BioGecko Vol 12 Issue 03 2023

ISSN NO: 2230-5807

GENERALIZED ASSOCIATION RULE MINING ON FUZZY MULTIPLE DATASETS FOR BRAIN INJURY PATIENTS

Gurpreet Singh Popli
C K Birla Hospital, Delhi, popligurpreet@gmail.com
Praveen Arora
Jagan Institute of Management Studies, Sector 5, Rohini, Delhi, praveen@jimsindia.org
Deepti Chopra
Jagan Institute of Management Studies, Sector 5, Rohini, Delhi, deepti.chopra@jimsindia.org

Abstract
Brain injury is a serious medical condition that can be caused by a variety of factors including trauma, stroke, or disease. Diagnosing and treating patients with brain injury requires identifying several factors that contribute to the injury, including the patient's medical history, imaging results, and other clinical data. This information is often collected from multiple sources, which can be difficult to integrate and analyze because of the imprecision and uncertainty of the data. Brain injury is a major public health problem that can result in significant morbidity and mortality. Patients with brain injury often have complex and heterogeneous clinical presentations, making it difficult to identify patterns and associations between different diagnostic and treatment variables. In recent years, association rule mining has emerged as a powerful technique for detecting interesting patterns and relationships in large datasets. However, conventional association rule mining algorithms are not well suited to handle imprecise and uncertain data. To address this limitation, this research paper presents a new approach for mining generalized association rules on fuzzy multiple datasets for patients with brain injuries. The proposed approach was developed to process imprecise and uncertain data by incorporating fuzzy logic. The research study uses data from multiple sources, including medical records, imaging, and patient reports, to identify patterns and associations between different factors contributing to brain injury. To improve outcomes for patients with brain injuries, it is important to identify patterns and associations between different diagnostic and treatment variables. Fuzzy multiple data sets are common in health research due to the uncertainty and imprecision inherent in clinical data. The proposed approach was developed to deal with fuzzy data, which in some cases can be more informative and accurate than crisp data.The work addresses the extraction of fuzzy association rules in a given database designed by entity-relationship models (ER) at multiple levels.The contribution of the study is an attempt to standardize algorithms to find the most appropriate result from tables of fuzzy data.Our results have important implications for improving the diagnosis and treatment of patients with brain injury.

Additional Keywords and Phrases: Association Rule Mining, Data warehousing, E-R modelling, Fuzzy item sets

1 INTRODUCTION

The study in this paper focuses on exploring knowledge from databases based on fuzzy data by extracting rules from databases with multiple tables. The work uses the EAS (Extended Apriori Star) and the newly discovered algorithm AJS (Apriori Join Star).

Traumatic brain injury is one of the most dangerous injuries, as severe bleeding and serious

complications are common. Therefore, prompt medical treatment is essential to increase the survival rate of patients with brain injuries. However, it is often difficult to make treatment decisions because cases are complex and have similar patterns. Using association rules in a trauma support system is the most efficient way to achieve effective decision making. The discovered algorithm forms rules that provide a reliable method for making predictions/recommendations about status and exact outcome. The resulting rules can help physicians make quick and accurate decisions, and they will also develop a diagnostic system to detect the severity of traumatic brain injury using fuzzy logic.
With the extended Apriori algorithm that finds fuzzy association rules, the associated higher levels can be discovered based on the strong association rules. In the extended Apriori Star algorithm, each leaf element in taxonomic structures is inserted into the transaction set T to form a so-called extended transaction set T'. In the case of fuzzy taxonomic structures, T' is generated not only by adding all the ancestors of each leaf element in fuzzy structures to T, but also by representing the degrees to which the ancestors are indicated by the transactions in T. This is called a transaction set. However, it also cannot handle multiple tables. The Extended Apriori Star algorithm, which can handle multiple tables, finds the non-redundant fuzzy association rules.

In the Extended Apriori Star algorithm, the degree between each leaf node and its ancestor must be computed from more than one table, while this is not required in the classical algorithm. Second, the counting operation is replaced by the counting operation. And third, the Extended Apriori Star algorithm uses join and entity support in determining frequent element sets. By taking entity support into account, it does not exclude from the result entity sets that are frequent with respect to their entity table but not with respect to the relation table, and it also allows calculating the correct support and confidence for rules that exist between attributes of the same entity table.

2 EXISTING STUDIES

Several algorithms have been proposed for searching association rules and fuzzy generalized taxonomic structures, such as Apriori [12] and AprioriTid [1]. These algorithms have been shown to outperform others such as AIS [2] and SETM [9]. Mining association rules is becoming increasingly important in decision making processes [4]. Mining association rules at multiple concept levels [8] and generalized association rules [13, 14, 22, 23] can lead to more specific knowledge discovery. Subalgorithms have been developed to handle fuzzy association rules involving collections of fuzzy sets, which provide a natural and abstract knowledge representation [3][7]. In the proposed study, algorithm [10] is used to manage quantitative attributes associated with multiple fuzzy sets. This new algorithm addresses the limitations of COMA patient diagnosis by improving the accuracy of medical data. The study uses fuzzy taxonomy structures [5], association rules in entity-relationship models (ER) [6], and extended algorithms from [15, 16, 20]. In this way, the need to compute a linkage of tables is eliminated, resulting in higher efficiency and lower cost. The algorithm AQ [11] for mining quantitative data was also used in the study. The study in [17] proposes an algorithm that identifies patterns and associations between different diagnosis and treatment variables to improve outcomes for patients with brain injury. The previous studies show the potential of fuzzy logic based systems for medical diagnosis. By representing complex medical data as fuzzy sets and using fuzzy rules and reasoning to make diagnoses, these systems can provide accurate and reliable diagnostic results. However, it is important to note that fuzzy logic-based systems should not be used as a substitute for medical expertise and diagnosis, but rather as a tool to assist medical professionals in their diagnostic and treatment decisions. The studies listed in [19,20,21] were aimed at understanding post-traumatic seizures and their early treatment as well as the assessment of coma patients.

