
5846	600	2	Bathroom Products
5846	600	3	Beer and Wine
5846	600	1	Baking Goods
5846	720	15	Frozen Desserts
5846	720	18	Hardware
5847	439	34	Vegetables
5847	439	34	Vegetables

4.2. Mining results via PSO algorithm

In this experiment, every transaction record in the Food-Mart2000 has 1–13 items. After calculating the IR values, It findthat R(1→6) = 5.86822 is the largest. Therefore, It can generatefive different dimensions of encoding types for the particleswarm. They are two dimensions, 1→2, three dimensions, 1→3 and 2→3, four dimensions, 1→4, 2→4, and 3→4, and five dimensions,1→5, 2→5, 3→5, and 4→5, and six dimensions, 1→6,2→6, 3→6, 4→6, and 5→6. According to these five dimensions, it can implement the PSO mining process. An example of twodimensions is illustrated in Fig. 6.Since the computational results are different for each replication, a total of 30 replications are conducted in order to get thefinal experimental results. The results show that there are threepossibilities as listed in Table 1.

Next, it can conduct the three-dimensional PSO associationrule mining. The population size is 20. The computational results can be found in Appendix A. The results indicate that the maximal setup value is 0.03652 for the minimal support thresholdvalue. Because the support threshold value is smaller than 0.05, the numberof high-frequency item sets mined is too large. This means thatmore meaningless rules are generated. In the current experiment, since the maximal setup values of minimal support for three, four,five, and sixdimensions are all smaller than 0.05, none of the mining results are used.

TABLE: 1 THE RESULTS OF THE TWO-DIMENSIONAL PSO ASSOCIATION RULES FOR FOODMART2000

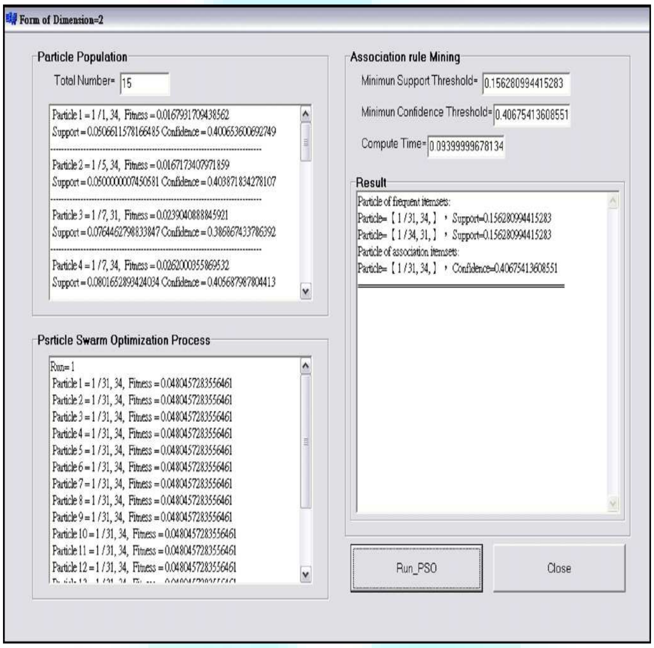
Possible results	High-frequency itemset(≥minimal support)	Association itemset (≥minimal confidence)
1	Minimal support = 0.15628 {Snack Foods, vegetables}	Minimalconfidence = 0.40675 {Snack Foods→Vegetables}
2	Minimal support = 0.15628 {Snack Foods, Vegetables}	Minimal confidence = 0.38497 {Snack Foods→Vegetables} {Vegetables→Snack Foods}
3	Minimal support = 0.09578 {Dairy, Vegetables} {Snack foods, Vegetables}	Minimal confidence = 0.40709 {Dairy→Vegetables}
Final result	{Snack Foods, Vegetables} {Dairy, Vegetables}	Snack Foods→Vegetables} {Vegetables→Snack Foods} {Dairy→Vegetables}

4.3. Performance evaluation analysis

4.3.1. Comparison of PSO and GA

Though Section 4.2 has shown that the proposed PSO algorithmcan provide very promising results, further investigation is still necessary.Thus, it cancompare with the genetic algorithm proposedin [9]. For both the PSO algorithm and genetic algorithm, the numberof product items and the number of transaction records are25 and 140,000, respectively. Basically, the two algorithms were implemented under the same conditions.

FIGURE: 8 A DEMONSTRATION OF THE IMPLEMENTATION OF PSO ASSOCIATION RULE MINING USING THE FOODMART2000 DATABASE



It attempt to ascertain whether or not computation time is related to the population sizeor number of evolutions, which are illustrated in Figs. 9 and 10, respectively. Figs. 9 and 10 clearly indicate that the proposed PSO algorithm out performs the genetic algorithm both in population size andnumber of evolutions. Table 2 also shows a similar outcome, showingthat PSO can converge faster. This is very important as the database is very large.

