Gravitational Search-Based Effective Knowledge Discovery Process for Spatio-Temporal Databases

Rainu Nandal[1]*　　　　　**Rahul Rishi[2]**

[1]*University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, India*
[2]*University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, India*
* Corresponding author's Email: rainunandal0880@gmail.com

Abstract: The rapidly rising and widespread use of database technology - including heterogeneous, geo-referenced and multidimensional data - there is a growing interest in developing new techniques for extracting knowledge from data. These techniques are the subject of the emerging field of Knowledge Discovery in Databases (KDD). In this paper, an algorithm to mine the intelligence rules from the input data using gravitational search (GS) algorithm is proposed. Initially, the significant patterns are selected from the input database with the use of measures like density ratio, sequence index, and density index. Subsequently, R-tree is modelled from the reduced dataset and is solved using the recently developed optimization algorithm. The rules obtained from the GS algorithm are useful for knowledge discovery. The experimentation is carried using the datasets to evaluate the performance in terms of mined rules and the computation time. The Performance metrics employed are a number of optimized rules, computation time and memory usage. From the analysis, it can be seen that the number of optimized rules decrease with increase in threshold and support.

Keywords: Sequential Rules; Data Mining; GS algorithm; R-Tree; Density index; knowledge discovery; significant pattern.

1. Introduction

Of late, the spatio-temporal data mining has been well-acknowledged as a unique investigation domain involving the data discovery from the datasets consisting of overt or implied temporal, spatial or spatio-temporal data [1]. The natural events include both the spatial and temporal components [2], diseases contagion [3], temporal geographic data techniques [4] and ecological technologies [5] carry out the function of processing the spatial, temporal and attribute data elements for the data discovery. The temporal objects are represented by two models of time such as the time points and time intervals, which are employed to register the facts and data on the spatial objects [6].

The spatiotemporal object concurrently encapsulates the spatial and temporal features of data and tackles the geometry fluctuating over a period of time [7]. In accordance with the modern temporal data mining methods [8] or spatial data mining approaches [9], we are able to locate the

spatial or temporal data from the spatiotemporal dataset [10]. Nevertheless, it is vital to extort novel types of the spatiotemporal data for the services by duly taking into account the temporal and spatial elements of the objects [11]. In addition, the Association Rule Mining (ARM) has set its elegant foot in the realm of data mining as the most well-acclaimed modern investigation approach for locating the fascinating association between the variables in the gargantuan databases. Till that time the Spatial Association Rules Mining (SARM) and Sequence Mining (SM) have ruled the roast as the two diverse subareas on spatial rules and temporal rules correspondingly [12].

The spatiotemporal data discovery constitutes the process of locating the spatiotemporal data set from the specified spatiotemporal data set. The spatiotemporal data mining [13, 14] brings in trains several thorny research hassles in the extortion of beneficial spatiotemporal data from the colossal spatiotemporal datasets [15]. A major chunk of the modern spatiotemporal data mining approaches

tends to toe the line of the raster-based spatial model, which is expanded with the temporal conception [14, 16, 17]. Further, the corresponding spatio-temporal sequential patterns are very vital to the exploration of the spatial and temporal evolutions of various events in a multitude of application realms.

The remaining segment of the document is orchestrated as follows: Section 2 effectively reviews the modern research works related to the subject. A 360-degree account of the novel method finds its august presence in Section 3. The contents of Section 4 colourfully carve out the colourful outcomes with the attendant discussion thereon. The concluding Section 5 characteristically carries out the customary chore of the conclusion of the document after presenting a concise portrayal of the captivating document.

2. Related Works

The literary domain is flooded with various fantastic techniques spearheaded with the intention of promoting the knowledge discovery process of the spatio-temporal database. Recounted below are some of the modern noteworthy works related to the captioned topic.

Y. Huang et al. [11] had their Eureka Day when they effectively brought to limelight the novel Mining Sequential Patterns from Spatio-Temporal Event Data Sets. In the document, they intelligently brought in the index as the relevance measure for the spatio-temporal sequential models, which were incredibly significant thanks to the striking qualities of interpretability through the medium of the spatial statistics. The efficiency assessments deploying the artificial and the factual data sets went a long way in proclaiming the fact that the Slicing-STS-Miner gained a clear edge over the STS-Miner by an order of magnitude faster in the case of capacious data sets.

A. Salas et al. [12] amazingly launched an innovative data discovery process for the spatio-temporal data by employing an Application to the river water quality watch. In the backdrop of such scenarios, they brilliantly brought forward a data discovery procedure applied to the hydrological data. With an eye on realizing the target, they intelligently integrated the consecutive technique to extort the knowledge on data gathered at several stations situated along a plethora of rivers. At the outset, data was pre-processed so as to attain the diverse. Subsequently, they deftly deployed a touchstone technique to extort the sequential patterns. In the long run, they brilliantly brought in a bewitching blend of the two methods (1) to filter

patterns in accordance with the interest measure, and (2) to cluster and illustrate them pictorially, with the intention of lending a helping hand to the specialists. The corresponding segment was invariably employed to estimate the spatialized indicators to facilitate the analysis of environmental and river surveillance pressure data.