3 PROPOSED STUDY

Algorithm- Apriori Join Star

```
Sub_algorithmOJ_Support (OJ, min_sup, T, Ei)
forallitemsets I Є OJ
        compute ∑count // sum of all the degrees that are associated with the transaction
                  in T
        If (count>= (min_sup x |T|)
                    call Entity_Support (OJ, T,Ei)
                    call Join_Support (OJ, T)
        endif
                    Ck=I
endfor

Sub_algorithmEntity_Support (OJ, T, Ei)
Select count (*) from OJ where OJ.Itemset=T.Itemset and Ei.Entity Key is unique.// i=1, 2, 3…, n
Divide the count value with the distinct entity keys of Ei.

Sub_algorithmJoin_Support (OJ, min_sup, T)
Select count (*) from OJ where OJ.Itemset=T.Itemset
Divide the count value with number of tuples in OJ.

Input: Outer Join of all the tables, minimum support value, fuzzy taxonomy.
Output: list of frequent itemsets.
Determine the degree to which leaf item belongs to its ancestor.
µxy =     (      µle)
          ∀l: x → y  ∀eonl
call degree (LNi, INj)
Set K=1; Ck = 1-Itemsets (E1, E2,…., En)
Call OJ_Support (OJ, min_sup, T,Ei)
Frequent F = ( if c. Entity_Sup || c. join_sup  Є Ck>= min_sup) // c = candidates  of Ck
AllFrequent AF = (c.E.entity_Sup || c.J.join_sup Є Ck>= min_sup) // c.E =Entity
Ck=apriori_gen (F, min_sup) // [5]
If Ck = Ø then Exit.
Go to Step 4.
```

This algorithm works for the outer join table of all the entity and relationship tables.The results of extended apriori star and apriori join star comes out to be same.

It can be observed that Extended Apriori Star keeps the entity itemsets that are frequent with respect to the entity tables but not frequent with respect to the relation tables, and generates on the fly the linkage of only those tables whose attributes are present in the candidate collection, and ignores all the tables whose attributes are not present in the candidate collection, i.e. The extended Apriori Star algorithm avoids performing a join scan of all tables and scans only some of the entity tables, while the Apriori Join Star algorithm does not work with individual entity tables; it first performs an outer join of all entity and relationship tables and then works entirely with the outer join table, i.e., it scans all entity tables, making its data larger than the data generated on the fly by the extended Apriori Star algorithm.

4 ANALYSIS AND INTERPRETATION

The results show that the discovered method is able to find interesting generalized fuzzy association rules from several tables. In this paper, the parameter settings used to perform the experiments for the study are described, and the performance of the newly discovered algorithm is evaluated on several datasets.

4.1 Performance

After the discovered algorithm worked correctly for a small sample database, the algorithms were run for an increasing number of transactions with different parameter settings to verify the algorithms' performance.

The algorithms were tested against three different scales. These are listed below:

(a) Execution time required - This metric was chosen because the time required by the algorithms determines how well the algorithms perform. Do the discovered algorithms generate the rules in the given amount of time or do they require more time? What is the impact on execution time as database size increases and minimum support decreases? Do the algorithms have good scalability? Evaluating the execution times is one of the ways to find out which of the two discovered algorithms is better in terms of time requirements.

(b) Memory Utilization - In the design and implementation of the newly discovered algorithm Apriori join star, the study chose speed for most of the decisions that require a tradeoff between spacerequirements and speed. Therefore, the main memory usage in the discovered algorithms is certainly higher compared to algorithms such as Extended Apriori [7], Apriori Join, and Apriori Star [6]. This measure was taken to verify that the algorithms are efficient on computers with less main memory. It can be observed that the main memory consumption for EAS and AJS is proportional to the size of the output and does not overflow when the size of the input increases.

(c) Redundancy - The main factor that hinders the application of association rules is the large number of rules returned by the mining process. In this study, an effective solution to eliminate the redundancy in the set of generated association rules is presented. An algorithm for this solution is presented in this study. The goal of the study is not to efficiently derive the full rule set under certain constraints, but to generate a compact but high-quality rule set. Experiments on some datasets show that the number of rules in the rule set could be greatly reduced.

-

Apart from the experimental comparisons, efficient implementation is required to achieve better performance and quality of results. The following two choices were made in this study.

Dividing continuous values into intervals, where a large number of intervals would generate a large number of candidates and too many candidates would lead to degradation of performance, so 8-10 intervals are chosen for numerical attributes in this study.

Support, Confidence, Entity Support, and Join Support are four rule measures supported by Extended Apriori Star and APRIORI JOIN STAR. a) Support is simply the number of transactions that contain all elements in the antecedent and consequent parts of the rule. Confidence is the ratio between the number of transactions containing all elements in both the consequent and antecedent parts of the rule and the number of transactions containing all elements in the antecedent part. Entity Support is calculated as the number of rows that contain search item sets with unique entity keys, relative to the number of unique entity keys. Join Support is calculated as the number of rows containing itemset from the join table, related to the cardinality of the join table.

The experiments for the study include a number of datasets from Kaggle. Table 1 shows the characteristics of the datasets used in the experiments. As shown in the table, D1 and D2 are two datasets with different parameter settings. Join (D1) and Join (D2) are again datasets formed by merging the entities and relationships of the original datasets D1 and D2. Join (D1) is formed by joining D1 and Join (D2) is formed by joining D2.

Database	Parameters	Description	D1	D2	Join(D1)	Join(D2)
Entity 1	|D|	No. of Transactions	100K	100K	200K	200K
	|T|	Avg. Size of Transaction	8	14	6	10
	|I|	Avg. Size of maximal potentially large itemsets	4	6	4	6
	|L|	No. of maximal potentially large	2000	2000	2000	2000

		itemsets				
	|N|	No. of Items	1000	1000	1000	1000
Entity 2	|D|	No. of Transactions	10K	10K	20K	20K
	|T|	Avg. Size of Transaction	4	7	5	8
	|I|	Avg. Size of maximal potentially large itemsets	2	4	4	6
	|L|	No. of maximal potentially large itemsets	1000	1000	1000	1000
	|N|	No. of Items	300	300	300	300
Relationship R	|NR|	No. of Relationships	105690	100400	100779	109000
	σ	Average and standard deviation of no. of relationships for each tuple of entity 1	8,6	10,5	10,5	10,5
		Mean of second Entity 2	8	5	16	18
Outer Join		Number of Transactions	105690	100400	107898	109890

Table 1: characteristics of the Datasets Execution Time required

The first experiment examines how the runtime changes when the minimum support value and the database size change. Tables 2 and 3 below show the execution times of Extended Apriori Star and APRIORI JOIN STAR under different minimum supports.

