International Journal of Science and Engineering Applications
Volume 8–Issue 07,231-235, 2019, ISSN:-2319–7560

Making the Appointment at OPD by Using Clustering Algorithm

Than Win
Department of Information Technology
Technological University (Mawlamyine)
Myanmar

Ei Ei Phyo
Department of Information Technology
Technological University (Mawlamyine)
Myanmar

Abstract: Life is becoming too busy to get medical check and treatment in person. The main idea is to provide ease and comfort to patients while taking appointment from doctors and it also resolves the problems that the patients has to face while making an appointment. Enhancing patient care management is one of the aims of healthcare industry to improve the healthcare system worldwide. Making Appointment at OPD System is an important component of scheduling and managing appointments. This paper advances with the facilities that eliminate the chaos of traditional appointment service and offers the appointment booking, to view doctors list, to cancel and update appointments with a user form to manage all the sections. It provides advanced functionality the process that easy to access personal hospital services that help organizations to stay connected with their customers, the most importantly patients, the result in significant time and time savings. In this paper, K-Means clustering algorithm is implemented to analyze the OPD system.

Keywords: data mining, clustering, k-means clustering, initial centroid, Euclidean distance

1. INTRODUCTION

Data mining is an active research area and research is ongoing to bring statistical analysis and artificial intelligence (AI) techniques together to address the issues. Data mining tools are then used to uncover useful patterns and relationships from the data capture. Currently, data mining techniques, tools, and researches are being expanded to the various field [1]. Clustering is one among the oldest techniques used in Data mining [2]. Clustering is a standard procedure in multivariate data analysis. It is designed to explore an inherent natural structure of the data objects, where objects in the same cluster are as similar as possible objects in different clusters are as dissimilar as possible. Clustering is unsupervised learning and do not rely on predefined classes. In clustering we measure the dissimilarity between objects by measuring the distance between each pair of objects. These measure include the Euclidean, Manhattan and Minkowski distance [3]. Scheduling appointments is one of the most important administrative responsibilities performed in the medical check. The most important criteria the medical assistant must take into consideration when scheduling appointments are exhibiting good interpersonal skills and reducing the amount of time a patient has to wait to see the physician .People went to a medical office expecting to wait as long as an hour or more. People have little tolerance for waiting in a medical check. Lifestyles have changed and people have busy lives. Many have to take personal time away from work to go to the medical check, and they feel that their time is as valuable as the physician's time. Scheduling appointments correctly and efficiently is crucial to the smooth operation of the medical check. Many factors must be taken into consideration when scheduling appointments. The patient who has made an appointment weeks or even months in advance wants to be seen within 15 to 20 minutes after arriving at the medical check. The physician wants a smooth flow of patients during the time scheduled for seeing patients. Patients who are ill or have accidents want to be able to see their physician on the day of the illness or injury. They prefer to be given a specific time, even if it is later in the day, rather than come into the medical check and wait for an open moment. A patient appointment system or appointment schedule for health care center started long time ago .Management of patients' appointments has earlier works and has developed simplified queuing models and fairly static scheduling conditions. Another attempt was made to calculate the waiting time between patient and doctor using the mathematical queuing models to minimize waiting time .However; traditionally the appointment system has considered that the doctor time is more important than patient time .So an appointment system was designed to minimize the doctor idle time but current designing of an appointment system is based on decisive factors with respect to both the patient and doctor [4]. The patient appointment system has complex structures because it represents the patient appointment time in the healthcare center and controls the patient waiting time based on the type and the period of patient appointment. Moreover, a patient appointment system is meant for: managing doctor's time, reducing patient's waiting time, reducing doctor's idle time, reducing nurse's idle time, and improving the quality of service in the health care.

2. AIMS AND OBJECTIVES

This paper is to provide patients full access to manage their hospital appointments which, facilitates with a service for appointment reservation, updating and canceling management minimizing customer inconvenience and assuring a better healthcare. The appointments system is a means for health care center and hospital that allows for quick booking and managing of patient's appointments while eliminating the possibility of reiteration of the same time slot for different patients. Specific objectives include:

- To understand how to use data mining techniques and methods
- To know the importance of the basics methods
- To establish of a paperless environment
- To advance utilization of medical resources at the medical center
- To increase efficiency of medical care outcome between hospital and customer time saving

3. METHODOLOGY

3.1 Data Mining

Data mining is the process of applying these methods to data with the intention of uncovering hidden patterns. Data mining or data mining technology has been used for many years by many fields such as businesses, scientists and governments. It is used to sift through volumes of data such as airline passenger trip information, population data and marketing data to generate market research reports, although that reporting is sometimes not considered to be data mining [5].

One of the most important tasks in Data mining is to select the correct data mining techniques. Data mining technique has to be chosen based on the type of business and the type of problem. A generalized approach has to be used to improve the accuracy and cost effectiveness of using data mining techniques. There are basically seven main Data Mining techniques. There are a lot of other Data Mining techniques but these seven are considered more frequently used by business people.

- Statistics
- Clustering
- Visualization
- Decision Tree
- Association Rule
- Neural Networks
- Classifications [6]

3.2 Clustering

Clustering is a process in which a group of unlabeled patterns are partitioned into a number of sets so that similar patterns are assigned to the same cluster, and dissimilar patterns are assigned to different clusters. This will help to understand the differences and similarities between the data. There are two goals of clustering algorithms: determining good clusters and doing so efficiently. Clustering has become a widely studied problem in a variety of application domains, such as in data mining and knowledge discovery, statistical data analysis, data classification and compression, medical image processing and bioinformatics [7].

Clustering techniques are classified into supervised and unsupervised methods. The unsupervised clustering method is used to detect the underlying structure in the data set for classification. Supervised clustering method involved with the human interaction. The unsupervised clustering techniques are most popular due to the minimal knowledge about the dataset. Similarity is fundamental to the definition of a cluster, and various clustering techniques use different similarity definitions and techniques. The very famous distance measure may be the Euclidean distance. But, it has taken much time to cluster the data's compare to other distance metrics like Minkowski. In this paper, a new Minkowski distance based K-means clustering algorithm for clustering the data is proposed. In addition, the performance analysis is compared with our proposed algorithm.

3.3 K-Means Clustering Algorithm

K-Means clustering is one of the methods of cluster analysis which partitions or divides the number of observations into k clusters. Every observation belongs to the cluster with the nearest mean. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. The main purpose of K-means clustering is to classify the data [8].

The K-means clustering algorithm is a data mining and machine learning tool used to cluster observations into groups of related observations without any prior knowledge of those relationships. By sampling, the algorithms attempts to show in which category, or cluster, the data belong to, with the number of clusters being defined by the value k [9].

The K-means algorithm is one of the simplest clustering techniques and it is commonly used in medical imaging, biometrics, and related fields [19]. K-means is used in various topics, ranging from market segmentation, computer vision, statistics and astronomy to agriculture. It is also used as a preprocessing step for other algorithms. There are a lot of applications of the K-means clustering like Pattern recognitions, Classification analysis, Artificial intelligent, image processing, machine vision, etc. [8]

K-means clustering intends to partition n objects into k clusters in which each objects belongs to the cluster with the nearest mean. This method produces exactly k different clusters of greatest possible distinction. The best number of clusters k leading to the greatest separation (distance) is not known as a priori and must be computed from the data.

K-means Algorithms

- Clusters the data into k groups where k is predefined.
- Select k points at random as cluster centers.
- Assign objects to their closet cluster center according to the Euclidean distance function.
- Calculate the centroid or mean of all objects in each cluster.
- Repeat steps 2, 3 and 4 until the same points are assigned to each cluster in consecutive rounds [10].

4. EXPERIMENTAL RESULTS

Table 1. Assign Value for Department Timetable (Heart)

Department Timetable(Heart)	Values
09:00 AM-12:00 PM	2
10:00 AM-12:00 PM	2
08:00 AM-10:00 AM	8
02:00 PM-04:00 PM	5
07:00 AM-10:00 AM	7
03:00 PM-05:00 PM	6
08:00 AM-11:00 AM	1
07:00 AM-11:00 AM	4
08:00 AM-12:00 PM	8
09:00 AM-11:00 AM	3

Table 2. Assign Value for Patient Diseases Timetable (Heart)