M. Shaheen et al. [14] stunned the scientific society by significantly synthesizing the Context Based Positive and Negative Spatio-Temporal Association Rule Mining. They competently coined a captivating technique for the purpose of carrying out the spatial association rule mining from datasets projected at a temporal bar in which the contextual scenario was duly taken into account while engendering the positive and negative frequent itemsets. The novel approach for the positive and negative association rule mining was founded on the Apriori technique which was additionally expanded to embrace the context variable and replicate temporal series spatial inputs. The statistical appraisal illustrated with zero ambiguity that the innovative approach attained enhanced echelons of efficiency in engendering the precise, consistent and vigorous data vis-a-vis the performance of the conventional techniques.

N. B. Aoun et al. [15] naively brought to spotlight the novel Graph-based technique for the purpose of the human action detection by employing the spatio-temporal features. Thanks to the stellar advancement of the video data amassed and uploaded in the websites particularly the YouTube, it became highly essential to have an efficient assessment of the video actions. It was configured on the habitually employed Space–Time Interest Points (STIP) local features accompanied by a graph-based video version which modelled the spatio-temporal relations among these features. The investigation was accomplished on two critical datasets such as the Hollywood2 and UCF YouTube Action. The cheering test outcomes went a long way in authenticating the par-excellence of the proficient technique.

Bouquets of applause greeted B. Aydin et al. [16] when they amazingly brought to limelight the Spatiotemporal Indexing Techniques for Efficiently Mining Spatiotemporal Co-occurrence Patterns. They toiled hard to exclusively designate the spatio-temporal indexing approaches for extracting the co-occurrence patterns from the spatiotemporal datasets with sprouting polygon-based illustrated. They brilliantly brought in a novel structure for extracting the spatiotemporal co-occurrence models by deploying several indexing approaches for the purpose of effectively accessing data. Standing out

amongst them are the two well-investigated spatiotemporal indexing structures such as the Scalable and Efficient Trajectory Index (SETI) and Chebyshev Polynomial Indexing which were presently performed and offered in their structure.

In the available research articles dissimilar tress are taken into account for investigating procedure this comprises many disadvantages are declared below:

- Hierarchical clustering forms clusters to decrease energy dissipation. It has two stages: a setup phase to unify the clusters and a steady-state phase to permit all nodes to diffuse periodically at their time slots. The nodes send their information to the cluster-heads, and then the cluster-heads send the amassed information to the sink at the finish of each cycle
- PEGASIS Customs a chain from all sensor nodes so that each and every node conveys to and receives from its neighbours, and then the nodes take a turn as a cluster-head. It is an improved version of LEACH and keeps a momentous amount of energy associated with LEACH.
- The hierarchical tree creates a tree where data accumulation and information relay are achieved only at non-leaf nodes, and the accumulated information is sent to the root node selected as a sink. The algorithm is originated using the root node that broadcasts a control message to apprise the non-leaf nodes about the sensor ID, its parents, its residual power, its status, and hoping

- Prevailing sequential pattern mining approaches on market basket information may be unnatural because of the continuity of space and time. The major problem with transacting ionization in spatial and temporal domains is that the temporal, spatial, or spatio-temporal relationships across partition/transaction limitations may be lost in a disjoint partitioning, and associations may be over reckoned for an overlying partitioning.

3. Proposed Method

In our paper, we suggest an efficient technique for effective knowledge discovery process using a Spatio-Temporal Database. Initially, we take the data which is in the form of Spatio-Temporal model. The measures are employed so that only the significant patterns are selected. It also ensures non-important or fewer significance patterns are omitted. It is done so as to reduce the computation complexity too, and this process contains three measures such as density ratio, sequence index, and density index. The significant pattern undergo the event R-tree process to generate a tree pattern and redundancy is one of the crucial problems in event R-tree, in order to reduce the redundancy and optimize the rules, we are using an optimization algorithm named as Gravitational Search optimization algorithm (GS).

