isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 6 Issue: 3 Jun,2018,ISSN_NO: 2321-3337

APPLICATIONS OF BIG DATA TECHNIQUES TO USE IN SMART CITIES DEPLOYMENT

KAMAL.N
Professor – Department of CSE, GRT Institute of Engineering and Technology, Tiruttani
drkamalnatarajan@gmail.com

Abstract - The project presents the main foundations of Big Data applied to Smart Cities. General Internet of Things based architecture is proposed to be applied to different smart cities applications. In our project we are trying to predict and provide solution to improvise the railway infrastructure and train services. Indian Railways (both local & state) is a mode of transport service where thousands of transit-card transactions can be processed every minute. Thus our proposed system involves smart transit cards which can monitor the user behavior. Thus these transmit cards would produce huge amount of data which can be used for analytics and prediction. From the transit card, passenger age, source and destination can be analyzed for providing solution for improvisation.

1. INTRODUCTION

A Smart City emerges when the urban infrastructure is evolved through the Information and Communication Technologies (ICT). The paradigm of Internet of Things (IoT) has enabled the emergence of a high number of different communication protocols, which can be used to communicate with commercial devices using different data representations. In this context, it is necessary an IoT-based platform to manage all interoperability aspects and enable the integration of optimal Artificial Intelligence (AI) techniques in order to model contextual relationships.

There are numerous mobile data sources like smart phones, smart-cards, wearable sensors and, in the case of vehicles, on-board sensors. All these sensors provide information that makes possible to detect urban dynamic patterns[3]. Nonetheless, most existing management systems of cities are not able to utilize fully and effectively this vast amount of data and, as a result, there is large volumes of data which is not exploited. This trend is known as Big Data.

2. Background

In this section, we discuss the about the overview of Big Data. Specifically, the important architectural aspects of IoT are provided, key characteristics and approaches for CPS and autonomous systems are presented, and we also elaborate on concerns and solutions for modern sensor-based architectures.

2.1. Over view of Big Data

Big data stored large amount of data whether it is structured or unstructured and disorganized [20] Big data is a form data that is processing in the form of traditional database [2]. Big data systems will store huge amount of data, store information in efficient manner to produce information to enhance smart city services. By this information big data will help administrator to plan for any expansion in smart cities, areas and resources.

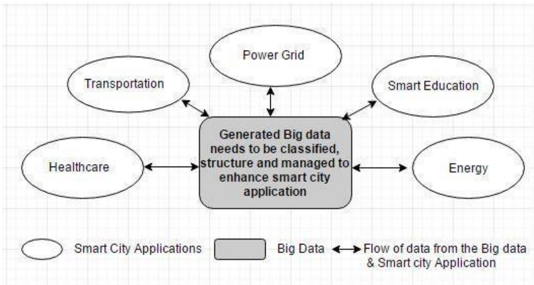

Fig 2: Smart Cities related to Big data

Some characteristics of big data

Volume: The size of data that has been created from data sets.

Velocity: The speed, at which data is stored, processed, analyzed, generated. It will support in real-time Big data management systems.

Variety: The different types of data being generated that most data is unstructured and cannot be easily tabulated or categorized.

However, bounded by the tools and the technologies available for services. For Big data to achieve its services and goals in smart cities, it needs the right tools and methods to classified and analyzed efficiently and effectively. By understanding available limitations and capabilities, we can capture many for better application and services for smart cities using big data.

2.1. IoT Paradigm

Imagine a world where billions of IP-connected objects are sensing, communicating and sharing information. Imagine these objects regularly collecting data, analyzing it and initiating action –

unleashing a new wealth of intelligence for planning, management and decision making. If you can envision this place, you've understood the concept called Internet of Things (IoT)[4].

