ISSN: 0974-8571 Vol.13 No. 2 December, 2021

International Journal of Computational Intelligence in Control

An Association Rule Based Model for Discovery of COVID19 Patients

[1] **K Manikandan,** Ph.D Research Scholar, PG & Research Department of Computer Science, Christhu Raj College (Affiliated to Bharathidasan University, Tiruchirappalli), Panjapur, Trichy, Tamilnadu. Email: km.kathirai@gmail.com

[2]**Dr. Ramalingam sugumar**, Professor & Director, PG & Research Department of Computer Science,Christhu Raj College (Affiliated to Bharathidasan University, Tiruchirappalli), Panjapur, Trichy, Tamilnadu. Email:rsugusakthi1974@gmail.com

Received: 26th September 2021
Revised: 19th October 2021
Accepted: 17th November 2021

Abstract- Association rule mining is a data mining technique in which pattern of occurrences of one set of items with another set of items in databases of transactions are discovered as rules of implications with certain measures of interestingness. Support or the frequency of occurrences of sets of items and confidence are the most widely used measures of interestingness of association rules of the form X-Y where X and Y are disjoint sets of items. Though the problem of association rule mining emerged from analysis of market basket data in supermarket there are numerous areas of applications of association rule mining technique. In this paper, an association rule based model for identifying the covid19 patients and their symptoms by the doctor, giving the treatment to the patients. For this the symptoms carried by the patients of the given treatment are converted to a set of transactions and then a data base of such transactions is prepared for discovering the association rules in such a way that the antecedent of a rule represents the affected patients of the covid19 treatment. Such rules once discovered can be used for various purposes for the society. It will helpful for health people for proper planning related to identify the covid19 patients and creation of need based treatment respectively.

Keywords- **Mining, Association, technique, treatment covid 19**

1 Introduction

To data growth creasing in many application areas such as health care, business intelligence, finance, retail, bioinformatics, telecommunication, Web search engines, social media, and digital libraries. Data mining chances the needs for scalable, flexible and efficient data analysis in today's information age. Data mining also known as knowledge discovery (KDD) (1 from data it can be considered as a natural evolution of information technology and a confluence of several disciplines they are machine learning, statistics, information retrieval, pattern recognition, and bioinformatics and application fields. It is the process of discovering interesting patterns that represent knowledge and are novel, potentially useful, and easily understandable by humans from large amounts of data. Data mining, as a knowledge discovery process, usually includes an iterative sequence of the following steps: data cleaning, data integration, data selection, data transformation, pattern discovery, pattern evaluation, and knowledge presentation. It can be performed by any kind of data, transactional data, data warehouse data, database data, advanced data types that have versatile forms, structures and rather different semantic meanings, such as Web data, engineering design

data, data streams, hypertext and multi-media data, graph and networked data, spatial and spatiotemporal data, time-related or sequence data as long as the data are meaningful for the target application

There are numerous techniques to specify the kinds of patterns to be discovered in data mining tasks [2]. In general, such data mining tasks are classified into two major groups: descriptive and predictive. Descriptive tasks characterize the properties of data in a target dataset. Predictive tasks carry out induction on the available data to make predictions. The main techniques used in data mining tasks are classification and regression; characterization and discrimination; cluster analysis; the mining of frequent patterns, associations, and correlations; association rules; outlier detection; sequence analysis; time series analysis, sentiment analysis, social network analysis; prediction; and text mining [3]. As a highly application-driven discipline, data mining incorporates many techniques from the other research areas such as artificial intelligence, machine learning, database systems, data warehouses, information retrieval, high-performance computing, statistics, pattern recognition, and visualization. Consequently, the interdisciplinary nature of data mining development significantly contributes to the success of data mining and its extensive application domains.

