JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Big Data Analytics for Processing Real-time Unstructured Data from CCTV in Traffic Management

Kevin Sheth

Dept. of Computer Engineering Savitribai Phule Pune University,

Pune, Maharashtra, India

Abstract:-Today, many devices generate data Anytime, anywhere. The data has increased significantly. The handling becomes complicated. It is difficult to process and consists of unstable attributes. In traffic management, install CCTV Monitor specific locations on the highway. CCTV Create unstructured data for images and videos format. Due to the complexity of the data, it is difficult to process this data. This survey proposes to implement big data analytics to process unstructured real-time data from CCTV into displayed on the dashboard. It uses the Yolo V4 architecture and the COCO dataset to implement the YOLO framework for traffic flow counting and detection of illegal parking that is classified as anomalous situations. Next, the unstructured data from CCTV was converted to JSON semi-structured format. The data can also be visualized in real-time to help local governments understand highway conditions. Historical data is stored in the NoSQL database to give you a deeper understanding of vehicle traffic patterns and more. This system requires an ROI drawing line as a trigger to count the passing vehicles.

Keywords:- Real-time ,CCTV, unstructured data ,big data, Traffic management.

I. INTRODUCTION:-

Big data is used to describe a large collection of data. Data grows rapidly and has complex types. Consists of several characteristics such as volume, speed, and variety. Types of data are i) structured ii) unstructured iii)semi-structured. Structured data is fixed format data that can be easily processed using traditional databases such as RDBMS. A semi-structured form is a structured form, but it has dynamic attributes. JSON or XML format. Unstructured data is an unfamiliar format such as images and videos. Many applications today generate large amounts of data. Produces big data phenomena everywhere. Big data phenomena occur in many places Includes industry, banking, media, tourism, health care, transportation, and more. Since the data is in an unstructured format, it produces a lot of data, but I don't know how to handle it.

One of the main areas is traffic management, where businesses or governments collect data from images. And video from CCTV. Many CCTVs are installed on highways to monitor traffic. Generated data From CCTV, is one example of unstructured data. The data is large in size and grows very rapidly in video format. To process unstructured data, you need a big data solution. At the same time, you need to understand the complexity of traffic as soon as possible. This means that unstructured data from CCTV needs to be processed in real-time.

There are two ways to process data: 1) batch processing and 2) real-time processing. More real-time data. It's more complicated than a batch because it needs to be processed in a short amount of time or in near real-time. Real-time processing requires processing several domains, one of which is in traffic management. Real-time processing helps authorities instantly understand current traffic

conditions. One example is to understand traffic density. Big data analysis helps you analyze CCTV traffic in real-time and return structured reports to active executives.

II. LITERATURE SURVEY:-

SR NO	PAPER TITLE , PUBLICATION DETAILS	PRE-PROCESSING	FEATURE EXTRACTION AND CLASSIFICATION	ACCURACY	POST-PROCESSING	RESEARCH GAP IDENTIFIED
1	Nada Elgendy and Ahmed Elragal "Big Data Analytics: A Literature Review Paper" Springer International Publishing Switzerland August 2014	Due to the rapid pace of data growth, solutions must be discovered and provided to manage and extract value and knowledge from these data sets.	Big Data Analytics and Decision Making, Customer Intelligence, Supply Chain and Performance Management, Quality Management and Improvement, Risk Management and Fraud Detection	90%	By using various methods we can handle the data	In future research can focus on providing road map or framework.
2	D. P. Acharjya Kauser Ahmed P "A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools" International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016	This big data analysis requires efforts on multiple levels to extract the knowledge needed for decision making.	Hadoop, Mahout, Spark,etc	-	It provides a platform to explore big data at multiple stages	efficient tools to be developed must have provision to handle noisy and imbalance data, uncertainty and inconsistency, and missing values
3	Saurabh Malgaonkar, Sanchi Soral, Shailja Sumeet , Tanay Parekhji ," Study on Big Data Analytics Research Domains"	Data analytics is the trending field of data analysis to observe patterns and predict future outcomes.	Cloud Systems, Data Analytics And Interoperability, Data Analysis, Machine Learning And Neural Networks	91%	The final part of the article covers machine learning algorithms and neural networks to train a dataset to recognize patterns from the modeled data and to predict outcomes based on training and pattern recognition.	More research can be done in the same domain.
4	Pichaimuthu Mohankumar "BIG Data Analytics: A Framework	Most companies have an unstructured	HBase using Hadoop,	89%	The paper finds an efficient way	By using text mining algorithms ,we