Table 2: Time Required for Execution of EAS

MinSup	D1 (minutes.seconds)	D2 (minutes.seconds)	Join(D1) (minutes.seconds)	Join(D2) (minutes.seconds)
Extended Apriori Star results				
0.5	5.2	20.44	5.7	21.1
0.4	6.8	30.45	7.2	30.9
0.3	7.4	31.33	7.8	32.00
0.2	10.5	38.67	10.9	39.21
0.1	18	72.9	18.9	73.3

Table 3: Time Required for Execution of AJS

MinSup	D1 (minutes.seconds)	D2 (minutes.seconds)	Join(D1) (minutes.seconds)	Join(D2) (minutes.seconds)
APRIORI JOIN STAR results				

0.5	5.1	20.23	5.5	21.00
0.4	6.47	27.23	6.89	27.9
0.3	7.2	28.04	7.7	29.00
0.2	10.00	37.28	10.9	37.9
0.1	17.6	68.49	18.00	70.4

It can be observed that the runtime increases linearly with the size of the database [2][3]. This linear relationship shows that the proposed algorithm scales well. More time was required to compute the membership degrees. Since these were fuzzy taxonomies, which are more complicated than the original exact taxonomies, the time required to generate the extended transaction set was higher and the average length of each extended transaction was much longer than that of the Crisp case. From Table 2 and Table 3, the execution times of the Extended Apriori Star (EAS) and APRIORI JOIN STAR (AJS) algorithms increase as the minimum support decreases because the total number of candidate itemsets increases and the density of the distribution of frequent itemsets is sparse at high supports, resulting in few frequent itemsets with supports near minsup. The diagrams for this are shown in Fig. 1 and Fig. 2.

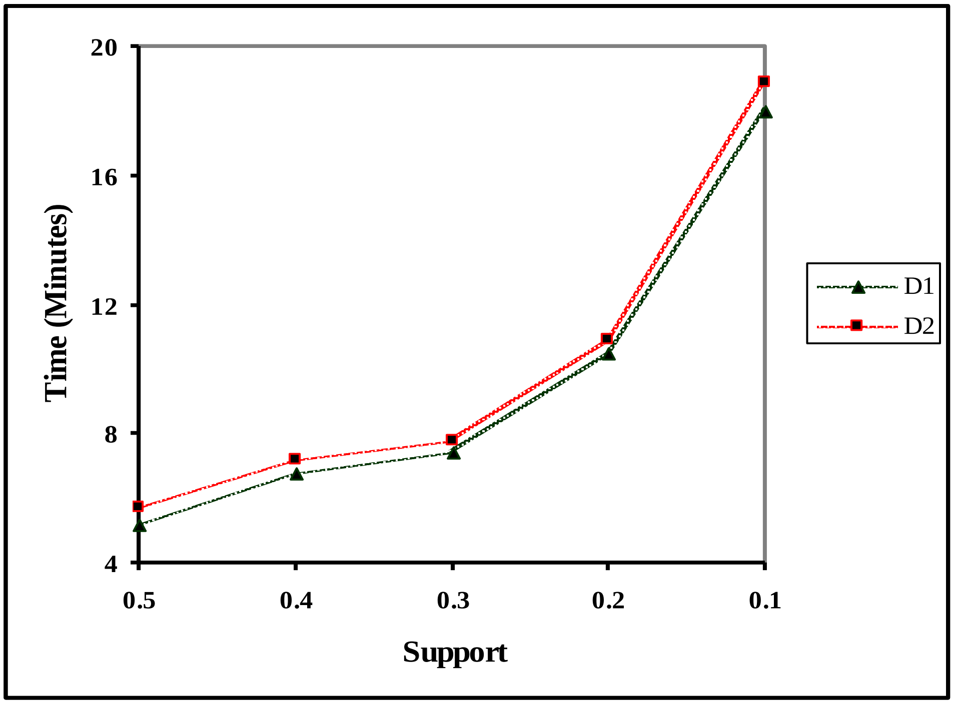

Fig 1 Time Requirement for EAS as the database increase and minimum support decrease

As can be seen from the diagram in Fig. 1, the minimum support varies between 0.5 and 0.1, and the extended Apriori star algorithm works with two data sets D1 and D2, where D2 is the concatenation of data set D1. From the diagram, it can be seen that the extended Apriori Star algorithm took 5 minutes and 2 seconds when the minimum support was 0.5 and the dataset D1 was used, and as the minimum support decreased by 0.1% at each level, the execution time continued to increase. With a minimum support of 0.1, the EAS algorithm took 18 minutes to generate the rules. Similarly, for data set D2, the time required was 5 minutes and 7 seconds at 0.5 and 18 minutes and 9 seconds at 0.1.

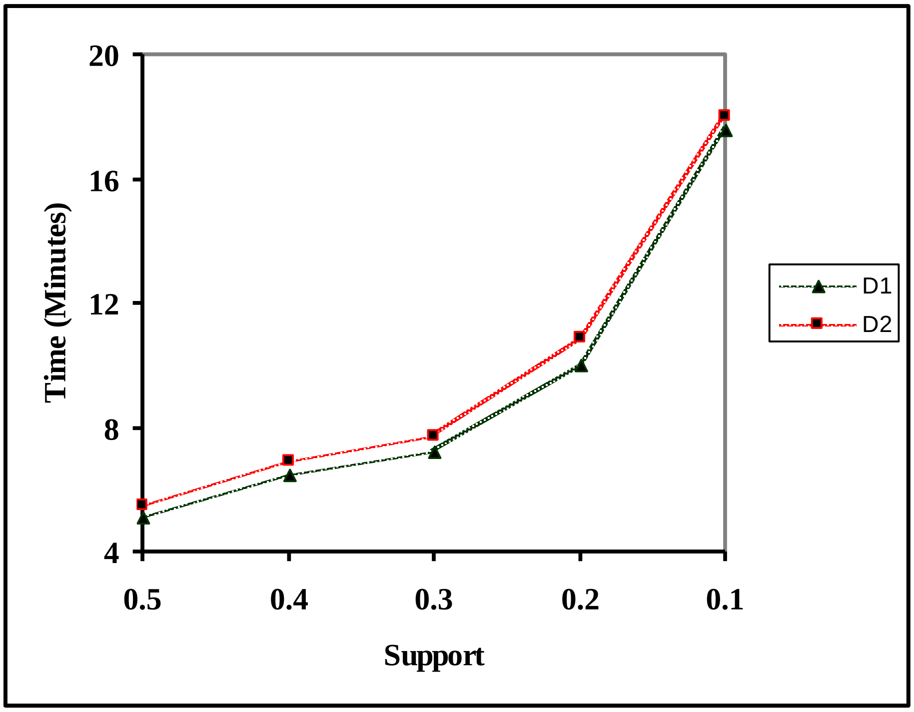

Fig 2 Time Requirement for AJP as the database increase and minimum support decrease

APRIORI JOIN STAR works only on the outer join of the entity tables, which are much larger than the join table of the Extended Apriori star, so the difference in I/O operations of both algorithms is very small

.

