A Framework for Multi-Sensor Data fusion in the Context of IoT Smart City Parking Data

Monika Saxena¹, Vaibhav Vyas¹ and C.K. Jha¹

¹Department of Computer Science, Banasthali Vidyapith, Banasthali, Niwai Rajasthan 304022 India

E-mail: muskan.saxena@gmail.com

Abstract. Now a day, things are communicating using the Internet. In the context of a smart city all things are interconnected and they are transferring the data to each other. Smart city sub-applications like parking, pollution, traffic, transportation are some examples. By using the internet, we can combine with communication, the big challenge in a multi-sensory environment to fetch meaning full data from sensors. There are many of the methods are used to analyse and mine IoT data like classification, clustering, association rule, and regression, etc. The data is collected on the Internet of Things (IoT) from heterogeneous sources in different formats. Data Fusion is a method to operate on fused data that are collected from different sources. In the case of IoT, the concept of distributed data fusion work well. One of the popular IoT applications is Smart City, where we can collect data and analysis them according to need. In this research paper, the researcher introduced the new framework of data fusion where the data is collected from a different type of smart city sub-application. and framework deployment is done by parking data set of smart cities.

1. Introduction

The Internet of Things and data on IoT needs complex models with many of the different parameters. The conventional methods are giving a limited performance on Big IoT Data. The performance is calculated in form of Scalability, throughput, timeliness, overhead, computational cost, and heterogeneity. The managing and mining of Real-world applications in Big IoT Data are challenging yet very compelling tasks [1]. To analysis, the complexity of data over the IoT, the researcher has reviewed data, model, and system-level models. For the heterogeneous data mining the powerful computing processing and platforms are required [1]. Researchers get the following challenges to analyze Big data analytics in IoT.

1.1 Scaling: The data is growing rapidly day by day. Traditional algorithms like HACE, K-Means, etc. used to improve performance, are not giving better solution if data scaled up [2].

1.2 Data Heterogeneity: The IoT application sources like Smart Cities, Social Media, etc. are of different types of data sets. To filter out useful data from a large set of data is a challenge in now a day.

1.3 Time Evolving Data: The size of data increasing day by day as well as the time complexity of the algorithm is increased. The data filtered from 1000 GB take more time than data filtered from 1000MB. Due to heterogeneity and noisy data time complexity is increases [3].
1.4 Compression: It is better to store large data in the compressed form. So, it will be easy to transfer over the internet.

1.5 Visualization: Visualization of Big Data is a fundamental challenge. We can't major how much data available in social media or other Big Data resources like a multi-sensor.

1.6 Hidden Data: We need to extract useful knowledge and behavior that must process from the Big complex Data [2].

1.7 Multisensory Data Fusion: The Big data can be categorized in to Multi-source, Heterogeneous and Multi-Model. The basic data is divided in form of spatial and temporal data. the major challenge is to analyze the spatial and temporal data when the data comes from different heterogeneous sources [4].

![Characteristics of Multisensory Data Fusion](image)

Figure 1 is showing the different characteristics of multisensory data fusion. The data can be defined by Multi source, as it is collected from various different sources. The data can be heterogeneous, because it is containing the various types of data like spatial, temporal, static, dynamic etc. The data collected from different sources can be of different type, so it need to model by Multi modeling process, here we can model the different type of data together.

2. Related Work

The basic data that is collected from different sources can be three types: Temporal, Spatial and Spatial-Temporal. The properties of temporal data are periodicity and moment for example: Alarm beeps at 9:00 AM, temperature of room is at 3:00 PM and 4:00 PM. Spatial data contains the properties like durability it comes like position, shape and size. If we talk about heterogeneous sources then collected data contains the properties of both spatial and temporal data [5].

Many of the data fusion models are available to calculate and analyse the data. They are categorized in to feature based, stage based and semantic based. Some of the categories are shown in below figure 2.
Figure 2. Categorization of Data Fusion Models

The common data fusion methods giving excellent performance for the spatial and temporal data. But when we need to analyze data from multi-sensory environment, where the data is collected from heterogeneous data sources, no above category of data fusion methods will work [6].

The analytics of multisensory data and Big data requires the Deep learning. Up to now no direct method or model is available to analyze spatial-temporal data fusion. This paper is representing a framework with proposed distributed multi-sensory data fusion framework and model (Algorithm) for big data and multi-sensory heterogeneous data analytics.