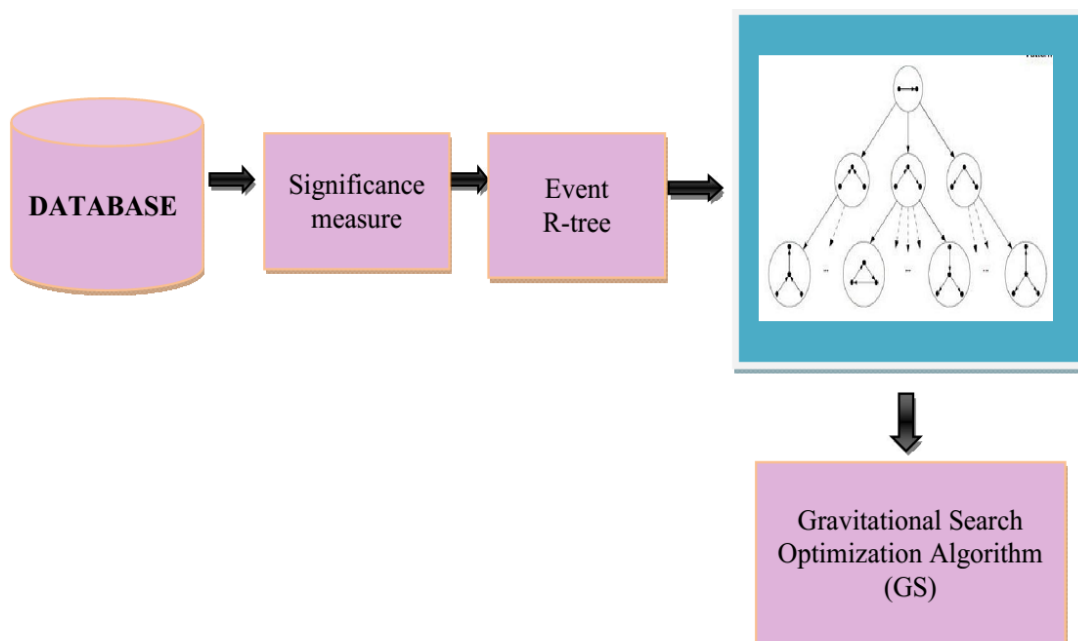figure 1 shows the block diagram of the proposed method.

Figure.1 Block diagram of proposed method

3.1. Significance Measure

So as to characterize the significance of a spatio-sequence of event types, we are using three measures such as density ratio, sequence index, and density index. Some other option to describe the significance measure of a pattern x1→x2 is to calculate the number of events of typex2 that follow events of type x1. Though, it ignores the distribution of event x2 in the overall space. It leads to spurious sequences in the significant sequential patterns, particularly when the event counts in the event types are large. Due to this issue, we propose a sequence index and density ratio as significance measures for spatio-temporal sequential patterns. In addition to this, the sequence index provides interpretability using spatial statistics. At first, we suggest density ratio as the significance and sequence index to measure sequential patterns with an arbitrary number of events.

3.1.1. Density Ratio

The density of an event set in a selected spatio-temporal space D is the average number of events in a unit of D, which is the total number of events of the event set divided by the volume |D| of the selected spatio-temporal space [D] or a given spatio-temporal space D, the density of an event set E located in each unit of D, so, it can be defined by,

$$Density(E, D) = \frac{|\{e \mid e\varepsilon\, E^\wedge e\, D\}|}{|D|} \quad (1)$$

Where, $|D| \rightarrow$ Volume of the space D

$|\{e \mid e\, \varepsilon\, E^\wedge e\, D\}| \rightarrow$ Cardinality of the set of events in E and D

As per the density concept, the density of an event type x in space D can be depicted by, density $(x.\varepsilon, v)$, where, $x.\varepsilon$ is the event of x.

Density Ratio for two events

For two event sets, E and E' and a given follow (→) predicate defined by a neighborhood function H(e) for any event e ε E, the density ratio of E and E' can be represented by,

$$Density\ ratio(E-> E') = \frac{avg_{e\varepsilon E}(densityE, H(e)))}{|density(E', D)} \quad (2)$$

Where,

D→spatio temporal embedding space

The follow (→) relationship is an overloaded predicate.

The event E' follows another event set E along with the density ratio. In simple, as per the above eqn, the numerator indicates the average density of

the event set E' which follows some event in the event set. The denominator represents the base density of event set E' in the embedding space.

In the event sets of two event types x and x', the density of events x' follows the event of x is higher than overall density of x' which results, the density ratio is greater than 1, the event type x' follows x to some degree. In another case that, if the density of event type x' following the event of x is lower than the overall density of x', which results, the density ratio is less than 1,the event type x repels the event type x' to some degree. Typically, the density ratio is close to 1, the event types x and x' are equally distributed.

3.1.2. Sequence Index

We consider L event types in a spatio-temporal sequence which are also known as (L)-sequence. Here, we refers the subsequence contains the *m*th and *n*th event types of (L) sequence S' as S'[*m:n*], with *m*[1,L-1],*n*[*m*+1,L], the *m*th event type in the sequence S' as S'[*m*].The sequence S' to be significant, the significance measure needs two vital requirements:

➤ The events of event type S'[*m*] in S' needs to follow an event sequence of the sequence S'[1:*m*-1].

