Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

Design of Big Data Analytics using Unified Data Modelling Systems in Mobile Cellular Networks

[1]Mr. V. Ramakrishan
M.E, (Ph.D) Assistant Professor
Dept. of CSE Saveetha Institute of Medical and Technical Sciences.

[2] Dr. Anbalagan
Assistant Professor
Dept.of CSE , Annamalai University.

[3]Dr. M.S. Saravanan
Professor Dept. of CSE,
Saveetha Institute of Medical and Technical Sciences.

***Abstract*: Mobile cellular networks have become both the generators and carriers of massive data. Big data analytics can improve the performance of mobile cellular networks and maximize the revenue of operators. In this paper, we introduce a unified data model based on the random matrix theory and machine learning. Then, we present an architectural framework for applying the big data analytics in the mobile cellular networks. Moreover, we describe several illustrative examples, including big signalling data, big traffic data, big location data, big radio waveforms data, and big heterogeneous data, in mobile cellular networks. Finally, we discuss a number of open research challenges of the big data analytics in the mobile cellular networks.**

INDEX TERMS :- Big data analytics, mobile cellular networks.

I. INTRODUCTION

Recent years have witnessed tremendous advances in wireless cellular networks. With recent advances of wireless technologies and ever-increasing mobile applications, mobile cellular networks have become both generators and carriers of massive data. When geo-locating mobile devices, recording phone calls, and capturing mobile applications' activities, an enormous amount of data is generated and carried in mobile cellular networks. Historically, the massive data in mobile cellular networks hasn't been paid much attentions. With data constantly accumulated in the database and the technologies of *big data analytics* rapidly developed, the great value hide behind data has gradually been revealed. It is desirable to make good use of this precious resource, big data, to improve the performance of mobile cellular networks and maximize the revenue of operators. Traditional data analytics shows its in-adequateness when encountered with the big cellular data. First, traditional data analytics deals with structured data. The large amount of App-based data is, however, generally unstructured. Second, the implementation of data analysis is traditionally confined within a department, or a business unit. The analytical conclusions come from very limited, local angles, rather than global perspectives. Third, the analytics mainly aims at transaction data, and pays less attention to the operational data, due to its incapability to make real-time decisions.

In this paper, we introduce a unified data model based on random matrix theory and machine learning. Then, we present a framework that enables big data analytics in mobile cellular networks. In addition, we discuss several case studies of big data analytics in mobile cellular networks, including big signalling data, big traffic data, big location data, big radio waveforms data, and big heterogeneous data. Moreover, we present a number of challenges that need to be addressed for the deployment of big data analytics in mobile cellular networks. To the best of our knowledge, the interrelationship between big data and mobile cellular networks has not been well addressed in the existing works. In essence, it is the unique characteristics associated with big data and mobile cellular networks that present interesting challenges that have not been fully tackled in the literature. We believe that the works we have done here help understand how to make full use of big data analytics to improve the performance of mobile cellular networks. The rest of the article is organized as follows. In Section II, we present an overview of big data analytics. A framework of big data analytics is then presented in Section III. Following that, we discuss several case studies that leverage big data analytics in cellular networks. Section IV discusses several open research challenges. We conclude this study in Section V.