Association rule mining is a procedure which is meant to find frequent patterns [4], correlations, associations, or causal structures from data sets found in various kinds of databases such as relational databases, transactional databases, and other forms of data repositories. Association rule mining process basically consists of two steps (1) finding all the frequent item sets that satisfy minimum support threshold and (1) generating strong association rules from the derived frequent item sets (5] by applying minimum confidence threshold Association rules are if and then statements that used to reveal relationships between uncorrelated data in a database, relational database of other information repository. It is used to extract the relationships between the objects data which are frequently used together. Applications of association rules are basket data analysis, storage planning etc. For example, if the customer buys milk then he/she may also buy bread. There are two significant measures that association rules uses, support and confidence. It describes the relationships and rules created by studying data for frequently used if and then patterns. Association rules are generally required to satisfy a user-defined minimum support and a user-defined minimum confidence.

Support: Support defines the transactions that contains item set. If p, q are two item sets, then the support can be defined as the transaction T which defines p/q.

Confidence: Confidence is defined as the percentage of transactions where the item sets are most likely to occur. If p, q are two item sets, then, the probability p U q is a subset of transaction, T is called as the confidence.

Frequent patterns algorithms are Apriori algorithm Frequent [6] pattern growth algorithm and Eclat algorithm. These algorithms are used to generate rules on associated attributes.

Apriori algorithm is a two stage process. First, the candidate item set generation and second, the rule generation. Before starting the working procedure of apriori algorithm, the minimum support Pis defined by user. Apriori algorithm starts by scanning the complete database, D and find all the frequent items from the database D. First scan the complete database only for 1-item sets, and then successive iterations deals the 2-itemset. Thus new list of frequent items are created. The process continues until all the frequent item sets are extracted from D. Only those frequent items whose minimum support is greater than or equal to P is taken for rule generation.

II Related Work

The objective of this approach is to discover designs that were not clearly known. Old-style data mining techniques have focused largely on noticing the statistical correlations between the items that are more frequent in the transaction databases. Generally, more than a few applications are using data mining in different fields like medical, marketing and so on. Many methods and techniques have been developed for mining the information from the databases. In this work, [7] an efficient method for item set mining on utility and frequency based model and association rule mining.

Mining frequent item sets (Fls) is an significant problem [8] in the field of data mining, and thus there have been many different approaches planned to solve this problem. However, mining frequent item sets usually works on binary databases and has a constraint that is only concerned with the appearance of items regardless of their importance. In practical applications, items often have different importance depending on their meanings, and that

leads to the emergence of weighted databases. In this work the author introduce a new method for mining frequent weighted item sets (FWIs) from a weighted.

The method initially extracts frequent item sets for each zone using existing distributed frequent pattern mining algorithms. It also associates the time efficiency of Map Reduce based frequent pattern mining algorithm with Count Distribution Algorithm and Fast Distributed Mining algorithms. It presents new approach to identify consistent and inconsistent association rules from sales data located in distributed environment and overcomes the main memory bottleneck and computing time overhead of single computing system by applying computations to multi node cluster. Now the association generated from frequent item sets are also large that it becomes complex to analyze it. Thus, the Map Reduce based consistent and inconsistent rule detection (MR-CIRD) algorithm [9] is to detect the consistent and inconsistent rules from big data and provide useful and actionable knowledge to the domain experts.

This online generated data is too big that it becomes very complex to process and analyze it using traditional systems which consumes more time. This work overcomes the main memory holdup in single computing system. There are two major goals of this paper. Big sales dataset of AMUL dairy is pre processed using Hadoop Map Reduce that convert it into the transactional dataset. Then, after removing the null transactions; distributed frequent pattern mining algorithm MR-DARM -Map Reduce based Distributed Association Rule Mining 101 is used to find most frequent item set. Lastly, strong association rules are generated from frequent item sets. In this work also compares the time efficiency of MR-DARM algorithm with existing Count Distributed Algorithm (CDA) and Fast Distributed Mining (FDM) distributed frequent pattern mining algorithms.

