

Study on Predicting Various Mining Techniques Using Weighted Itemsets

J.Jenifa, Dr.V. Sampath Kumar.

ME (CSE), Shivani Engineering College Tiruchirappalli
Department of CSE, Shivani Engineering College, Tiruchirappalli.

Abstract: Pattern mining is one among datamining technique meant to reveal the existing patterns from data. Datamining is helpful for retrieving the hidden and most interesting patterns. Frequent mining can be obtained with and without candidate generation schemes. From the study of literature, complexity of data has been reduced by the utilization of Infrequent Pattern Mining method, which represents rare data correlations among data. The proposed approach is very helpful to handle the problem of predicting Infrequent and Weighted itemsets, which can be termed as Infrequent Weighted Itemset (IWI) mining problem. There are two methods have been used for handling IWI mining problem namely IWI and minimal IWI. To enhance the potential application such as medical science , biological datas by using IWI approach. How to extract the infrequent items from biological database depending on its weights and use the characteristics of the data how to develop some application. Experimental results show efficiency and effectiveness of the proposed algorithm.

Index Terms: Clustering, Classification, Weighted Association rules and Infrequent pattern mining, Weighted support

I. INTRODUCTION

Datamining is the process discovering useful patterns and searching enormous amount of information to find very important data. This approach has been used when searching for complex multiple dependencies among data without the help of human brain. Pattern mining is one of the most important researches in data mining. The basic idea is to extract relevant *data* of knowledge describing parts of a database. The number of discovered patterns is larger than the size of the original database! To tackle this problem, many techniques have been developed in recent years for finding not all, but *useful sets of patterns* like frequent pattern detection.

Frequent Pattern Mining is the process of predicting item which is mostly occurs in database. Itemset mining is the process of discovering the correlation among data. This itemset mining process has the interest to deal with frequent itemset. Frequent itemsets mining process is used in number of real-life application. Datamining process is based on KDD.¹ It is a useful technique for retrieving useful data and develops techniques and methods based on that data. Pattern discovery and extraction is based on KDD.

The database contains large number of records and fields. In medical application the database containing even $d=10^3$ fields. So manual extraction of pattern is very complex, some cases needed automated analysis of database is important. The KDD is a multi-step process for discovering information from database, which involves a)Data preparation, b)Searching for patterns, c)Knowledge evaluation and d)Refinement.

Fig 1.illustrates the Frequent itemset detection process takes much more time to implement a specific application at the same time cost of implementation is very high because, find the occurrence of data in the database will be taken a repeated search of database. To tackle this problem weights are preassigned with the items and detect the significance of that item. There are lot of problems during the prediction of frequent items.

¹.Knowledge Discovery from Databases

Problem in Frequent Pattern Mining

- Huge Number of frequent itemsets
- Hard to analyze
- Most of them are similar

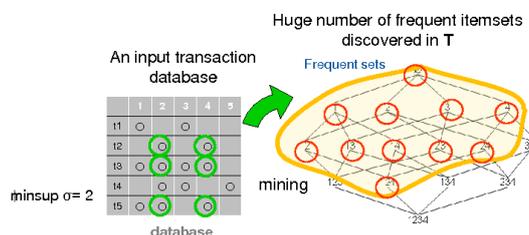


Fig1.Shows the problem of frequent pattern mining

In Infrequent Pattern Mining, patterns are rarely found in database are often considered to be interesting and are eliminated using the support measure. Such patterns are known as infrequent patterns. An infrequent pattern is an itemset or a rule whose support is less than the minimum sup threshold. Although a vast majority of infrequent patterns are uninteresting, some of them might be useful to the analysis, particularly those that correspond to negative correlations in data. For example, the sale of DVDs and VCRs together is low because any customer who buys a DVD will most likely not buy a VCR and vice versa. Such negative correlated patterns are useful to help identify competing items, which are items that can be substituted for one another. Examples of competing items include tea versus coffee, butter versus margarine, regular versus diet soda, and desktop versus laptop computers. If the actual support for any one of these pairs is less than their expected support, then the infrequent patterns is interesting otherwise no need to detect that infrequent pattern.