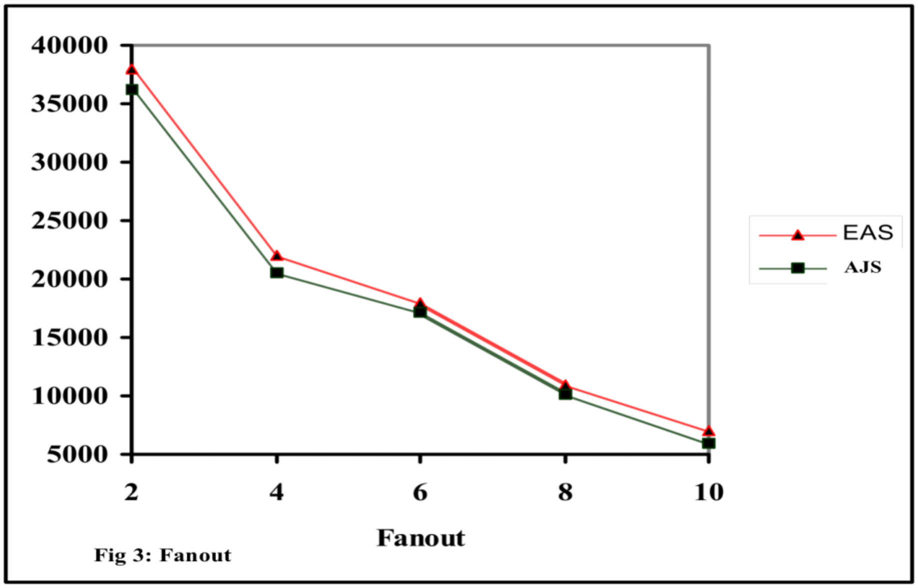

Fig 3: Fanout

In Figure 2, the Apriori Join Star algorithm is applied to the same two data sets, D1 and D2, again varying the minimum support from 0.5 to 0.1. Again, it can be seen that AJP takes less time with higher support and more time with lower support.

APRIORI JOIN STAR operates slightly faster than Extended Apriori Star in terms of execution time required. The reason is that Extended Apriori Star algorithm scans each entity table separately and also forms the concatenation of all tables and then scans them. In the experiments, all the entity tables are involved in the relation table, so the Extended Apriori Star algorithm finds frequent join itemsets in addition to the candidate 1 itemsets in all the other steps and therefore has to create the join of all the tables very frequently on the fly, so the I/O operations are higher in the case of Extended Apriori Star.

The fanout was varied from 2 to 10, which resulted in a decrease in the number of levels of taxonomic structures and also decreased the computation time for taxonomic structures. As the fanout increases, the computation of degrees of membership decreases for both algorithms as the number of levels of taxonomic structures decreases. The number of frequent itemsets generated by both algorithms also decreases sharply as the fanout increases, reducing the total time required to generate the final rules of interest, as shown in the diagram in Figure 3.

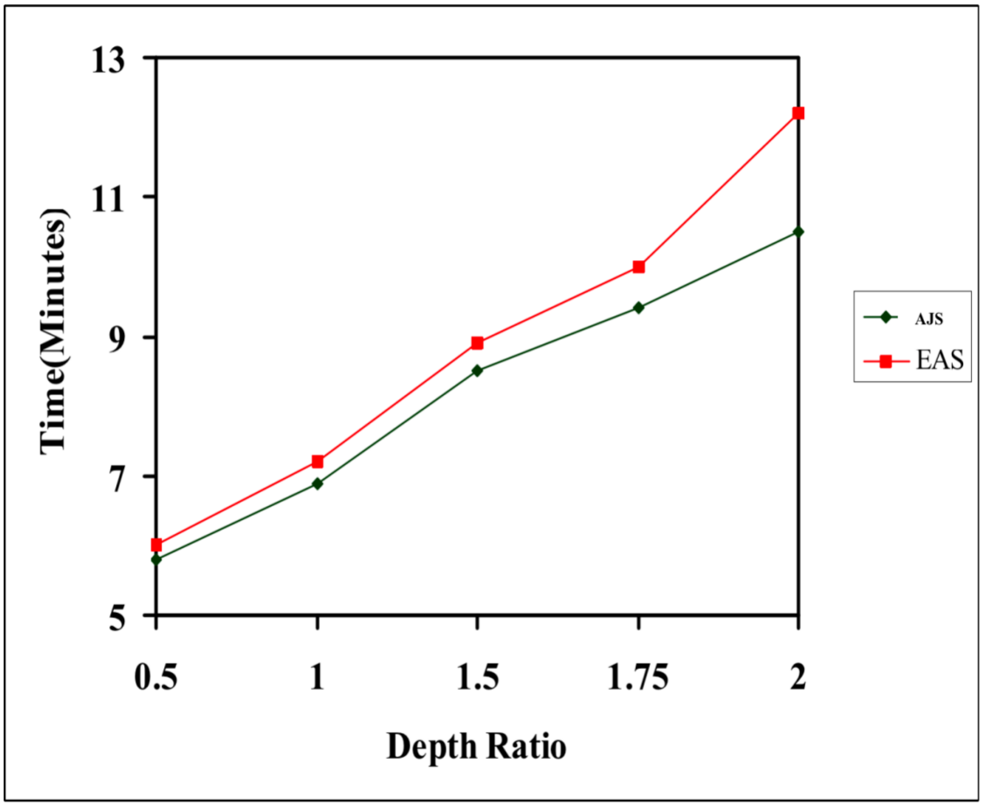

Fig 4: Depth Ratio

Depth ratio: the depth quotient was changed from 0.5 to 2. At a high depth quotient, items in frequent itemsets were selected from the leaves or lower levels of the concept hierarchy, while at a low depth quotient, items were selected from higher levels of the concept hierarchy. At a depth ratio of 2, APRIORI JOIN STAR performed about 5% better than Extended Apriori Star. The reason is that APRIORI JOIN STAR was able to remove a larger number of itemsets from the candidate collection at high depth ratios. The corresponding diagram is shown in Figure 4.

The graphs shown in Figs. 1 and 2 clearly indicate that the discovered algorithm successfully generates the rules from the synthetically generated datasets and that the algorithms scale linearly with the size of the input database. The graphs also show that the algorithm APRIORI JOIN STAR is a clearer winner than Extended Apriori Star in terms of temporal performance in the case of entity-relationship models, although no large difference in performance is observed.

This is also seen in Fig 3 where APRIORI JOIN STAR again performs better than Extended Apriori Star when the fanout is increased from 2 to 10.