The doubt in database can be handle with the assistance of U-Prefix Span algorithm [11]. This algorithm is functional on the database for finding out the frequent sequential patterns by providing the minimum support. As Database is very large, the number of frequent sequential patterns is also large which is difficult to study. Hence there is essential to apply the Top K Rule Mining algorithm on the generated frequent sequential patterns to obtain the top K rules. According to the generated rules the prediction can be done. The result and analysis demonstration that the U-Prefix Span takes more time for execution than Top K rule mining algorithm.

The growing E-tourism systems provide [12] intelligent tour recommendation for tourists. In this sense, recommender system can make personalized proposals and provide satisfied information associated with their tour cycle. Data mining is a proper tool that extracting possible information from large database for making strategic decisions. In the study, association rule analysis based on FP-growth algorithm is applied to find the association relationship amongst scenic spots in different cities as tour route recommendation, In order to number out valuable rules. The scheme was evaluated on Wangluzhe cultural tourism service network operation platform (WCTSNOP), where it could verify that it is able to quick recommend tour route and to rapidly enhance the recommendation quality.

This work is based on [13] automobiles study and will help the sellers and customers in making decisions. The objective is to find the important selling factors that affect the relevant sale vehicles by using the association rule mining algorithm. Most famous algorithm of association rule mining is Apriori is used for knowledge discovery. Research work will improve the existing Apriori algorithm and will reduce some of the drawbacks of the existing algorithm.

This work proposes FDM, a new algorithm based on FP-tree and DIFF set data structures for efficiently discovering frequent patterns in data. FDM [14] can adapt its characteristics to efficiently mine long and short patterns from both dense and sparse datasets. Several optimization techniques are also outlined to increase the efficiency of FDM. An evaluation of FDM against three frequent item set data mining algorithms, d Eclat, FP-growth, and FDM (FDM without optimization), was performed using datasets having both long and short frequent patterns. The experimental results show significant improvement in performance compared to the FP-growth. D Eclat, and FDM* algorithms.

In this paper improved Apriori algorithm 1151 which will help in reducing multiple scans over the database by cutting down unwanted transaction records as well as redundant generation of sub-items while pruning the candidate item sets. The performance of this algorithm is analyzed against the FP Growth algorithm in which there is no generation of candidate set.

This paper presents a load balancing technique designed specifically for parallel publications applications running on multicore applications. This [16] architecture provides a hardware parallelism through cores inside the CPU. It increased performance low cost as compare to single core machines attracts HPC high performance computing connectivity.

A distributed association rule mining algorithm on Spark named as Adaptive-Miner which uses adaptive approach for finding frequent patterns with higher accuracy and efficiency. Adaptive-Miner uses an adaptive strategy based on the partial processing of datasets. [17] Adaptive-Miner makes execution plans before every iteration and goes with the best suitable plan to minimize time and space complexity Adpative-Miner is a dynamic association rule mining algorithm which change its approach based on th nature of dataset. Therefore, it is different and better than state-of-the-art static association rule mining algorithms and conduct in-depth experiments to gain insight into the effectiveness, efficiency, and scalability of the Adaptive-Miner algorithm on Spark.

III. Proposed Method

In this work association rule mining technique is applied for mining of covid19 [18] patients for treatment in sets of data about symptoms. The proposed model for applying association rule mining technique is described below. The data regarding the symptom condition for various patients is represented as treatment (transactions). Then the frequent item sets (symptoms) and the association rules of the form X-Y where X represents the set of symptoms for the treatment represented by Y are discovered with respect to the pre specified minimum confidence and minimum support. Thus the association rules are discovered in a selected manner with additional constraint on the consequents. The item sets appearing in the antecedent represent the symptoms of the patients for the treatment appearing in the consequent. The support for the corresponding association rule gives the frequency of occurrence of the rule containing the symptoms and the corresponding treatment in the whole database.