A. Biological Data

Biological datas are data or measurements collected from biological sources, which are often stored or exchanged in a digital form. Biological data are commonly stored in files or databases. This includes bioinformatics, biological database, biological model, datamining, datamodels and DNA sequence.

B. DNA

DNA is the prime genetic molecule, carrying all the hereditary information within chromosomes, immediately focused attention on its structure.

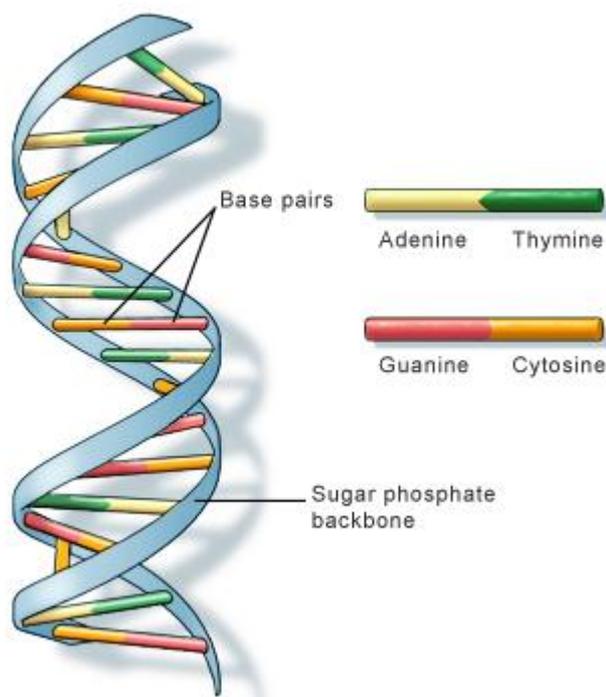


Fig 2.Shows the structure of DNA

The fundamental DNA structure was found to be the double helix. It told us that all genes have roughly the same three-dimensional form and that the differences between two genes reside in the order and number of their four nucleotide building blocks along the complementary strands. For example, the chromosome of some small viruses has single-stranded, not double-stranded, molecules. Moreover, the precise orientation of the base pairs varies slightly from base pair to base pair in a manner that is influenced by the local DNA sequence. Some DNA sequences even permit the double helix to twist in the left-handed sense, as opposed to the right-handed sense originally formulated for DNA's general structure. And while some DNA molecules are linear, others are circular. Still additional complexity comes from the super coiling (further twisting) of the double helix, often around cores of DNA-binding proteins.

C. RNA

RNA, which differs from DNA in three respects First, the backbone of RNA contains ribose rather than 2deoxyribose. Rather, RNA functions as the intermediate, the mRNA, between the gene and the protein-synthesizing machinery. Another function of RNA is as an adaptor, the tRNA, between the codons in the mRNA and amino acids. RNA can also play a structural role as in the case of the RNA components of the ribosome. Yet another role for RNA is as a regulatory molecule, which through sequence complementarity binds to, and interferes with the translation of, certain mRNAs. Finally, some RNAs (including one of the structural RNAs of the ribosome) are enzymes that catalyze essential reactions in the cell