Memory Utilization

In the experiments, the total number of elements, N, was set to 100K. This environment presents an extremely stressful situation for the Extended Apriori Star and APRIORI JOIN STAR algorithms in terms of memory utilization due to the very large number of elements. Figure 5 shows the memory usage of Extended Apriori Star as a function of support for N = 100K for the D1 database and N=200K for the outer join database (D1).

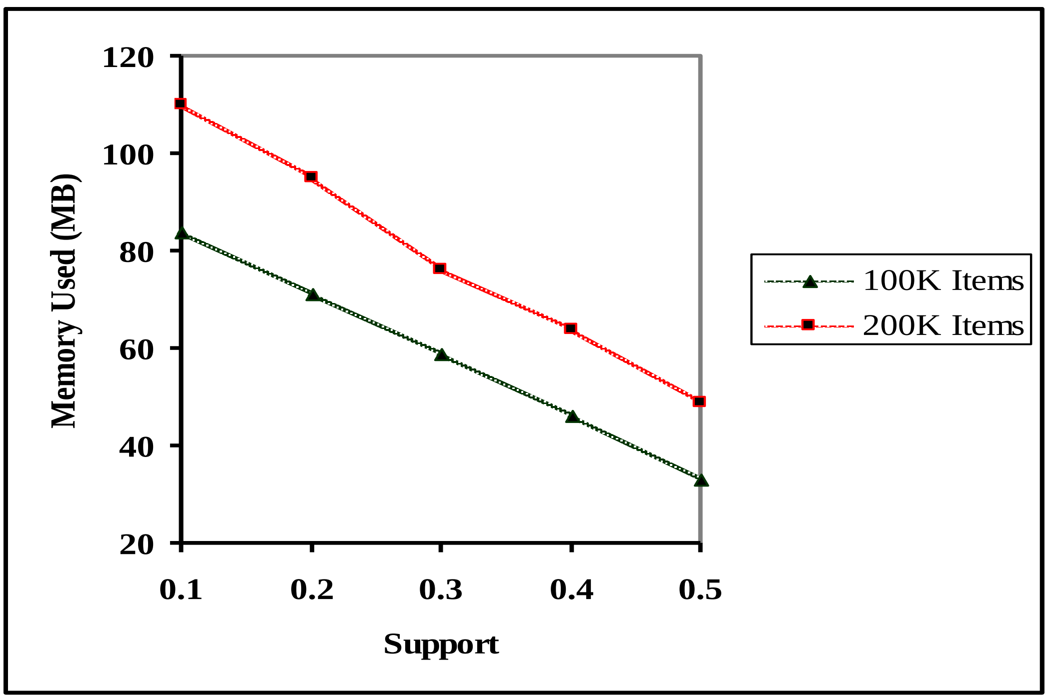

Fig 5: Memory utilization in Extended Apriori Star

The main memory usage of both algorithms scales well with the number of elements. At the threshold of 0.1% support, the memory consumption of APRIORI JOIN STAR for N = 100,000 elements was 84 MB, while for N = 200,000 elements it was 110 MB; an increase of less than 26% when the number of elements is doubled. The reason for this is that Extended Apriori Star's main memory usage does not depend directly on the number of elements, but only on the size of the output.

4.2 Performance on Real World Data Set

To verify whether the results from the synthetic data also apply to real data sets, the two newly discovered algorithms were applied to real data sets from a supermarket.

Supermarket data: This is data on grocery purchases by customers. A total of 5000 items are involved. The taxonomy has 3 levels with 29 roots. There are about 10000 tuples in entity 1 and 1000 tuples in entity 2 and 10523 tuples in the relationship table, with an average of 6 items per transaction. Figure 6 shows the time required by both algorithms when the minimum support is reduced from 3% to 0.5%. These results are similar to those obtained with synthetic data, with APRIORI JOIN STAR being a little faster than Extended Apriori Star.

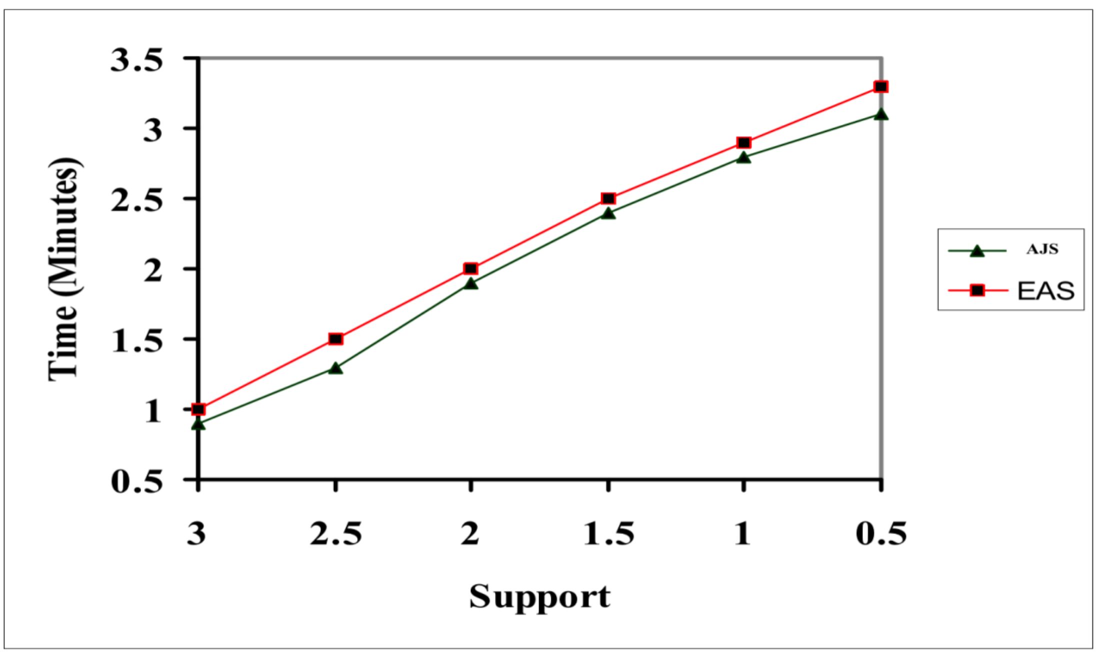

Fig 6: Performance on real datasets

Aiming to address the uncertainty in multi-level association rule mining from entity-relationship models, the chapter has focused on analyzing the performance of the two discovered algorithms EAS and AJS on synthetically generated data as well as on real datasets. Although uncertain data exists in practical databases, no mining algorithms have been developed to discover multilevel rules from uncertain data that exist in star schemes. To address this problem, the proposed algorithms were tested on datasets with different parameter settings and with different sample sizes. The results of the two algorithms showed that Extended Apriori Star was computationally more complex than APRIORI JOIN STAR, but not significantly, and that fuzzy rules from multiple tables were more efficient with APRIORI JOIN STAR than with Extended Apriori Star. Both algorithms scale linearly with the size of the input database and efficiently process the fuzzy data of the star schema by generating multi-level linguistic rules from multiple tables designed using an entity-relationship model, which was the main objective of the study. The experimental results also show that the algorithms are efficient even on computers with less main memory and prove the feasibility of the proposed mining algorithms.

