International Journal on Applications in Engineering and Technology
Volume 3: Issue 2: April 2017, pp 8 –10. www.aetsjournal.com ISSN (Online) : 2455 - 0523

SMART CITY DATA ANALYTICS

Nanda Kumar.G.S , Vimala.O , Vidhya Shree.R

***Abstract*—**Now the government is interested in adopting the smart city concept in their cities and implementing big data applications that support smart city components to reach the required level of sustainability to improve the living standards. Smart cities utilize multiple technologies to improve the performance of health, transportation, water services, education and other services leading to higher levels of comfort to their citizens. This involves reducing costs and resource consumption in addition to more effectively and actively deal with their citizens. One of the recent technologies that have a huge potential to enhance smart city services is big data analytics. As digitization has become an integral part of everyday life, data collection has resulted in the accumulation of huge amounts of data that can be used in various beneficial areas. The general requirements that can be applied to all big data applications are Big data management, this can be done through Smart network infrastructure, Advanced algorithm, security and privacy, and government role. Smart city in data analysis include the intensive use of information technology, this integration of the physical and social components of the City to implement advanced monitoring and control tool applications.

Keywords — SmartCity, Dataset, Analysis

I. INTRODUCTION

A new scope opens up for what can be controlled in cities through Big Data. From the commercial world, it has been shown that organisations that effectively use the data to obtain decisions in their organisations out. This approach equally applies to the city managers at the management of their operations. In developing cities, the reality is that operations are uncoordinated and data capture is still a heavy manual process. Modern environments benefit from having consolidated services across areas such as asset, fleet and workforce management where pertinent information is available from these departments to facilitate shorter turnarounds in service delivery, overall customer satisfaction and the understanding of bottlenecks. This is enabled by the implementation of data warehouses and business intelligence that developing cities still lack. The best practice for this is taking the form of establishing an operations centre. Here, city managers are able to access data in real-time streams and city operators are able to form dynamic communities of interest around incidents as they occur. Data is tapped into to provide collaboration, situational aware decision making (CSADM) capabilities to fulfil the objective of servicing the citizen. And by this we can obtain and fulfil the actual need of the citizen. This will help the government to improve the performance and if it works correctly the people will also respond to the maximum extent and which in turn make the city smarter.

Nanda Kumar.G.S , Associate Professor, Department of COMPUTER SCIENCE AND ENGINEERING , Kumaraguru College of Engineering and Technology, Coimbatore. (Email : nandakumar.gs.cse@kct.ac.in)

Vimala.O , UG Final year COMPUTER SCIENCE AND ENGINEERING Kumaraguru College of Engineering and Technology, Coimbatore.
(Email : Vimalavimal296@gmail.com)

Vidhya Shree.R , UG Final year, COMPUTER SCIENCE AND ENGINEERING , Kumaraguru College of Engineering and Technology,Coimbatore.
(Email : vidhyashreevidhu95@gmail.com)

II. MISSION STRATEGY

- Pan-city initiative in which at least one smart solution is applied city wide
- Develop areas step by step – three models of area based developments
    - Retrofitting
    - Re development
    - Greenfield

Figure 1.Smart City overview

III. ALGORITHM

The algorithm used here is k – means clustering algorithm which analyses and group by clusters. K-means is one of the simplest unsupervised learning

algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centres, one for each cluster. These centres should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centre. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycentre of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centre. A loop has been generated. As a result of this loop we may notice that the k centres change their location step by step until no more changes are done or in other words centres do not move any more. Finally, this algorithm aims at minimizing an objective function knows as squared error function.

This is given by

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\left\| x_i - v_j \right\| \right)^2$$

Where

'$||x_i - v_j||$' is the Euclidean distance between x_i and v_j.

'c_i' is the number of data points in i^{th} cluster.

'c' is the number of cluster centres.

IV. RELATED WORKS

For Smart Cities Big Data, data has become a subtle force of its own and has affected their daily lives. Our daily activities are recorded and are used to support transactions we have with all types of organisations. To meet the growth in data sizes, new technology solutions are appearing that are able to harness this new data source, such as back-end systems that use in-memory computing and massively parallel processing systems. Techniques and methods of managing its integration have also appeared, such as data virtualisation, efficient processing of unstructured data and advanced machine learning that is able to identify patterns across large data sets. [1]. „ Role of Big Data and Analytics in Smart Cities typically aims at decision maker that are used in Big Data Analytics as a tool for making Smart City. The paper covers how Internet of Things, Machine to machine, Big Data and Smart Cities Linkages can help in doing predictive analytics which can be helpful to human wellbeing. [2]. „

„ Trace Analysis and Mining for Smart Cities: Issues, Methods, and Applications Traces of moving objects in a city, which depict lots of semantics concerning human mobility and city dynamics, are becoming increasingly important. In this article it gives a brief introduction to trace data; then we present six research issues in trace analysis and mining, and survey the state-of-the-art methods finally, five promising application domains. Trace data implies the underlying patterns and laws of mobility in human society. It is also embedded with rich information on humans (e.g., human activities, social events, and social relationships) and cities (e.g., the semantics of regions and the dynamics of a city). [3]. „ An Efficient K-Means Clustering Algorithm involves a novel algorithm for performing k-means clustering. It organizes all the patterns in a k-d tree structure such that one can find all the patterns which are closest to a given prototype efficiently. All the prototypes are potential candidates for the closest prototype at the root level. However, for the children of the root node, we may be able to prune the candidate set by using simple geometrical constraints. This approach can be applied recursively until the size of the candidate set is one for each node. [4]. Research Issues on K- Means Algorithm: An Experimental Trail Using Mat lab will Clustering problems arise in many different applications: machine learning data mining and knowledge discovery, data compression and vector quantization, pattern recognition and pattern classification. This paper presents the results of an analysis of the representative works related to the research lines of k-means algorithm devoted to overcome its shortcomings. [5].

