Topic Paper

Human Computer Interaction for Data Mining in Medical Registers

By:Supriya Bachal (Graduate Student 15)

INDEX

Abstract:

This paper projects a proposal for building a system which applies the KDD-HCI integrated approach for visualization of data stored in medical registers.The KDD-HCI approach is one that tries to rea benefits from both these very vast domains and combine them together to get plausible results.There is a dire need for computational intelligence in the field medicine because of the sheer complexity of data that it produces.KDD brings to the table data mining and machine learning algorithms,while HCI helps to visualize the results of these algorithms and present them to the user in a simplified lucid way.The proposed system uses WEKA environment .The objective of this proposal is to bridge the gap between the medical practitioners and analyst using Human Computer interaction in the form of visualization of data.

Index Terms

Human Computer Interaction(HCI), Knowledge Discovery in Databases(KDD), ,temporal visualization,volumetric image visualization.

1.Introduction:

Medical Registers are records of patients that are maintained for patient aftercare,preventive medical purposes,risk listings etc.With the advancement in technology these medical records are now stored digitally and are called Electronic Health Records(EHR) or Electronic Medical Records(EMR).These records form a huge historic database in a medical organization and can be mined to recognise patterns and perform diagnosis,they can also be used for the study of effect of drugs on a patients and monitor the general health of the community. It is no secret then that these databases have potential for knowledge.However these databases contain raw, unstructured, complex, multidimensional data which can be referred to as big data. The process of treating and analyzing this data such as to extract valuable information from it is called Knowledge Discovery in a Database This process is often undertaken by a statistician or a data scientist who is trained to analyze data. The end user of this analysis is however usually a technical novice from the medical field or the patient.

2.Background and related work:

Medicine has always been a major area of research.The importance of medical research cannot be justified in words.Medical research saves lives and is an undying field.Change in atmospheric conditions and environment as well as evolution of human beings leads to new diseases in the society. To treat these diseases is is necessary to identify the cause of the disease as well as to identify drugs that can treat the disease.It is also evident from many studies that some diseases

are called tropical diseases and are prevalent in only some parts of the world. Since there is a lot of data that medical science has about various cases it is possible to identify patterns in this data through analysis.Many statisticians and data analysts are working towards discovering trends in data from medical registries. However to interpret the trends that they see and to know whether any valuable information can be explored from the trend or pattern , they need knowledge in the field of medicine, on the other hand medical practitioners have knowledge about the interpretation of the trends but they do not know how to analyze such a huge amount of data.To solve this problem KDD-HCI integrated approach is used. The KDD part is done by the statististician or the data analyst, using different algorithms then the results of the analysis is converted into a simplistic user friendly visual Interface using HCI so that the medical practitioner can interpret the analysis.

There are a few terms that are used in this system multiple times this section defines the context of these terms;

Datamining:

Data Mining is defined as the process of working with data, it is a process which makes data useful and mines information from it to reach logical solid conclusions using various methods as shown below.Data for the system can be taken from the Swedish cancer register can be used for the pilot study.

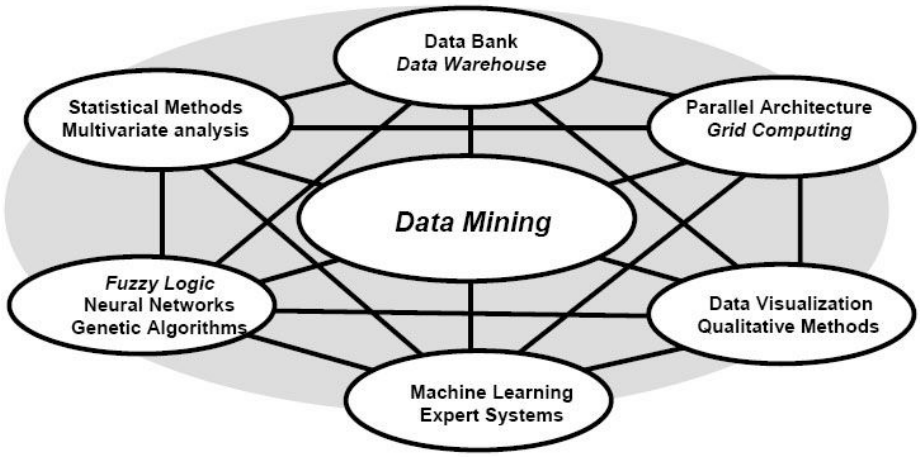


Fig. 2. Data mining process.