A. LINKING DEEP LEARNING WITH RANDOM MATRIX THEORY: A NOVEL VISION

Artificial intelligence and machine learning are main techniques for big data analytics. In the future, it is believed that they will account for 80-90% of the global computing resources. Deep learning may account for 40-50% of this market. Artificial intelligence is an area of computer science that deals with giving machines the ability to seem like they have human intelligence. One example is a robot. Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data. The process of machine learning is similar to that of data mining. Both systems search through data to look for patterns. However, instead of extracting data for human comprehension _ as is the case in data mining applications _ machine learning uses that data to improve the program's own understanding. Machine learning programs detect patterns in data and adjust program actions accordingly. The algorithms for big data analytics need data and engineering to validate. Since 2006, deep learning has been the most active new area of machine learning research. The major remaining challenges arise from the lack of explicit mathematical expressions: it is unclear that the designed algorithms are reproducible, extendable, theoretically provable, and interpretable. In the context of big data analytics for cellular networks, deep learning is of great interest. Deep learning the mathematics in data science and a basic methodology the state-of-the-art for images and speech. The depth of images can be extracted. On the other hand, the big data in cellular networks is distributed across different domains, e.g., space, time, codes, and antennas. Distributed machine learning is among the important areas in machine learning. Given the current explosion in the size and amount of data, a central challenge in machine learning is to design efficient algorithms for solving large-scale problem instances. In matrix neural networks, which takes matrices directly as inputs. Therefore, the input layer neurons form a matrix, for example, each neuron corresponds to a pixel in a grey scale image. The upper layers are also but not limited to matrices. Matrices are passing through each layer without vectorization that is previous algorithms. To achieve this, each neuron senses summarized information through bilinear mapping from immediate previous layer units' outputs plus an offset term. The basic model of a layer is the following bilinear mapping write formula

$$Y = \sigma (UXV^T + B) + E$$

where **U**; **V**; **B** and **E** are matrices with compatible dimensions, **U** and **V** are connection weights, **B** is the offset of current layer, σ(.) is the activation function acting on each element of matrix and **E** is the error. This bilinear mapping certainly connects matrix neural networks to matrix or tensor factorization type of algorithms such as principal component analysis (PCA). Random matrix theory is naturally linked with PCA.

B. BIG DATA ANALYTICS FOR MOBILE CELLULAR NETWORKS

Here, we highlight the connection between big data analytics and mobile cellular networks, and in a bigger picture, the link between data science and wireless networks. Big data analytics is not the replacement, but rather a supplement to all in the gap between today's higher requirement for deciphering more potential information and the traditional data analytics. In mobile cellular networks, the sources of traditional analytics are basically *centralized*, such as from charging and billing systems, operation systems, etc. In practice, however, the huge amount of data is scattered across the organization, like the device data, cell site data, network data, back office data, etc. Higher dimensionality of data implies better inference, as mentioned in Section II-A, $O(1=n)$; the convergence rate of the empirical eigenvalue distribution to its limit. Big data analytics supports an global point of view, making it possible to integrate the distributed collected data to extract correlation, as mentioned in the discussion of big distributed data in Section II-G. Semi-structured, unstructured data cannot be processed until big data analytics comes to the scene. Speaking of data volume, besides its suitability for magnitude of TB, even PB, big data analytics has a remarkable feature, called scalability, i.e., the ability to analyse the data with ever increasing scale and complexity. Mobile cellular networks' operational decisions have usually been made manually or depending on the hardware inside. With the advances of big data analytics, the operations of mobile cellular networks can be lower-error, higher-precision, dynamic, and importantly in real-time. Real-time reactions bring in better decisions, not only for the optimization of the network, but also the quality of user experience. Big data analytics has many striking advantages. Nevertheless, at this point, its analysis tools are generally complex and programming-intensive, not the same friendly as the traditional ones.

By effectively applying big data analytics, nearly every department involving sales and marketing, customer support, operation and maintenance, network construction, etc. can achieve significant benefits. Many opportunities are right over there, for instances, tailored marketing campaigns and recommendations can be carried out for a specific group of subscribers; more initiative concerns or solutions can be delivered to customers rather than waiting for the complains or claims; real-time monitoring of the operation and maintenance systems can prevent the fraudulent behaviours, or warn the congestion conditions; pinpointed network coverage analysis can facilitate the construction of network layout and further enhance subscribers' quality of experience. Through the various innovative schemes, mobile cellular operators can enhance customers' loyalty and thus lower the customer churn, simplify business operations, develop new services, reduce expenditure and increase revenue.

II. AN ARCHITECTURAL FRAMEWORK TO SUPPORT BIG DATA ANALYTICS IN MOBILE CELLULAR NETWORKS

In this section, we present an architectural framework to support big data analytics in mobile cellular networks.

A. DATA COLLECTION

Mathematically speaking, data collection is the process of forming data matrix **X**; shown in (1). It is worthwhile to point out that the dimensionality of data vector is very high. We take a view of treating data matrix **X** as a large random matrix. This view has many consequences,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

leveraging new mathematical results such as free probability.