IoT refers to everyday objects that are readable, recognizable, locatable, addressable, and/or controllable via the Internet – irrespective of the communication means – whether it's RFID, wireless LAN or wide-area networks. With IoT, "things" can be People, Location (of objects), Time Information (of objects), Condition (of objects). IoT changes not only the definition of "things" but their function and application, because objects from the "real" world can now seamlessly integrate with the "virtual" world, enabling anytime, anywhere connectivity[5].

With more physical objects and smart devices connected in the IoT landscape, the impact and value of IoT on our daily lives is quickly becoming immeasurable[5]. At the simplest level, people can make better decisions such as taking the best routes to work or choosing their favorite restaurants. New services are emerging for consumers such as pay-as-you-use services, and societal challenges can be easily met, for example, making remote health monitoring for elderly patients a reality.

For enterprises, IoT provides range of tangible business benefits – from improved management and tracking of assets and products, to new business models, operational efficiencies and cost savings achieved through optimized equipment and resource usage. Across industry verticals, IoT presents an abundance of opportunities for innovation. With real-time data and potentially cross-domain data sharing, new business models can be created. And IoT can address both industrial and consumer needs. For example, the potential of IoT as a service enables new markets and value chains to build up a competitive edge[4]. In the future, cognitive applications or systems in the context of IoT will play an even larger role.

3. Related work

The Railways carries more than 23 million passengers over a route of 66,000 km, passing through more than 7,100 railways stations and employing over 1.3 million people. In Existing system, the user prediction is not at all taken into account for improvisation. Railway facilities have not improved very substantially over the past few decades which lead to congestion and infrastructure damage. Also there is no initiative to improve the quality of service[1].

Modern railway has a high speed, heavy load and intensive development trend. It not only brings the opportunities of railway transport capacity and volume, but also makes the data of various types of large-scale continuous growth. In the stage of the development of the railway by the production enterprise to the service enterprise, it is necessary to vigorously promote the application of cloud computing, and to plan the cloud computing framework. We study a parallel support vector machine model based on multi-level SVM, and realize the parallel algorithm in cloud computing environment. The algorithm divides the large training data set into a number of small training sets by Map, and then a new SVM is combined with these small training sets. Finally, the data is trained to be a new SVM by Reduce[2]. At the end of the paper, we use the SVM parallel forecasting method to predict the passenger flow of China Railway, and compare the performance of the distributed with that of non-distributed algorithms. Experimental results show that the proposed algorithm has better effect than single machine algorithm in terms of time consumption and classification accuracy. With the increase of nodes, the time consumption is significantly shortened[4].

The high costs for the management of the modern and complex industrial control systems make it necessary to enhance the current maintenance processes. Therefore, the need arises to clearly define the goals of the maintenance, in order to evolve and continuously enhance the management methods, to efficiently integrate the maintenance activities with the ones related to the production, the service provisioning, and the operation, and to use smart computer-based maintenance systems[3]. This paper proposes a general maintenance approach and its specific application to the railway sector.

As Big Data becomes part of railroad data analysis, there are many challenges which need to be addressed by the railway industry. This extended abstract highlights some of the challenges from specific examples in railway engineering. This work does not present the challenges of dealing with Big Data in general which is beyond the scope of this paper[5]. The examples provided in this extended abstract cover both the engineering and the management of railroad applications.

Railway infrastructure monitoring is a vital task to ensure rail transportation safety. A rail failure could result in not only a considerable impact on train delays and maintenance costs, but also on safety of passengers. In this article, the aim is to assess the risk of a rail failure by analyzing a type of rail surface defect called squats that are detected automatically among the huge number of records from video cameras. We propose an image processing approach for automatic detection of squats, especially severe types that are prone to rail breaks[6]. We measure the visual length of the squats and use them to model the failure risk. For the assessment of the rail failure risk, we estimate the probability of rail failure based on the growth of squats. Moreover, we perform severity and crack growth analyses to consider the impact of rail traffic loads on defects in three different growth scenarios. The failure risk estimations are provided for several samples of squats with different crack growth lengths on a busy rail track of the Dutch railway network[7]. The results illustrate the practicality and efficiency of the proposed approach.