➤ The density ratio between any two consecutive event types has to be significant.*2.*

3.1.3. Density Index

In addition to the two measures, we proposed density index, it is defined as the product of density ratio and sequence index, it is given by,

$$Density\ Index = Density\ ratio \times Sequence\ Index \quad (3)$$

3.2. Event R-Tree

The process of discovering the sequences which frequently occurs in a spatio-temporal database is the ultimate aim of the sequential pattern mining. Let E be a set of event types and it is in the form of $E=\{E_1, E_2...E_k\}$, where K is the number of event types. Each event type consists of a set of events denoted as $E.e = \{e_1, e_2.....e_n\}$, where n denotes the number of events.

In a sequential pattern, the spatio-temporal event type may appear for multiple times. For an instant, consider the sequential pattern $E_1 \rightarrow E_2 \rightarrow E_3 \rightarrow E_1$, as per this example the event type E_1 occurs more than one time.

Table 1. Sample spatio_ temporal data

Event Type	Event Set
E_1	e1(1.05,4), e2(2.1, 1.55), e3(4.2,3.2), e25(4.9, 6.54), e27(8, 4.1)
E_2	e4(6.4, 7.2), e5(0.7, 5.3), e7(3.5, 4.6), e9(4.2, 7.7), e10(2.55,2), e11(2.6,4), e14(7, 5.1), e16(4.2 ,2.6), e17(4.9, 3.8), e24(7.8, 9.5)
E_3	e6(3.7, 7.4), e13(5.1,2.6), e15(2.6, 4.9), e18(6.1, 6.6), e19(9.55, 1.3), e20 (5.6, 3.6), e30(7.8, 2.5) e31(1.3, 6.3)
E_4	e8 (5.2, 7.6), e12(6.3,3.9), e21(4.3, 3.8), e22(7.9, 4.45), e23(3.7, 2.8), e26(5.4, 3.8), e28(8.3,5.3), e29(5.1, 5.8), e32 (2.6, 8)

The length of a sequential pattern is not bounded by the number of event types due to the repetition of event types in the sequential patterns.

However, the length of the sequential pattern is bounded by n, since n is the maximum number of events. One of the issues of mining spatio-temporal sequential pattern is to find all the significant event type sequence of the form $E_1 \rightarrow E_2 \rightarrow E_3 ...... \rightarrow E_n$, where n is the length of the sequential pattern.

3.2.1. Adaptive Neighborhood Computation

The steps followed in neighborhood calculation are:
- At first, consider the region with user given space and time limits.
- Note the set of all events which occurs within this region.
- Keep the time as constant and vary the space to some extent.
- Calculate all the set of events for a new region.
- Set a threshold value (condition for minimum bounding rectangle).

- Continue the process by varying the space limits with constant time limit until reaches the threshold value.
- Apply the spatial distance to left and right directions in space, here, the temporal interval can be only applied to an upward direction due to the unidirectional property of time.
- If computed value < MBR (the minimum bounding condition), extend the unit in the neighborhood in both direction of space.
- While satisfying the conditions such as,
  - The minimum bounding condition is satisfied
  - It reaches some iteration
- Terminate the process.

3.2.2. Event R-Tree Generation

The spatial data retrieval is an effective topic in artificial intelligence, databases and information systems, data mining, and computational geometry. Generally, the spatial data are represented in the form of points, lines, and rectangles. Also, it is mostly applicable in the spatio-temporal database.

Figure 2 depicts the event R-tree and its structure. The above event R-tree contains the parent node and leaf node. The MBR is generally defined as the Maximum Bound Rectangle with a number of elements in it. The parent node contains two MBRs which are R1 and R2. So the number of the element present in MBR R1 is assigned as the leaf node of R1. Also, the number of the element present in MBR R2 is assigned as the leaf node of R2. Likewise, a rectangle in the data space the leaf node and the intermediate nodes can be built. One of the vital issues for efficient retrieval is the overlapping node and the MBR which means the rectangle with minimum elements or otherwise called as the half filled nodes. The overlapped nodes are partitioned into separate rectangle. Along with this, the half filled nodes are reduced by a process called joining nodes.
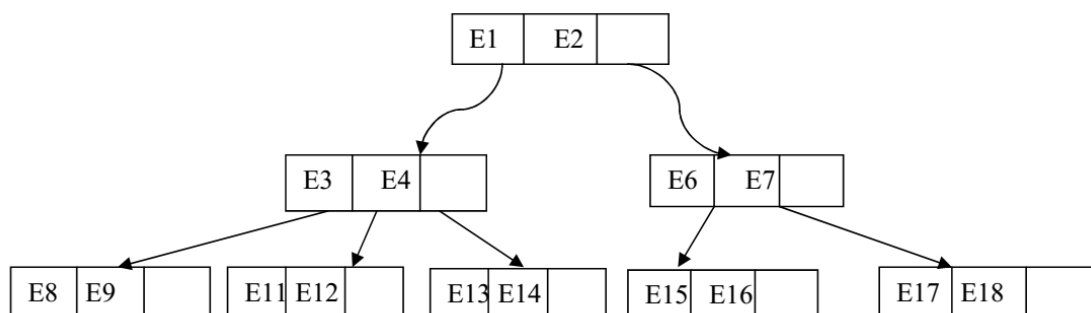

Figure.2 R-tree structure

Some of the operations in E-tree are given below:
  (i)   Search
  (ii) Insertion
  (iii) Deletion
  (iv) Splitting the nodes

(i) Search

Initially, the search operation breaks the search space into sub-regions. Then the search operation is done in search space for each sub-region. This search process begins from the top of the tree which is called as the root node to all the node interconnected with it and is called as the leaf node.