Big data in mobile cellular networks can be gathered from either internal or external sources. The external data comes from the state/local statistical bureau, market research agencies, customer complaint departments, and etc. The internal sources usually refer to the operational systems, business systems and other supporting systems. Operational systems have been but not well explored by the traditional data analytics, and enormous values are expected to be excavated by the powerful big data analytics. Therefore, the data collection in operational systems is the focus of this section. Data collection methods can be divided into two categories: through data sources, and through auxiliary tools. Mobile devices themselves are data collection tools. For examples, they can: (1) collect audio information through microphones; (2) collect pictures, videos, and other multimedia information through cameras; (3) collect geological locations through GPS, Wi-Fi or Bluetooth, etc. Network data can be acquired through some package capture technologies or certain specialized software's, such as Com View, Smart Sniff, etc. Furthermore, professional staff can put some probes into the network interfaces, such as air interfaces, A/Gn interfaces, and etc. to collect the signalling data.

B. BIG DATA ANALYSIS AND PREPROCESSING

From a statistical analysis point of view, we extract correlation contained in the data matrix **X**. We present two basic approaches. One is the sample covariance matrix. The other is, through a matrix transform, to use the fundamental Singe Ring Theorem. The large-scale collected datasets reside at different locations with different formats. Thus, the gathered datasets are usually at a raw state with much redundancy, inconsistency or useless information. Meanwhile, the involved vast amount of semi-structured and unstructured data make it impossible to _t into the relational database with neat tables of columns
and rows. NoSQL databases are becoming the alternative technology for big data. To avoid unnecessary storage space, and ensure the processing efficiency, the data should be pre-processed to be ready for data analysis, before it is transmitted to the storage systems. For example, the data collected from mobile APPs, geolocation sensors, video cameras, network logs, CDRs, weblogs will not be stored directly in a storage system until it is pre-processed to correlate and normalize the data. Three common data pre-processing techniques are: integration, cleaning and redundancy elimination.

C. BIG DATA ANALYTICS PLATFORMS AND TOOLS

When it comes to big data, the most frequently mentioned word is Hadoop [37]. Apache Hadoop is an open-source software framework for distributed storage and distributed processing of large-scale datasets. The power of clusters enforces Hadoop to store and process data at an amazing speed. Initially, Hadoop is developed for such routine functions as keyword classification on search engines. To cater for the requirements of various business applications, Hadoop gradually turns into a general-purpose big data operating platform,
where different data manipulations and data analytical operations can be plugged into. All the features make Hadoop peculiarly adapt to processing or analysing the data in mobile cellular networks, such as CDRs, GPS data, web clickstream, network logs, etc. which is needed to be pulled out from storage systems for analysis frequently. Meanwhile, various suitable data analysis methods for mobile cellular networks can be integrated into Hadoop platforms. Apache Spark is a popular open-source platform for largescale data processing that is well-suited for iterative machine learning tasks. By allowing user programs to load data into a cluster's memory and query it repeatedly, Spark is well-suited to machine learning algorithms. In contrast to Hadoop's two-stage disk-based MapReduce paradigm, Spark's multi-stage in-memory primitives provides performance up to 100 times faster for certain applications.

D. BIG DATA ANALYTI APPLICATIONS

The applications of big data analytics in mobile cellular networks can be divided into two categories: internal business supporting applications and external innovative business model development. The internal business supporting applications mainly include the operational efficiency, subscribers' experience enhancement, tailored marketing, etc. As mobile cellular networks have vast amount of useful data, innovative business models can be promoted, such as the third-party data providers for various enterprises without infringement of subscribers' privacy.

II. CASE STUDIES OF BIG DATA ANALYTICS IN MOBILE CELLULAR NETWORKS

The following case studies provide an illustration of introducing big data analytics into mobile cellular networks. The focus is on improving network performance and deriving valuable insights. The application scope covers different scenarios, from current deployed mobile cellular networks to upcoming 5G, from network operational optimization to push for emerging research topics.

