

Incremental Algorithm For Extraction, Data Visualization And Measurement Of Data Abstraction Of DICOM Image Tags

Manpreet Singh Dhillon
M.TECH (ECE Department)
Chandigarh Group of Colleges,
Landran (Punjab), India

Abstract

Large number of medical images in digital format is generated by hospitals and clinics every day. Such DICOM images constitute an important source of anatomical and practical information for diagnosis of diseases. Now the main task is to interpret the data in such a way that it provides perspicacity, trends and easy visualization by the variations options dynamically. In our investigation, we shall try to make visualization of DICOM images. Various abstraction techniques are used in visualization system to facilitate analysis. Also, it is possible to compare different abstraction techniques by using Bhattacharyya distance measure and HDM to select an abstraction method that meets the requirement of analytic task.

1. Introduction

Health care has become one of the most important services. Hospitals, physicians, insurers, and managed-care firms are networking, merging, and forming integrated organizations to finance and deliver health care. With the technological advancements, the application of computers has grown to a very large extent in almost every walk of life, especially the medical sciences. Whether it is the medical research to develop healing techniques or clinical diagnosis & treatment of peculiar diseases, computer provides fast and effective solutions with high accuracy. Hospitals, doctors, and other healthcare centers around the world require the ability to send and receive healthcare data, including patient information and various lab reports means that vast amounts of healthcare information are

exchanged on a daily basis. However, medical data can be extremely complicated due to the abundance of clinical terminology, as well as the structural complexity in the formation of the presented information. Thus, this information must be presented in a standardized format in order to ensure that the data is universally understood and organized. [1] All healthcare information must be sent in a specialized healthcare language

Digital Imaging and Communications in Medicine (DICOM) is a standard for handling, storing, printing, and transmitting information in medical imaging with help of communication protocol. It includes a file format definition and a network communications protocol. DICOM files can be exchanged between two entities that are capable of receiving image and patient data in DICOM format. [2] The National Electrical Manufacturers Association (NEMA) holds the copyright to this standard. DICOM enables the integration of scanners, servers, workstations, printers, and network hardware from multiple manufacturers into a picture archiving and communication system.

1.1. Data Extraction

It is act or process of retrieving data out of (usually unstructured or poorly structured) data sources for further data processing or data storage. The term data extraction is applied when data is first imported into a computer from primary sources like measuring or recording devices.

Electronic Health Record (EHR) and Electronic Medical Record (EMR) are used to keep the record of patient's medical information. The only difference is that EHR is a global concept and EMR is localized record. The terms EHR and EMR are used interchangeably. An EHR system is also often

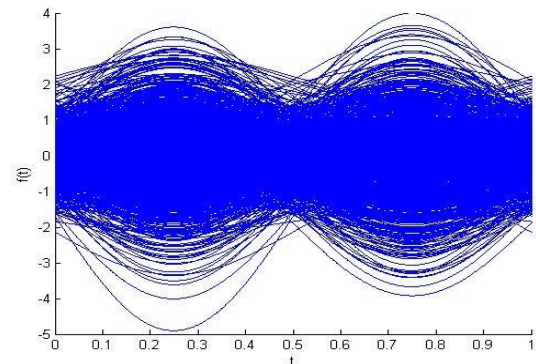
abbreviated as EHR or EMR. EHR is an electronic record of patient health information generated by one or more encounters in any care delivery setting. EHR includes the information of patient demographics, progress notes, and problems, medications, past medical history, immunizations, laboratory data and radiology reports. The EHR automates clinician's workflow. The EHR has the ability to generate a complete record of a clinical patient encounter as well as supporting other care related activities directly or indirectly via interface, including evidence-based decision support, quality [3] management, and outcomes reporting. EHR, technology represents a vast improvement over paper-based systems, and is changing the way healthcare is administered in medical practices.

1.2. Data Visualization

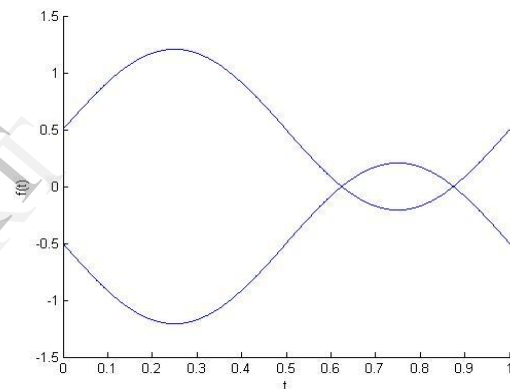
Data visualization is the study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information". [4] According to Friedman (2008) the "main goal of data visualization is to communicate information clearly and effectively through graphical means. Data visualization is closely related to information graphics, information visualization, scientific visualization, and statistical graphics.