(ii)Event R-Tree Insertion

Inserting a new rectangle in R- tree is a vital process. Due to inserting the new rectangle as a separate leaf node will increase the height of the tree which makes the query retrieval process complex. So, we use the best way in inserting the new node is inserting it to the leaf node with half filled node. If there is no any half filled node insert the new rectangle as the new leaf node. The best way is to find the MBR with least data objects.

(iii)   Event R-tree Deletion

In deletion, the initial stage is to locate the rectangle to be deleted in the leaf node of the tree. After locating the node it is removed from the leaf node. For deleting the particular rectangle from the leaf node the search undergoes overall intermediate nodes. The complete structure of the tree is rearranged after deleting the particular node.

(iv)   Splitting R-tree Nodes

The Splitting of the node in the event R- tree is one of the important processes. Here while inserting a new node to the MBR there is a chance of overflow. So partitioning the overflowing node inefficient way is an important task. Here the overflow node is separated into two regions. Initially, the distance between each points and the inserted point in the MBR is calculated. Finally, the point with minimum distance is partitioned as a separate MBR.

3.3. Gravitational Search Optimization Algorithm (GS)

After the event r-tree process, we obtain a tree structure, in this, redundancy is one of the crucial problems in event R-tree, in order to reduce the redundancy and optimize the rules, and we are using an optimization algorithm named as Gravitational Search optimization algorithm (GS).

The gravitational search algorithm is the one of the latest nature-inspired algorithm to solve the optimization problems based on the Law of gravity. Many researchers have applied the gravitational search algorithm on large numbers of problems because it requires only two parameters and having ability to find near the global optimum solution and provides better results as compared to other nature inspired algorithms.

The gravitational force between two particles is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. The particles are considered as objects and their performance has evaluated with their masses. In GSA, each particle has associated with four specifications: particle position, its inertial mass, active gravitational mass and passive gravitational mass. The position of particles provides the solution of the problem while fitness function is used to calculate the gravitational and inertial masses. In gravitational search, the initial population is generated which is evaluated using the fitness function. The gravitational constant is later found out. Then, compute the gravitational masses, acceleration and the force. Update the particle velocity and position. Continue the process till the stopping criteria is met. figure 3 shows the step by step process of gravitational search optimization algorithm.

At first, we take the rules from the R-tree as the initial solution. Here, the rules are considered as an individual, the fitness function is derived to evaluate the feasibility of each rule. The fitness for the functions is designed based on the support, confidence, frequency and lift parameters of the proposed   approach.
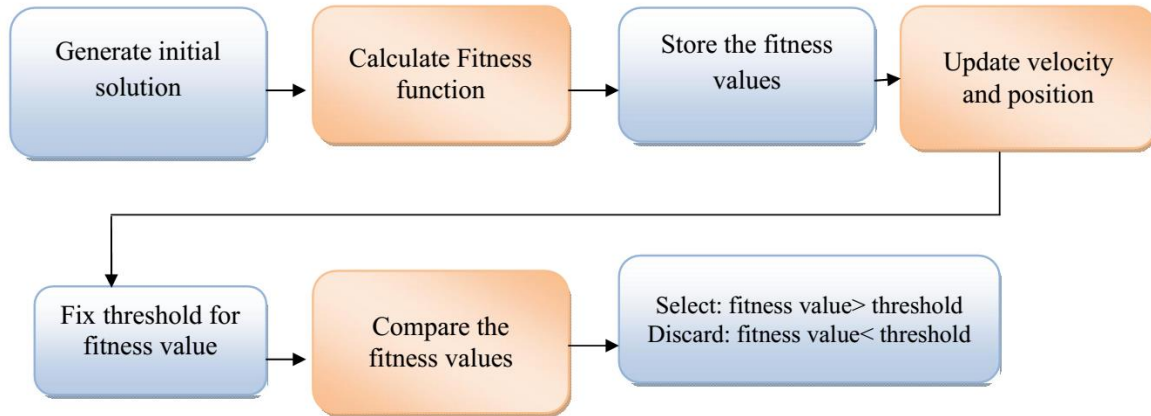
Figure.3 Architecture of Gravitational Search optimization algorithm

Apply the fitness function and it is computed by the formula,

$$F = conf(Rt) + \log(S(Rt) \times (m\_S(lt) + m\_S(Rt)) \times lift(Lt)) \quad (4)$$

