ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

A HETEROGENEOUS DATA FUSION FRAMEWORK FOR TRAFFIC AND INCIDENT PREDICTION

Gunnala Umesh [1], Donthi Devendar [2],
Kondabathini Keerthana [3], Dr. E. John Alex[4]
[1,2,3] UG Student, Dept. of ECE, CMR Institute of Technology, Hyderabad
[4] Professor, Dept. of ECE,
CMR Institute of Technology, Hyderabad

ABSTRACT

This paper introduces DataFITS (Data Fusion on Intelligent Transportation System), an open-source framework that collects and fuses traffic-related data from various sources, creating a comprehensive dataset. We hypothesize that a heterogeneous data fusion framework can enhance information coverage and quality for traffic models, increasing the efficiency and reliability of Intelligent Transportation System (ITS) applications. Our hypothesis was verified through two applications that utilized traffic estimation and incident classification models. DataFITS collected four data types from seven sources over nine months and fused them in a spatiotemporal domain. Traffic estimation models used descriptive statistics and polynomial regression, while incident classification employed the k-nearest neighbors (k-NN) algorithm with Dynamic Time Warping (DTW) and Wasserstein metric as distance measures. Results indicate that DataFITS significantly increased road coverage by 137% and improved information quality for up to 40% of all roads through data fusion. Traffic estimation achieved an R^2 score of 0.91 using a polynomial regression model, while incident classification achieved 90% accuracy on binary tasks (incident or non-incident) and around 80% on classifying three different types of incidents (accident, congestion, and non-incident).

INTRODUCTION

DATA availability is a critical aspect in the design of modern Intelligent Transportation Systems (ITSs), which implement models to understand better various patterns of the transportation system [1], thus improving mobility and safety for people and goods. With modern society depending heavily on efficient and reliable transportation, the importance of these systems has seen a rapid increasein significance over recent years. In Germany alone, both the number of registered cars and the number of carried passengers using public transportation have

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

shown a substantial increase, reaching their all-time

highs of 48.5 million cars (2022) and 12.7 billion carried passengers (2019, before the pandemic) [2], [3]. As a result, urban areas experience an increasing number of traffic-related incidents (e.g., congestion and accidents), increasing time delays, emissions, and fuel consumption [4].

For this reason, academia and industry have driven efforts to create the next generation of transportation systems that are eco-friendly, cost-efficient, and powered by data analysis and communication technology. We hypothesize that a heterogeneous data fusion framework can enhance the coverage and quality of information serving as input for traffic models, thus increasing the efficiency and reliability of ITS applications. Therefore, we propose the Data Fusion on Intelligent Transportation System (DataFITS) framework, providing a spatiotemporal fusion of data used to train models for two ITS applications, traffic estimation, and incident classification. DataFITS collects and combines real heterogeneous data (e.g., weather, traffic, incident) from various sources (e.g., open databases, map applications), preparing them by fixing errors, adapting the data structure, and finally fusing them in the exact location and point in time. Our hypothesis is verified using data characterization to quantify the benefits of combining heterogeneous data sources and the proposal of two ITS applications. The performance of the two applications ratifies the benefits of larger data coverage/quality while estimating traffic and classifying incidents. Thus, the main contributions of this investigation are:

- An open-source framework DataFITS for heterogeneous spatiotemporal data fusion, covering the acquisition, processing, and fusion of data, available in a public code repository.[1]
- The characterization of a heterogeneous dataset combining real traffic data from two cities in Germany, collected from seven sources over nine months and provided together with the repository.
- Two traffic estimation models, one using descriptive statistics and another using polynomial regression with different parameters such as time, road type, and weather, and a comparison between single and fused datasets.
- An incident classification model trained and evaluated on heterogeneous fused data using k-nearest neighbors

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

(k-NN), with Dynamic Time Warping (DTW) and Wasserstein as distance methods.

LITERATURE REVIEW

"Big data analytics in intelligent transportation systems: A survey,"