4.3.2. Performance evaluation of the PSO algorithm's searchingcapability

This subsection intends to discuss the searching performanceof the PSO algorithm. The first issue is the relationship betweenpopulation size and speed. The parameter setup is as follows:

(1) Number of dimensions: 2.
(2) Number of data: 12,100.
(3) Number of replications: 20.
(4) Population size: 5, 10, 20, 30, 40 and 50 Particles.

The average running times for different population sizes areillustrated in Fig. 11.In addition, if most of the parameter setups are the same exceptthat the size of population is 20, termination condition is 100 generationsand the number of experimental replications is 50, thenthe relationship between the number of evolutions and runningspeed is presented in Fig. 12. In summary, though the proposed PSO algorithm for associationmining requires more computation time with increasing population size, the increase is not significant. In addition, the numbersof evolutions mostly fall within the range from 1 to 10. This meansthat the convergence of computation is very fast. Thus, only a smallnumber of evolutions are good enough for real application.

Furthermore, in regard to the selection of threshold value setup, this study can provide the most feasible minimal support andconfidence. This dramatically decreases the time consumed bytrial-and-error. Thus, the proposed PSO algorithm is better than the traditional Apriori algorithm since it does not need to subjectivelyset up the threshold values for minimal support and confidence. This can also save computation time and enhance performance.

FIGURE:9 RELATIONSHIP BETWEEN POPULATION SIZE AND COMPUTATION TIME FOR PSO AND GA

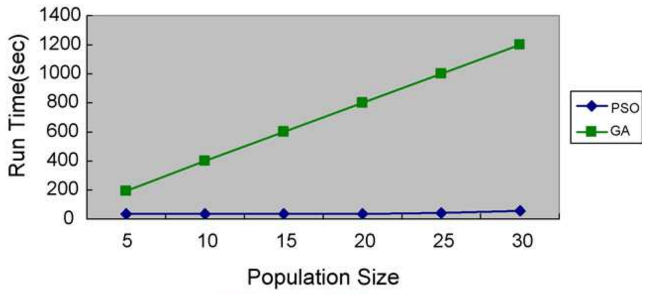


FIGURE:10 RELATIONSHIP BETWEEN NUMBER OF EVOLUTIONS AND COMPUTATION TIME FOR PSOAND GA.

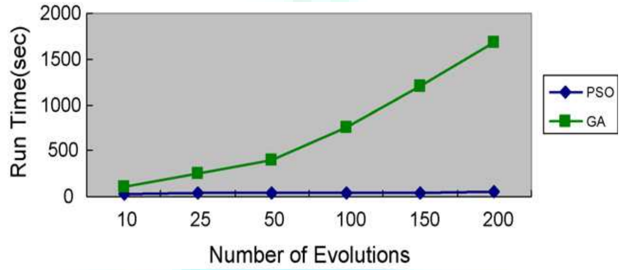


FIGURE: 11 RELATIONSHIP BETWEEN POPULATION SIZE AN COMPUTATION TIME USING THEFOODMART2000 DATABASE

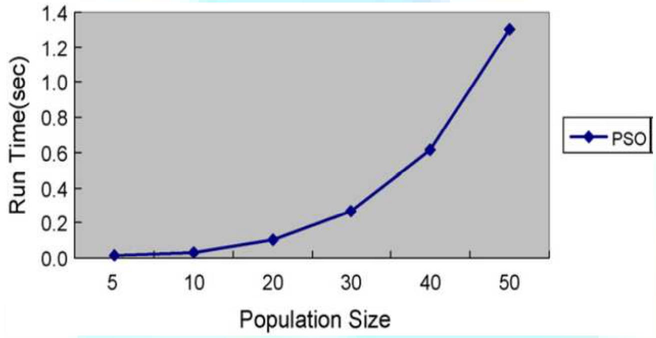
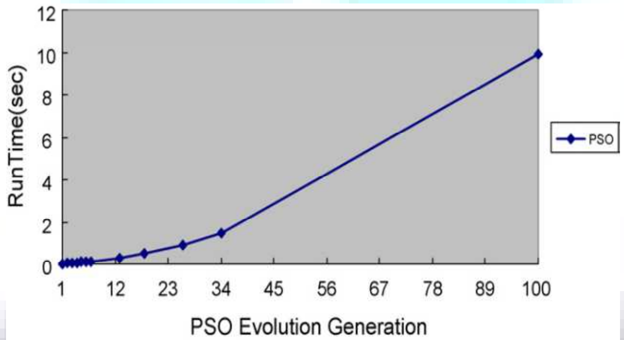


FIGURE: 12RELATIONSHIP BETWEEN NUMBER OF EVOLUTIONS AND COMPUTATION TIME.