Where, $F \rightarrow$ Fitness, $S \rightarrow$ support, $m\_S \rightarrow$ minimum support

The minimum support is the support required to sustain a rule regarding the multi-relational database. The minimum support is represented as min_support and is defined as,

Where, $T(X,Y) \rightarrow$ number of transactions, which contain the rules, $T_n \rightarrow$ total number of transactions

The parameters confidence and lift are derived from the parameter support. The parameter can be derived as,

$$confidence = \frac{Support(X \cup Y)}{Support(X)} \quad (5)$$

$$lift = \frac{Support(X \cup Y)}{Support(X) \times Support(Y)} \quad (6)$$

After the fitness calculation process, update the velocity and position. Compute gravitational constant and masses given by:

$$G_i(t) = \frac{Fit_i(i) - worst(t)}{Best(t) - worst(t)} \quad (7)$$

$$M_i(t) = \frac{G_i(t)}{\sum_{i=1}^{N} G_i(t)} \quad (8)$$

Compute the acceleration using the formula,

$$a_i(t) = E \times G_i(t) \quad (9)$$

Where,
$a \rightarrow$ acceleration, $E \rightarrow$ Euclidean distance
Compute the updated velocity and position:

$$V = rand \times v + a_i(t) \quad (10)$$

$$X = x + V \quad (11)$$

Where, $V \rightarrow$ velocity, $rand \rightarrow$ random value
The fitness value specifies the rules to be important for getting the optimized rules which result in reduced redundancy. The fitness values are selected to distinguish between the rules, those with the high chance of getting maximum values and those with low chance of getting maximum values. The distinction is made by applying a threshold value to the fitness values, the fitness value, which is higher than that of threshold value are selected and rest are discarded.

$$opt\_rules = \begin{cases} select, if \ fitness > threshold \\ discard, if \ fitness < threshold \end{cases} \quad (12)$$

Where, opt-rules represent the intelligent rules which are getting optimized according to the GSO algorithm.

4. Results and Discussion

The experimental results and analysis of the proposed technique are presented in this section. The proposed system is implemented in java programming language with JDK 1.7.0 in a windows machine containing configurations Intel (R) Core i5 processor, 1.6 GHz, 4 GB RAM, and the operation system platform is Microsoft Wnidow7 Professional. We have used the publically available Spatio-Temporal Database for the knowledge discovery process.

4.1. Performance Analysis

Recently, spatio-temporal data mining has been identified as a distinct research area concerned with knowledge discovery from datasets containing explicit or implicit temporal, spatial or spatio-temporal information.

Table 2 shows the overall outcome obtained from the gravitational Search-based Effective knowledge discovery process.

Table 2. overall outcome of the proposed method

Threshold	Before Reduce (count)	After Reduce (count)	Time (sec)	Memory (byte)
0.000001	1554	1000	27858	12060496
0.0000025	1554	632	11174	10381448
0.000005	1554	447	9774	10165696
0.0000075	1554	365	8747	7977480
0.00001	1554	316	8457	7486976

Running time

The running time in the Java program is squarely rooted on based on the memory and database values. The time complication is habitually represented in such a way as to ignore the coefficients and lower order terms and in most cases, the time duration is computed in millisecond (ms).

Memory Usage

The Java effectively organizes the memory for use. New objects are generated and positioned in the heap. When your application does not contain any reference to an object, the Java garbage collector takes up the work of elimination of the related object and removes it from the memory in order that your application is competent to employ the relative memory further.

This section depicts the graphical representation of the proposed method along with their corresponding values in the form of tabulation.

In gravitational Search-based Effective knowledge discovery process, we used the spatio-temporal database for the execution purpose. It clearly depicts the time consumption and memory usage various threshold values.

Figure 4 shows the time for various threshold values, the time gradually decreases while the threshold values are increasing. We have obtained the maximum time of 27858ms when the threshold value is 0.000001, likewise, we got minimum time of 8457ms while the threshold value is 0.00001.

Comparison Graph for Different Techniques

This section shows the Time and memory values for the proposed technique and other three models like the R-Tree with GSO, Multi-valued decision system approach [22] and Slicing STS miner algorithm [11].