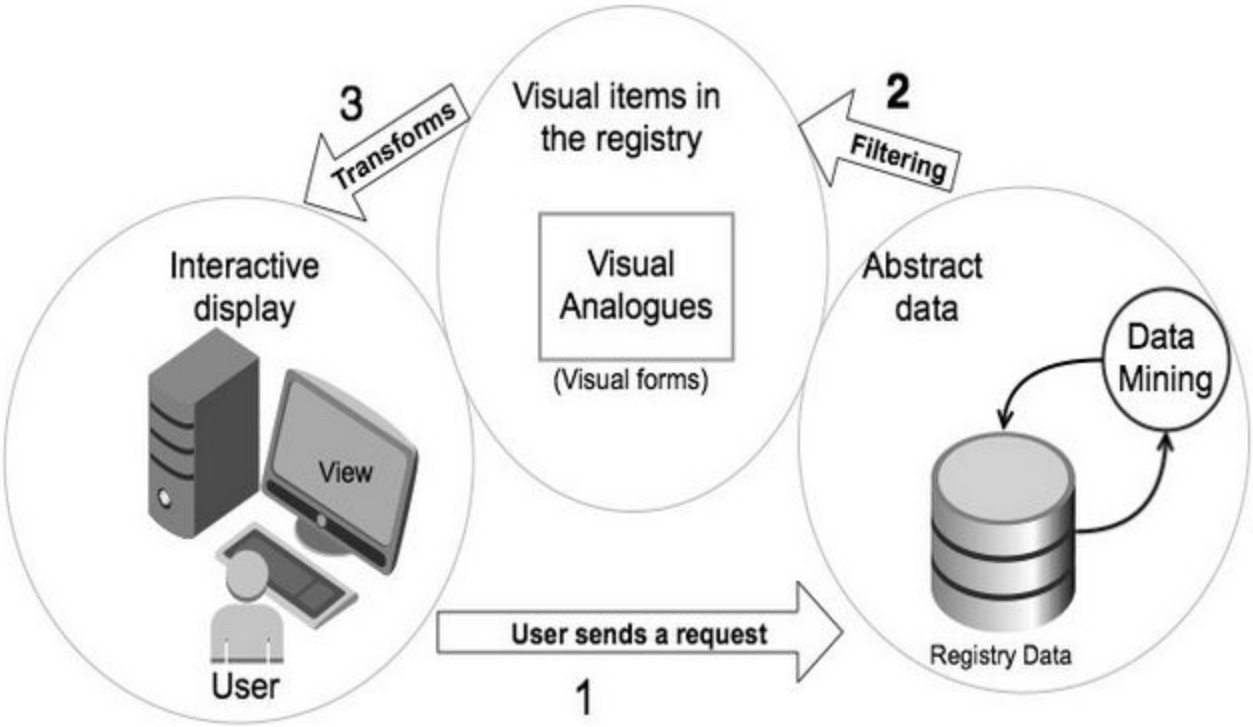
Data visualizations:

Data visualizations make it easier for the person to understand what the analysis of the data has derived.Under the domain of cognition, the term visualisation is referred to as the construction of visual image in the mind aiming at forming a mental model of the data that is being analysed

(Ware, 2000).Elementary visualizations include pixel based visualizations, graph based visualization and geometric based visualizations.In this system apart from these temporal visualization, volumetric image visualization are types of data visualizations are also used .For the visualization of temporal data there are many techniques such as wall diagrams,tabular representation,star representation, spiral representation,concentric circles technique,lifeline technique etc.Temporal representations are done at every stage of the data mining process so that for future reference they are available.In the case of Volumetric Image Viewing techniques we have a few standard techniques such as ;multiplanar reformation (MPR), surface rendering (SR), and volume rendering (VR).

