A Novel RFID Data Mining System: Integration Of Effective Sequential Pattern Mining And Fuzzy Rules Generation Techniques

[1]Barjesh Kochar and [2]Rajender Singh Chhillar

GJCST Computing Classification
I.2.3 ,H.2.8

Abstract- **Data warehousing and Data mining find enormous applications; RFID technology is one among them. A RFID data warehousing system with novel data cleaning, transformation and loading technique has been proposed in the previous work. The system has been dedicatedly implemented in one of the significant RFID applications tracking of goods in warehouses. The warehoused RFID data is in specific format and so an effective mining system is required to mine the needed information from the database. The existing mining algorithms are inefficient in extracting the information from the warehoused RFID data. In this paper, a novel data mining system is proposed, which effectively extracts the information regarding the nature of movement of the RFID tags. The proposed mining system generates an intermediate dataset (Idataset) from the warehoused dataset. From the I-dataset,sequential patterns are mined with different pattern length combinations. From the mined sequential patterns, fuzzy rules are generated, which depicts the nature of movement of the RFID tags. The implementation results show that the proposed mining system performs well by extracting the significant RFID tags and its combinations and the nature of movement of the tags.**

Keywords- Data mining system, sequential pattern mining, RFID data, I-dataset, fuzzy rules, RFID readers, RFID tags.

I. INTRODUCTION

Enormous data collection has been possible throughout the previous decades as a result of the recent development of information, and availability of low-priced storage. The ultimate intention of this huge data collection is for the utilization of this information to realize competitive benefits, by analyzing the data i.e., determining formerly unidentified patterns in data that can direct the process of decision making [1] [17]. Conventional data analysis methods are generally based on direct handling of the data by a person and are not extendable to large data sets.Fundamental tools are available in database technology for efficient storage and searching of large data sets. But,helping humans to analyze and understand large masses of data continues to be a challenging and unresolved issue. New methods and intelligent tools offered by the upcoming data mining field assure to meet these challenges. The term data mining, which is also called Knowledge Discovery in Databases (KDD), is defined as ―The non-trivial extraction of implicit, previously unknown, and potentially useful information from data" [2] [15].

Data mining [3] as a multidisciplinary united effort from databases, machine learning, and statistics, is winning in turning masses of data into small valuable pieces. The ultimate objective of a data mining task in a real-world application might be e.g. to permit a company either to enhance its marketing, sales, and customer support operations or to identify a fraudulent customer through better understanding of its customers. Data mining methods have been effectively applied in various fields including marketing [22], manufacturing, process control, fraud detection [21], bioinformatics, information retrieval,adaptive hypermedia, electronic commerce and network management [4]. Broadly, the data mining tasks can be categorized into two types: Descriptive mining and Predictive mining [11]. The descriptive mining denotes the method in which the fundamental characteristics or common properties of the data in the database are portrayed. The method of predictive mining figures out patterns from the data so that predictions can be made. The predictive mining methods include tasks like Classification, Regression and Deviation detection.

Mining of information from a huge database finds many latest and emerging applications. Radio Frequency Identification (RFID) is one among the fields that incorporate the sequential pattern mining in RFID database. Radio Frequency Identification (RFID) is a high-speed, realtime, precise information gathering and processing technology, which distinctively identifies object by radiofrequency signal [14]. RFID technology can help a broad variety of organizations and individuals, for instance, hospitals and patients, retailers and customers, and manufacturers and distributors all through the supply chain to accomplish considerable productivity gains and efficiencies [23]. It is also extensively applied in logistics, traffic management, medical treatment, national defense,mining, and endlessly infiltrated to new fields [12]. The heart of an RFID system is the ―RFID tag" which encloses an integrated circuit and a transponder for data storage and transmission [25]. The tag can be tied to an object and be read out with reader hardware either in the form of a handreader or a RFID gate [24]. The tags are radically different from printed barcodes in their capability to hold data, at which range the tags can be read, and the nonexistence of line-of-sight constraints [26].

The RFID technology and its applications identified as one of the most hopeful technologies of the century are in a slow maturation period. Warehousing the RFID data and mining information from the warehoused RFID data are the major

---

1Barjesh Kochar and 2Rajender Singh Chhillar

issues that persist in the technology. In the previous work, we have proposed a novel data cleaning, transformation and loading technique to warehouse the RFID data efficiently. The warehoused data is in a specific format and so mining the required information needs an effective mining system.But, the existing mining algorithms are ineffective in their way of mining the information from the warehoused RFID data. To overcome this issue, in this paper, we have proposed an effective data mining system to mine the warehoused RFID data. In the proposed system, firstly, an intermediate dataset is generated from the warehoused RFID dataset. Secondly, a sequential pattern mining technique is proposed to mine the reader pattern from the intermediate dataset. Finally, fuzzy rules are generated based on the mined sequential pattern to make the decisions from the path followed by the RFID tags. The rest of the paper is organized as follows: Section 2 gives a brief introduction about the Sequential Pattern Mining and Section 3 makes A short review about the recent literary works. Section 4 details the proposed RFID data mining system with required mathematical formulations and tabulations. Section 5 discusses about the implementation results and Section 6 concludes the paper.

II. Sequential Pattern Mining

Mining Frequent Itemsets from transaction databases is A basic task for several types of knowledge discovery such as association rules, sequential patterns, and classification [5].Sequential pattern mining, which finds out frequent patterns in a sequence database, is a significant issue among the various data mining issues [7]. Discovering sequential patterns from a huge database of sequences is a significant issue in the field of knowledge discovery and data mining [20]. The sequential pattern mining problem was first presented by Agrawal and Srikant [19]. Mining Sequential Patterns in huge databases has become a significant data mining task with wide range of applications [9] [18].The aim of sequential pattern mining is to discover all frequent sequential patterns with a user-specified minimum support [10].The sequential pattern is a prearranged list of frequent itemsets in a customer‘s sequence and discloses the customer‘s purchasing manners in transactional databases, from which the policy-maker may make a knowledgeable business decision [16]. It extracts patterns that occur more frequently than a user-specified minimum support while preserving their item occurrence order [9]. It is helpful in various applications for enhancing the quality of analysis, including the analysis of web surfing, transactional customers‘ behaviors, network alarm patterns, business analysis, web mining, security and bio-sequences analysis [16]. For instance, supermarkets frequently gather customer purchase records in sequence databases in which A sequential pattern would signify a customer‘s purchasing behavior. In such a database, every purchase would be denoted as a set of items purchased, and a customer sequence would be a sequence of such itemsets [8]. The issue of sequential patterns discovery was motivated by retailing industry problems. Nevertheless, the results apply to several scientific and business domains, like stocks and markets basket analysis, natural disasters (e.g. earthquakes), DNA sequence analyses, gene structure analyses, web log click stream analyses, and so on [13]. In this task, time is the most significant factor, particularly when the results are required in a limited period of time [6].