For our experiments with brain injury patients, the study uses the Glasgow Coma Scale (GCS) [27] and injury severity score [15][29], which are commonly used in injury severity assessment. The GCS is a neurological scale designed to reliably and objectively assess a person's state of consciousness at both initial and follow-up assessment. The patient is assessed using specific criteria on the scale, which ranges from 3 (for the state of profound unconsciousness) to 14 (for the original scale) or 15 (for the revised scale).

The Glasgow coma scale is shown in figure 8. Electroencephalography (EEG) is also used in the study. Figure 7 shows the brain images recorded by EEG in the form of a graph.

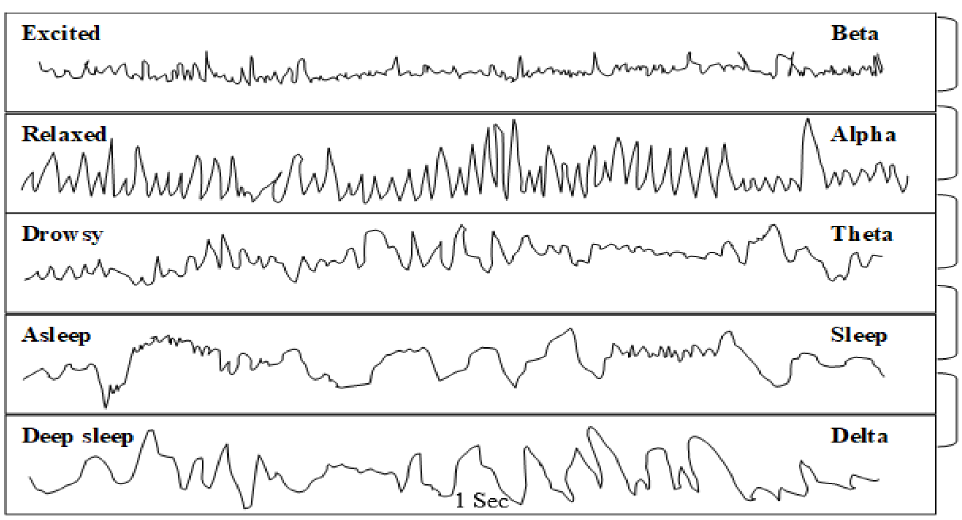

Fig 7: EEG Images

Although medical knowledge, concerning the symptom-disease relationship constitutes one source of imprecision and uncertainty in the diagnostic process, the knowledge concerning the state of the patient constitutes another. The physician generally gathers knowledge about the patient from the past history, physical examination, laboratory tests results and other investigative procedures such as X-ray, ultrasonic, CT scan and MRI. The knowledge provided by each of these sources carries with it varying degrees of uncertainty. The past history offered by the patient may be subjective, exaggerated, under estimated, or incomplete. Mistakes may be made in the physical examination and symptoms may be overlooked. To provide better precision of laboratory tests, X-rays and other similar procedures require a correct interpretation of the results.

Activity	1	2	3	4	5	6
Eyes	Does not open eyes	Opens eyes in response to painful stimuli	Opens eyes in response to voice	Opens eyes spontaneously	N/A	N/A
Verbal	Makes no sounds	Incomprehensible sounds	Utters inappropriate words	Confused, disoriented	Oriented, converses normally	N/A
Motor	Makes no movements	Extension to painful stimuli (decerebrate response)	Abnormal flexion to painful stimuli 2 (decorticate response)	Flexion / Withdrawal to painful stimuli	Localizes painful stimuli	Obeys commands

Fig 8: Performance on real datasets

The physician's medical knowledge is represented as a fuzzy relation between symptoms and diseases [20][21]. Given a fuzzy set A of symptoms observed in the patient and the fuzzy relation R representing the medical knowledge linking the symptoms in the set S to the treatment in the set T, the set C of possible conditions of the patient can be derived using the composition rule of inference for all $t \in T$.

$C = A^{\circ} R$ or $\mu C(t) = \max[\min(\mu A(s), \mu R(s,t))] s \in S$

This max-min composition corresponds to the fuzzy conditional statement when A represents C by R. The degrees of membership of the observed symptoms in the fuzzy set A, as shown in Figure 9, can represent the degree of certainty of the presence of the symptom or its severity. The set C denotes the degree of certainty with which a particular condition of the patient can be analyzed.

It should be noted that not all symptoms, treatments, and the conversion of symptoms and the conversion of the physician's medical judgment into fuzzy quantities are included in the study. A very small part of these things is included in the study, which will help to explain the role of the study in a simple way.