3. The Proposed system:

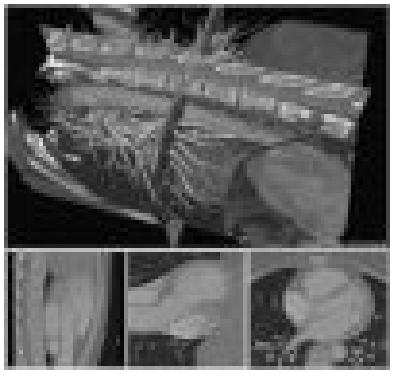
The proposed system may use Waikato Environment for Knowledge Analysis (WEKA) for data mining or apply the algorithms directly to the dataset using a programming language such as java. WEKA is an open source software and can be installed through GUI or command line.The data for the proposed system can be loaded by uploading files or using Universal Resource Locators(URLs). The proposed system is an integration between KDD and HCI providing a platform for optimum utilization of data and ease of use of the results. The overview of the system is given below.



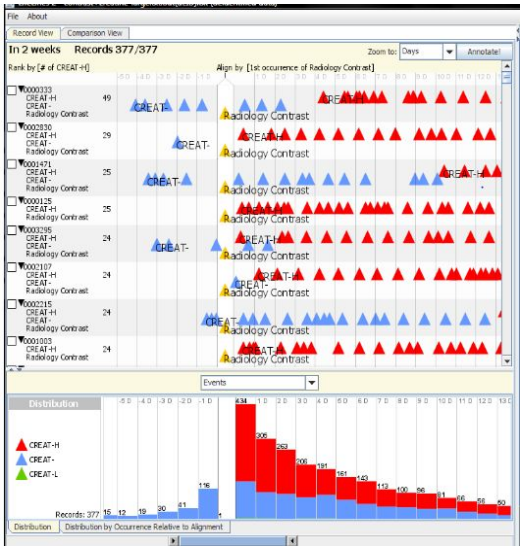
The proposed system highlights the following functionalities:

1.It is always advisable to use a large data from varied sources for statistical analysis to gain acceptable results.To increase the accuracy of the results medical data from several centers nationally and internationally is analyzed.Since data is collected from various sources data preprocessing is necessary in order to address problems like probability, uncertainty, incompleteness, vagueness, noise, etc.Hence the automated approach cannot be implemented directly by the medical practitioners. The complexity of machine learning algorithms keeps the medical practitioners away from implementing them in their research workflow.They however have expert domain knowledge and thus can interpret the patterns from the data and use this information for their research.Currently the data is sent away to the statistician who runs the analysis on the data and reports the analysis in the form of a file having rows and columns which is cognitively tiresome for the medical practitioner to look at.If the domain expert can be made a part of the knowledge discovery process it would lead to better results.Unlike other applications analysis of medical data, detailed and explicit medical expert knowledge and knowledge of clinical processes is urgently required.The clinical expert is provided with a report of the analysis. Designing a simple user friendly interface for the domain expert to get involved and interact with the KDD process in real-time is necessary.

2.The two most popular imaging techniques for 3D imaging for diagnostics are Magnetic Resonance Imaging(MRI) and X-ray Computer Tomography (CT) .Traditionally the results from both these techniques is converted to sliced 2D images and displayed on a flat screen. However now through the development of new visualization techniques and Interfaces it is possible to project these images as 3D images. The dataset produced from the imaging is a high-resolution isotropic three- or four-dimensional dataset. The visualization of such data can be implemented using volumetric image visualization techniques which are supported by the GPU(graphical processing unit)..Image interaction plays a major role in this application, navigation (scrolling through stack of images in different planes) and manipulation (adjusting contrast setting, rotating 3D models) are the two most important applications of HCI in this context.To rotate or scroll the image a 3D mouse can be used as an input device. The 3D mouse has 6 DOF, 3 translational and 3 Rotational . Computed tomography (CT) and magnetic resonance imaging (MRI) are viewed as a stack of images (volumetric image) instead of single images printed next to each other.It will also help medical students understand the human anatomy better and make it convenient for the medical practitioner to look at the data.This specific feature can be used for making decisions related to surgeries such as whether a endoscopic surgery is sufficient or it is necessary to perform an invasive surgery.HCI thus helps surgeons make important decisions.