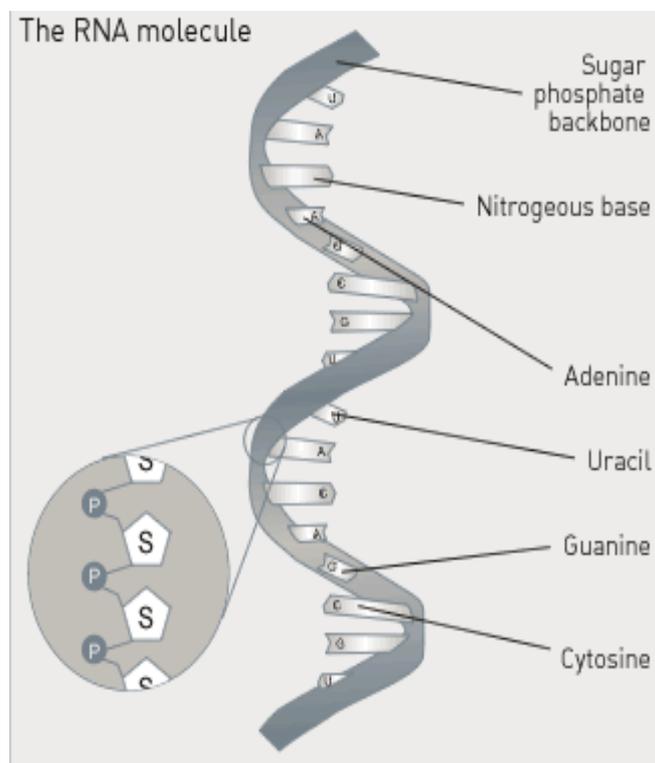


Fig 3. The structure of RNA

RNA, which at first glance appears to be very similar to DNA, has its own distinctive structural features. It is principally found as a single-stranded molecule. Yet by means of intra-strand base pairing, RNA exhibits extensive double-helical character and is capable of folding into a wealth of diverse tertiary structures. These structures are full of surprises, such as non-classical base pairs, base-backbone interactions, and knot-like configurations. Most remarkable of all, and of profound evolutionary significance, some RNA molecules are enzymes that carry out reactions that are at the core of information transfer from nucleic acid to protein. was at first.

D. Protein

Protein is the essential part of all organisms. They are the linear polymers combination of 20 amino acids. In our body 50% of dry weight cells are protein. Remaining amount of protein molecules are generated from living organism. It is a special molecular tool. Proteins perform thousands of tasks to sustain life. Protein is the important component of body tissues. It is the particular nutrient for growth. The body is a combination of protein and other nitrogenous compounds being degraded and regenerated continuously. Proteins are generated

from body than the amount of food consumed. Proteins are large molecule which one is made up of combination of amino acids and bond together using covalent bonds. Protein defined as molecule with more number of amino acids. It has one or more polypeptide chains. A protein has different size and shapes. Proteins are the most abundant organic molecules in animals, playing important roles in all aspects of cell structure and function.

Fig 4.The elements of amino acids

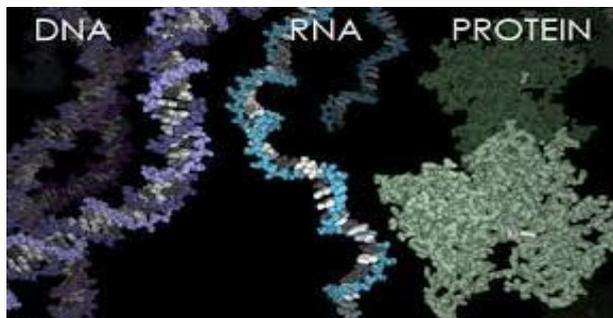


Fig 5. The general view of DNA, RNA and Protein

Proteins protect the body against disease. When the body detects invading antigens, it manufactures antibodies, giant protein molecules designed specifically to combat them. The antibodies work so swiftly and efficiently that in normal, healthy individual, most diseases never have a chance to get started. The body's fluids are contained within the cells (intracellular) and outside the cells (extracellular). Extracellular fluids are found either in the spaces between cells (interstitial) or within blood vessels (intravascular). Wherever proteins are, they attract water and this helps to maintain the fluid balance in their various compartments. In addition, proteins help maintain the balance between acids and bases within the body fluids by accepting and releasing hydrogen ions. Even though proteins are needed for growth, maintenance and repair, they will be used to provide glucose when the need arises.