	for Unstructured Data Analysis" International Journal of Engineering and Technology (IJET) March 2013	model. Information retrieval and extraction is necessary and important work in the fields of the Semantic Web.			to store unstructured data and an appropriate approach to data retrieval	would try to get more insights
5	Suyash Mishra, Dr Anuranjan Misra "Structured and Unstructured Big Data Analytics" (ICCTCEEC-2017)	Data is generated in Various formats so its difficult to analyze the data. There is no fix format	Real-time Online or stream Processing, Batch Processing		They describe about various techniques and software used to Manage, process unstructured big data in efficient manner, and increases the performance of complexity analysis.	By using MapReduce unstructured data can transformed and converted into structured data
6	Jaein Kim, Nacwoo Kim, Byungtak Lee, Joonho Park, Kwangik Seo, "RUBA: Real-time Unstructured Big Data Analysis Framework" October 2013	The recommendation framework provides dynamic modification and real-time analysis for unstructured big data analysis	Big Data, Unstructured Data, Real-time System, CEP, CQL.	88%	The object monitoring system is implemented as a test system applied to our framework and we have confirmed all the functionality and usability of our framework.	We would try to use RUBA Framework for real time applications
7	Rubal, Sheetal Kalra "Real-Time Applications of Big Data- A Survey" (IJERT) Vol. 5 Issue 03, March-2016	A large amount of data is generated from different sources that can be structured or unstructured. This type of data is difficult to process and manage and contains millions of records information including social media, web sales, etc.	Big data, Hadoop, HDFS, MapReduce, NoSql, Realtime data analytics	90%	To improve the quality of information and decision making it is important to effectively analyze this large volume of data to answer new challenges.	We can upgrade to Analysis as a Service (AaaS).
8	Sasan Amini, Ilias Gerostathopoulos, Christian Prehofer "Big Data Analytics Architecture for Real-Time Traffic Control" Conference Paper · June 2017	There is a critical need to develop new tools and systems to keep pace with the rise of big data.	Kafka, HDFS	80%	They proposed a bendy structure. The structure is primarily based totally on a scientific evaluation of	We can use Kafka streams or spark streams to perform complex projects.

					the necessities of the domain	
9	N. Naga Lakshmi and T.Asha Latha "AUTOMATED TRAFFIC MANAGEMENT SYSTEM USING BIG DATA TECHNOLOGY" (IJLTET)	large scale challenges to perform large scale calculations are very difficult and large amounts of data are generated from sensors.	Captures Snapshot of vehicle, Send SMS to owner	89%	Smart sensors are used for identifying drivers who ignore traffic laws.	We can implement this system to reduce accidents.
10	Mauricio Perez, Alex C. Kot, Anderson Rocha "DETECTION OF REAL-WORLD FIGHTS IN SURVEILLANCE VIDEOS" IEEE 2019	Previous work was either too superficial or unrealistic. No one has done real time detection on Long-duration CCTV recording.	3D-CNN, local interest-points	92%	Datasets containing 1000 videos of real fights, More than 8 hours of CCTV material with comments	spatial features, which have not demonstrated positive complementary to the temporal information
11	Payal Saha, Mohit Mittal, Shreya Gupta "Big Data Trends and Analytics: A Survey" (IJCA) 2018	t is envisioned that the Big Data concept will ensure that huge chunks of data are reduced to a manageable form.	Big data, Hadoop, Mapreduce, Data analytics, Big data tools.	95%	Discussed concepts of Big Data and challenges	More research is required because data is increasing day by day
12	Subramaniyaswamy , Vijayakumar, Logesh R and Indragandhi V "Unstructured Data Analysis on Big Data using Map Reduce" (ISBCC'15)	Social networking sites like Facebook, Twitter have discovered that data growth will get out of control in the future	Hadoop, MapReduce, Collaborative Filtering, Mahout, Maven, Sentiment Analysis	85%	It processes data in parallel as fractions in distributed clusters and aggregates all data between clusters to get final processed data.	New analysis is done
13	Jeffrey Dean and Sanjay Ghemawat ," MapReduce: Simplified Data Processing on Large Clusters" OSDI 2004	MapReduce is related programming and implementation model for processing and generating datasets	Large-Scale Indexing	98%	The MapReduce programming model has been successfully used at Google for various purposes.	Large datasets can be implemented through MapReduce Technique
14	Joao Ricardo Lourenco, Veronika Abramova, Bruno Cabral, Jorge Bernardino "NoSQL in practice: a write-heavy enterprise application" 2015 IEEE International Congress on Big Data	Current benchmarks evaluate database performance by running specific queries on mostly aggregated data	NoSQL, Big Data, Enterprise, Write-Heavy, MongoDB, Couchbase, Cassandra, SQL Server	85%	A homogeneous cluster using four machines with similar hardware was chosen to host the databases.	We should use in real time
15	Ankit Parag Shah, Jean-Bapstite Lamare, Tuan Nguyen-Anh, and Alexander Hauptmann "CADP: A Novel	Addresses the lack of public data for the study of automated spatiotemporal	CN,ACM, R-CNN	80%	demonstrated the performance of accident forecasting in	We should implement in the future