V. CONCLUSION

Following the development of information technology, searchingfor meaningful information in large databases has becomea very important issue.That explains why association rules miningis the most popular technique in data mining. However, the traditional Apriori algorithm has a very critical drawback in thatthe minimal support and confidence is determined subjectively.

This study has demonstrated that using the PSO algorithm candetermine these two parameters quickly and objectively, thus, enhancing mining performance for large databases by applying theFoodMart2000 database.

Future studies can focus on testing different updating rules.Moreover, different product items may have different importance.A weighted PSO mining algorithm could be further investigated inorder to provide more practical approaches for industries.

REFERENCES

1. AhamedNickabadi,Mohammed mjchdiEbadzadhe,Rezasafabakhsh "A novel particle swarm optimation algorithm with adaptive inertia weight" Applied soft computing 2011.
2. Chi-Yang Tsai, I-Wei Kao "Particle Swarm Optimization with selective particle regeneration for data clustering". Expert systems with Application 2010.
3. Cornelia Gyorodi,RobertGyorodi, T.Cofeey&S.Holban- Mining association rules using Dynamic FP-trees",-in proceedings of The Irish signal and systems conference university of Limerick,Ireland , 30th June 2ndjuly 2003. ISBNO 0-9542973-1-8.
4. D.Lin and Z.M.Kedam, "Pincer Search: A new algorithm for discovering the maximum frequent set", In proceedings of Vi intl –Conference.
5. G.jayalakshmi, Dr.K.NageswaraRao "Mining association rules for large transactions using new support and confidence measures" Journal of Theoretical and Applied Information Technology, 2005
6. J.Han, J.Pei, Y.yin. "Mining Frequent Patterns without candidate Generation ",Proc.of ACM-SIGMOD2000

7. J.Kenndy, and R,Eberhart, ”Particle Swarm Optimization”, Proceedings of the IEEE Conference on Neural networks-ICNN, 95, vol, iv, perth, Australia, 1995, pp.1942-1948
8. Ming –Syan Chen, Tiawei Han, Philip S Yu. Data mining: An Overview from a Database perspective[J].IEEE Transactions on knowledge and Data Engineering,1996, 8(6):866-833
9. PankajKandpal,”Association Rule mining In Partitioned Database: Performance Evaluation and Analysis”, A Dissertation.
10. R.J.Kuoa,CMChaob, Y.T.chiu” Application of particle swarm optimization to association rule mining”, Applied soft computing 2011
11. S.Brin, R.Motwani, J.D.Ullam and S.Tsur,” Dynamic item set counting and implication rules for market basket data. In Proc.of the ACM SIGMOD. Int’l conf.on Management of data, 1997.
12. S.H. Kung, “Applying Genetic Algorithm and Weight Item to Association Rule,” Master Thesis, Department of Industrial Engineering and Management, YuanZe University, Taiwan, 2002.
13. S.Y.Wur,Y.Leu,”An effective Boolean algorithm for mining association rules in large database “, in proceeding of the 6th International Conference on Database system for Advanced Application s,1998,pp.179-186

APPENDIX

APPENDIX A. THREE-DIMENSIONAL EXPERIMENTAL RESULTS OF FOODMART 2000

Possible result	High-frequency item set (≥minimal support)	Association item set (≥minimal confidence)
1	Minimal support = 0.03652 {Dairy, Snack Foods, Vegetables}	Minimal confidence = 0.41424 {Dairy, Snack Foods → Vegetables}
2	Minimal support = 0.03239 {Can, snack foods, vegetables} {Dairy, Snack Foods, Vegetables}	Minimal confidence = 0.4237 {Can, Snack Foods →Vegetables}
3	Minimal support = 0.02950 {Can, Snack Foods, Vegetables} {Dairy, Snack Foods, Vegetables} {Fruit, Snack Foods, Vegetables} {Jams and Jellies, Snack Foods, Vegetables} {Meat, Snack Foods, Vegetables}	Minimal confidence = 0.4204 {Can, Snack Foods →Vegetables} {Fruit, Snack Foods →Vegetables} {Jams and Jellies, Snack Foods → Vegetables}
4	Minimal support = 0.025785 {Can, Snack Foods, Vegetables} {Beverages, Snack Foods, Vegetables} {Dairy, Snack Foods, Vegetables} {Fruit, Snack Foods, Vegetables} {Jams and Jellies, Snack Foods, Vegetables} {Meat, snack Foods, Vegetables}	Minimal confidence = 0.41106 {Can, Snack Foods →Vegetables} {Beverages, Snack Foods, Vegetables} {Dairy, Snack Foods, Vegetables} {Fruit, Snack Foods →Vegetables} {Jams and Jellies, Snack Foods → Vegetables}

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application and Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you tosupply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail**infoijrcm@gmail.com** for further improvements in the interest of research.

If youhave any queries please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals