INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND CREATIVE ENGINEERING (ISSN:2045-8711)
VOL.7 NO.08 AUGUST 2017, IMPACT FACTOR: 1.04

Wireless Sensor Network using Monitoring the Environmental Activities

S.Kalaivani
M.Phil Research Scholar, Department of Computer Science,
Vellalar College for Women(Autonomous),Erode, Tamil Nadu, India.
Email: kalaivkl26@gmail.com
Dr.P.Radha
Assistant Professor, Department of Computer Science,
Vellalar College for Women(Autonomous),Erode, Tamil Nadu, India.
Email:radhasakthivel68@gmail.com

Abstract- Wireless sensor network plays an important role in monitoring environmental activities. Many sensor devices are used to collect the spatial or temporal data. The data sets that are collected may have irregularities, missing values inconsistent data. To handle these data, data preprocessing is performed to remove, unwanted data and to fill in the missing values .Various clustering algorithm is performed on those data for cluster formation. This project analyses the two major clustering algorithms: K-means clustering and Fuzzy C-means clustering .The clusters are formed using both the algorithms and their performance is analyzed .The performance of these clusters are analyzed based on the inter and intra cluster distance. Based on the result, it is proved that the Fuzzy C means algorithm is efficient than K-means algorithm.

Keywords-Wireless sensor, clustering, data preprocessing, Fuzzy C, K-means algorithm.

1. INTRODUCTION

Wireless sensor network grows and rapidly improves, this enable the new communication services. Sensor networks are the most useful way to collect the various parameters and information. A wireless sensor network is a collection of nodes organized into a cooperative network. Each node consist one or more microcontrollers, CPUs or DSP chips. Each node communicates with each other. Most of the wireless sensor networks are bi-directional in nature and they control all the sensor activity. The development of wireless sensor networks was motivated by military applications such as battlefield surveillance, industrial process monitoring and control, machine health monitoring, and so on. A sensor node may vary in size and the cost.

Sensor nodes consist of processing unit with limited computational power and limited memory, sensors, communication device, power source in the form of battery. The base stations are the main components of wireless sensor network with more computational power, energy and resources. They act as a gateway between sensor nodes and the end user and they forward data from the wireless sensor network to a server. Sensor network basically consist of large amount of sensor nodes that are deployed to large physical area to monitor and detect the real time environmental activities. These sensor nodes works together to collect the data like temperature, humidity, acceleration etc from surroundings. As sensor network is useful in application like in habitat monitoring, health monitoring, traffic, weather, pollution etc and in all such real life application sensor nodes generate large amount of data so mining data is really a fruitful task. Due to advancement in the wireless sensor networks the networks have ability to generate a large amount of data, and to find out the useful knowledge regarding the sensor network we apply data mining techniques. Her the wireless sensor network is linked up with environment monitoring and this link helps in various areas like fire detection in forest areas, saving wild life, and in other tropical conditions by analyzing temperature, humidity etc.

In this paper, the data mining techniques like data pre-processing and cluster analysis were processed and analyzed.

2. RELATED WORKS

Data in the real world is dirty. Real world data is often incomplete and noisy, say wrong values or duplicate records. This results in poor quality data which in turn results in poor quality mining results. Quality decisions are based on quality data and data warehouses needs consistent integration of quality data, which has no missing or noise data. In order to get quality data, the data in the database need to be checked for accuracy, completeness, consistency, timeliness, believability, interpretability and accessibility. The data preprocessing tasks were as follows:

Data Cleaning: Filling in missing values, smooth the noisy data identify or remove outliers and resolve inconsistencies.
Data Integration: Integration of multiple databases or files.
Data Transformation: Normalization and aggregation.
Data Reduction (Feature Selection): Obtains reduced representation in volume but produces the same or similar analytical results.
Data Discretization: Part of data reduction but with particular importance, especially for numerical data.

The clustering problem is defined as follows: For a given set of data points, it's proposed to partition them into one or more groups of similar objects. The similarity of the objects with one another is typically defined with the use of some distance measure or objective function. The clustering problem has been widely researched in the database, data mining and statistics communities. The nature of the clusters may vary with both the moment at which they are computed as well as the time horizon over which they are measured. For example, a user may wish to examine clusters occurring in the last month, last year, or last decade. Such clusters may be considerably different.

Therefore, a data stream clustering algorithm must provide the flexibility to compute clusters over user-defined time periods in an interactive fashion.

3. DATASET

The dataset for proposed work is downloaded from website link http://daac.ornl.gov/LBA/guides/CD04_Meteorology_Fluxes.html. The data is presented as value measured at 30 minute interval over 3.5 years and compiled at the km 83 tower site. This data includes the variables relate to Meteorology, soil moisture, fluxes of momentum, heat, water vapor and carbon dioxide beneath the flux sensors.