In the following, discovery of the association rules are shown with the symptoms criteria on the antecedents and the corresponding treatment on the consequents of the rules. A sample data set is prepared by using the list of treatment and names of symptoms are shown in table 1 and table 2 respectively

Table 1. is the transaction database which have sample 10 Patients treatment details (transaction) with symptoms(items)

Table 1 Transaction Database

Transaction (treatment)	Items (symptoms)
T1	I1, I3, I7
T2	I2, I3, I7
T3	I1, I2, I3
T4	I2, I3
T5	I2, I3, I4, I5
T6	I2, I3
T7	I1, I2, I3, I4, I6
T8	I2, I3, I4, I6
T9	I1
T10	I1, I3

By using the above table 1 we calculate the size of the transaction. It is available on table 2.

Table 2 Symptoms Database

Items (symptoms)	Symptoms Details
I1	Cough
I2	Fever
I3	Cold
I4	Wheezing
I5	Throat pain

I6	Blood Presser (BP)
I7	Sugar

By using the above table 2 have the symptoms with explanation for the patients. It is available on table 2

Table 3 Transaction Database with size

Transaction (treatment)	Items (symptoms)	Size
T1	I1, I3, I7	3
T2	I2, I3, I7	3
T3	I1, I2, I3	3
T4	I2, I3	2
T5	I2, I3, I4, I5	4
T6	I2, I3	2
T7	I1, I2, I3, I4, I6	5
T8	I2, I3, I4, I6	5
T9	I1	1
T10	I1, I3	2

The table 3 gives the information about the number of items scanned to get 1 frequent itemsets.

Table 4 Transaction Database with support

Items (symptoms)	Transaction (treatment)	Support
I1	T1, T3,17,T9,T10	5
I2	T2,T3,T4,T5,T6,T7,T8	7
I3	T1,T2,T3,T4,T5,T6,T7,T8,T10	9
I4	T5,T7,T8	3
I5	T5	1
I6	T7,T8	2
I7	T1,T2	2

The table 4 contains items, item sets whose support min sup are eliminated or removed from the database

Table 5 Final Transaction Database

Transaction (treatment)	Items (symptoms)	Size
T5	I2, I3, I4, I5	4
T7	I1, I2, I3, I4, I6	5
T8	I2, I3, I4, I6	5

The table 5 contains the frequent symptoms of the patients affected by the disease and its respective support count of the treatment (transactions) from the database.

Table 6 Frequent 3 items

Transaction (treatment)	Items (symptoms)	Size
T5	I2, I3, I4, I5	4
T7	I1, I2, I3, I4, I6	5
T8	I2, I3, I4, I6	5

Based on the above process to find the frequently affected symptoms of the treatment given to the patients from the transaction database. The table 6 contains frequent symptoms and its size, association rules are generated from non-empty subsets which satisfy minimum support value. The results informs the doctors, the following symptoms of 12, 13 and 14 the patients affected (fever, cold and Wheezing) by the covid19 disease.

IV Conclusion

In this work, an association rule based model for discovering the symptoms criteria for treatment is proposed from related large datasets as part an information system based covid19 data analytics. The dataset for the treatment and its symptoms are designed as transactions. The symptoms are represented in the antecedents of the association rules while the relevant treatment is represented in the consequent. Further, criteria for classification of rules are defined for analyzing the discovered rules based on an input set of treatment for finding covid19 and also for finding people with different sets of symptoms for the treatment.

References

[1] Wensheng Gan, Jerry Chun-Wei Linn, Philippe Fournier-Viger, Han-Chieh Chaoa, Justin Zhan, "Mining of frequent patterns with multiple minimum supports", Engineering Applications of Artificial Intelligence 60 (2017) 83_96, 2017.

[2] Addi Ait-Mlouk , Tarik Agouti, Fatima Gharnati, "Mining and prioritization of association rules for big data: multi-criteria decision analysis approach", J Big Data (2017) 4:42., 2017.

[3] Yuan Mana. "Feature Extension for Short Text Categorization Using Frequent Term Sets", 2nd International Conference on Information Technology and Quantitative Management, ITOM 2014. Procedia Computer Science 31,2014.

[4] Arthur Zimek, Ira Assent and Jilles Vreeken, "Frequent Pattern Mining Algorithms for Data Clustering". J. Han (eds.), Frequent Pattern Mining, DOI 10.1007/978-3-319-07821-2_16, © Springer International Publishing Switzerland 2014.