III. Related Works

Jigyasa Bisaria et al. [27] have put forward a rough set orientation to the problem of constraint driven mining of sequential pattern. They have divided the search space of sequential patterns utilizing indiscernibility relation from the theory of rough sets and they have put forward an algorithm that permitted pre-visualization of patterns and imposition of diverse types of constraints in the mining task. The algorithm C-Rough Set Partitioning was more than ten times faster than the naive algorithm SPRINT that was founded on different types of regular expression constraints. Shigeaki Sakurai et al. [28] have detailed that all frequent sequential patterns included in sequential data are efficiently found out by the sequential mining methods. Those methods assessed the frequency by utilizing the support, which is the previous criterion that satisfied the Apriori property.Nevertheless, since the patterns were general and the analysts cannot acquire new knowledge from the patterns, the discovered patterns do not always correspond to the interests of the analysts. To find out sequential patterns that are more appealing for the analysts, they have put forward A criterion, namely, the sequential interestingness. They have illustrated how the criterion is related to the support and that the criterion fulfilled the Apriori property. Moreover,founded on the proposed criterion, they have put forward an efficient sequential mining method. Finally, by applying the method to two types of sequential data, they have demonstrated the effectiveness of the proposed method. Themis P. Exarchos et al. [29] have offered a methodology with two process for sequence classification that utilizes sequential pattern mining and optimization. In the first stage, a sequence classification model, founded on a set of sequential patterns was defined and two sets of weights were set up, one for the patterns and the other for classes. In the second stage, the weight values were assessed by utilizing an optimization technique to accomplish optimal classification accuracy. By changing the number of sequences, the number of patterns and the number of classes, extensive appraisal was done on the methodology, and it has compared with similar sequence classification approaches. Chen [30] has found out a data structure, Up Down Directed Acyclic Graph (UDDAG), for efficient sequential pattern mining. Along both ends of detected patterns, UDDAG has permitted bidirectional pattern growth. Therefore, in $\log 2(k+1)$ levels of recursion, a length-k pattern can be identified, which has resulted in fewer levels of recursion and faster pattern growth. When minSup is large such that the average pattern length is close to 1, UDDAG and PrefixSpan have similar performance because the problem degrades into frequent item counting problem. Nevertheless, UDDAG scales up much better. In scalability tests, UDDAG often surpassed PrefixSpan by almost one order of magnitude. UDDAG was also significantly faster than Spade and LapinSpam. Barring extreme cases, UDDAG has utilized memory comparable to that of PrefixSpan and less than that of Spade and LapinSpam. Moreover, extension of UDDAG towards applications that involve searching in large spaces has been made possible due to its special feature.Jirachai Buddhakulsomsiri and Armen Zakarian [31] have offered a sequential pattern mining algorithm which has enabled the extraction of hidden knowledge from a huge automotive warranty database by product and quality engineers. Elementary set concept and database manipulating methods were utilized by the algorithm to look for patterns or relationships among instances of warranty claims over time. IF-THEN sequential rules represent these patterns, where the IF part has included one or more instances of warranty problems at one time, and the THEN part has included warranty problem(s) that happen at a later time. When the sequential patterns were generated, the algorithm utilized rule strength parameters to filter out insignificant patterns so that only significant rules were supported. The Significant patterns have provided knowledge of one or more product failures that leads to future product fault(s). Warranty data mining application from the automotive industry has demonstrated the effectiveness of the algorithm. For the automotive case, A discussion on the sequential patterns created by the algorithm and their interpretation were also given.Ya-Han Hu et al. [32] have illustrated multi-time-interval sequential patterns, a variation of time-interval sequential patterns, which can be disclosed the time-intervals between all pairs of items in a pattern to offer more time-related knowledge. As a result, two efficient algorithms, called the MI-Apriori and MI-PrefixSpan algorithms were developed by them to solve this problem. Experimental results have illustrated MI-PrefixSpan algorithm was faster than the MIApriori algorithm, however in long sequence data, MIApriori algorithm has better scalability.Ya-Han Hu et al. [33] have concentrated on mining sequential patterns in the business-to-business (B2B) environment. Utilizing traditional methods in the B2B environment may end up in numerous uninteresting and meaningless patterns and extensive computational time due to very long sequences of customers, and every customer often purchasing majority of the items. They have set up three conditions (constraints) to solve those problems, namely, compactness, repetition, and recency and consider them together with frequency in choosing sequential patterns. To find out frequent sequential patterns which have fulfilled the conditions, an efficient algorithm has been built up. Computational efficiency and effectiveness of their proposed method in the B2B environment in extracting useful sequential patterns has been confirmed by observed results.