Data visualization is also defined as "the set of techniques used to turn a set of data into visual insight. It aims to give the data a meaningful representation by exploiting the powerful discerning capabilities of the human eye a computer cluster consists of a set of loosely connected computers that work together so that in many respects they can be viewed as a single system. The components of a cluster are usually connected to each other through fast local area network, each node (computer used as a server) running its own instance of an operating system. Computer clusters emerged as a result of convergence of a number of computing trends including the availability of low cost microprocessors, high speed networks, and software for high performance distributed computing [5]. A cluster is two or more interconnected computers that create a solution to provide higher availability, higher scalability or both. The advantage of clustering computers for high availability is seen if one of these computers fails; another computer in the cluster can then assume the workload of the failed computer. Users of the system see no interruption of access. The advantages of clustering computers for scalability

include increased application performance and the support of a greater number of users.



“Figure1. Data Visualization before Clustering Algorithm”



“Figure2. Data Visualization after Clustering Algorithm”

“Table1. Different types of Clustering Algorithms for better Visualization”

No.	Name	Description	Why
1.	K-Means Clustering Algorithm	K-means (Macqueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. This algorithm aims at minimizing an objective function, in this case a squared error function.[6]	The k-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centers. The k-means algorithm can be run multiple times to reduce this effect.
2.	Fuzzy c-Means Clustering	Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunin in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition. [7]	It is based on minimization of the following objective function: $J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \ x_i - c_j\ ^2$ $1 \leq m < \infty$ where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j , x_i is the i th of d -dimensional measured data, c_j is the d -dimension center of the cluster, and $\ \cdot\ $ is any norm expressing the similarity.

1.3. Data Abstraction

Data Abstraction techniques are widely used in multiresolution visualization systems to reduce visual clutter and facilitate analysis from overview to detail. However, analysts are usually unaware of how well the abstracted data represent the original dataset, which can impact the reliability of results gleaned from the abstractions. Abstraction is the process of recognizing and focusing on important characteristics of a situation or object and leaving/filtering out the unwanted characteristics of that situation or object. [10] Qingguang Cui, define two data abstraction quality measures for computing the degree to which the abstraction conveys the original dataset: the Histogram Difference Measure and the Nearest Neighbor Measure.

1.3.1. Bhattacharyya Distance

[8] It measures the similarity of two discrete or continuous probability distributions. It is used to determine the relative closeness of the two samples being considered.

For discrete probability distributions p and q over the same domain X , it is defined as:

$$D_B(p, q) = -\ln(BC(p, q))$$

$$\text{Where: } BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)}$$

Is the Bhattacharyya coefficient

1.3.2. Histogram Difference Measure

The HDM is derived based on the average relative error of aggregation used in approximate query processing of databases as well as image similarity measures used in image retrieval [9]. In this technique, a histogram is computed from all of the pixels in the image, and the peaks and valleys in the histogram are used to locate the clusters in the image. A refinement of this technique is to recursively apply the histogram-seeking method to clusters in the image in order to divide them into smaller clusters. This is repeated with smaller and smaller clusters until no more clusters are formed.

Thus, if we let n be the total number of observations and k be the total number of bins, the histogram m_i meets the following conditions:

$$n = \sum_{i=1}^k m_i.$$

2. Related Work

Several researchers have proposed the effective use of data visualization. Data visualization techniques deal with the concept of perception. The main issue with perception is "how humans attach meaning to the sensory information they receive. Stephen few (2007) discussed regarding data visualization which represents the use of images to represent information.

It provides a powerful means both to make sense of data and to then communicate what we've discovered to others. [11] Daniel A. Keim (2002) regarding information visualization and visual data mining can help to deal with the flood of information. The advantage of visual data exploration is that the user is directly involved in the data mining process. Fernanda B. Viégas et al discussed information visualization which is traditionally viewed as a tool for data exploration and hypothesis formation. With the ability to create visual representations of data on home computers, artists and designers have taken matters into their own hands and expanded the conceptual horizon of information as artistic practice.