E. Amino acids

Amino acids are the basic building blocks of proteins and the fundamental significance to all organisms. It plays a very big role in the areas of nutrition, medicine and plant production. L-amino acid is essential in life science. It enhances the fermentation and cell cultivation process. Biological applications are successfully implemented with the help of *pure amino acids*. Additional specification of L-amino acids used to develop some application in research and development field such as *BIO-CHEMISTRY*.

There are two groups of amino acids

1. Common amino acids
 - Polar
 - Non-polar
 - Acidic
 - Basic
2. Uncommon amino acids
 - Polar
 - Non-polar
 - Basic

Proteins are biopolymers of **α - amino acids**, so named because the amino group is bonded to the carbon atom, next to the carbonyl group. The physical and chemical properties of a protein are determined by its constituent amino acids. The individual amino acid subunits are joined by amide linkages called **peptide bonds**.

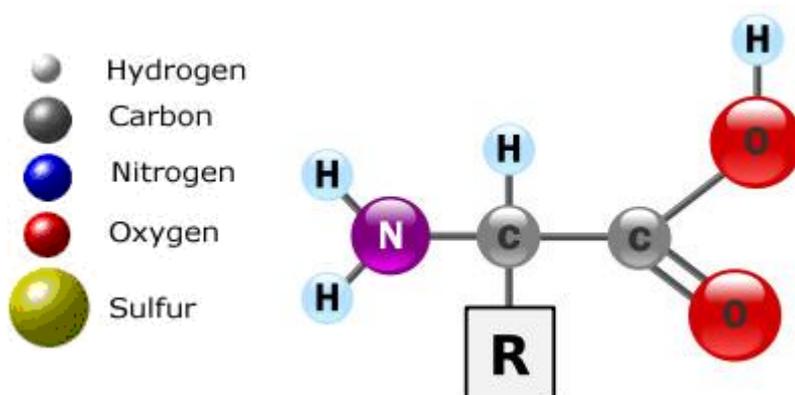


Fig 6. Shows the general structure of Amino acid

Infrequent pattern mining detection based on its weights can be predicted from the BIOLOGICAL database. Protein is a combination of amino acids. Some dominating amino acids will be create some genetic disorder problem with the help of IWI mining we can extract that pattern from database. Table1 indicates the number of amino acids available in proteins and its alphabetic indication. Table2 represents the Specific Characteristics of Common Amino Acids. Table3 shows the Specific Characteristics of Uncommon Amino Acids.

TABLE 1: AMINO ACIDS WITH ITS ALPHABETS

Amino acid	Three letter Indication	One letter Indication
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

TABLE2: COMMON AMINO ACIDS

Common Amino Acids				
Types	Amino acids	Mol formula	Exact mass	Mol wt
Non polar	1.alanine	C ₃ H ₇ NO ₂	89.05	89.09
	2.isoleucine	C ₆ H ₁₃ NO ₂	131.09	131.17
	3.leucine	C ₆ H ₁₃ NO ₂	131.09	131.17
	4.methionine	C ₅ H ₁₁ NO ₂ S	149.05	149.21
	5.phenylalanine	C ₉ H ₉ NO ₂	165.08	165.19
	6.proline	C ₅ H ₉ NO ₂	115.06	115.13
	7.tryptophan	C ₁₁ H ₁₂ N ₂ O ₂	204.09	204.23
	8.valine	C ₅ H ₁₁ NO ₂	117.08	117.15
polar	1.asparagine	C ₄ H ₈ N ₂ O ₃	132.05	132.12
	2.cysteine	C ₃ H ₇ NO ₂ S	121.02	121.16
	3.glutamine	C ₅ H ₁₀ N ₂ O ₃	146.07	146.14
	4.glycine	C ₂ H ₃ NO ₂	75.03	75.07
	5.serine	C ₃ H ₇ NO ₃	105.04	105.09
Types	Amino acids	Mol formula	Exact mass	Mol wt
Acidic	1.aspartic acid	C ₄ H ₇ NO ₄	133.04	133.10
	2.glutamic acid	C ₆ H ₁₁ NO ₄	161.07	161.16
basic	1.arginine	C ₆ H ₁₄ N ₄ O ₂	174.11	174.20
	2.histidine	C ₆ H ₉ N ₃ O ₂	155.07	155.15

