Efficient Centralised System for Biomedical Data Collection and Prediction

Shubham Lungare[1], Parul Singh[2], Shubham Saxena[3], Sameer Keluskar[4], Prof. Darshika Lothe[5]
[1, 2, 3, 4, 5] Department of Computer Engineering
[1, 2, 3, 4, 5] JSPM's Imperial College of Engineering and Research Wagholi Pune Maharastra

Abstract- *Over the past two decades, the epidemic of various diseases has challenged the public health community to fundamentally rethink and redesign the framework for calculating preventive measures regarding these infectious diseases. This paper is special issues which is focusing on the collection of biomedical information (in the terms of specimens, measurements and questionnaire responses) from members across the globe and provide them a system which is capable of giving various precise representation of information.*

Keywords- Data Mining Concepts, Information and Communication Technology, Security.

I. INTRODUCTION

As the field's data mining progresses each day, we can apply data mining tools and techniques in various fields such as marketing, ecommerce etc. In the same way Biomedical Field is not far away from this technology. Advancement in Biomedical Informatics and Biomedical Engineering provides the foundation for modern patient-centred health care solutions, health care systems, technologies and technique. We are going to see such system in this paper.

II. SYSTEM ARCHITECTURE

The proposed architecture for the system consists of three different levels, which are described as follows:

- Database: A database is nothing but an organized collection of data. It is the collection of schemas, tables, queries, reports, views and other objects. In this system we are having two different databases at two different levels.
- Local Database: This Database is used to store medical records (like age, diseasediagnosed, medicine prescribed, number of sittings required effect and side-effects
  of the medicine) of the patients received hospital wise
- Global Database: This Database is going to aggregate the data coming from all other local databases and store them.
- Centralised Processing Cum Data Mining Engine: This level is like an intermediate between the database level and user level. Due to the data sensitivity and privacy issues of the system, we need to have some safety precautions before we publish the data for the users. Therefore we need to integrate health records without revealing individual identities of patients for other users except their personal doctor without compromising with the quality of information provided.
- User Level: The front end or the user end cannot be unified as it contains different end-user hardware as well as access rights. The amount of information a scientist or a researcher will see will be quite different from the doctor in charge of the patient.

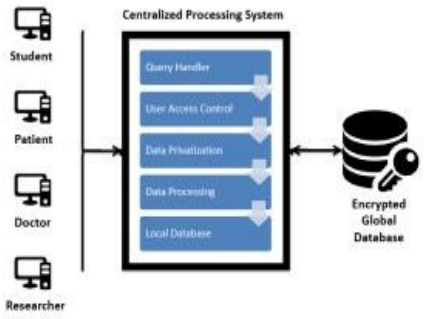

Fig. 2.1 Proposed System Architecture

The access levels of reading or editing the information from the website will also be controlled by the centralized administration. The appearance of the website over computer or mobile screens will also have to be looked after. Local system security like antivirus or antimalware security and scheduled system updates on devices are supposed to be looked after as well.

More Detailed perspective of the module which helps in the prediction of various kinds of trends in the system. As a result it produces various kinds of charts, pie diagrams and histograms.

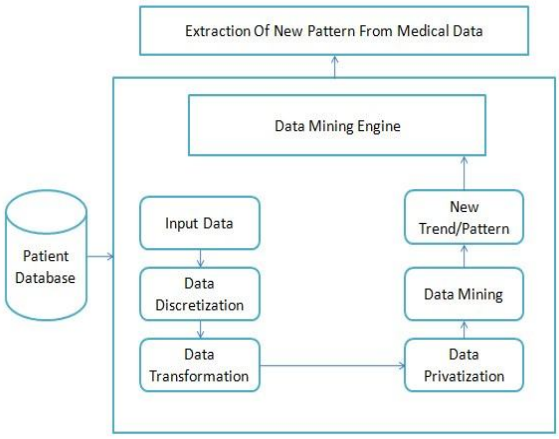

Fig. 2.2 Detailed View of Data Mining Engine

Data Mining Engine helps in finding new trends from the existing system by applying various data mining algorithms such as Apriori and FP Growth. The figure given below depicts the various internal modules of this engine.

III. ALGORITHM APPLIED

While Development of this system various algorithms are used to predict data in various formats such as donut charts, pie charts, line graphs and histograms etc.