A. BIG SIGNALING DATA

In mobile cellular networks, the transmission of voice and data is accompanied by control messages, which are termed as signalling. The signalling works according to the predefined protocols and ensure the communication's security, reliability, regularity and efficiency. Signalling monitoring plays an important role in appropriate allocation of network resources, improving the quality of network services [39], real-time identifying network problems, and etc. With the rapid development of various mobile cellular networks, the volume of signalling data grows tremendously and the traditional signalling monitoring systems have too many problems to deal with. In Fig. 2, we describe a signalling data monitoring and analysing system architecture with big data analytics. This

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

architecture mainly consists of three components: data collecting, data analysing and applications. In data collection, various signalling protocols are copied from multiple network interfaces without interrupting normal operations. Afterwards, these copies are gathered and filtered through the protocol processor and then sent to the analyser.

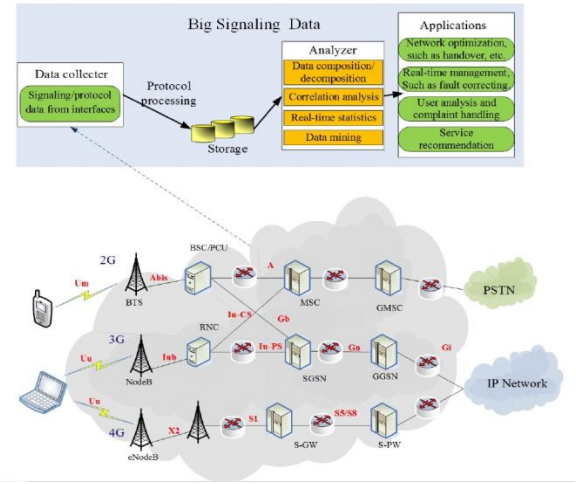

FIGURE 2. The big signalling data in cellular networks.

In the analyser, the data is processed using various algorithms, such as decomposition, correlation analysis, etc. Finally, the analysis results can be used by various applications. For example, Celibi *et al.* [40] analysed the BSSAP messages from *A* interface in a Hadoop platform to identify handovers from 3G to 2G. The simulation results show that the identified 3G coverage holes are consistent with the drive test results.

B. BIG TRAFFIC DATA

With the widespread usage of mobile Internet, the volume of traffic data increases at an unprecedented rate. Acting as a carrier of the traffic data, cellular operators have to manage the network resource appropriately to balance network load and optimize network utilization. Traffic monitoring and analysing is an elementary but essential part for network management, enabling performance analysis and prediction, failure detection, security management, etc. Traditional approaches to monitor and analyse the traffic data seem, however, straightforward and inadequate in the context of big traffic data, as illustrated in Fig. 3. The interrelationship between big data and software-defined networking (SDN) has been studied. Liu *et al.* [3] proposed a novel large-scale network traffic monitoring and analysis system based on a Hadoop platform. The system is practically deployed in a commercial cellular network with 4.2 Tbytes input volume every day. The evaluation results indicate that the proposed system is capable of processing big network-generated data and revealing certain traffic and user behaviour phenomenon. Since understanding the traffic dynamics and usage condition is of significance for improving network performance, the topic of traffic characteristics becomes a hot focus. In [5], the authors investigated three features of network traffic, namely network access time, traffic volume, and diurnal patterns from the perspective of device models. All the above results are beneficial for cellular network operators to make corresponding adjustments for network capacity management and revenue growth.

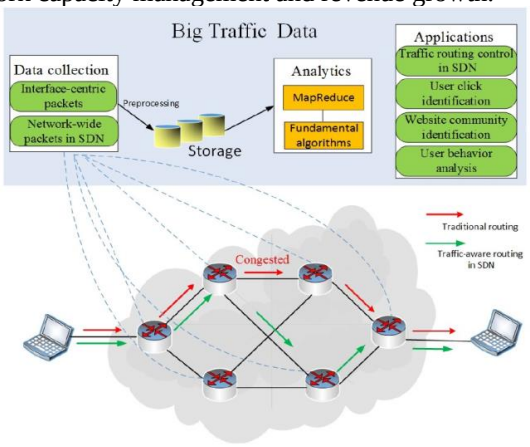

FIGURE 3. Big traffic data in mobile cellular networks.