	3.lysine	C ₆ H ₁₄ N ₂ O ₂	146.11	146.19
--	----------	--------------------------------------------------------------	--------	--------

TABLE3: UNCOMMON AMINO ACIDS

Uncommon Amino Acids				
Types	Amino acids	Mol formula	Exact mass	Mol wt
Non polar	1.aminobutanoic Acid	C ₄ H ₉ O ₂	103.06	103.12
	2.aminohexanoic Acid	C ₆ H ₁₃ NO ₂	131.09	131.17
	3.cyclohexyla Lanine	C ₉ H ₁₇ NO ₂	171.13	171.24
	4.allylglycine	C ₅ H ₉ NO ₂	115.06	115.13
	5.vinylycine	C ₄ H ₇ NO ₂	101.05	101.10
Polar	1.citrulline	C ₆ H ₁₃ N ₃ O ₃	175.10	175.19
	2.homoserine	C ₄ H ₉ NO ₃	119.06	119.12
	3.hydroxyproline	C ₅ H ₉ NO ₃	131.06	131.13
Basic	1.diaminobutanoic Acid	C ₄ H ₁₀ N ₂ O ₂	118.07	118.13
	2.diaminopropionic Acid	C ₃ H ₈ N ₂ O ₂	104.06	104.11

F. Chapter Organization

As earlier discussed in Section I indicate the types of amino acids and its groups. Section II indicates the existing system. Existing system deals with how the frequent itemset were predicted, how the weights are assigned to the data's and what was the way for predicting infrequent items. Existing papers have lots of drawbacks in terms of computation time and cost. The proposed approach extracts the infrequent weighted items from biological database and based their characteristics to develop some application with amino acid datasets. Going to find some dominating amino acid which causes diseases based their weights and analysis the algorithm performance.

II. LITERATURE SURVEY

The various infrequent item detection algorithms have been used for detecting correlation among data. The several different approaches used to detect the frequent and infrequent items with its weights can be discussed below:

A. Weighted Association Rules (WAR) ²

The author [14] has proposed Weight associated with each item in a transaction; it shows the interest of that item within the transaction. This paper explained the concept of associate the weight parameter with each item in the resulting association rule. This called as weighted association rule. Market basket application is enhanced with the help of weighted association rule concept because the volume of purchase by a customer can be easily extracted. The weights are generated during rule generation process. The customers are segmented based on their volume of purchasing. WAR used a two-fold approach where the frequent itemsets are generated through standard association rule mining algorithms without considering weight. Post-processing is then applied on the frequent itemsets during rule-generation to derive the maximum WARs. WAR doesn't interfere with the process of generating frequent itemset. Rather, it focuses on how weighted association rules can be generated by examining the weighting factors of the items included in generated frequent itemsets.

WAR is proposed for association rule problem. By the presence of this problem the mining process has two steps

1. Generating frequent itemsets
2. Deriving WAR from each frequent itemset

During the WAR derivation process,

1. The concept of density is employed to separate transaction concentrated regions from the rest.
2. An efficient ordered shrinkage algorithm is proposed to derive WARs from a high density region through shrinkages to meet the confidence requirement.

R = transaction set
Minsup= Minimum support
Minconf=Minimum confidence
d=Minimum density

² WAR-Weighted Association Rule

n=Dimensional space

F=Frequent itemset

The aim of this paper to retrieve the maximum WAR so the support, confidence, and density all are greater than the user defined minsup, minconf, and minimum density. (1)The Space partition and counter generation technique is used to identify the frequent items by pruning the extra grids.(2)WARs generation is done by shrinking the dense boxes and meets its confidence requirement. The Quantitative Association Rule³ has been proposed for numerical attribute values that are associated with the every database records.QAR is the another best approach to mine WARs.