IV. THE RFID DATA MINING SYSTEM

The RFID data has been warehoused effectively by means of a novel data cleaning, transformation and loading technique, which was dedicatedly proposed for RFID data. The warehoused data has been compressed and loaded in A specific format and so extraction of required knowledge from them is difficult. The proposed novel RFID data mining system effectively mines the required knowledge from the warehoused RFID data. The proposed system is comprised of three stages, namely, Generation of I-dataset, sequential pattern mining and generation of fuzzy rules. In the first stage, a dataset is generated from the warehoused RFID data. In the second stage, sequential patterns are mined using the proposed mining technique and in the third stage, fuzzy rules are generated from the mined sequential patterns. Prior to detail the proposed mining system, the format of the warehoused RFID data is briefed in the following sub-section for easy understanding of the proposed mining system.Let R_i ; $0 \le i \le N_R -1$be the number of RFID readers present and $_j T$; $0 \le j \le N_T -1$ be the RFID tags which are in movement, where, N_R is the number of readers and N_T is the number of Tags present in the warehouse. The tags may enter into any reader at any time while in process. Each RFID tag has its own Electronic Product Code (EPC), which represents the class to which $T\ j$ belongs. The Tag representation with EPC code can be given asT_{kj} ;$k \in [EPC_1, EPCN\mathrm{c}\ ]$, where,Nc is the number of product classes available.Generally, the tags are given I.D. only by their EPC code.But, here the tags are given numerical IDs along with the EPC code. As an example, assuming the Tag representation is T_{10} ,T_{11} ,T_{21} ,T_{32} ,T_{42} ,T_{52} , …., $-1\ NcNT\ T$, then,T_{10} represents the tag that belongs to the code class EPC_1 and code ID ‗0‘ and, T_{11} represents the tag that belongs to the code class $EPC1$ and code ID ‗1‘and so on. The EPC of a tag defines the product to which the tag is attached. For example, EPC_1 may be the code for the product ‗pen‘;EPC_2 may be the code for the product ‗calculator‘, and so on. The RFID readers may be present in any of the chambers of a warehouse and they monitor the movement of the RFID tags (i.e. every product present in the warehouse).Hence, when the tag moves from one location to the other, the Reader present in that location makes an entry in the database by its tag ID and the time of presence of the tag.Thus obtained RFID data has been cleaned, transformed and loaded effectively in the data warehouse by our novel data cleaning, transformation and loading technique, which was done in the previous work. An exemplary view of the warehoused data is given in the Table

Table I: An exemplary view of warehoused RFID data with three readers

S.No	Reader R_1		Reader R_2		Reader R_3	
	Tags IN	Tags OUT	Tags IN	Tags OUT	Tags IN	Tags OUT
1	[3,20,30]	[5,10]	[6,12]	[8,14]	[31,21]	[15,40]
2	[11,1,3,8]	[20]	[13,7,20]	[6,12,17]	[6,12,17]	[31, 19]
3	[27,25]	[30, 3, 11]	[3,11]	[13]	[30,1β]	[21]
4	[24,9]	[1,27]	[27,30]	[3,7]	[1]	[6,12,30]
5	[15]	[8,24,9]	[9,24]	[27, 11]	[27, 8]	[27]

The RFID data for only three readers are given in the table. In the Table I, the tags that entered into the corresponding reader‘s range and the tags that left out from the corresponding reader‘s range for some five interrogations are given. The interrogation period, the frequency of occurrence and the availability index are given in the Table

Table 2: The warehoused RFID data with interrogation period, frequency of occurrence and availability index for three readers

S. No	Reader R_1			Reader R_2			Reader R_3		
	Transformed Interrogation Time	Frequency of Occurrence	Availability Index	Transformed Interrogation Time	Frequency of Occurrence	Availability Index	Transformed Interrogation Time	Frequency of Occurrence	Availability Index
1	1.08	7	[3,5,6]	0.09	6	[41,34]	0.06	7	[14,15]
2	2.1	5	[8,10]	0.01	5	[21,12,13]	1.2	6	[23,1]
3	1.3	3	[16,3,9]	0.08	3	[1,4,5]	1.0	4	[1,7,10]
4	1.9	2	[15,17]	1.4	2	[4,2,7,8]	1.9	2	[35,16]
5	4.5	1	[24,20,1,4]	4.1	2	[23,15]	3.8	1	[9]

The proposed data mining system extracts the information of the flow of path for every tag. It generates rules based on the mined data so that the nature of each product (in terms of its RFID tag) can be understood well. To accomplish this, the transformed data present in the Table I is sufficient for the proposed system.

A. GENERATION of I-dataset

In the first process of the proposed data mining system, the warehoused data is processed and an I-dataset is generated.The I-dataset is generated by querying the warehoused dataset. The proposed system intends to mine the information in the dataset given in the Table I. Firstly, the dataset given in the Table I is mathematically represented for further querying process. This can be accomplished by representing the set of Tag IDs that are entering into the Reader‘s range and the set of Tag IDs that are leaving out from the Reader‘s Range from Table I. The set representation is given as $\{T^{IN}\}_{il}$ and $\{T^{OUT}\}_{il}$; $0 \le \le -1$ l N_1 where, N is the number of interrogations performed by the reader, such that $\{T^{IN}\}_{il} \subseteq \{R^{IN}\}_i$ and $\{T^{OUT}\}_{il} \subseteq \{R^{OUT}\}_i$. Hence, two sets, $\{R^{IN}\}_i$ and $\{R^{OUT}\}_i$, are generated, where, each set is comprised of N_1 subset. But, the cardinality of each subset present in $\{R^{IN}\}_i$ (here, $i \in (0, N_R - 1)$) need not be equal to each other. Also, the cardinality of $\{R^{OUT}\}_i$ need not be equal to each other. Each subset is comprised of the Tag IDs that entered and left out from the Reader. As $\{R^{OUT}\}_i$ $\{R^{IN}\}_i$ stated earlier, the set is queried to generate the Idataset Querying can be performed either over the set or over the set Here, querying is performed over the set, $\{R^{OUT}\}_i$ The I-dataset is comprised of all the RFID tags and the readers that cover the tags in their way. This can be obtained as follows

$$\{P(l)\}_j = \begin{cases} R_i\,; & if\ T_j \in \{T^{OUT}\}_{il} \quad \forall i \\ -1\,; & otherwise \end{cases} \tag{1}$$

$$P_j = P_j - \alpha \tag{2}$$

In Eq. (1), the condition is checked by varying for all i and the Path set {P} is determined. For every Tag, a set {P} is determined (i.e. *Pj for Tj*). Generally, some tags may be present in a reader‘s range for a long time (i.e. the tags may not appear in Tout for many interrogation cycles). In such circumstances, the set {P} holds an element of value -1 in the corresponding interrogation cycle. The filtered path set is obtained in the Eq. (2), where, $\alpha = \{-1\}$, in which the readers that covered the tags are present and then sorted in ascending order based on the Tag IDs.
Thus, by querying, (mathematically represented in Eq. (1) and Eq. (2)) the readers under which each tag has covered are collected and so the set {P} is determined and the Idataset is generated from the set {P}. For instance, the {P} is assumed *to be as:* $P_0 = \{R_1, R_3, R_5, R_{10}, R_{13}\}$,

$P_1 = \{R_2, R_4, R_1, R_6\}$, $P_2 = \{R_{11}, R_3, R_5, R_1, R_3\}$, $P_3 = \{R_7, R_9, R_5, R_7, R_{12}, R_4\}$ and $P_4 = \{R1, R9, R4\}$. The exemplary I-dataset determined from the aforesaid set $\{P\}$ is tabulated below.