4. Proposed System

Our proposed system provides improvisation solution for our railway system using real time data. The real time data are collected using transit cards. In this project we are trying to predict and provide solution to improvise the railway infrastructure and train services. Predicting the consumers count who uses the railway services are solved through the application of big data technique (R Programming).The information relevant to trip profiling must include data about: time (in terms of day of the week and time of the day), origin and destination stations and approximate age of the traveller.This information is being continuously recorded in different databases of the train service. Collected data is made available to the railway department.

4.1. Implementing and Interfacing Smart card

Radio Frequency Identification (RFID) is a system that facilitates the tracking of objects. FID is a wireless system that works in conjunction with an organization's information technology infrastructure to improve business processes. The development of RFID was done by to enhance tracking and access applications in the 1980's in manufacturing and other hostile environments. The reader emits a radio signal that activates the tag and reads and writes data to it. As products are shipped, received or stored, the information (encoded on a bar code like tag) can be read and received by the reader, which is attached to a computer. RFID has been integrated into the EPC (Electronic Product Code). The EPC is a unique number that identifies a specific item in the supply chain. The EPC is stored on a RFID tag, which combines a silicon chip and a reader.

4.2. Programming Microcontroller

The Arduino Uno is a microcontroller board based on the ATmega328(datasheet). It has 14 digital input/output pins(of which can be used as PWM outputs), 6 analog inputs, a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started.

The Uno differs from all preceding boards in that it does not use the FTDI USB-to-serial driver chip. Instead it features the Atmega8U2 programmed as a USB-to-serial converter. "Uno" means one in Italian and is named to mark the upcoming release of Arduino 1.0. The Uno and version 1.0 will be the reference versions of Arduno moving forward. The Uno is the latest in a series of USB Arduino boards and the reference model for the Arduino platform, for a comparison with previous versions, see the index of Ardino boards.

4.3. Data Collection using Java

API by Java developers, it still provides all the necessary functionality for proper serial communications. In order to make API portable across platforms, the API defines an abstract Serial port class. This class is then subclassed and platform specific functionality is implemented in the subclassed object. This concrete class then interacts with a dynamic link library (DLL) file through the Java Native Interface (JNI). Once a Serial Port has been created, communications through the physical port are conducted through standard InputStream and OutputStream objects. These streams send and receive information as bytes, integers or arrays of bytes.

4.4. Encryption

The Data Encryption Standard (DES) is an outdated symmetric-key method of data encryption. DES works by using the same key to encrypt and decrypt a message, so both the sender and receivcer must know and use the same private key. Once the go-to, symmetric-key algorithm for the encryption of electronic data, DES has been superseded by the more secure Advanced Encryption Standard (AES) algorithm. The Dta Encryption Standard is a block cipher, meaning a cryptographic key and algorithm are applied to block of data simultaneously rather than one bit at a time. To encrypt a plain test message, DES groups it into 64-bit blocks. Each block is enciphered using the secret key into a 64-bit cipher text by means of permutation and substitution.

4.4. Cloud Storage

CloudMe is a file storage service operated by CloudMe AB that offers cloud storage, file synchronization and client software. It features a blue folder that appears on all devices with the same content, all files are synchronized between devices. The CloudMe service is offered with a free business model and provides encrypted SSL connection with SSL Extended validation certificate. CloudMe features a Cloud storage that allows the users to store, access and share their content, both with each other and with people outside the service.

4.5. R Programming

R is a programming language and software environment for statistical analysis, graphics representation and reporting. A huge amount of multidimensional data has been collected in various fields such as marketing, bio-medical and geo-spatial fields. Mining knowledge from these big data becomes a highly demanding field. However, it far exceeded human's ability to analyze these huge data. Unsupervised Machine Learning or clustering is one of the important data mining methods for discovering knowledge in multidimensional data.