The datasets have been changed for further experimental operation at the same time we can see the worthiness of WAR algorithm.

B. Weighted Association Rule Mining Using Weighted Support and Significance Framework

The process of discovering significant binary relationships in transaction datasets in a weighted setting is one of the issues. The traditional model of association rule mining is adapted to handle WAR mining problems where each item is allowed to have a weight. The goal is to mining significant relation instead of insignificant once. When generating the large itmsets using weights in the iterative process is very difficult. The problem of invalidation of the *downward closure property* in the weighted setting is solved by using an improved model of weighted support measurements and exploiting a ⁴*weighted downward closure property*. A new algorithm called Weighted Association Rule Mining is developed based on the improved model. The algorithm is both scalable and efficient in discovering significant relationships in weighted settings as illustrated by experiments performed on simulated datasets.

The *downward closure property*, which is used to justify the efficient iterative process of generating and pruning large itemsets from its subsets. In order to tackle this challenge, we made adaptation on the traditional association rule mining model under the “significant – weighted support” metric framework instead of the “large – support” framework used in previous works. In this new proposed model, the iterative generation and pruning of significant itemsets is justified by a *weighted downward closure property*. The threshold support is expressed in terms of ⁵*minimum item supports* of the items that appear in the rule.

Item weight: *Item weight* is a value attached to an item representing its significance. We denote it as $w(i)$.

Itemset weight: Based on the item weight $w(i)$, the weight of an itemset, denoted as $w(is)$, can be derived from the weights of its enclosing items.

Transaction weight: *Transaction weight* is a type of itemset weight. It is a value attached to each of the transactions.

Weighting space

(1) **Inner-transaction space WSt:** This space refers to the host transaction that an item is weighted in.

(2) **Item space WSI:** This space refers to the space of the item collection that covers all the items appears in the transactions.

(3) **Transaction space WST:** This space is defined for transactions rather than for items.

Weighted support – significant framework vs. support – large framework

An itemset is declared as *large* if its support is above a predefined minimum support threshold. In the WARM method, we say an itemset is *significant* if its *weighted support* is above a pre-defined minimum weighted support threshold. The threshold values specified by the user from any point of view.

The problem has been found the limitation of the traditional Association Rule Mining model, in particular, its incapacity for treating units differently. We proposed that weight can be integrated in the mining process to solve this problem. We identify the challenge faced when making improvement towards using weight, in particular the invalidation of downward closure property. A set of new concepts are proposed to adapt weighting in the new setting. Among them is the proposal of using *weighted downward closure property* as a replacement of the original *downward closure property*. This is proved as valid and justifies the effective mining strategy in the new framework of *weighted support significant*. The new framework is designed to replace the original support large framework in order to tackle the problem in weighted settings. Through studying the simulation of the lattice building, conclusion is drawn that weight can be used to steer the mining focus to those important itemsets with high degree of significance. This is further proven by experiments on synthetic datasets. The experiments show that the mining results in the weighted setting conform to the expected hypothesis. The experiments also show that the algorithm is scalable.

³QAR-Quantitative Association Rules

⁴WARM-Weighted Association Rule Mining

⁵MIS-Minimum Item Support

C. Mining Weighted Association Rules without Preassigned Weights

In datamining the association rule mining is one of the issues. Traditional method has the weight associated with the items in each transaction. But the author has proposed a new measure **w-support**⁶, which does not require preassigned weights. Using link based model the quality of transaction can be measured. This approach is verified by the use of experimental results. w-support is a new measure of item sets in databases with only binary attributes. The weights of transactions are different so there is no problem about the occurrence of frequent itemset. The weights are completely derived from the internal structure of the database based on the assumption that good transactions consist of good items. This concept extending Kleinberg's HITS model and algorithm to bipartite graphs. w-support is distinct from weighted support in weighted association rule mining, where item weights are assigned. With the help of w-support interesting patterns can be discovered without any overhead. The process needs the following steps

1. The datasets are collected and declare the threshold value.
2. Find all the significant item sets with w-support above the given threshold.
3. Compare the w-support and w-confidence with support and confidence based on the same threshold value.
4. The link based model is useful in adjusting the mining results.
5. Evaluate the performance of algorithm.