Figure 4 shows the time for various threshold values, the time gradually decreases while the

threshold values are increasing. We have obtained the maximum time of 27858ms when the threshold value is 0.000001, likewise, we got the minimum time of 8457ms while the threshold value is 0.00001. This proposed work compared to other existing techniques Multi-valued decision system approach and Slicing STS miner algorithm is 15 to 26% varied. Subsequent to the integration of the internal and external utility values to the web access sequence the tree is generated. In accordance with this, the run time is determined by both the techniques.
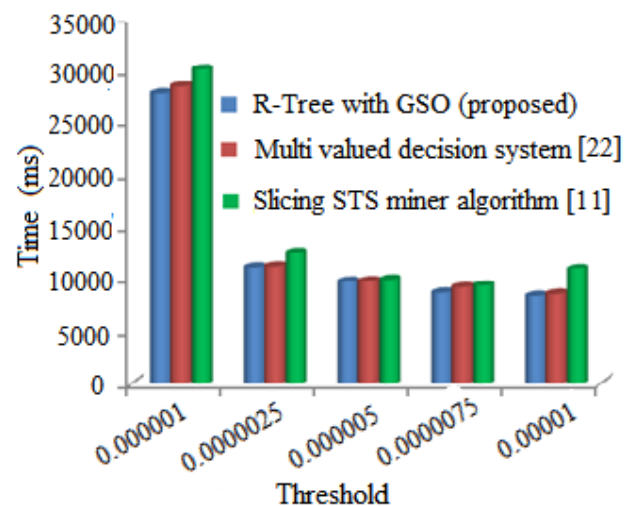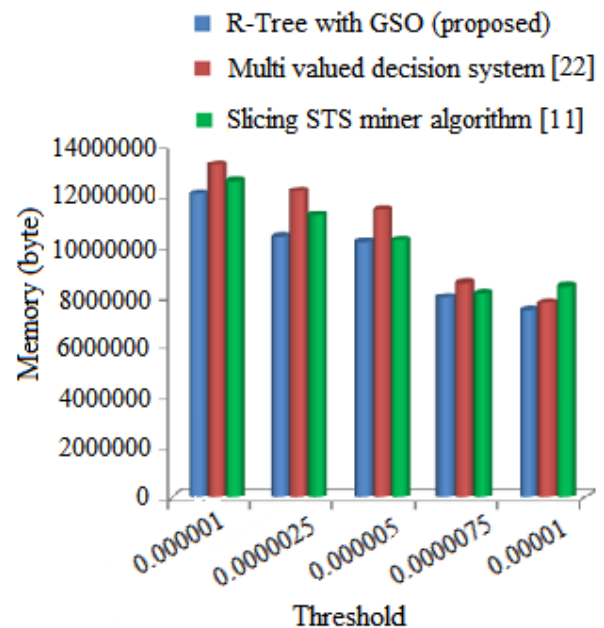

Figure.4 Time Comparison
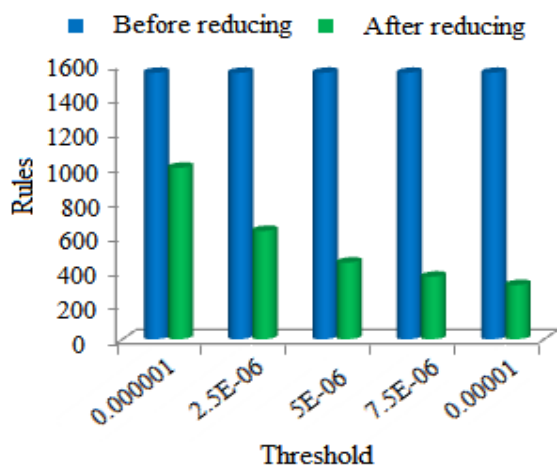

Figure.5 Memory Comparison

Figure.6 comparisons of the optimum rules for various threshold values

Figure 5 represents the memory usage for various threshold values, the memory usage decreases when the threshold value goes on increasing. We have achieved less memory usage 7486976 while the threshold value is 0.00001; similarly, we obtained high memory usage 12060496 while the threshold value is 0.0000001. The memory allocation of both the trees is modified after integrating the utility values of the web access sequence.

Figure 6 depicts the comparisons of the optimum rules for various threshold values, as per the analysis, the graph clearly shows that the rules which are obtained after the reduction are high while compared with the rules which are obtained before reducing process. It is crystal clear from the graph that the rules which are generated after the decrease are extremely high when compared with the rules achieved prior to the reduction procedure.

Analysis by varying the support

Analysis of finding the optimized rules based on varying the support represented in below table 3.

Table 3. Number of Rules by Varying Support

Support	Optimized Rules
2	909
4	803
6	689
8	556
10	498

Inferences from table 3

o The analysis is made by varying the support and finding out the optimized rules.
o From the results, we can infer that optimized rules increase with a decrease in a threshold.

5. Conclusion

Recent research literature has explored the sequential patterns on transaction data and trajectory analysis on moving objects. We have made a tree pattern using event r-tree method. Redundancy is one of the crucial problems in event R-tree, in order to reduce the redundancy and optimize the rules; we utilized an optimization algorithm called as Gravitational Search optimization algorithm (GS). The techniques have been implemented using JAVA and evaluated under standard metrics. The Performance metrics employed are a number of optimized rules, computation time and memory usage. From the analysis, it showed that the number of optimized rules decrease with increase in threshold and support. Further research may include studying the effects of using different tree algorithms like High Utility Web Access Sequence (HUWAS) are used to optimize the rules along with newly inspired optimization techniques to improve the performance.