5. Experimental Results

This paper discusses about the implementation of the big data techniques to smart cities deployment regarding the public transport service of a city and it also collects the real time data and process using big data analytics tool.

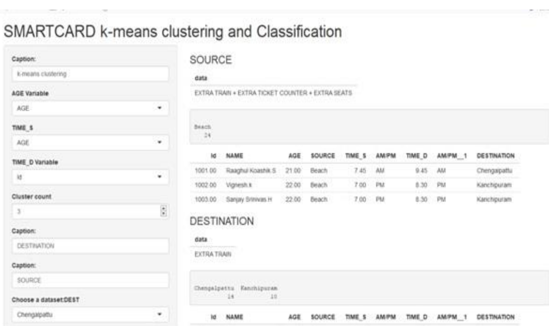

Fig.5.1. Passenger details

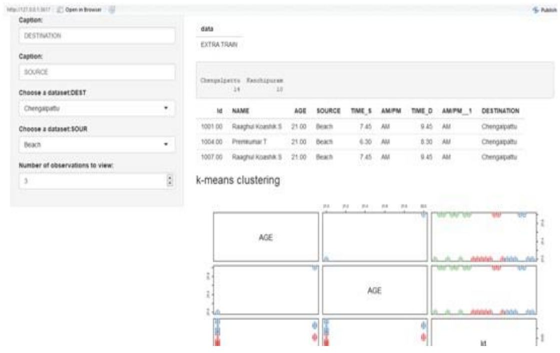

Fig.5.2. Clustering passenger data

Fig.5.3. Analyzing passenger smartcard 1

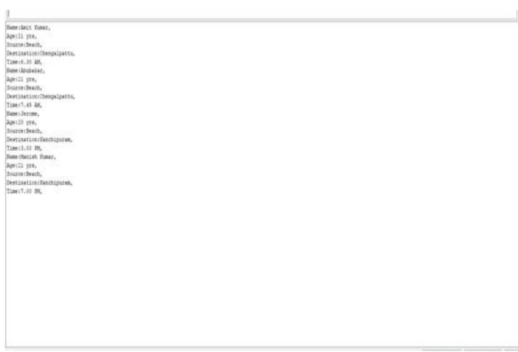

Fig.5.4. Analyzing passenger smartcard 2

7. CONCLUSION:

This paper displays the benefits of applying big data techniques over data originated by IoT-based devices deployed in smart cities. Especially for the public transport service of a city. Thus this project collects the real time data and process using big data analytics tool. The predicted result would provide an efficient solution for railway department to improvise the service and infrastructure.

8. REFERENCE

[1] N. Komninos, "Intelligent cities: variable geometries of spatial intelligence," Intelligent Buildings International, vol. 3, no. 3, pp. 172–188, 2011.

[2] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[3] L. Da Xu, W. He, and S. Li, "Internet of things in industries: a survey," Industrial Informatics, IEEE Transactions on, vol. 10, no. 4, pp. 2233– 2243, 2014.

[4] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, "Big data analytics: Computational intelligence techniques and application areas," Int. J. Inf. Manage, pp. 10–15, 2016.

[5] Z. Yan and D. Chakraborty, "Semantics in mobile sensing," Synthesis Lectures on the Semantic Web: Theory and Technology, vol. 4, no. 1, pp. 1–143, 2014.

[6] A. Carroll and G. Heiser, "An analysis of power consumption in a smartphone." in USENIX annual technical conference, 2010, pp. 1–14.

- [7] M. Kantardzic, Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, 2011.

About the Author

Dr.Kamal.N received his B.E Degree in Computer Science & Engineering from Madras University and M.E Degree in Computer Science & Engineering from Sathyabama University. He has completed his PhD from St.Peter's University. He is working as a Professor and Head in the Department of Computer Science and Engineering in GRT Institute of Engineering and Technology, Tiruttani. His research interests include Mobile Ad-hoc networks, Vehicular Ad-hoc networks, Sensor Network