A. Apriori Algorithm

Apriori [1] is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database.

Similarly in this system we are going to predict disease that the patient may have in forth coming years by analysis the previous transaction/records of the database. For example if we need to predict a patient that he/she may going to get suffer from diabetes in next four years the we need to consider following concepts.

While Apriori taken into consideration we need to focus on two different concepts such as confidence and support.

1. Support is an indication of how frequently the items appear in the database.
2. Confidence indicates the number of times the if/then statements have been found to be true.

Support can be calculated by counting the number of patients suffering from diabetes out of total number of patients. If we are getting a high support count then we proceed further for calculating its confidence. That can be done by comparing the patient symptoms with the previous supported patients.

Pseudo-code:

C_k: Candidate itemset of size k
L_k : frequent itemset of size k

L_1 = {frequent items};
for (k = 1; L_k !=∅; k++) **do begin**
  C_{k+1} = candidates generated from L_k;
  for each transaction t in database **do**
    increment the count of all candidates in C_{k+1} that are contained in t
  L_{k+1} = candidates in C_{k+1} with min_support
  end
return $\cup_k L_k$;

Fig.3.1 Pseudo-code for Apriori Algorithm

If the system get a confidence of more than 75%, then system is able to conclude that the person is likely to have this disease in next 3 years provided that his/her symptoms persists.

B. FP-Growth

The FP-growth algorithm [3] is right now, one of the quickest approach to mine frequent item set s. It relies on a prefix tree representation of the transactions in the given database (called an FPtree), which is economically feasible for storing the transactions.

The basic idea of the FP-growth algorithm can be explained as a recursive elimination scheme: in the pre-processing step, delete all items from the transactions that are not frequently occurring independently, i.e., do not appear in a user specified minimum count of transactions. Then choose all transactions that contain the least frequent item among those that are frequent & delete this item from them. Repeat this process over the obtained reduced (also known as projected) database, notice that the item sets found in the recursion share the omitted item as a prefix. After that, remove the processed item also from the database of all transactions and restart,

i.e., process the next most frequent item and so on. In these steps the prefix tree, which grows by links between the branches, is studied to quickly look for the transactions containing a given item and also to omit this item from the transactions after processing.

In this paper we describe a simple implementation of this algorithm for mining of different age groups that are being affected by *frequent diseases*. Pseudo-code of FP-growth is as follows:

Input: constructed FP-tree Output: complete set of frequent patterns Method: Call FP-growth (FP-tree, null).

Procedure FP-growth (Tree, α) {

1) If Tree contains a single path P then
2) For each combination do generate pattern β α with support = minimum support of nodes in β.
3) Else for each header a in the header of Tree do {
4) Generate pattern β = a α with support = ai.support;
5) Construct β.s conditional pattern base and then β.s conditional FP-tree Tree β.
6) If Tree β = null
7) Then call FP-growth (Tree β, β)} }

Now we are going to discuss about some access control policy that is going to be used in this system for assigning various access policies to various users.

A. RBAC: Role Base Access Control

With RBAC, roles can be well understood by their names, and they determine the sets of permissions to be granted to users. In addition, it's easy to audit which users have access to a given permission and what permissions have been granted to a given user. A limited number of roles can represent many users or user types, and roles can be assigned to users by non-expert personnel. However, roles must be engineered before RBAC can be used. Furthermore, RBAC must be constrained to handle dynamically changing attributes, such time of day and location. Core RBAC can't handle such attributes.

IV.EFFICIENCY OVER OTHER SYSTEMS

As per our survey those system which is out there supporting biomedical fields are working independently for their organisation and just providing simple facilities to various clients such as scheduling appointment and collection of report once the job is done.

But this system is connected to a distributed network of hospitals which is handling numerous users such as patients, doctors, researcher, pharmacists, and other medical agencies which are benefited with this system.

This system does not rely only on providing appointments to the patients but it does following things:

- Patients Registered over this network can gain access to hospitals all over the globe for their treatment.
- Patients also can see what other chronic disease they may have in next 3 years.
- Doctor can study about various other medical cases to enhance their knowledge without getting any exposed data of patient such as his/her identity.
- Researchers can view variety of research models as per the historic database transaction and able to conclude results for their research.
- Pharmacists are able to view trends in medicines, which encourage them for building new pathological solutions.