Patient Diseases Timetable (Heart)	Values
08:00 AM	10
11:00 AM	5
09:00 AM	4
06:00 PM	8
08:30 AM	5
03:00 PM	4
10:00 AM	2
12:00 PM	9
07:00 AM	11
09:00 AM	6

Table 3. Assign Value to Department and Patient Diseases for Dataset

No	Department Timetable (Heart)	Patient Diseases Timetable (Heart)
1	2	10
2	2	5
3	8	4
4	5	8
5	7	5
6	6	4
7	1	2
8	4	9
9	8	11
10	3	6

Euclidean distance equation:

$$d(x, y)(a, b) = \sqrt{(x - a)^2 + (y - b)^2}$$

K = 2, Initial cluster centers are (2, 10), (1, 2).

$$\text{Cluster1 } (2, 10) = \sqrt{(2-2)^2 + (10-10)^2} = 0$$

$$\text{Cluster2 } (1, 2) = \sqrt{(1-2)^2 + (2-10)^2} = 9$$

Table 4. Calculation Result of an Iteration 1

Point	(2, 10) Distance Mean 1	(1, 2) Distance Mean 2	Cluster
(2,10)	0	9	1
(2, 5)	5	4	2
(8, 4)	12	9	2
(5, 8)	5	10	1
(7, 5)	10	9	2
(6, 4)	10	7	2
(1, 2)	9	0	2
(4, 9)	3	10	1
(8,11)	7	16	1
(3, 6)	5	6	1

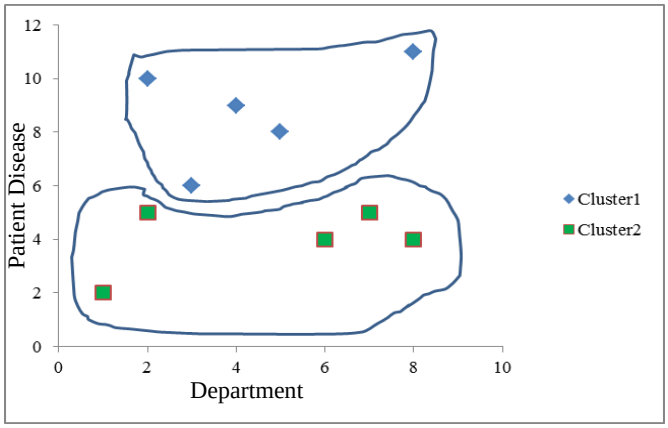

Figure 1. Graph of Iteration 1

Cluster1 (2,10),(5,8),(4,9),(8,11),(3,6)

Cluster2 (2,5),(8,4),(7,5),(6,4),(1,2)

For Cluster1,

((2+5+4+8+3)/5, (10+8+9+11+6)/5) = (4.4, 8.8)

$$\text{Cluster1 } (4.4,8.8) = \sqrt{(4.4-2)^2 + (8.8-10)^2} = 3.6$$

$$\text{Distance from Cluster2 } (4.8,4) = \sqrt{(4.8-2)^2 + (4-10)^2} = 8.8$$

Table 5. Calculation Result of an Iteration 2

Point	(4.4,8.8) Distance mean1	(4.8,4) Distance mean2	Cluster
(2,10)	3.6	8.8	1
(2, 5)	6.2	3.8	2
(8, 4)	8.4	3.2	2
(5, 8)	1.4	4.2	1
(7, 5)	6.4	3.2	2
(6, 4)	6.4	1.2	2
(1, 2)	10.2	5.8	2
(4, 9)	0.6	5.8	1
(8,11)	5.8	10.2	1
(3, 6)	4.2	3.8	2

(3.1)

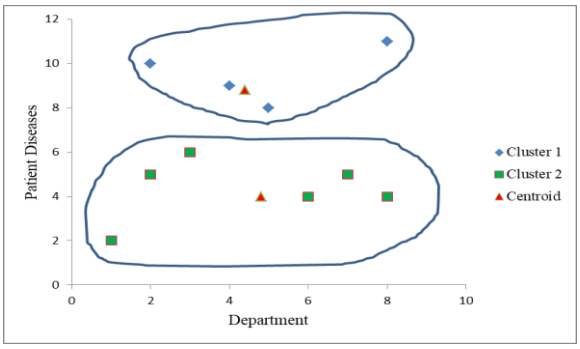

Figure 2. Graph of Iteration 2

Cluster1 (2, 10),(5,8),(4,9),(8,11)