Table 3: The exemplary I-dataset for four RFID tags

S.No	Tag IDs	Readers
1	T_0	$R_1, R_3, R_5, R_{10}, R_{13}$
2	T_1	R_2, R_4, R_1, R_6
3	T_2	$R_{11}, R_3, R_5, R_1, R_3$
4	T_3	$R_7, R_9, R_5, R_7, R_{12}, R_4$
5	T_4	R_1, R_9, R_4

The field ‗Tag IDs' present in the I-dataset, which is given in the Table III, is comprised of all the Tag IDs and the field ‗Readers' for a record named *T*0 is comprised of all the Reader IDs that covered the Tag *T*0 in its course. The Idataset is subjected to sequential pattern mining so that the required path information is extracted from the I-dataset.

B. Mining Sequential RFID Data Patterns

In the I-dataset and the set P, a generalized Tag ID has been used. But, practically, the Tag IDs are represented with EPC code. So, from this instant, the Tag IDs are grouped as per the product with which the tags are attached. To accomplish this, it is assumed that the tags which has first NP number of IDs belongs to *the EPC* code *EPC*1, the tags which has second NP number of IDs belongs to the EPC code EPC2 , and so on. Hence, a new tag ID representationTab ; $EPC_1 \le a \le EPC_{N_c}$ And $0 \le b \le N_{P-1}$ is given for $T_j;\ 0 \le j \le N_{T-1}$, where, b = j%NP and a= 1+ (j / NP) flow extracted from the set $\{P\}$ is given as

$$T_{ab} = \left(R_h^{(0)} \longrightarrow R_h^{(1)} \longrightarrow R_h^{(2)} \longrightarrow \cdots \longrightarrow R_h^{(|T_{ab}|-1)}\right) \quad (3)$$

where, hÎ[0, NR −1] and | Tab | is the size of the pattern length.. Then, from Tab , a single length (L = 1) sequential pattern is mined. The pseudo code for mining the single length sequential pattern is given in Fig. 1.

Initialize S, sup
for $a = 1$ to $N_c - 1$
    set sup_{a0} to 1
    set sup_{a1} to 0
    set x to 1
    Assign T_{a0} to S_{a0}
    for $b = 1$ to $N_P - 1$
        if $T_{ab} \in S$
            for $y = 1$ to $|S|-1$
                if S_{ay} is T_{ab}
                    increment sup_{ay}
                end if
            end for
        else
            Assign T_{ab} to S_{ax}
            increment sup_{ax}
            increment x
            set sup_{ax} to 0
        end if
    end for
end for

Fig 1: Pseudo code for mining single length sequential patterns from the Idataset

By mining as per the given pseudo code, the sequential pattern with L = 1(i.e. for every individual EPC code) is obtained from the pattern set {S}. Each pattern has its own support, i.e. frequency of occurrence of the pattern, which is obtained in the support set sup . From the obtained S_l (as $L = 1$), the sequential patterns with support greater than supmin ($sup \ge supmin$)are selected for further mining operations and the remaining patterns that doesn't satisfy supmin are neglected. Then, the sequential patterns with $L > 1$are mined. The pseudo code for mining the sequential patterns with $L > 1$ is given in the Fig. 2. The mined sequential patterns using the algorithm, which is illustrated in the pseudo code, are checked for the minimum support.The sequential patterns those have support greater than supmin is selected for further process.

for $L = 2$ to N_p
    for every L - length combination
        $S^{(L)}_{x_1x_2\ldots x_L} \leftarrow S^{(1)}_{x_1} \cap S^{(1)}_{x_2} \cap \cdots S^{(1)}_{x_L} : \{x_i;\ 1 \le i \le L\} \in (0, N_c + 1)$
        for $y = 0$ to $|S_{x_1x_2\ldots x_L}|-1$
            $sup^{(L)}_{x_1x_2\ldots x_L}(y) \leftarrow \min\left(sup^{(1)}_{x_1}(z), sup^{(1)}_{x_2}(z) \cdots sup^{(1)}_{x_L}(z)\right);$
                $S^{(1)}_{x_i}(z) = S^{(L)}_{x_1x_2\ldots x_L};\ z \in [0, |S_{x_i}|-1]$
        end for
    end for
end for

Fig 2: Pseudo code for mining sequential patterns with L > 1 from the Idataset

Hence, sequential patterns that are commonly present in the path flow are mined with different combination of their support. The obtained maximum length sequential patterns have strong support as it survives all the lower length combinations.

C. Generation of Fuzzy Rules Set

The sequential patterns mined by the mining algorithm with different combinations are subjected to generate fuzzy rules. The sequential patterns, $S^{(L)}_{x_g} : x_g \in (0, N_c + 1),\ 1 \le g \le L,\ 1 \le L \le N_c$, which are obtained from the mining algorithm are mathematically detailed as

$$S^{(1)}_{x_1} = \left(R_h^{(0)} \to R_h^{(1)} \to \cdots \to R_h^{\left(\left|S^{(1)}_{x_1}(z)\right|-1\right)}\right);\ 0 \le z \le \left|S^{(1)}_{x_1}\right|-1 \quad (4)$$

$$S^{(2)}_{x_1x_2} = \left(R_h^{(0)} \to R_h^{(1)} \to \cdots \to R_h^{\left(\left|S^{(2)}_{x_1x_2}(z)\right|-1\right)}\right) \quad (5)$$

$$S^{(3)}_{x_1x_2x_3} = \left(R_h^{(0)} \to R_h^{(1)} \to \cdots \to R_h^{\left(\left|S^{(3)}_{x_1x_2x_3}(z)\right|-1\right)}\right) \quad (6)$$

$$S^{(N_c)}_{x_1x_2\cdots x_{N_c}} = \left(R_h^{(0)} \to R_h^{(1)} \to \cdots \to R_h^{\left(\left|S^{(N_c)}_{x_1x_2\cdots x_{N_c}}(z)\right|-1\right)}\right) \quad (7)$$