Vol 12 Issue 03 2023

ISSN NO: 2230-5807

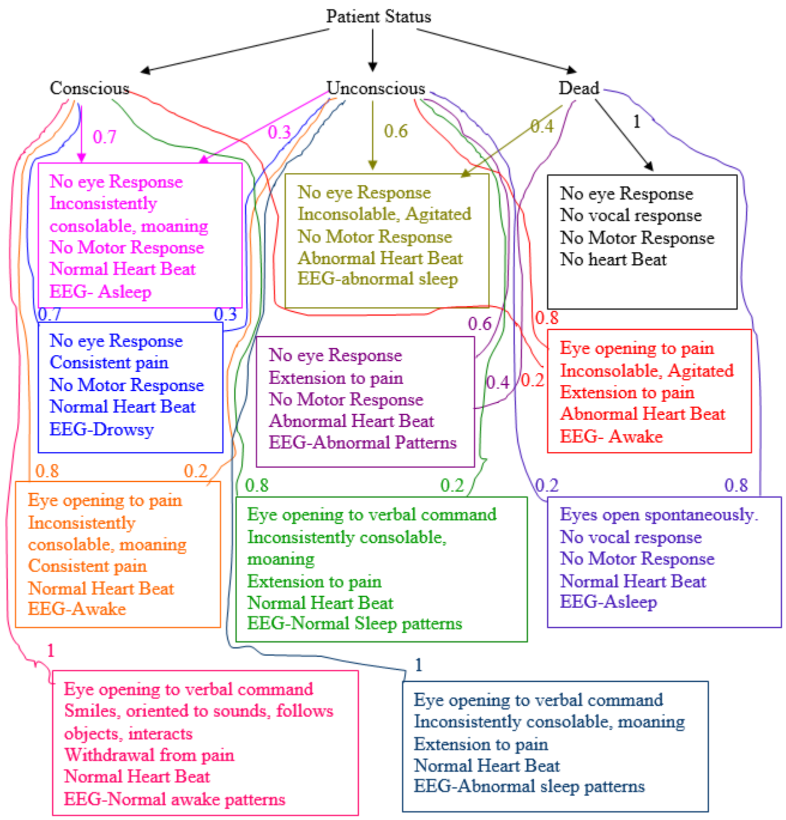

Fig 9: Fuzzy taxonomic structure over status of the Coma Patient

A= {Conscious, Unconscious, Dead}

S= {electrolyte imbalance, hydrocephalus, skeletal or muscular deformities, seizure, obtundation, stupor}

T= {renal dialysis for electrolyte balance, ICP, diuretics, surgery, cardiovascular support, hydration, antibiotics, TMS}

Fuzzy membership values for the representation of different symptoms are established and used to form the relationships. The taxonomic structure resulting from the set A is shown in Figure 9.
For the implementation of the discovered algorithm, a database is prepared for the study, which contains entity tables - patient, symptoms, medical examination, treatment, Glasgow score - and a relationship table patient status. All tests - e.g., all medical tests and procedures performed at any stage - are included in the Medical Tests table. It should be noted that there are two categories of symptoms: one that can be seen and another that can only be observed through the medical reports. Fuzzy set A

ISSN NO: 2230-5807

consists of the symptoms that can be seen and set B consists of the symptoms that can only be observed after receiving the medical reports. It is assumed that except for PID, age, and tag, all attributes in all tables are of character type.

The values under each character type attribute are descriptive and the description can be different under each tuple depending on the type of attribute. For example, ICP level for a more serious patient will be more than a patient who is less serious.

For the experiment, the details of 50 patients were taken. After forming the fuzzy sets and membership values in the fuzzy taxonomic structure, the Extended Apriori star and the newly discovered algorithm Apriori Join star algorithm is applied on the above mentioned database and the clinical spectrum of the patients is used in the experiment. A sample record of one patient is shown in Fig 10

The series of steps followed in summarized way to mine rules from the medical example are:

1) Fuzzy taxonomies structures formed as shown in Fig 9.
2) Degrees of membership calculated to form extended tables.
3) join of table to form table Join and compute entity and join support.
4) filter out itemsets whose either entity or join support is greater than minimum support threshold.
5) This process keeps on repeating until the candidate set is empty.
6) The degree of confidence is then calculated as mentioned above and the itemset whose degree of confidence is greater than or equal to minimum confidence finally forms the fuzzy generalized association rule for multiple table.

PID	Age/ Gender	Diagnosis	Cause	Complication	Initial GCS	Duration	Worst Physical Findings during Coma	Work-up Results During coma	Rate of recovery	Final GCS	Ultimate GOS	Treatment
1	31/Male	Encephali tis	Cytom calovir us	Polyradiculo neuritis	6 (2-3-1)	1 year	Spontaneous Breathe, No eye Response Inconsolable, Agitated No Motor Response Abnormal Heart Beat decortication, Sensorineural hearing loss, pH<3, BP-100/60, age-36, nocontrol-17 days	Diffuse theta on EEG, cerebrospi nal fluid, cells 25/3, protein 98 mg/dl, lymphoma originating from retroperito neum, cerebral deformity on MRI	Weak recovery	8	2	Ganciclovi r, High anticy-Tomegalo-virustiter Immuno-globulin

Fig 10: Sample record of Coma Patient

The rules that are given in Table 1, shows the condition analyses of a patient. The condition is picked up from Fig 10 which is the description of the attribute rating of table 1. For this example the minimum support threshold was taken as 30% and minimum confidence threshold was taken as 60%.

Coma patients Exact Outcome Rules	DSupport	DConfidence	Method
{Hydrocephalus, No eye Response Inconsolable, Agitated No Motor Response Abnormal Heart Beat EEG-abnormal sleep, pH<3, BP-100/60, age-36, nocontrol-17 days}→ Severe disability	66.6%	78%	Apriori Join star(AJS) with the support of EEG, GCS
{Seizure,Eye opening to verbal command Inconsistently consolable, moaning Extension to pain Normal Heart Beat EEG-Abnormal sleep patterns, pH-5, BP-110/70, age-59, control-30 days} → Moderate Disability	62%	76%	AJS with the support of EEG, GCS
{No hyperventilation, Eye opening to verbal command Inconsistently consolable, moaning Extension to pain Normal Heart Beat	63.4%	71%	AJP with the support of EEG, GCS

EEG-Normal Sleep patterns, pH-7, BP-11/60, age-44, control-15 days}→ Good Recovery			

Table 1 Outcome

The results were compared with the findings of neurologists. It was found a significant relationship between the findings of neurologists and systems output for normal, mild and severe electroencephalography tracing data. Getting this system in routine use will facilitate to make a rapid decision for the degree of trauma with electroencephalography.

Conclusion

This paper is devoted to the association rules mining, which belongs to one of the most frequently used data mining technique. It is very important to see that even after many years of the introduction of association rules, new algorithms continue to appear in this area. The paper is focused on the extension of association rules mining technique beyond the original Boolean rules, quantitative rules, multilevel rules, fuzzy rules and multiple table rules.The main goal of the paper was to propose and describe an innovative algorithm that can mine rules from databases containing multiple tables with fuzzy data with concept hierarchy.

The study deals with fuzzy relational data models, with an objective to provide a generalized approach for treating multi level precise, as well as imprecise, data from star schema. Since one of the major objectives of fuzzy logic is to represent approximate reasoning used in natural languages, it is expected that in a database environment, appropriate blending of a relational data model and fuzzy logic will enhance the capabilities of the existing database systems. Studies of some of the existing proposals for using fuzzy logic in a relational database environment, studies of rule mining from multiple tables using ER models has also been presented. For a successful blending of fuzzy logic and databases using ER models containing data with concept hierarchy, it was, however, essential to develop a suitable design technique for such systems.