V. ANALYSIS WORK

Data analysis tools and software are typically used to sort through enterprise data in order to identify patterns and establish relationships.

The analysis tools for machine learning and other techniques to extract data available are Rapid Miner (formerly known as YALE) written in the Java Programming language, this tool offers advanced analytics through template-based frameworks. It also provides functionality like data pre-processing and visualization, predictive analytics and statistical modeling, evaluation, and deployment. Weka, With the Java-based version, the tool is very sophisticated and used in many different applications including visualization and algorithms for data analysis and

predictive modeling. R – programming, it's a free software programming language and software environment for statistical computing and graphics. Orange, Python is picking up in popularity because it's simple and easy to learn yet powerful.

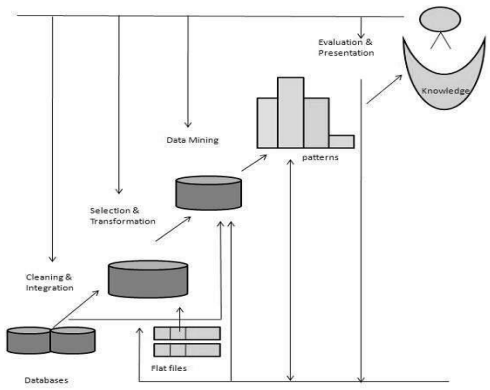

Figure 2.Knowledge Discovery Process for Data Analysis.

Discovery Process for Data Analysis.

Analysis can be done using various tools which has all these advantages,

- Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform.
- A comprehensive collection of data pre-processing and modelling techniques.
- Ease of use due to its graphical user interfaces.

VI. PROPOSED SYSTEM

. The proposed system is specially designed for any governmental organization to maintain their area smarter i.e. to reduce our problem in day to day life. The proposed system deals with collecting the data from the people in that locality and these data will be stored in the database. This system deals with the area by making use of Datasets and analysis tool which is developed for getting feedbacks from the peoples. Later these data need to be converted to csv format and in this format the data will be sending to the analysis tool and by applying the k-means clustering the result will be obtained. The system involves software implementation

A. Software Implementation

The website used in this system is for collecting various data from various peoples in the city. Website is used for effective and fast communication, since everyone has a smart phone. In the website, the user can submit their feedbacks using the links that has given already. Peoples who are ready to give their feedbacks about society can have an opportunity. The filled data are collected and arranged for the analysis purpose. Here the website is created using the php and HTML; we have used these languages because this language is easy to understand. And the created website is hosted using the online tool hostinger, now the data that is filled in the website will be stored in the database that is created in the hostinger. Later whenever we need to analyse the data we want to convert the data in the database to csv format and then we can give the csv file as an input to the analysis tool and then apply the k-means clustering algorithm. Once the algorithm is applied clusters will be formed based on the patterns given by the maximum people.

VII. CONCLUSION AND FUTURE WORK

For the development purpose, we need to make the city smarter. Many governmental organisations are taking initiatives to make the city smarter. With the improvement in technologies, we can collect the data from the citizen in a particular area through the feedback forms and by analysing the data we can decide what step needs to be taken to make the place smart.

The results of the analysis for our work, allow us to establish a framework and analyse the theoretical study of the k-means algorithm. According the tests conducted in Mat lab, this laboratory showed that it is actually very conducive to experimental testing, the implementation of k-means, allows monitoring the performance of the algorithm through the information that can be deployed at runtime, such as result of the objective function and the number of points exchanged in iteration.

The number of iterations required can vary in a wide range from a few to several thousand depending on the number of patterns, number of clusters, and the input data distribution. Thus, a direct implementation of the k-means method can be computationally very intensive. This is especially true for typical data mining applications with large number of pattern vectors.

REFERENCES

[1] Parkash, Prabu V, "IoT Based Waste Management for Smart City",International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 2, February 2016.

[2] Giffinger R., Gudrun H.: Smart cities ranking: an effective instrument for the positioning of cities? ACE: Architecture. City & Environ, 4 (12), pp. 7–25 (2010)

[3] A. Boustani et al., "Investigation of the Waste-Removal Chain through Pervasive Computing," IBM J. Res. Dev. vol. 55, no. 1.2, Jan. 2011

[4] C. Song et al., "Limits of Predictability in Human Mobility,"Science, vol. 327, no. 5968, Feb. 2010, pp. 1018–21.[5] N. Eagle et al., "Inferring Friendship Network Structureby Using Mobile Phone Data," P. Natl. Acad. Sci. USA,vol. 106, no. 36, July 2009

[5] An efficient k-means clustering algorithm: Analysis and implementation, T. Kanungo, D. M.ount, N. Netanyahu, C. Piatko,R. Silverman, and A. Y. Wu, IEEE Trans. PatternAnalysis and Machine Intelligence, 24 (2002), 881-892