3. Methodology

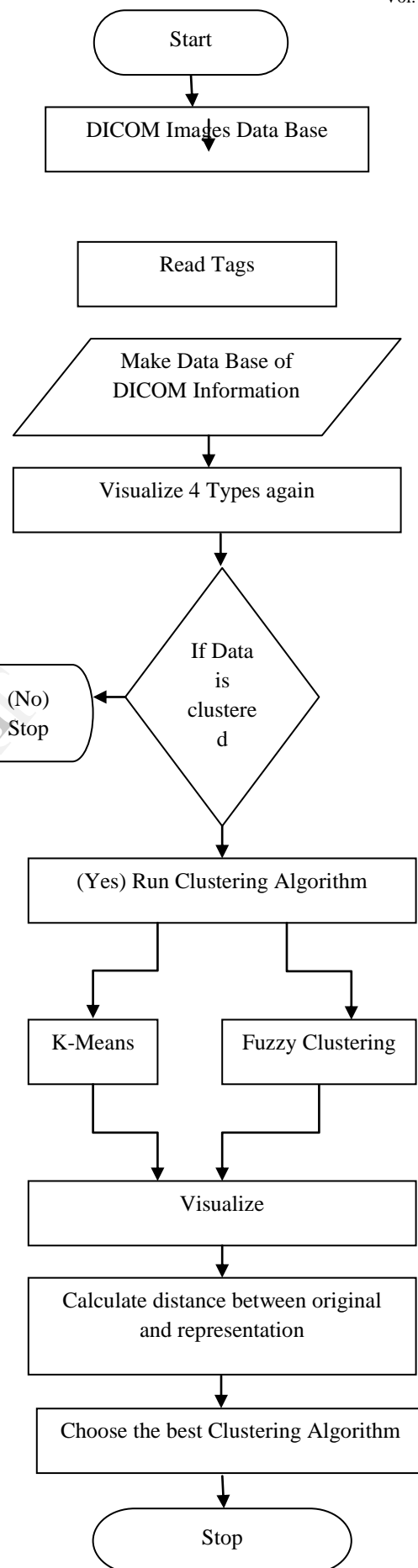
First the data base of DICOM images is collected, than tags are extracted from the DICOM images. From these tags the data base is developed of DICOM information. After this, whole data is visualized in different forms of graphs (Plot Matrix, Andrew Plot, Scatter Plot, and Parallel Cords). If visualized data is clustered than clustering algorithm is used (K- Means and Fuzzy). After clustering visualize the four types of graph again, than the distance is calculated between original and representation and evaluation of performance of clustering algorithm is made based on the three parameters

3.1. Time

3.2. Distance

3.3. Dimension

If distance close to zero that algorithm is regarded as best among the others.



4. Conclusion

While the above approaches deal with semantic level of biomedical data, none discuss the extraction and compilation of data from these DICOM tag information, consolidating and presenting it in a form which provides some information that can be used to develop an insight into disease patterns. The present study aims at extraction and compilation of data from DICOM image tags information, consolidating and presenting it in a form which provides information that can be used to develop an insight into disease patterns. Acquiring medical data is no longer a problem. It's everywhere in all departments of medical sciences, and we're already quite adept at hoarding it in databases. The issue now is making sense of all those signals and finding technical correlations in the medical stream which would be helpful in diagnostics of an individual and masses. That's where data abstraction and visualizations come in. Whether it is a static graph or a real-time data wave, the act of seeing data unlocks much of its utility. However, it must also be noted that it is a data problem, which requires programming and design, rather than as a design problem that has a data component, or a data problem that has a design component. There is an urgent need to take advantage of data that might otherwise never be useful due to lot of clutter in its visualization and need to find the right kind of display for medical diagnostic purposes. There is considerable room for improving the visualization of most the medical data set results. First, at the algorithmic level, the scale and complexity of the graphs produced from the data mining stage still have the potential to embarrass available graph layout algorithms. On the human side, further thought is needed on the mapping from data attributes to visual attributes, in particular where the visualization is superimposing access properties above the basic site structure. Part of this work can and should be based on known characteristics of perception and principles of visualization design, however, the ultimate utility of the representation will only become apparent once it is assessed through controlled experiments, and this will require time and a more polished version of the user interface with less reduced clutter.

References

- [1] Cogniz Consultancy Health level 7.
- [2] DICOM brochure nema.org.
- [3] Health Information and Management systems society EHR: electronic health record.
- [4] Michael Friendly (2008) "Milestones in the history of thematic cartography, statistical graphics, and data visualization".
- [5] Boder David, Robert Pennington (June 1996) "Cluster Computing Applications" Georgia Tech College of Computing. Retrieved 2001-07-13
- [6] J.B.Maqueen (1967) "Some Methods for classification and Analysis of Multivariate observations, Proceedings of 5th Berkeley Symposium on Mathematical Statistics and probability, Berkley, University of California.
- [7] J.C.Dunn (1973) "A Fuzzy Relative of the Isodata Process and its use in Detecting Compact well Separated Clusters" Journal of Cybernetics.
- [8] Bhattacharyya, A (1943) "On a measure of divergence between two statistical populations defined by their probability distributions bulletin of the Calcutta Mathematical Society 35:99-109.
- [9] Shelza and Balwinder Singh "A Novel System for Abstraction and Visualization of CAD Images" IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 Vol. 2, No.2, April 2012.
- [10] Qingguang cui .Mathew O. wards, Elke A. Rundensteiner & Jing jang. Measuring data abstraction quality in multiresolution visualizations. IEEE transactions on visualization and computer graphics, vol. 12, NO. 5, pages 709-716, 2006.
- [11] Alexander Hinneburg, Daniel A. Keim and Markus Wawryniuk. Hd-eye: visual mining of high – dimensional data. IEEE computer graphics and applications, pages 22-31, 1999