3. Research Goal and Objectives
A lot of solutions and modifications are required for variety, volume, and velocity of data. The basic problems are state following, that researchers are going to overcome using the new proposed data fusion framework.

- Process Mining of heterogeneous Multi-Sensory Data
- Scheduling of Process Multi-Sensory Data for high performance
- To find the hidden pattern in IoT data set

In analyzing and processing Multi-Sensory Data, researchers have performance challenges in terms of high performance, process scheduling, pattern mining and decision making. The variety and volume of dynamic big data will easily be analyzed by distributed data fusion [7].
3.1 Research Approach
The proposed technique run in a process of many phases like:

a) Collecting the IoT data: The Multisensory data, from different sources like parking dataset, transportation dataset, pollution dataset, and IoT dataset [6].

b) Data preparation or preprocessing: there are preparation processes were run on the collected Big Data to make them ready for processing by the algorithms [8].

c) Implementation of the new fusion algorithm: design the logic of the new algorithm and writing the code of the new technique.

d) Evaluation of the developed algorithm: The new algorithm run on the test data which is prepared in the previous phases and compare the data and results, in terms of speed of processing and space of storage in order to evaluate the improvements that has been achieved by the proposed technique [9].

e) Analyze algorithm on tools like MATLAB to compare the results of the new technique with the traditional methods to show the improvements in terms of processing time & storage.

3.2 Proposed Framework
The new proposed method and framework is designed to remove the limitations of the available algorithms for heterogeneous data. The proposed algorithm and framework are easy to implement and it lower the storage requirement as well. furthermore, it reduces the time of execution because the whole process executes in single pass [2].

The proposed method is using the concept of the multilevel hashing. The hashing method based on defining the transaction weight of data through the database transactions. The weight of data is used in searching in between thousands of transactions it reduces the time of searching and accelerate the process. In a multilevel hashing scheme, a hash table is created level wise, using the following approach each item on data I is associated with H hash values. The all of the items I will placed in to a bucket. that bucket is arranged in form levels and each item are placed there with its location [3].

To get the result one hash function is required, the hash function for each bucket on each level is managed by fully functional or Edge devices. When the data is collected from different Edge devices the semantic based data fusion algorithm is applied on it. according to the semantic based fusion the data is classified in the same semanticity [4]. The proposed fusion algorithm is containing the concept of semantic based fusion algorithm on Edge devices or on Controller part.

The proposed framework specified in below figure 3, The static and dynamic data is collected from different sources the data forms are both static and dynamic. The data cleaning is applied for different data sets. The proposed algorithm is applied on collected data sets; the proposed algorithm is based on Multilevel hashing scheme. The proposed algorithm is based on semantic based fusion algorithms. According to semantic algorithms the context aware data is collected from different sources. Each reduced functional constrained device (Sensors) transfers their data to Edge devices. The integrate function integrates all the data in final model.

Figure 4 is showing the authority, expertise and total value of collected context data, which is applied on final model integration.
3.3 Evaluation Method

As mentioned above the newly proposed method does not generate candidates instead it works in one pass as. Pass 1: It starts with creating the item sets directly from transactions. it reads the transaction and form the item set and utilizes the multilevel hash table techniques based on its weight to check if it already exists in the item set table it increments the count of the frequent by 1 otherwise if it does not exist it add the item set into the itemset table and moves to the next transaction and repeat the process.
until it reaches to the last transaction. At this time, it completes the table of the items set having every item set with its frequent value [1].

Each item that is collected from edge devices are provided one transaction ID, using transaction ID the item is allotted into bucket. If the item is already available there then it contains only the location of sensor from where it comes from. By this the hash function and patterns are created. In the next occurrence the pattern is matched again with the available pattern. The same pattern shows the same behaviour and transfer scheme of sensors. Using the multi-level hash scheme, the execution of same pattern is faster than other available algorithms [10].

The below figure 5 flowchart is showing the whole procedure of the new proposed algorithm. Firstly, some parameters are initialized, which is iteration, transaction id, level of the hash table, pattern size, and bucket. The whole process will perform according to the value of the support count [11].