Hence, Nc kind (length) of pattern sets are obtained from the mining algorithm. Each pattern set consists of $\left|S^{(L)}_{x_g}\right|$

number of sequential patterns. For every kind of pattern set, a set of support values are also obtained, which is comprised of the support values for every sequential pattern present in the pattern set. To generate fuzzy rules, R_{TH}, termed as rule threshold, is set. Probably, the R_{TH} is generated within the interval (0,1) . Then,

N_e elements are taken from the pattern in sliding fashion and the rules are generated. Generally, the fuzzy rules are of the form, *if* $IN_1 = i_1, IN_2 = i_2, \cdots, IN_n = i_n$ then OUT = i_{out}. The same format is followed here and the fuzzy rules are generated. The step-by-step procedure of the generation of fuzzy rules is given below.

Step 1-Initialize L = N_c

Step 2-Get the L^{th} pattern set and initialize $z = 0$

Step 3-Get the z^{th} sequential pattern from the L^{th} pattern set and initialize $p = 0$

Step 4- Select a sub-pattern as follows.

$$IN_q = R_h^{(p+q)},\ 0 \le q \le N_e - 1 \tag{8}$$

Step 5-Find all the sub-patterns as IN from $S^{(L)}_{x_1 x_2 \cdots x_L}$ similar to that of the selected pattern and select the following sub-pattern as OUT from the corresponding sequential pattern. For example, a sequential pattern of length L = 3 with a combination of the pattern set that corresponds to EPC code 1, 2 and 4 are given as

$$S^{(3)}_{124} = \begin{pmatrix} R_1 \to & R_2 \to & R_3 \to & R_4 \\ R_3 \to & R_1 \to & R_2 \to & R_4 \\ R_1 \to & R_2 \to & R_5 \to & R_6 \end{pmatrix} \tag{9}$$

The support values for every sequential pattern, which is present in the above pattern set is given as

$$\sup^{(L)}_{x_1 x_2 \cdots x_L} = \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} \tag{10}$$

From the above pattern set, the selected IN and OUT for fuzzy rules are given as

$$IN = R_1 \to R_2 \tag{11}$$

$$OUT = \begin{pmatrix} R_3 \to R_4 \\ R_4 \\ R_5 \to R_6 \end{pmatrix} \tag{12}$$

Step 6-Sort the selected patterns based on the support values and generate fuzzy rules by considering the IN subpattern for _if' statement, OUT sub-pattern for _then' statement and the corresponding support value is given asfuzzy score. The fuzzy rule for the aforesaid example can be generated as

if IN = $R_1 \to R_2$ then OUT = $R_5 \to R_6$ with fuzzy score = 7

if IN = $R_1 \to R_2$ then OUT = R_4 with fuzzy score = 6

if IN = $R_1 \to R_2$ then OUT = R_3 ® R_4 with fuzzy score = 5

Step 7-Increment p and go to Step 4 until p < Ne

Step 8-Increment z and go to Step 3 until $z < \left| S^{(L)}_{x_1 x_2 \cdots x_L} \right|$

Step 9-Decrement L and go to Step 2 until L > 0

Step 10-Get all the fuzzy rules separately for every L^{th} pattern set. From the fuzzy rules, the nature of every tag (based on EPC code) as well as the significant combination of different tags can be understood. The significance of the combination of the tags is decided by the combination which has support greater than supmin .

V. RESULTS AND DISCUSSION

The data mining system, which is proposed in this paper, is implemented in the working platform of JAVA (version JDK 1.6). The proposed system has been evaluated in the RFID application, tracking of goods in warehouses. Here, it is assumed that the warehouse holds some six stationery goods such as Pencil, Pen, Notebook, Diary, Paper clips, Memo holder and so $Nc = 6$. The RFID tags are affixed with each product and so each product has its EPC code. The warehouse is considered to have 200 products/per class (NT = 200) and eight RFID readers (i.e. NR = 8) are available. From the warehoused data, the I-dataset is generated in the format, as given in the Table III. Then, the sequential patterns are generated with a minimum support supmin = 1. Some of the sequential patterns, which are generated from the I-dataset, are given in the Table IV. With the aid of the sequential patterns, the fuzzy rules are generated and the rules corresponding to the pattern are given in the Table V. The number of sequential patterns with different lengths and the number of fuzzy rules with different pattern combinations are depicted in the Fig. 3 and Fig. 4.

Table 5: Generated Sequential Patterns and its support with different possible lengths

Length	Pattern Id	Pattern	Support
	S_1	$R_5 \to R_8 \to R_4$	2
	S_1	$R_5 \to R_1 \to R_8 \to R_6 \to R_5$	1
	S_2	$R_5 \to R_5 \to R_1$	2
1	S_2	$R_4 \to R_3 \to R_1$	1
	S_3	$R_5 \to R_2 \to R_1 \to R_4$	2
	S_3	$R_4 \to R_3 \to R_2$	1
	S_4	$R_1 \to R_3 \to R_6$	2
	S_4	$R_5 \to R_2 \to R_3 \to R_4 \to R_7$	1
	S_5	$R_4 \to R_1$	2
	S_5	$R_7 \to R_1 \to R_2 \to R_3$	1
	S_6	$R_4 \to R_8 \to R_2$	2
	S_6	$R_5 \to R_8 \to R_1 \to R_6$	1
2	S_{12}	$R_5 \to R_8 \to R_4$	1
	S_{12}	$R_5 \to R_1 \to R_6$	1
	S_{13}	$R_6 \to R_5 \to R_4$	1
	S_{13}	$R_5 \to R_8 \to R_3 \to R_7$	1
	S_{14}	$R_7 \to R_6 \to R_8 \to R_5$	1
	S_{14}	$R_5 \to R_8 \to R_5 \to R_7$	1
	S_{15}	$R_5 \to R_1 \to R_6$	1
	S_{15}	$R_7 \to R_8 \to R_4 \to R_1$	1
	S_{16}	$R_7 \to R_5 \to R_3 \to R_6$	1