HITS model and algorithm are used to derive the weights of transactions from a database with only binary attributes. Based on these weights, a new measure w-support is defined to give the significance of item sets. It differs from the traditional support in taking the quality of transactions into consideration. Then, the w-support and w-confidence of association rules are defined in analogy to the definition of support and confidence. An Apriori-like algorithm is proposed to extract association rules whose w-support and w-confidence are above some given thresholds. Experimental results show that the computational cost of the link-based model is reasonable. At the expense of three or four additional database scans, we can acquire results different from those obtained by traditional counting-based models. Particularly for sparse data sets, some significant item sets that are not so frequent can be found in the link based model. Through comparison, we found that our model and method address emphasis on high-quality transactions. The link-based model is useful in adjusting the mining results given by the traditional techniques. Some interesting patterns may be discovered when the hub weights of transactions are taken into account. Moreover, the transaction ranking approach is precious for estimating customer potential when only binary attributes are available, such as in Web log analysis or recommendation system.

D. On Minimal Infrequent Itemset Mining

The author has proposed new algorithm MINimal Infrequent iTemset⁷ to discover the infrequent itemset from database. Infrequent itemset detection is used to enhance the potential application such as statistical disclosure risk assessment, bioinformatics, and fraud detection. This is the first algorithm designed specifically for finding these rare itemsets. The properties of itemset used in this algorithm. NP-complete problem is shown during the outcome of experimental results. The infrequent item detection is important by the following reasons:

1. Statistical disclosure risk assessment where rare patterns in anonymized census data can give the chance to statistical disclosure;
2. Bioinformatics where rare patterns in microarray data may lead to genetic disorders; and
3. Fraud detection where rare patterns in financial or tax data may suggest unusual activity associated with fraudulent behavior.

MINimal Infrequent iTEmsets algorithm is based on SUDA2 algorithm which is used to find the minimal unique itemset⁸. MINimal Infrequent iTEmsets used to find minimal τ -infrequent or minimal τ -occurrent itemsets. The amount of computation time is calculated based on the correlation between the Minimal Infrequent Itemsets⁹. With the help of pruning technique we can improve the running time.

E. Finding Interesting Rule Groups in Microarray Datasets

Microarray datasets typically contain large number of columns but small number of rows. Association rules have been very useful in analysing such datasets. Some exiting association rule mining algorithms are not suitable for handling datasets with large number of columns.

⁶ W-support-Weighted support

⁷ MINIT-MINimal Infrequent iTEmsets

⁸ MUI-Minimal Unique itemset

⁹ MII-Minimal Infrequent Itemsets

Numbers of Association rules are generated from large datasets which contain large number of column combination. The author has proposed new algorithm called FARMER which is specially designed for discovering association rules from microarray datasets. Instead of finding individual association rules, FARMER algorithm finds interesting rule groups which are essentially a set of rules that are generated from the same set of rows. Unlike conventional rule mining algorithms, FARMER searches for interesting rules in the row enumeration space and exploits all user-specified constraints including minimum support, confidence and chi-square to support efficient pruning. Several experiments on real bioinformatics datasets show that the FARMER algorithm is very efficient and magnitude faster than the previous association rule mining algorithms.

Association rules, Interesting rule groups¹⁰, Enumeration, pruning technique have been implemented for effective outcome of analysis. This algorithm is very useful in gene network. Because, easily predict the relationship between the genes with the help of association rule mining technique. Biological database contains enormous amount of information in the form of rows and columns. Genes are belong to many rules by the use of clustering we can group them. The existing association rule mining algorithm adopt the column enumeration process, therefore they are very time consuming process on microarray datasets. But the FARMER discovers how to randomly pick a rule for each rule group.