C. BIG LOCATION DATA

Human activities are based on locations, and location data analysis is informative. As illustrated in Fig. 4, the locationbased big data arising from GPS sensors, WiFi, Bluetooth through mobile devices, have become precious strategic resources. These resources would provide support for government administration, such as public facility planning, transportation system constructions, demographic trends, risk warnings for crowed people, rapid emergence responses, crime hot spots analysis, etc. It can also gain amazing business insights, such as mobile advertising and marketing. An end-to-end Hadoop-based system was developed with a number of functional algorithms operated on call record details (CRDs). With the information about subscribers' habits and interests, it is capable of providing invaluable information about when, where and how a category of individuals (e.g., sports fans, music lover, et.) move.

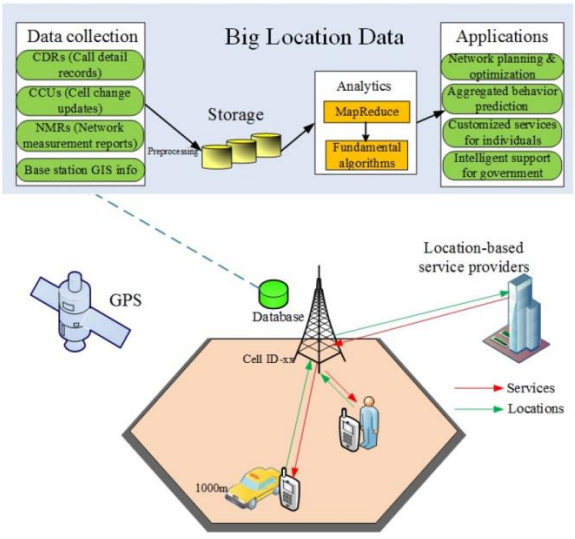

FIGURE 4. Big location data in mobile cellular networks.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

D. BIG RADIO WAVEFORMS DATA

Using large random matrices as building blocks to model the big data arising from a 5G massive MIMO system that is implemented using software-defined radios, as illustrated in Fig. 5. They exploited the fact that all data processing is done at CPU so all the modulated waveforms are stored at the RAMS or at the hard drives. On the other hand, big data analytics based on the random-matrix theory is applied to the collected data from their testbed, where a mobile user communicates with the massive MIMO base station while moving. The experimental results can estimate the user's moving speed, whether motionless, at a nearly constant speed, at a slow speed or at a higher speed. These analytics is also implemented to reflect the correlation residing in the transmitted signals. These applications validate the fact that the massive MIMO system is not only a communication system, but also a massive data platform which can brings tremendous values through big data analytics.

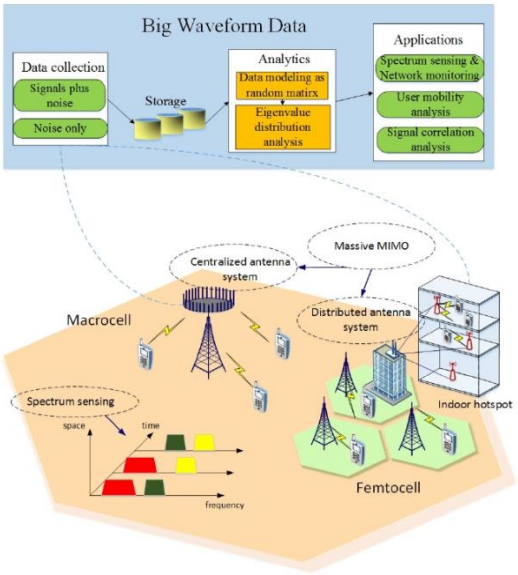

FIGURE 5. Big data captured for modulated radio waveforms in a massive MIMO cellular network.

E. BIG HETEROGENOUS DATA

One critical task of big data analytics in mobile cellular networks is the integration of very heterogenous data: correlation mining in massive database. Data sources are rich in types such as data rate, packet drop, mobility, etc. Different base stations host these data over time. They need be aggregated across space and time to obtain big data analytics. For example, for cyber security, there are many different heterogeneous sources, such as ``numerous distributed packet sniffers, system log _les, SNMP traps and queries, user profile databases, system messages, and operator commands." Essentially, data fusion is a technique to make overall sense of data from different sources that commonly have different data structures.