	S_{16}	$R_6 \rightarrow R_5 \rightarrow R_4$	1
	S_{23}	$R_4 \rightarrow R_8 \rightarrow R_7 \rightarrow R_5$	1
	S_{23}	$R_6 \rightarrow R_5 \rightarrow R_2 \rightarrow R_1$	1
	S_{24}	$R_5 \rightarrow R_3 \rightarrow R_1 \rightarrow R_4 \rightarrow R_6$	1
	S_{24}	$R_5 \rightarrow R_3 \rightarrow R_1$	1
	S_{25}	$R_6 \rightarrow R_1 \rightarrow R_3 \rightarrow R_2$	1
	S_{25}	$R_5 \rightarrow R_1 \rightarrow R_6 \rightarrow R_8$	1
	S_{26}	$R_4 \rightarrow R_1 \rightarrow R_3 \rightarrow R_5$	1
	S_{26}	$R_6 \rightarrow R_4 \rightarrow R_8 \rightarrow R_5$	1
	S_{34}	$R_7 \rightarrow R_5 \rightarrow R_3$	1
	S_{34}	$R_6 \rightarrow R_1 \rightarrow R_3 \rightarrow R_5$	1
	S_{35}	$R_7 \rightarrow R_5 \rightarrow R_3$	1
	S_{35}	$R_7 \rightarrow R_5 \rightarrow R_1$	1
	S_{36}	$R_6 \rightarrow R_5 \rightarrow R_4 \rightarrow R_8$	1
	S_{36}	$R_5 \rightarrow R_8 \rightarrow R_1 \rightarrow R_6$	1
	S_{45}	$R_4 \rightarrow R_6 \rightarrow R_1 \rightarrow R_5 \rightarrow R_7$	1
	S_{45}	$R_4 \rightarrow R_1 \rightarrow R_3 \rightarrow R_7$	1
	S_{46}	$R_6 \rightarrow R_4 \rightarrow R_2 \rightarrow R_7$	1
	S_{46}	$R_8 \rightarrow R_2 \rightarrow R_5 \rightarrow R_7 \rightarrow R_6$	1
	S_{56}	$R_8 \rightarrow R_5$	1
	S_{56}	$R_1 \rightarrow R_4 \rightarrow R_7 \rightarrow R_6 \rightarrow R_5$	1
	S_{123}	$R_5 \rightarrow R_1 \rightarrow R_6$	1
	S_{124}	$R_5 \rightarrow R_8 \rightarrow R_3 \rightarrow R_7$	1
	S_{126}	$R_6 \rightarrow R_5 \rightarrow R_4$	1
3	S_{235}	$R_4 \rightarrow R_8 \rightarrow R_5$	1
	S_{345}	$R_7 \rightarrow R_5 \rightarrow R_3$	1
	S_{356}	$R_7 \rightarrow R_5 \rightarrow R_1$	1
	S_{456}	$R_5 \rightarrow R_6 \rightarrow R_2 \rightarrow R_7$	1
	S_{456}	$R_8 \rightarrow R_5$	1

Table 5: Fuzzy rules with fuzzy score generated from the extracted sequential patterns