V. VIEWS OF PROPOSED SYSTEM

Screenshots of proposed System are as follows:

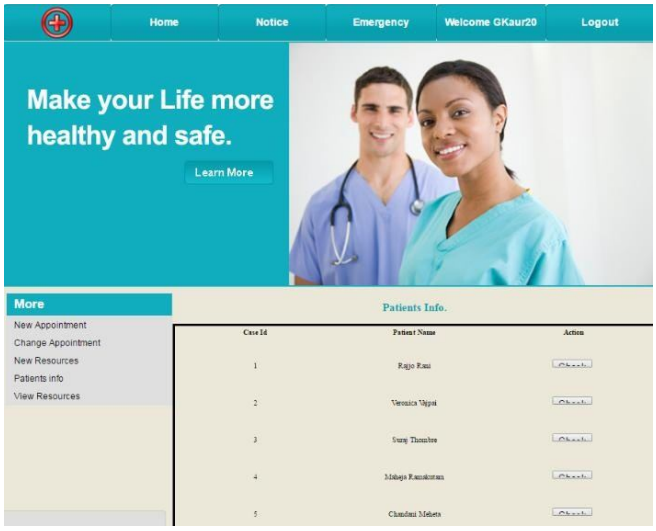

Fig. 5.1 Profile Management

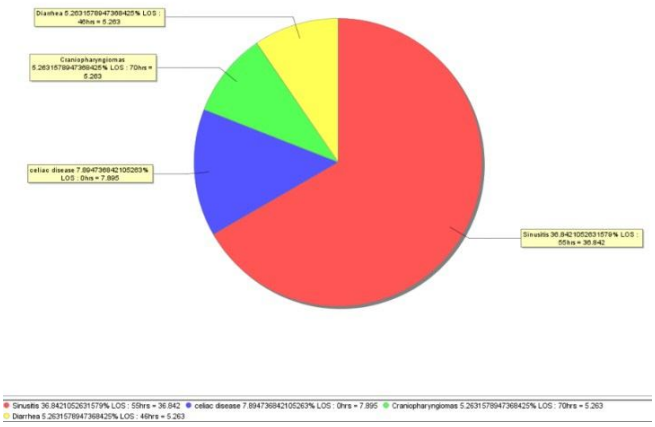

Fig 5.2 Disease Prediction

Fig 5.3 Appointment Handler

VI. CONCLUSION

Here, We have proposed a centralized database system, which is having algorithms like frequent pattern tree (FP-tree), for keeping compact, important information about frequent patterns of diseases trending among different age groups, and have introduced algorithms for confidentiality by which identity of any person will not be revealed to others. The system as described will provide services to various users, medical practitioners, researchers.

VII. FUTURESCOPE

Mobile has grown significantly in recent years, with Google now saying that mobile search has surpassed desktop search. Migrating this framework to mobile technologies helps in finding greater possibilities to Biomedical Field.

REFRENCES.

[1] Srinivasa K G * , Venugopal K R 1 and L M Patnaik ction using Fuzzy C - Means Clustering for Data Mining Systems". IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.3A, March 2006 Alexey Tsymbal 1,3 , Seppo Puuronen 1, Mykola Pechenizkiy 2, Matthias Baumgarten 3 , David Patterson 3, "Eigenvector-based Feature Extraction for Classification". FLAIRS2002

[2] Yu Tao, Vallipuram Muthukkumarasamy, Brijesh Verma and Michael Blumenstein, "A Texture Feature Extraction Technique Using 2D–DFT and Hamming Distance". Compuatational intelligence and multimedia appliactions ,2003.

[3] Fabian M¨orchen ∗, "Time series feature extraction for data mining using DWT and DFT (November 5, 2003)".

[4] Ying Zhao and George Karypis C, "Comparison of Agglomerative and Partitional Document Clustering Algorithms ".

[5] Lu-Chou Huang , Huri-Chung Chu,Chung-Yueh Lien, Chi-Hung Hsiao,Tsair Kao,Privacy preservation and information security protection for patients portable electronic health records , Computers in Biology and Medicine,vol.39.

[6] Personal information protection and Electronic Documents Act, Second Session , Thirty-six Parliament , 48-49 Elizabeth II ,1999-2000 , Statutes of Canada 2000.