4. IMPLEMENTATION

4.1 Data Preprocessing

Data Pre-processing involves cleaning the data by putting in missing values and removing uninteresting data. It may also include summarization and aggregation of the data. This step basically involves preparing the data for analysis. Hence first and foremost this process can detect the irregularities in the sensor data and apply pre-processing technique.

4.2 K-means clustering algorithm

The k-means clustering algorithm consists of two separate phases: the first phase is to define k centroids, one for each cluster. The next phase is to take each point belonging to the given data set and associate it to the nearest center. When all the points are included in some clusters, the first phase is completed and an early grouping is done. At this point it's necessary to recalculate the new centroids, as the inclusion of new points may lead to a change in the cluster centroids. Once we find k new centroids, a new binding is to be created between the same data points and the nearest new center, generating a loop. As a result of this loop, the k centroids may change their position in a step by step manner. Eventually, a situation will be reached where the centroids do not move anymore.

Algorithm 1: The k-means clustering algorithm

Input:

D = {d1, d2...dn} //set of n data items.
K // Number of desired clusters

Output: S

A set of k clusters.

Steps:

1. Arbitrarily choose k data-items from D as initial centroids;
2. Repeat
2.1 Assign each data item di to the cluster which has the closest centroid;
2.2 Calculate the new mean of each cluster;
Until convergence criterion is met

4.3 Fuzzy C-means clustering algorithm

Fuzzy C-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. Here, this method is used in clustering of the network data. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m}\left\|x_i - c_j\right\|^2, \quad 1 \le m < \infty$$

Where m is any real number greater than 1, uij is the degree of membership of xi in the cluster j, xi is the ith of d-dimensional measured data, cj is the d-dimension center of the cluster. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership uij and the cluster centers cj by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}, \quad c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

This iteration will stop when

$$max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \varepsilon$$

Where ε, a termination criterion between 0 and 1 and k is are the iteration steps. This procedure converges to a local minimum or a saddle point of Jm.

The algorithm is composed of the following steps:

1. Initialize $U = [u_{ij}]$ matrix, $U^{(0)}$
2. At k-step: calculate the centers vectors $C^{(k)} = [c_j]$ with $U^{(k)}$ $c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$
3. Update $U^{(k)}$, $U^{(k+1)}$ follows $u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$
4. If $\|U^{(k)} - U^{(k+1)}\| < \varepsilon$ then STOP; otherwise return to step 2.

5. PERFORMANCE EVALUATION

In this module the results of Fuzzy C means and K means algorithm is compared to analyse the efficient perfomance . Inter-cluster distance measured within-cluster sum of squares. The Intra cluster distance, is the distance between All pairs of points in the cluster or between the centroid and all points in the cluster. The performance has been analysed based on the inter and intra cluster distance . In K-means clustering , the intra cluster distance is greater when compared with Fuzzy C-means. Thus the K-means clustering algorithm will take more time to compute than Fuzzy C-means clustering algorithm.

6. CONCLUSION

In this paper, the implementation of data set is done in both Fuzzy C means and K means. Later it is compared with the performance of two clusters that has been generated by the above said algorithm. As a result it is proved that Fuzzy C-means clustering algorithm is better for monitoring environmental activities than K-means clustering algorithm.

REFERENCES

[1] C.C.Aggarwal, J.Han, J.Wang, P.S.Yu. A Framework for Clustering Evolving Data Streams. VLDB2003. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol.2.Oxford: Clarendon, 1892, pp.68–73.

[2] I.F.Akyildiz, W.Su, Y.Sankarasubramaniam, and E.Cayirci. Wireless Sensor Networks: A Survey. Computer Networks, Vol.38, No.4, pp.393-422, 2002.

[3] S.Bandyopadhyay and E.J.Coyle. An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks. IEEE INFOCOM 2003.

[4] D.Barbara, W.Dumouchel, C.Faoutsos and P.Haas. The New Jersey Data Reduction Re-port. IEEE Data Engineering Bulletin,Vol.20, No.4, pp.3-45, 1997.

[5] D.Estrin, R.Govindan, J.Heidemann, and S.Kumar. Next Century Challenges: Scalable Coordination in Sensor Networks. MobiCOM 1999

[6] G.J.Pottie and W.J.Kaiser. Wireless Integrated Network Sensors.Communications of the ACM. Vol.43, No.5, pp.51-58, May 2000.

[7] Agrawal, R.; Imielinski, T.; Swami, A. IEEE\z Transactionon Knowledge and Data Engineering; 1993; p 6.

[8] Han, J.; Kamber, M. Data Mining: Concepts and Techniques; Morgan Kaufmann: 2000.

[9] Dongqing Yang, Shiwei Tang, Qiong Luo ,Xiuli Ma, Dehui Zhang, Shuangfeng Li. Online Mining in Sensor Networks. NPC 2004, LNCS 3222, pp. 544-550, 2004.

[10] Agrawal, R.; Srikant, R. Mining Sequential Patterns.ICDE 1995, 3- 14