[5] Miss Jainee Patel, Mr. Krunal Panchal, "Frequent Item-sets Based on Document Clustering Using k-means Algorithm", Vol-2 Issue-3 2016 IJARIIE-ISSN(Q)-2395-4396, 2016.

[6] Walaa N. Ismail and Mohammad Mehedi Hassan, "Mining Productive-Associated PeriodicFrequent Patterns in Body Sensor Data for Smart Home Care", Sensors 2017, 17, 952; doi:10.3390/s17050952 www.mdpi.com/journal/sensors, 2017.

[7] Paul P Mathai, R.V. Siva Balan and IerinBabu, "An Efficient Approach for Item Set Mining Using Both Utility and Frequency Based Methods", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 12 (2017) pp. 3470-3473, 2017.

[8] Huong Bui, Bay Vo, Ham Nguyen, Tu-Anh Nguyen-Hoang, Tzung-Pei Hong , "A Weighted NList-Based Method for Mining Frequent Weighted Itemsets", An International Journal of Expert Systems With Applications, 15 October 2017.

[9] R Rampriya, Nivetha, SwethaSri , "Mapreduce Based Pattern Mining Algorithm In Distributed Environment", International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2017 IJSRCSEIT | Volume 2 | Issue 5 | ISSN : 2456-3307, 2017.

[10] Dinesh J. Prajapati , " Comparative Study Of Distributed Frequent Pattern Mining Algorithms For Big Sales Data", International Journal of Advanced Research in Engineering and Technology (IJARET) Volume 8, Issue 1, January- February 2017.

[11] JayshriBanpurkar, Amreen Khan, "Mining Frequent Sequential Patterns and Top Rules from Large Uncertain Database", International Research Journal of Engineering and Technology (IRJET) C-ISSN. 2395-0056, Volume: 03 Issue: 05 May-2016,

[12] FANG Huia, CHEN Chongcheng, LIN Jiaxiang, LIU Xianfeng, FANG Dong, "Association Rule Analysis For Tour Route Recommendation And Application To Wctsnop", The International Archives of the Photo grammetry, Remote Sensing and Spatial Information Sciences, VolumeXLII-2/W7, 2017 ISPRS Geospatial Week 2017, 18-22 September 2017, Wuhan, China, 2017.

[13] Dr. Gurpreet Singh, Er. Sonia Jassi, "Implementation and evaluation of optimal algorithms for computing association rule learning", International Journal of Engineering And Computer Science ISSN: 2319-7242 Volume 6 Issue 7, Page No. 22128-22133 July 2017.

[14] George GATUHA, Tao JIANG, "Smart frequent itemsets mining algorithm based on FP-tree and DIFF set data structures", Turkish Journal of Electrical Engineering & Computer Sciences, 2017 Implementation and

[15] Sangita Chaudhari, Mayur Borkhatariya, Apurva Churi, Mohini Bhonsle, "Analysis of Improved Apriori Algorithm", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 5, Issue 2, March 2016

[16] PrantikPancholi, ShitalKhairnar, Jyotikamble, AmolJadhao, " MACH: Performance Enhancement in Multi-core Processor using Apriori Algorithm with file Chunking", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056, Volume: 03 Issue: 04 | April-2016.

[17] Sanjay Rathee, Arti Kashyap, "Adaptive-Miner: an efficient distributed association rule mining algorithm on Spark", J Big Data (2018) 5:6 https://doi.org/10.1186/s40537-018-0112-0, 2018.

[18] Seyed Mohammad Ayyoubzadeh, Seyed Mehdi Ayyoubzadeh, Hoda Zahedi, Mahnaz Ahmadi, Sharareh R. Niakan Kalhori, "Predicting COVID-19 Incidence Using Google Trend and Data Mining Techniques: A case study of Iran (Preprint)", JMIR Public Health and Surveillance on: March 21, 2020