In this paper, the problem of mining fuzzy association rules in databases consisting of several tables organized in a schema, obtained from an entity-relationship design has been approached. This thesis has introduced the fuzziness in the underlying taxonomic structures and extended the classical algorithm in a way that a transaction may partially support a particular item. This has then led to reexamining the computation for the degree of support and the degree of confidence. Furthermore, the classical Apriori algorithm has been extended to incorporate the extended notions of Dsupport and Dconfidence. The fuzzy extensions presented in this paper enable us to discover not only crisp generalized association rules but also fuzzy generalized association rules when databases consisting of several tables organized in a schema within the framework of fuzzy taxonomic structures.

The EAS (Extended Apriori Star) and proposed AJS (APRIORI JOIN STAR) algorithm challenges some of the problems connected with fuzzy association rule mining for multiple tables. The study analyzes how the attributes of several entities appear together and generate rules with respect to the relationships existing between the entities and their ancestors. Strong association rules between items of fuzzy nature existing in multiple tables can be calculated that will undoubtedly help the managers of the Supermarkets in designing their shelf space. Thus algorithm was tested on synthetically generated databases and the results of the experimental tests show that algorithms scale linearly with respect to the size of the input database. The findings from both the algorithms have revealed that Extended Apriori star was more computational complex than APRIORI JOIN STAR and that fuzzy rules from multiple tables with APRIORI JOIN STAR was more efficient than with Extended Apriori Star. The experimental results also show that the algorithms are efficient even on the computers having less main memory and prove the feasibility of the proposed mining algorithms.

The major contributions of the work are listed below:

- Developing algorithm as a modified version of Extended Apriori, Apriori Star and Apriori Join algorithms to mine multi level fuzzy rules from ER Models.
- The developed algorithms analyze how the attributes of several entities appear together.
- The developed algorithms analyze rules with respect to the relationships existing between the entities and their ancestors.
- Developed algorithms have the added feature of redundancy control.
- Proposing the algorithms that can work with computers having less main memory.
- Evaluating the performance of the proposed algorithms on synthetically generated datasets with less number of items in the transactions and with more number of items in the transactions, different depth ratios, different fanout and decreasing minimum supports.
- Evaluating the performance of the algorithms on real-life datasets with several metrics.

References

1. Agrawal, R., and Srikant, R. (1994, September). Fast algorithms for mining association rules. In Proc. of the VLDB conference, Santiago, Chile. Expanded version available as IBM Research Report RJ9839, June 1994.
2. Agrawal, R., Imielinski, T., and Swami, A. (1993, May). Mining Association Rules between sets of items in large databases. In Proc. of the ACM SIGMOD Conference on Management of Data, pp. 207-216, Washington, D.C.
3. Bakk Lucas Helm. Master Thesis on Fuzzy Association rules - An implementation in R. Vienna 2007.
4. Chen, G., and Wei, Q. (2002). Fuzzy association rules and the extending Mining Algorithms, Information Sciences: An International Journal, 147, pp.201-228.
5. Chen, G., Wei, Q., and Kerre E. (2000). "Fuzzy Data Mining: Discovery of Fuzzy Generalized Association Rules". In Bordagna and Pasi (eds.), Recent Research issues on Management of Fuzziness in Databases, Physica-verlag (Springer).
6. Cristofor, L., and Simovici, D. (2001-2002). "Mining Association Rules in Entity-Relationship Modeled Databases", Technical Report, UMB.
7. Gottwald, SeigFried. Universes of Fuzzy Sets and Axiomatizations of Fuzzy set theory. StudiaLogica volume 82, Number 2, March, Springer, 2006.

8. Han, J., and Fu, Y. (1995, September). Discovery of Multiple-level Association Rules from Large Databases. Proceedings of the 21st International Conference on VLDB, Zurich, Switzerland.
9. Houtsma, M., and Swami, A. (1993, October) Set-oriented mining of association rules. Research Report RJ 9567, IBM Almaden Research Center, San Jose, CA.
10. Kuok, C.H., Fu, A., and Wong, M.H. (1998). Mining Fuzzy association rules in databases ACM SIGMOD Record, 27(1), ACM Press.
11. Srikant R. and Agarwal R. (1996). Mining quantitative association rules in Large Relational Tables, in proceedings of the ACM SIGMOD International Conference on Management of Data, pp 1-12, Montreal, Quebec, Canada.
12. Srikant, R., and Agrawal R. (1995). Mining Generalized Association Rules, proceedings of the 21st VLDB Conference, Zurich, Switzerland.
13. P Arora, RK Chauhan, A Kush, " Association Rule Mining for Multiple Tables With Fuzzy Taxonomic structures ", International Journal of Computer Theory and Engineering 2 (6), 866
14. P Arora, P Gandhi, G Sharma, S Saxena, "Study about Rule Mining for Multiple Tables with Fuzzy Data", Novel Research Aspects in Mathematical and Computer Science Vol. 1, 38-47, 2022
15. P Arora, RK Chauhan, A Kush "Frequent Itemsets from Multiple Datasets with Fuzzy data", International Journal of Computer Theory and Engineering vol. 3, no. 2, pp. 255-260, 2011.
16. P Arora, RK Chauhan, A Kush "Mining fuzzy generalized association rules for ER models", International Journal of Information Technology and Knowledge Management (IJITKM), 2008, vol-1, issue-2, pp 191-198.
17. Özkan, T., &Baykara, A. (2010). Fuzzy logic-based medical decision system for diagnosis of pulmonary embolism. Expert Systems with Applications, 37(12), 8385-8391.
18. Praveen Arora, SanjiveSaxena, Deepti Chopra, "Generalized Association Rules for ER Models by Using Mining Operations on Fuzzy Datasets" Recent Progress in Science and Technology Vol. 6, March 2023, DOI: 10.9734/bpi/rpst/v6/5539A
19. Teasdale G, Jennett B. "Assessment of coma and impaired consciousness. A practical scale". Lancet 1974,2:81-84. PMID 4136544, 1974.
20. Teasell R, Bayona N, Lippert C, et al. Post traumatic seizure disorder following acquired brain injury 2007;21(2):101-14
21. Trouiller P, Mantz J. Early management of traumatic brain injury. La Lattre du Neurologue 2016;10:55-60
22. Arora, P., Saxena, S., Madan, S., & Joshi, N. (2022). Frequent Itemsets: Fuzzy Data from Multiple Datasets. Novel Research Aspects in Mathematical and Computer Science Vol. 5, 47–57. https://doi.org/10.9734/bpi/nramcs/v5/6710F
23. Arora, P. . (2023). Mining Rules for Head Injury Patients Using Fuzzy Taxonomic Structures. Research Highlights in Disease and Health Research Vol. 5, 146–156. https://doi.org/10.9734/bpi/rhdhr/v5/6007A