**Figure 5. Flow Chart of Proposed Method**

All the patterns will be created by column-wise comparison, if the item already exists then will check for pattern, if the pattern already exists then according to the multi-level hash table the item will be
added on next level; if it does not exist, then simply item or the pattern will be added in the same level of a hash table and value of bucket will be incremented. The process will be performed until the pattern size is equal to the number of items. For every item and pattern, the same procedure will perform, the weight of occurrences is also calculated.

4. Results & Simulation
The proposed process is simulated with a new modified fusion algorithm in context to check Input/output Load & Traffic Control.

4.1 Test Case: The Smart City Parking Data Set
In smart cities, parking can be logged by using a variety of sensors that logs hundreds of physical measurements to count the number of cars in each park at any timestamp [1]. The multi-sensory data have been collected from the below data source: http://iot.ee.surrey.ac.uk:8080/dataset.

4.1.1 Sample of Data
The data size is 2 GB with more than 2 million records. Below the sample of the raw data of the measurements of the parking parameters are listed in below table 1.

| Vehicle Count | Update Time | ID | Total Spaces | Garage Code | Stream Time |
|---------------|-------------|----|--------------|-------------|-------------|
| 0             | 09:04.1     | 1  | 65           | NORREPORT   | 11/3/2014 16:18 |
| 0             | 09:04.1     | 2  | 512          | SKOLEBAKKEN | 11/3/2014 16:18 |
| 869           | 09:04.1     | 3  | 1240         | SCANDCENTER | 11/3/2014 16:18 |
| 22            | 09:04.1     | 4  | 953          | BRUUNS      | 11/3/2014 16:18 |
| 124           | 09:04.1     | 5  | 130          | BUSGADEHUSET| 11/3/2014 16:18 |
| 106           | 09:04.1     | 6  | 400          | MAGASIN     | 11/3/2014 16:18 |

4.1.2 Data Preparation
The following encoding schemes are applied on collected data. After the encoding of Raw data researchers gets the data in following format of table 2.

| Update Time | Vehicle Count | Total Spaces | Garage Code | ID |
|-------------|---------------|--------------|-------------|----|
| 1000904     | 1000          | 11           | 6           | 101|
| 1000904     | 1000          | 55           | 9           | 101|
| 1000904     | 1869          | 88           | 8           | 101|
| 1000904     | 1022          | 77           | 1           | 101|
| 1000904     | 1124          | 22           | 2           | 101|

4.1.3 The Parking Test Results
The above-proposed fusion method is implemented in MATLAB. Parking dataset prepared data of approx. 2,11,55,265 records, the following summary table shows the results of the number of frequency counts for all minimum supports (1%, 5%, 10%, 15%, 20%) as shown in table 3.
The proposed framework satisfies better results in-memory storage than the data mining algorithms like FP-Growth as well as better convergence time than pattern mining algorithms like Apriori. Furthermore, the graphs are showing the difference between all three methods [3].

| Apriori                  | FP-Growth                  | The New Method          |
|-------------------------|----------------------------|-------------------------|
| for support 1%          | for support 1%             | for support 1%          |
| Max memory usage : 214.016 mb | Max memory usage: 312.02mb | memory usage: 85.1526 MB |
| Total time ~ 268ms      | Total time ~ 148ms        | Total time ~ 218ms      |
| for support 5%          | for support 5%             | for support 5%          |
| Max memory usage : 82.123595 MB | Max memory usage: 151.33764297 MB | Max memory: 119.5662 MB  |
| Total time ~ 212ms      | Total time ~ 100ms        | Total time ~ 180ms      |
| For Support 10%         | for support 10 %           | For support 10%         |
| Max memory usage : 142.9518 MB | Max memory usage: 104.04248 MB | Max memory usage: 81.122058 MB  |
| Total time ~ 132ms      | Total time ~ 79ms         | Total time ~ 108ms      |
| for Support 15%         | for support 15 %           | for support 15%         |
| Max memory usage : 104.009 MB | Max memory usage: 154.120 MB | Max memory usage: 89.100 MB  |
| Total time ~ 54ms       | Total time ~ 63ms         | Total time ~ 49ms       |

Figure 6 & 7 shows the data visualization, the graphs are showing the comparative results of proposed, FP growth and Apriori. In the comparison of fusion based proposed method and conventional mining algorithms the fusion algorithms are giving better performance in case of Internet of things data.
The above graph results figure 6 and 7 clearly shows that the novel method and framework has given a relatively better performance than conventional algorithms in terms of the time of processing and memory usage with different support count values [12][13][14][15][16][17].