References

[1] S. Kaufmann, "Data visualization techniques for knowledge discovery based on domain generalization", *M.Sc. Thesis, Department of Computer Science, University of Regina, Regina*, Canada, pp.1-221, 2002.
[2] T. Cheng and J. Wang, "Applications of spatio-temporal data mining and knowledge discovery for forest fire prevention", *ISPRS Commission VII Mid-term Symposium "Remote Sensing: From Pixels to Processes, Enschede, the Netherlands*, pp. 8-11, 2006
[3] D. R. Moreno, M. Pascual, M. Emch and M. Yunus, "Spatial clustering in the spatio-temporal dynamics of endemic cholera", *BMC Infectious Diseases*, Vol. 10, pp. 1-12, 2010
[4] Y. ping, T. Xinming and W. Shengxiao, "Dynamic cartographic representation of Spatio-Temporal data", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 1-6, 2008.
[5] Z. Obradovic, D. Das, V. Radosavljevic, K. Ristovski and S. Vucetic, "Spatio-Temporal characterization of aerosols through active use of data from multiple sensors", *ISPRS TC VII Symposium – 100 Years ISPRS*, Vienna, Austria, pp. 424-429, 2010.
[6] M. N. Doja, S. Jain and M. AfsharAlam, "SAS: Implementation of scaled association rules on spatial

multidimensional quantitative dataset", *International Journal of Advanced Computer Science and Applications*, Vol. 3, No. 9, pp. 130-135, 2012.

[7] N. Pelekis, B. Theodoulidis, I. Kopanakis, Y. Theodoridis, "Literature review of spatio-temporal database models", *The Knowledge Engineering Review*, Vol. 19, No. 3, pp. 235-274, 2004.

[8] S. A. Sarabjot, D. A. Bell and J. G. Hughes, "The role of domain knowledge in data mining", *In Proc. of the Int. Conf. on Information and Knowledge Management*, pp. 37-43, 1995.

[9] K. Koperski, J. Han and J. Adhikary, "Mining knowledge in geographical data", *to appear in Communications of the ACM*, pp. 1-8, 1998

[10] J. F. Roddick and M. Spiliopoulou, "Temporal data mining: survey and issues", *Research Report ACRC-99-007,* University of South Australia, pp. 750-767, 1999.

[11] Y. Huang, L. Zhang and P. Zhang, "A Framework for Mining Sequential Patterns from Spatio-Temporal Event Data Sets", *IEEE Transactions On Knowledge and Data Engineering*, Vol. 20, No. 4, pp. 443-459, 2008.

[12] H. A. Salas, J. Aze, S. Bringay, F. Cernesson, N. S. Folcher and M. Teisseire, "A knowledge discovery process for spatiotemporal data: Application to river water quality monitoring", *Ecological Informatics*, Vol. 26, No. 2, pp. 127-139, 2015.

[13] J. W. Lee and Y. J. Lee, "A Knowledge Discovery Framework for Spatiotemporal Data Mining", *International Journal of Information Processing Systems*, Vol. 2, No. 2, pp. 124-129, June 2006.

[14] M. Shaheen, M. Shahbaz and A. Guergachi, "Context Based Positive and Negative Spatio-Temporal Association Rule Mining", *Knowledge-Based Systems*, Vol. 37, pp. 261-273, 2013

[15] N. B. Aoun, M. Mejdoub and C. B. Amar, "Graph-based approach for human action recognition using spatio-temporal features", *Journal of Visual Communication and Image Representation*, Vol. 25, No. 2, pp. 329-338, 2014,

[16] B. Aydin, D. Kempton, V. Akkineni, S. R. Gopavaram, K. G. Pillai and R. Angryk, "Spatiotemporal Indexing Techniques for Efficiently Mining Spatiotemporal Co-occurrence Patterns", *IEEE International Conference on Big Data*, pp. 1-10, 2014.

[17] Yugalkumar and G. Sahoo, "A Review on Gravitational Search Algorithm and its applications to Data Clustering and Classification", *I.J. Intelligent Systems and Applications*, Vol. 6, pp. 79-93, 2014.

[18] A. R. Ganguly and K. Steinhaeuser, "Data Mining for climate change and impacts", *IEEE international conference on data mining workshops*, pp. 15-19, 2008

[19] J. Roddick and M. Spiliopoulou, "A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research", *ACM SIGKDD Explorations*, pp. 34-38, 1999.

[20] S. Chai, F. Su, W. Ma, "An approach of discovering spatial-temporal patterns in geographical process", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 38, pp. 80-86.

[21] S. M. Sarkar, I. Sikder, Vijay and Konangi, "Spatio-temporal Pattern discovery in sensor data: A multi valued decision systems approach", journal of Knowledge-Based Systems, Vol. 109, pp. 137-146, 2016.