3. The integration of temporal visualization techniques for KDD.Temporal visualization systems are systems where the periodicity, and dimensionality of data is taken into account.This type of analysis can be used to identify symptoms of an event before it occurs eg;heart attack or for clinical policy making.Even though temporal data is the most informative form of data it is overwhelming for medical practitioners to deal with it as it has many dimensions. It is thus evident that temporal data must be represented properly.The lifelines 2 is a tool that can be used to visualize temporal data it gives a point to point summary of the data that is available,sometimes however it is necessary to get an interval based summary of the data in that case another tool called event-flow can be used for visualization of data. Event-flow helps the user to find temporal patterns of interest and presents them using a graphical interface.It can also perform data transformations and summarize the data that is available.The User of this system can be provided with an option to see the data specific to a certain region or a disease.This feature helps since some diseases occur only in tropical countries.Diseases specific analysis helps to find how medicines act through the course at different times and what are the changes that occur in the body.



4.Human Interpretation Issues

While designing the visualization of data the understanding of human perception should be accounted for, this helps to convert the abstract data into physical attributes.The interfaces will be designed such that clinicians have an opportunity to view the most relevant attributes for the field. It is important to understand that even though different groups of people having different skills look at the same corpus of data hence while designing the interface it is necessary to take into consideration each of their point of views.The data should be presented clearly in such a way that it is intuitive to the users.

User groups:

There are three different groups which are identified as the users of this application.

Clinicians:

This group uses the data that is mined from the system for various research purposes.For the currently prevalent system they do not have direct access to the data but they obtain the data through a statistician,in the form of a file which contains multiple rows and columns. For example to see the possibility of an epidemic through the pattern of symptoms the clinicians will have to request the statisticians for data of the region for the time,the statistician will process this request and the data may or may not be sent immediately.Thus user friendly changes to the system interfaces would improve the interactions with the registry system.Direct access to the data for the clinician will ensure that the system is made convenient for them and they can focus their effort on the research in the field of their expertise.It is important to keep track of the data to illustrate this imagine a scenario where the research look through the mined data in order to predict trends or make a hypothesis ,then it is necessary for them to have an overview of the variables that may be impactful.

Medical students:

For students in the field of medical sciences this system could be used as a learning tool. The mined data could be used for studying the different types of diseases their interdependence and the treatment effects for drugs.They will also be able to understand the how different people have different reactions to different drugs User interface with a good visualization technique will allow a student to explore the relationships between variables and their dependencies.Th through the reports generated. 3D images will help the students understand the human anatomy better.3D imaging actually helps the students to understand the structure of the organs and the intricacies associated with each organ.

The Public:

Traditional systems allow the public to view data in the form of rows and columns. It is a tedious job for the public to make sense of the multiple rows and columns they see.The data should be presented in a form of multi coloured diagrams in reports such that it is easy for the public to understand. The user interface could also ask the user to scroll next and back using a click or the keyboard in order to view the data of the next month/year/period.Providing Search for this system could be beneficial for Public,if they might be interested in a certain kind of diseases or condition then the user can enter the name of the disease or condition and results only relevant to those words will be displayed to the user.The 3D data and temporal data can also be beneficial for the patient to understand how his treatment is works and to educate himself about the disease.

Research - based Personas:

1.Prasad Godbole:

Dr Godbole is a 52 years doctor and a medical researcher at king's hospital in London.He likes reading and is actively involved with social work. He is affable and friendly.