V. OPEN RESEARCH CHALLENGES

There are many open research challenges that are still not well studied and need to be tackled by future research efforts. We discuss some of these research challenges in this section. From viewpoints of practice, privacy may be among the most important challenges. More advanced algorithms are needed to extract correlations from the data, while allowing different levels of privacy. For example, for mobile phones, data associated with bank transactions are highly private information, which should be carefully handled in big data analytics in mobile cellular networks. In addition, government and corporate regulations for privacy and data protection play a fundamental and necessary role in protecting the sensitive aspects of big data in mobile cellular networks. How to _lter out un-useful data is another significant challenge, due to the scarce bandwidth available in mobile cellular networks. Mobile cellular networks can produce staggering amounts of raw data, a lot of which are not of interest. It can be _ltered out and compressed by orders of magnitude. One challenge is to de_ne these _lters in such a way that they do not discard useful information. Another challenge is to automatically generate the right metadata, to describe what data is recorded and how it is recorded and measured. This metadata is likely to be crucial to downstream analysis. Frequently, the information collected will not be in a format ready for analysis. We have to deal with erroneous data: Some quality reports for radio strengths, such as RSSI, frame error rate, and packet error rate, are inaccurate. Data analysis is considerably more challenging than simply locating, identifying, understanding, and citing data. For effective large-scale analysis, all of these have to happen in an *automated* manner. Mining requires integrated, cleaned, trustworthy, and efficiently accessible data, declarative query and mining interfaces, scalable mining algorithms, and big data computing environments. Today's analysts are impeded by a tedious process of exporting data from the database, performing a non-SQL process and bringing the data back. Having the ability to analyse big data is of limited value if users cannot understand the analysis. Ultimately, a decisionmaker, provided with the result of analysis, has to interpret these results.

VI. CONCLUSIONS AND FUTURE WORK

Big data analytics will be an indispensable part of the mobile cellular operators' consideration of network operation, business deployment, and even the design of the next-generation mobile cellular network architectures. In this paper, the connection Next, an architectural framework for the applications of big data analytics in cellular networks was presented. Moreover, several illustrative examples were provided. Finally, we discussed some research challenges and big data analytics' prospects for next-generation cellular networks. Future work is in progress to address these challenges.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

REFERENCES

[1]. http://www.sas.com/en_us/insights/big-data/internet-of-things.html
[2]. http://www.kdnuggets.com/2015/07/impact-iot-big-data-landscape.html
[3]. http://data-informed.com/the-impact-of-internet-of-things-on-big-data/

[4]. http://rdi2.rutgers.edu/sites/rdi2/files/img/Greer_Rutgers_BigData_Apr_2014.pdf
[5]. http://www.zdnet.com/article/the-internet-of-things-and-big-data-unlocking-the-power/
[6]. https://www.mapr.com/blog/what-internet-things-and-why-does-it-matter-big-data-0
[7]. http://leadwise.mediadroit.com/files/35428Advantech.pdf
[8]. http://www.sas.com/en_us/insights/big-data/internet-of-things.html between big data analytics and mobile cellular networks has been systematically explored. We provided a broad overview of big data analytics based on Random matrix theory.
[9]. http://www.kdnuggets.com/2015/07/impact-iot-big-data-landscape.html
[10]. http://data-informed.com/the-impact-of-internet-of-things-on-big-data/

[11]. http://rdi2.rutgers.edu/sites/rdi2/files/img/Greer_Rutgers_BigData_Apr_2014.pdf
[12]. http://www.zdnet.com/article/the-internet-of-things-and-big-data-unlocking-the-power/
[13]. https://www.mapr.com/blog/what-internet-things-and-why-does-it-matter-big-data-0
[14]. http://leadwise.mediadroit.com/files/35428Advantech.pdf
[15]. http://www.zdnet.com/article/is-the-internet-of-things-strategic-to-the-enterprise/
[16]. https://www.mapr.com/blog/14-benefits-and-forces-are-driving-internet-things