Length	Pattern combination	Fuzzy Rule	Fuzzy score
	S_{123}	if IN=R_5 ,then OUT= $R_1 \rightarrow R_6$	1
	S_{124}	if IN=$R_5 \rightarrow R_8$,then OUT= $R_3 \rightarrow R_7$	1
	S_{124}	if IN=$R_8 \rightarrow R_3$,then OUT= R_7	1
	S_{126}	if IN=R_6 ,then OUT= $R_5 \rightarrow R_4$	1
3	S_{235}	if IN=R_4 ,then OUT= $R_8 \rightarrow R_5$	1
	S_{345}	if IN=R_7 ,then OUT= $R_5 \rightarrow R_3$	1
	S_{345}	if IN=R_7 ,then OUT= $R_5 \rightarrow R_3$	1
	S_{456}	if IN=$R_5 \rightarrow R_6$,then OUT= $R_2 \rightarrow R_7$	1
	S_{456}	if IN=$R_6 \rightarrow R_2$,then OUT= R_7	1
2	S_{12}	if IN=R_5 ,then OUT= $R_8 \rightarrow R_4$	1
	S_{12}	if IN=R_5 ,then OUT= $R_1 \rightarrow R_6$	1
	S_{13}	if IN=R_6 ,then OUT= $R_5 \rightarrow R_4$	1
	S_{13}	if IN=$R_5 \rightarrow R_8$,then OUT= $R_3 \rightarrow R_7$	1
	S_{14}	if IN=$R_7 \rightarrow R_6$,then OUT= $R_8 \rightarrow R_5$	1
	S_{14}	if IN=$R_6 \rightarrow R_8$,then OUT= R_5	1
	S_{15}	if IN=R_5 ,then OUT= $R_1 \rightarrow R_6$	1
	S_{15}	if IN=$R_7 \rightarrow R_8$,then OUT= $R_4 \rightarrow R_1$	1
	S_{16}	if IN=$R_7 \rightarrow R_5$,then OUT= $R_5 \rightarrow R_6$	1
	S_{16}	if IN=$R_5 \rightarrow R_5$,then OUT= R_6	1
	S_2	if IN=R_5 ,then OUT= $R_5 \rightarrow R_6$	2
	S_{23}	if IN=$R_4 \rightarrow R_8$,then OUT= $R_7 \rightarrow R_5$	1
	S_{23}	if IN=$R_8 \rightarrow R_7$,then OUT= R_5	1
	S_{24}	if IN=$R_5 \rightarrow R_3$,then OUT= $R_1 \rightarrow R_4 \rightarrow R_6$	1
	S_{24}	if IN=$R_5 \rightarrow R_1$,then OUT= $R_4 \rightarrow R_6$	1
	S_{25}	if IN=$R_6 \rightarrow R_1$,then OUT= $R_3 \rightarrow R_2$	1
	S_{25}	if IN=$R_1 \rightarrow R_3$,then OUT= R_2	1
	S_{26}	if IN=$R_4 \rightarrow R_1$,then OUT= $R_3 \rightarrow R_6$	1
	S_{26}	if IN=$R_1 \rightarrow R_3$,then OUT= R_6	1
	S_{34}	if IN=R_7 ,then OUT= $R_5 \rightarrow R_3$	1
	S_{34}	if IN=$R_6 \rightarrow R_1$,then OUT= $R_3 \rightarrow R_5$	1
	S_{35}	if IN=R_7 ,then OUT= $R_5 \rightarrow R_3$	1
	S_{35}	if IN=R_7 ,then OUT= $R_5 \rightarrow R_1$	1
	S_{36}	if IN=$R_6 \rightarrow R_5$,then OUT= $R_4 \rightarrow R_8$	1
	S_{36}	if IN=$R_5 \rightarrow R_4$,then OUT= R_8	1
	S_{45}	if IN=$R_4 \rightarrow R_6$,then OUT= $R_1 \rightarrow R_5 \rightarrow R_7$	1
	S_{45}	if IN=$R_6 \rightarrow R_1$,then OUT= $R_5 \rightarrow R_7$	1
	S_{46}	if IN=$R_6 \rightarrow R_4$,then OUT= $R_2 \rightarrow R_7$	1
	S_{46}	if IN=$R_4 \rightarrow R_2$,then OUT= R_7	1
	S_{56}	if IN=R_8 ,then OUT= R_5	1
	S_{56}	if IN=$R_1 \rightarrow R_4$,then OUT= $R_7 \rightarrow R_6 \rightarrow R_5$	1
	S_1	if IN=R_5 ,then OUT= $R_8 \rightarrow R_4$	2
	S_1	if IN=$R_7 \rightarrow R_1$,then OUT= $R_4 \rightarrow R_5 \rightarrow R_3$	2
	S_2	if IN=R_5 ,then OUT= $R_5 \rightarrow R_1$	2
	S_3	if IN=$R_5 \rightarrow R_2$,then OUT= $R_1 \rightarrow R_4$	2
	S_3	if IN=$R_2 \rightarrow R_1$,then OUT= R_4	2
	S_4	if IN=R_1 ,then OUT= $R_5 \rightarrow R_6$	2
	S_4	if IN=$R_5 \rightarrow R_2$,then OUT= $R_5 \rightarrow R_4 \rightarrow R_7$	1
	S_5	if IN=R_4 ,then OUT= R_1	2
	S_5	if IN=$R_1 \rightarrow R_5$,then OUT= $R_4 \rightarrow R_6$	2
	S_6	if IN=R_4 ,then OUT= $R_8 \rightarrow R_2$	2
	S_6	if IN=$R_4 \rightarrow R_3$,then OUT= $R_6 \rightarrow R_1$	2

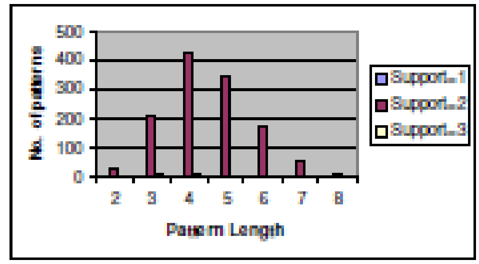

Fig 3: Number of sequential patterns with different pattern length and with the support value

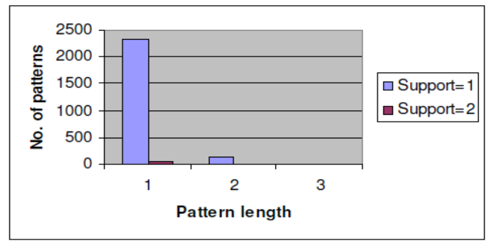

Fig 4: Number of fuzzy rules with different pattern combinations along with the support value

Rules Analysis-The rules define the nature of movement of every tag in terms of EPC and different tag combinations.For instance, in the given Table V, the tags with EPC code, 1, 2 and 5 commonly get into the range of R_1 and R_6 ,when they entered into the range of R_5 with a fuzzy score of 1. Hence, the nature of significant combination of the tags is well understood from the rules that are generated from the proposed mining system.

VI. Conclusion

In this paper, we have proposed a data mining system for mining the information that are relevant to the nature of movement of tags, which are affixed in the warehouse goods. The implementation results have shown that the proposed mining system mines the knowledge from the warehoused data by generating I-dataset, mining sequential patterns and then by generating fuzzy rules from the sequential patterns. The outcome of the system, fuzzy rules, has detailed the nature of the tag movement with a fuzzy score. Given a part of the tag (indirectly it refers to A product) movement, the fuzzy rules hold the remaining path of the tag (product). In this manner, different length combinations of the tags have been considered and the movement has been understood. The movements are not considered for all the tags and its combinations, but only for some significant tags and combinations. Hence, with the aid of the proposed data mining system, the tracking of goods in large warehouses can be performed effectively and also the extracted information may be helpful for the management of warehouse.

VII. References

1) Bin Li and Dennis Shasha, "Free Parallel Data Mining", ACM SIGMOD Record, Vol.27, No.2, pp.541-543, June 1998.