Cluster2 (2,5),(8,4),(7,5),(6,4),(1,2),(3,6)

For Cluster1,

((2+5+4+8)/4, (10+8+9+11)/4) = (4.75, 9.5)

For Cluster2,

((2+8+7+6+1+3)/6, (5+4+5+4+2+6)/6) = (5.4, 5.2)

Cluster1 (4.75, 9.5) = $\sqrt{(4.75-2)^2+(9.5-10)^2}$ = 3.25

Distance from Cluster2 (5.4, 5.2)

= $\sqrt{(5.4-2)^2+(5.2-10)^2}$ = 8.2

Table 6. Calculation Result of an Iteration 3

Point	(4.75,9.5) Distance mean1	(5.4,5.2) Distance mean2	Cluster
(2,10)	3.25	8.2	1
(2, 5)	7.25	3.6	2
(8, 4)	8.75	3.8	2
(5, 8)	1.75	3.2	1
(7, 5)	6.75	1.8	2
(6, 4)	6.75	1.8	2
(1, 2)	11.25	7.6	2
(4, 9)	1.25	5.2	1
(8,11)	4.75	8.4	1
(3, 6)	5.25	3.2	2

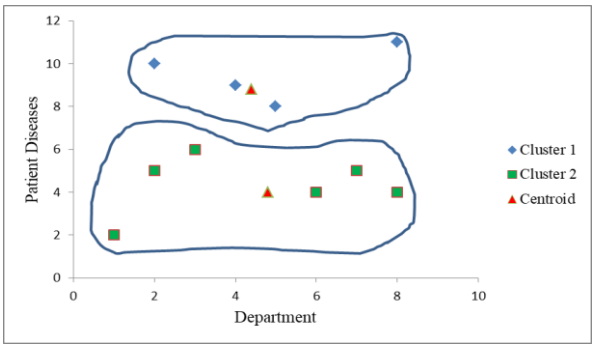

Figure 3. Graph of Iteration 3 (Final Result)

The calculated mid-points for each group as the initial 2 centroids, apply the K-means algorithm on the input data. After two iteration of the K-means algorithm, stability was achieved. The resulting clusters are showed as above.

5. CONCLUSION

The Scheduling appointment at OPD system using data mining. The whole system activities are divided into two major parts like patients and doctors. Each one has their own role to perform and system respond accordingly. Several patient have been created using department. The scheduling appointment at OPD has been collecting the medical data of patients. The volume of patient medical data at the various hospitals has been increasing. As a result, majority of outpatient do not have full medical record. With this situation, the specialist time is wasted since they have collect this information again and in addition, it becomes very difficult for them to keep track of the patients. This reduces the ability to carry out high quality clinical research in the hospitals and compromises the continuity as well as the quality of healthcare delivery in the hospital.

6. REFERENCES

[1] Sang Jun Lee and Keng Siau, “A review of data mining techniques”, University of Nebraska-Lincoin, Lincoin, Nebraska,USA.

[2] Jyoti Yadav and Monika Sharma, “A review of K-means Algorithms”, Assistant Professor, International Journal of Engineering Trends and Technology (IJETT), July, 2013.

[3] https://www.educha.com/7-data-mining-techniques-for-best-results/

[4] Adebayo Peter Idowa, Olajide Oluseaun Adeosun and Kehinde Oladipo Williams, “Dependable Online Appointment Booking System for NHJS Outpatient in Nigerian teaching hospitals”, International Journal of Computer Science & Information Technology (IJCSIT), August, 2014.

[5] Yihao Li, “Data Mining: Concepts, Backgrounds and Methods of Integrating Uncertainly in Data Mining”, Southeastern Louisiana University.

[6] https://www.researchgate.net/publication/265151471

[7] Bhojaraju Gunjal, “Database Management: Concepts and Design”, National Institute of Technology Rourkela, 5 February, 2014.

[8] https://ba-finance-2013.blogspot.com/2012/09/k-means-clustering.html?m=1

[9] https://www.lifewire.com/k-means-clustering-1019648

[10] https://www.sae.dsayad.com/clustering-kmeans.html