Scenario :

Dr Godbole is in his clinic and he has just examined a patient having severe neck pain.The patient was advised to get an MRI.The MRI machine in the hospital produces an isometric dataset that is used to produce a 3D image of the neck using the system. Dr Godbole rotates the image to find a cyst in the lower portion of the neck. Since the doctor knows where the cyst in located and it is not an area which has too many other sensitive organs ,nerves which could lead to fatal situations, The doctor decides to perform a non invasive procedure for the patient.The patient notifies the doctor that he has to travel Africa due to work commitments while discussing the date of the surgery. The doctor logs into the system and checks if the visual representations for the african medical registries for public health is available. It shows that yellow fever and malaria is most common in Africa for the past few years. The doctor educates the patient on preventive measures on yellow fever and malaria.

2. Zhan Huang:

Zhan Huang is a 32 year old medical student and is currently in India on a community welfare project. He likes to play polo and is interested in art and culture.

Scenario:

Zhang is in India and is dealing with patients having malnutrition. Zhang is treating patients with dietary supplements. It is not possible for the patients to run a test every now and then to test the level of certain nutrients in their body. Zhang cannot start the next phase of treatment until the body has reached a certain level varying for every nutrient. Zhang then logs into the system and loads the database for malnutrition(or whatever relevant database) and checks the temporal representation of the data to see at what time does the level of nutrients reach the level desired in Indian population using the Indian medical Registry.Zhang prescribes a test for the patient at around that time.

5.Challenges:

1. Big-Data

The main challenge for this project is that the data that we are dealing with is a large corpus of multi-dimensional data of various formats. Since the data is coming from various sources the class of data being the same the type of data may change. Thus the database has to be adaptable to store large amounts various types of data.The algorithms being used to preprocess the data have to be powerful enough to create good data for the mining to work well with. The data-mining and machine learning approaches should take into consideration the types of data that are present in the system.Visualization of big data is also a major concern in this proposal.

2. Security Of Data

Firstly not all countries and institutions make their registries public to find a suitable registry for the specified case is a task for the medical practitioner.The security of the data is also a major concern,since the data that we are mining is sensitive data it has be kept extremely secure. If the medical practitioner uploads a file of a patient's records in the database it is necessary that only the medical practitioner has access to this particular data which is not a publicly viewable.

6.Immediate Future scope:

1. Use of touch screens instead of 3D mouse:

For viewing the 3D images we could use a touch screen interactive display to rotate the 3D images,this would be a much better user experience.

2. Newer visualization techniques can be added to the reports generated:

For viewing the 3D images we could use a touch screen interactive display to rotate the 3D images,this would be a much better user experience.

3. Security of Data:

A new feature can be added to the system by means of which when a doctor uploads the patient data into the system, the identity of the patient is dropped but the data is visible to the medical students do that they can use it as a real-time case study.

7.Conclusion:

The use of HCI for medical registers requires in depth understanding of the type of data that has to be visualized as well as the the understanding of the user groups using the data. The system proposed would be beneficial for various sectors of the medical fraternity as well as the general public and helps to mediate between different groups having their own special skill set for research purposes and to make better medical decisions. The system also acts as an educational tool for medical students.

8.References:

1. HCI in Medical Visualization-Bernhard Preim
2. Human Computer Interaction Design for Data Mining in Cancer Registries-Grace B.Kanza
3. The Importance of Human–Computer Interaction in Radiology E-learning Annemarie M. den Harder,corresponding author Marissa Frijlingh, Cécile J. Ravesloot, Anne E. Oosterbaan, and Anouk van der Gijp
4. Interactive knowledge discovery with the doctor-in-the-loop: a practical example of cerebral aneurysms research Dominic Girardi,corresponding author Josef Küng, Raimund Kleiser, Michael Sonnberger, Doris Csillag, Johannes Trenkler, and Andreas Holzinger
5. Simplifying the Exploration of Volumetric Images: Development of a 3D User Interface for the Radiologist's Workplace by M. Teistler,corresponding author1 R. S. Breiman,2 T. Lison,3 O. J. Bott,3 D. P. Pretschner,3 A. Aziz,4 and W. L. Nowinski1.