2) Anand, Bell and Hughes, "EDM: A general framework for data mining based on evidence theory", Data and Knowledge Engineering, Vol.18, No.3, pp.189-223, 1996.
3) Agrawal, Imielinsk and Swami, ―Database Mining: A Performance Perspective", IEEE Transaction Knowledge and Data Engineering, vol. 5, no. 6, pp. 914-925, 1993.
4) Chen and Liu, "Data mining from 1994 to 2004: an application-oriented review", International Journal of Business Intelligence and Data Mining, Vol.1, No.1, pp.4-11, 2005.
5) Shankar and Purusothaman, " Utility Sentient Frequent Itemset Mining and Association Rule Mining: A Literature Survey and Comparative Study", International Journal of Soft Computing Applications, Vol.10, No.4, pp.81-95, 2009.
6) Mourad Ykhlef and Hebah ElGibreen, "Mining Sequential Patterns Using Hybrid Evolutionary Algorithm", WASET, Vol.60, No.149, pp.863-870, 2009.
7) Jia-Dong Ren and Xiao-Lei Zhou, "A New Incremental Updating Algorithm for Mining Sequential Patterns",Journal of Computer Science, Vol.2, No.4, pp.318-321, 2006
8) Hye-Chung Kum, Joong Hyuk Chang and Wei Wang, "Benchmarking the effectiveness of sequential pattern mining methods", Data & Knowledge Engineering, Vol.60, pp. 30–50, 2007.
9) Jiawei Han, Hong Cheng, Dong Xin and Xifeng Yan, "Frequent pattern mining: current status and future directions", Data Mining and Knowledge Discovery, Vol.15, No.1, pp.55-86, January 2007.
10) Dunren Che1 and Wei Zheng, "Smart support functions for sequential pattern mining", Journal of Computational Methods in Science and Engineering, Vol.6, No.2, pp.255- 263, 2006.
11) Yashpal Singh and Alok Singh Chauhan, "Neural Networks In Data Mining", Journal of Theoretical and Applied Information Technology, Vol.5, No.6, pp.36- 42,2009.
12) Guangming Wang and Gonglian Jin, "Research and Design of RFID Data Processing Model Based on Complex Event Processing", in Proceedings of International Conference on Computer Science and Software Engineering, Vol.5, pp. 1396-1399, 2008.
13) Dhany Saputra, Dayang R.A.Rambli and Oi Mean Foong, "Mining Sequential Patterns Using I-PrefixSpan", International Journal of Computer Science and Engineering, Vol. 2, No.2, pp.49-554, 2008.
14) C.M. Roberts, "Radio frequency identification (RFID)", Computers & Security, Vol.25,pp. 18 – 26, 2006.
15) Hatim A. Aboalsamh, "A novel Boolean algebraic framework for association and pattern mining", WSEAS Transactions on Computers, Vol.7, No.8, pp.1352-1361,August 2008.
16) Kuen-Fang Jea, Ke-Chung Lin and I-En Liao, "Mining hybrid sequential patterns by hierarchical mining technique", International Journal of Innovative Computing,Information and Control, Vol.5, No.8, August 2009.
17) Sathiyamoorthi and Murali Bhaskaran, "Data Mining for Intelligent Enterprise Resource Planning System", International Journal of Recent Trends in Engineering, Vol. 2, No. 3,pp.1-5, November 2009.
18) Jian Pei,Jiawei Han, Behzad Mortazavi-Asl, Jianyong Wang, Helen Pinto, Qiming Chen, Umeshwar Dayal and Mei-Chun Hsum, "Mining Sequential Patterns by Pattern- Growth: The PrefixSpan Approach", IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 10, pp.1-17, October 2004.
19) Agrawal and Srikant, ―Mining Sequential Patterns," in proceedings of International Conference on Data Eng., pp.3-14, Mar. 1995.
20) Garofalakis, Rastogi and Shim, "Mining sequential patterns with regular expression constraints", IEEE Transactions on Knowledge and Data Engineering, Vol.14, No.3, pp.530-552, June 2002.
21) Jayanthi Ranjan and Vishal Bhatnagar, "A Review of Data Mining Tools In Customer Relationship Management", Journal of Knowledge Management Practice, Vol. 9, No. 1, March 2008.
22) Michael J. Shaw, Chandrasekar Subramaniam, Gek Woo Tan and Michael E. Welge, "Knowledge management and data mining for marketing", Decision support systems, Vol.31, No.1, pp.127-137, May 2001.
23) Asghar Sabbaghi and Ganesh Vaidyanathan, ―Effectiveness and Efficiency of RFID technology in Supply Chain Management: Strategic values and Challenges", Journal of Theoretical and Applied Electronic Commerce Research, Vol. 3, No. 2, p.p. 71-81, 2008, ISSN 0718–1876.
24) Lahiri S., RFID source book, IBM press, Upper Saddle River, NJ, 2006.
25) Henning Baars, Hans-Georg Kemper, Heiner Lasi and Marc Siegel, ―Combining RFID Technology and Business Intelligence for Supply Chain Optimization – Scenarios for Retail Logistics", Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008), Waikoloa, Big Island, Hawaii, ISBN: 0-7695-3075-3.
26) Asif, Z., Mandviwalla, M., "Integrating the supply chain with RFID: a technical and business analysis", Communications of the Association for Information Systems, Vol. 15, No. 24, pp.393-427, 2005.
27) Jigyasa Bisaria, Namita Srivastava and Kamal Raj Pardasani, "A Rough Set Model for Sequential Pattern Mining with Constraints", International Journal of Computer and Network Security, Vol.1, no.2, pp.16-22, November 2009.

28) Shigeaki Sakurai, Youichi Kitahara and Ryohei Orihara, "A Sequential Pattern Mining Method based on Sequential Interestingness", International Journal of Computational Intelligence, Vol. 4, No.4, pp.252-260, 2008.
29) Themis P. Exarchos, Markos G. Tsipouras, Costas Papaloukas and Dimitrios I. Fotiadis, "A two-stage methodology for sequence classification based on sequential pattern mining and optimization", Data & Knowledge Engineering, Vol.66, pp.467–487, 2008.
30) Chen, "An Up Down Directed Acyclic Graph Approach for Sequential Pattern Mining", IEEE Transactions on Knowledge and Data Engineering, Vol.21, No.12, pp.1-1, 2009.
31) Jirachai Buddhakulsomsiri and Armen Zakarian, "Sequential pattern mining algorithm for automotive warranty data", Computers and Industrial Engineering, Vol.57, No.1, pp.137-147, August 2009
32) Ya-Han Hu, Tony Cheng-Kui Huang, Hui-Ru Yang and Yen-Liang Chen, "On mining multi-time-interval sequential patterns", Data & Knowledge Engineering, Vol.68, No.10, pp.1112-1127, October 2009
33) Ya-Han Hu, Yen-Liang Chen and Kwei Tang, "Mining sequential patterns in the B2B environment", Journal of Information Science, Vol. 35, No. 6, 677-694 (2009).