	Dataset for CCTV Traffic Camera based Accident Analysis" 2018 IEEE	annotations for road safety.			dataset using Faster R-CNN and an Accident LSTM architecture	

III. ALGORITHMIC SURVEY:-

Sr No	Paper Title	Algorithm Used	Time Complexity	Space Complexity	Accuracy	Advantages/ Disadvantages
1	Big Data Analytics: A Literature Review Paper	MapReduce and HDFS, B-DAD framework	O(K), O(1)	O(M),	80%	Scalability, Cost-effective solution, I t's not always very easy to implement each and everything as a MR program.
2	Study on Big Data Analytics Research Domains	Clustering, Classification, Predictive Data Mining	O (kn2)	O(k+n)	85%	Probabilistic Approach, gives information about statistical significance of features. The assumptions of logistic regression.
3	BIG Data Analytics: A Framework for Unstructured Data Analysis	Hbase	O(log(e))	-	-	Can store large data sets. No support SQL structure
4	DETECTION OF REAL-WORLD FIGHTS IN SURVEILLANCE VIDEOS	CNN	O(N)	-	81%	it automatically detects the important features without any human supervision. CNN do not encode the position and orientation of object.
5	CADP: A Novel Dataset for CCTV Traffic Camera based Accident Analysis	R-CNN	O(nt∗(ij+jk+kl)).	-	95%	Very High accuracy in image recognition problems. CNN do not encode the position and orientation of object.

IV. CONCLUSION:-

Deep learning algorithms and NoSQL databases are big data technologies that can process unstructured data in real time. These are very helpful in understanding traffic conditions and for police officers to monitor the highway. The proposed prototype can recognize objects such as cars, trucks, and buses, and aggregate vehicle types. The YOLO v4 model and COCO dataset have been

trained to classify highway traffic objects. You can also analyze the normal and unusual status of real-time unstructured data. Powered by the Dell Inspiron GPU 1050 Ti, i7 Intel processor, this real-time system can monitor real-time surveillance with 10fps real-time CCTV streaming.

REFERENCES:-

1. Nada Elgendy and Ahmed Elragal "Big Data Analytics: A Literature Review Paper" August 2014 DOI: 10.1007/978-3-319-08976-8_16
2. D. P. Acharjya Kauser Ahmed P "A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools" International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016
3. Saurabh Malgaonkar, Sanchi Soral, Shailja Sumeet , Tanay Parekhji ," Study on Big Data Analytics Research Domains
4. Pichaimuthu Mohankumar "BIG Data Analytics: A Framework for Unstructured Data Analysis" International Journal of Engineering and Technology (IJET) March 2013
5. Suyash Mishra, Dr Anuranjan Misra "Structured and Unstructured Big Data Analytics" (ICCTCEEC-2017)
6. Jaein Kim, Nacwoo Kim, Byungtak Lee, Joonho Park, Kwangik Seo, "RUBA: Real-time Unstructured Big Data Analysis Framework" October 2013
7. Rubal, Sheetal Kalra "Real-Time Applications of Big Data- A Survey" (IJERT) Vol. 5 Issue 03, March-2016
8. Sasan Amini, Ilias Gerostathopoulos, Christian Prehofer "Big Data Analytics Architecture for Real-Time Traffic Control" Conference Paper · June 2017
9. N. Naga Lakshmi and T.Asha Latha "AUTOMATED TRAFFIC MANAGEMENT SYSTEM USING BIG DATA TECHNOLOGY" (IJLTET)
10. Mauricio Perez, Alex C. Kot, Anderson Rocha "DETECTION OF REAL-WORLD FIGHTS IN SURVEILLANCE VIDEOS" IEEE 2019
11. Payal Saha, Mohit Mittal, Shreya Gupta "Big Data Trends and Analytics: A Survey" (IJCA) 2018
12. Subramaniyaswamy , Vijayakumar, Logesh R and Indragandhi V "Unstructured Data Analysis on Big Data using Map Reduce" (ISBCC'15)
13. Jeffrey Dean and Sanjay Ghemawat ," MapReduce: Simplified Data Processing on Large Clusters" OSDI 2004
14. Joao Ricardo Lourenco, Veronika Abramova, Bruno Cabral, Jorge Bernardino "NoSQL in practice: a write-heavy enterprise application" 2015 IEEE International Congress on Big Data
15. Ankit Parag Shah, Jean-Bapstite Lamare, Tuan Nguyen-Anh, and Alexander Hauptmann "CADP: A Novel Dataset for CCTV Traffic Camera based Accident Analysis" 2018 IEEE