5. Conclusion
All the result has been evaluated with different support counts and different type of datasets. According to the proposed framework the knowledge base is integrated at the local model and using the proposed algorithm different patterns will find out at the final model level. In the final model, it is very interesting to see the pattern making between two different knowledge bases. Using the proposed framework number of patterns and relationships will be easy to find in between different knowledge bases. For example, we can relate Traffic and Parking datasets also for example if the Traffic knowledge base has always the same traffic rate at a particular time and at the same time Parking is available than it is very easy to determine when Parking is available according to traffic day. So basically, it is easy to relate two different knowledge bases and to create different patterns. It clearly shows that by seeing only one parameter we can easily find out the value of another parameter.

6. References
[1]. Saxena, Monika & Jha, Dr. (2019). A New Pattern Mining Algorithm for Analytics of Real-Time Internet of Things Data. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(1), 1178-1183, 10.35940/ijitee.A4506.119119.
[2]. Saxena, Monika & Jha, C. (2016). Heterogeneous Big Smart data in Perspective of Smart Cities Development, International Journal of Computer Science Engineering (IJCSE), 5(3), 137-141.
[3]. Saxena, Monika & Jha, C. (2018). Analytics of IoT Streaming Data using Modified New Pattern Mining Algorithm. International Journal on Future Revolution in Computer Science & Communication Engineering, 4(1), 265-270
[4]. Saxena, Monika & Jha, C. (2018). Analytics of Pollution Smart City Data using New Pattern Mining Algorithm. International Journal of Engineering Research in Computer Science and Engineering, 5(1), 5-10
[5]. Dhaka, V. S., R. C. Poonia, and L. Raja. “A literature review on dedicated short range communication for intelligent transport.” International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE) 3.9 (2013): 1066-1071.
[6]. Rathee, G., Sandhu, R., Saini, H. et al. A trust computed framework for IoT devices and fog computing environment. *Wireless Network* 26, 2339–2351 (2020). https://doi.org/10.1007/s11276-019-02106-3

[7]. He, Qing, et al. "Parallel implementation of classification algorithms based on Map Reduce." *Rough Set and Knowledge Technology* (2010): 655-662.

[8]. Verma, Manish, et al. "A comparative study of various clustering algorithms in data mining." *International Journal of Engineering Research and Applications* (IJERA) 2.3 (2012): 1379-1384

[9]. Lokeswari, Y. V., and Shomona Gracia Jacob. "A Comparative study on Parallel Data Mining Algorithms using Hadoop Map Reduce: A Survey." *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*. ACM, 2016.

[10]. Arora, Amandeep Singh, Linesh Raja, and Barkha Bahl. "Data centric security approach: A way to achieve security & privacy in cloud computing." *Proceedings of 3rd International Conference on Internet of Things and Connected Technologies* (ICIoTCT). 2018.

[11]. Chakraborty, Sanjay, and Naresh Kumar Nagwani. "Analysis and study of Incremental DBSCAN Cluster Algorithm." arXiv Preprint arXiv: 1406.4754 (2014).

[12]. Kobren, Ari, et al. "An Online Hierarchical Algorithm for Extreme Clustering." arXiv preprint arXiv:1704.01858 (2017).

[13]. Saxena M., Saurabh P., Verma B. (2012) A New Hashing Scheme to Overcome the Problem of Overloading of Articles in Usenet. In: Wyld D., Zizka J., Nagamalai D. (eds) Advances in Computer Science, Engineering & Applications. *Advances in Intelligent and Soft Computing*, vol 166. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-30157-5_96

[14]. Saxena, M., Saurabh, P., & Verma, B. (2011). USENET: Internet News Software, Security–Needs and Goals. *International Journal of Scientific and Engineering Research*, 2(3), 159–164.

[15]. Saxena, D., & Saxena, M. (2011). Methods Used to Handle Overloading of Information in Usenet. *International Journal of Scientific and Engineering Research*, 49.

[16]. Saxena M. (2020). Novel framework and service model for internet of things in context of smart cities. *International Research Journal of Modernization in Engineering Technology & Science*. 2(9), 1730-1735.

[17]. Saxena, Monika. (2021). IoT Dynamic Log File Analysis: Security Approach for Anomaly Detection In Multi Sensor Environment. *International research journal of modernization in engineering technology & Science* Volume 3. 2582-5208