The FARMER algorithm is very essential to handle the microarray datasets. Each experimental results are needed a better outcome. If the datasets are large means it take some time to execute the result. Based on the magnitude of dataset the operations get delay. The conventional methods use the column enumeration but the FARMER overcomes that approach by introducing smart row enumeration method and pruning technique.

Biological data wants the exact result for its research. That time FARMER algorithm can be produced better result. The experimental result shows the efficiency and effectiveness of that algorithm.

III. CONCLUSION AND FUTURE WORK

From the study of literature, the existing itemset mining method has been used particular potential application development. Biological datas are very helpful for research areas because it has large datasets. Normally the biological database consists of thousands and millions of sequence. There are lots of problems how to handle those sequence in efficient way. In traditional approach the weights are associated with each item in a transaction. Proposed approach is about to discover infrequent weighted itemsets from biological data. The weights are denoted as the number of times the occurrence of that item in dataset. The normal sequence of amino acids get infected when the presence of diseases. Such situation infrequent item detection is very helpful for extracting that type of amino acids and discovers the suitable drug to cure that disease. The performance of algorithm is evaluated by comparing the normal and infected sequence of amino acids of different datasets with the help of IWI mining.

References

- [1] A. Frank and A. Asuncion, "UCI machine learning repository," 2010. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [2] Cong, A. K. H. Tung, X. Xu, F. Pan, and J. Yang, "Farmer: finding interesting rule groups in microarray datasets," in *ACM SIGMOD 04, Paris, France*, 2004.
- [3] D. J. Haglin and A. M. Manning, "On minimal infrequent itemset mining," in *Proceedings of the 2007 International Conference on Data Mining, DMIN'07*. CSREA Press, 2007, pp. 141–147.
- [4] F. Tao, F. Murtagh, and M. Farid, "Weighted association rule mining using weighted support and significance framework," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'03*, 2003, pp. 661–666.

¹⁰ IRG-Interesting Rule Groups

- [5] Gupta, A. Mittal, and A. Bhattacharya, "Minimally infrequent itemset mining using pattern-growth paradigm and residual trees," in *COMAD*, 2011, pp. 57–68.
- [6] G.lakshmi Priya, Shanmugasundaram Hariharan, "A study on predicting patterns over the protein Sequence datasets using association rule mining" in *Journal of Engineering Science and Technology* Vol. 7, No. 5 (2012) 563 - 573 © School of Engineering, Taylor's University.
- [7] IBM, "IBM Quest Synthetic Data Generation Code," 2009. [Online]. Available: <http://www.almaden.ibm.com/>
- [8] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, 2000, pp. 1–12
- [9] K.-S. Leung, C. L. Carmichael, and B. Hao, "Efficient mining of frequent patterns from uncertain data," in *Proceedings of the Seventh IEEE International Conference on Data Mining Workshops*, ser. ICDMW '07, 2007, pp. 489–494.1993, pp. 207–216
- [10] Luca Cagliero and Paolo Garza, "Infrequent Weighted Itemset Mining Using Frequent Pattern Growth" 2013
- [11] M. L. Antonie, O. R. Zaiane, and A. Coman, "Application of data mining techniques for medical image classification," in *MDM/KDD'01*, 2001.
- [12] R. Agrawal, T. Imielinski, and Swami, "Mining association rules between sets of items in large databases," in *ACM SIGMOD 1993*,
- [13] Sun and F. Bai, "Mining weighted association rules without preassigned weights," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 4, pp. 489–495, 2008.
- [14] W. Wang, J. Yang, and P. S. Yu, "Efficient mining of weighted association rules (WAR)," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'00*, 2000, pp. 270–274.
- [15] X. Dong, Z. Zheng, Z. Niu, and Q. Jia, "Mining infrequent itemsets based on multiple level minimum supports," in *Second International Conference on Innovative Computing, Information and Control, ICICIC'07*, 2007, pp. 528–531.