Big data is becoming a research focus in intelligent transportation systems (ITS), which can be seen in many projects around the world. Intelligent transportation systems will produce a large amount of data. The produced big data will have profound impacts on the design and application of intelligent transportation systems, which makes ITS safer, more efficient, and profitable. Studying big data analytics in ITS is a flourishing field. This paper first reviews the history and characteristics of big data and intelligent transportation systems. The framework of conducting big data analytics in ITS is discussed next, where

the data source and collection methods, data analytics methods and platforms, and big data analytics application categories are summarized. Several case studies of big data analytics applications in intelligent transportation systems, including road traffic accidents analysis, road traffic flow prediction, public transportation service plan, personal travel route plan, rail transportation management and control, and assets maintenance are introduced. Finally, this paper discusses some open challenges of using big data analytics in ITS Big data is becoming a research focus in intelligent transportation systems (ITS), which can be seen in many projects around the world. Intelligent transportation systems will produce a large amount of data. The produced big data will have profound impacts on the design and application of intelligent transportation systems, which makes ITS safer, more efficient, and profitable. Studying big data analytics in ITS is a flourishing field. This paper first reviews the history and characteristics of big data and intelligent transportation systems. The framework of conducting big data analytics in ITS is discussed next, where

the data source and collection methods, data analytics methods and platforms, and big data analytics application categories are summarized. Several case studies of big data analytics applications in intelligent transportation systems, including road traffic accidents analysis, road traffic flow prediction, public

transportation service plan, personal travel route plan, rail transportation management and control, and assets maintenance are introduced. Finally, this paper discusses some open challenges of using big data analytics in ITS Big Data is an emerging paradigm and has currently become a strong attractor of global interest, specially within the transportation industry. The combination of disruptive technologies and new concepts such as the Smart City upgrades the transport data life cycle. In this context, Big Data is considered as a new pledge for the transportation industry to effectively manage all data this sector required for providing safer, cleaner and more efficient transport means, as well as for users to personalize their transport experience. However, Big Data comes along with its own set of technological challenges, stemming from the multiple and heterogeneous transportation/mobility application scenarios. In this survey we analyze the latest research efforts revolving on Big Data for the transportation and mobility industry, its applications, baselines scenarios, fields and use case such as routing, planning, infrastructure monitoring, network design, among others. This analysis will be done strictly from the Big Data perspective, focusing on those contributions gravitating on techniques, tools and methods for modeling, processing, analyzing and visualizing transport and mobility Big Data. From the literature review a set of trends and challenges is extracted so as to provide researchers with an insightful outlook on the field of transport and mobility.

"Smart city data platform for real-time processing and data sharing,"

The concept of a smart city comes with the need to support a data platform that can gather, process and export the data of millions of sensors, coming from different sources, with information in different formats, in a scalable approach for real-time and historical data visualization, processing, and actuation in the city.

This paper proposes a Data Platform for the Aveiro Tech City Living Lab, to gather, process, visualize and actuate on mobility, environmental and network data. The architecture of the platform provides a real open platform that is accessible for third-parties to collect data and to experiment their own solutions, through a secure and open data platform at their disposal. The platform was also designed to be scaled and modified when necessary, being composed by distributed solutions across the existent modules. The results with respect to the amount of data gathered, the ingestion and query performance comparison between the persistence solutions considered, and examples of data, show how this platform can be used to efficiently develop new applications and use both real-time and

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

historical data for future predictions and actuations in a smart city.

The current cities still have several social problems, such as road traffic congestion, environmental pollution, among others. Moreover, as time goes by, the cities' population tends to grow, which will worsen those problems. Smart cities are a promising solution to these problems [1], [2], by providing infrastructures to gather information from different sensors and devices, enabling the knowledge of the city status, and the detection of the problems and possible solutions to them. The possibility to predict the city status through the intelligence on the current and historic gathered data enables the support of mechanisms to improve the city management, and ultimately, the cities' overall quality of life.

"On the design of vehicular virtual sensors,"

Physical sensors are an important part of control systems, especially vehicular control systems. Sensor readings help drivers to control their vehicles as well as their internal systems while keeping a vehicle stable and running. Currently, a modern luxury car carries hundreds of diverse and precise sensors and not all of them are visible to the driver. However, there are phenomena and aspects for which there are no physical sensors available. Virtual sensors combine readings from multiple sensors in order to develop their own output values based on conditions and models, and, eventually, substitute and monitor failing physical sensors, as well as sense complex variables. Designing a virtual sensor is usually a difficult process due to the complexity of the different processing stages it comprises. This work studies the process of creating and prototyping vehicular virtual sensors, describing the development stages and presenting examples of virtual sensors created with a framework developed to facilitate the design process.

"City data hub: Implementation of standard-based smart city data platform for interoperability,"

Like what happened to the Internet of Things (IoT), smart cities have become abundant in our lives as well. One of the smart city definitions commonly used is that smart cities solve city problems to enhance citizens' life quality and make cities sustainable. From the perspective of information and communication technologies (ICT), we think this can be done by collecting and analyzing data to generate insights. The City Data Hub, which is a standard-based city data platform that has been developed, and a couple of problem-solving examples have been

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

demonstrated. The key elements for smart city platforms have been chosen and they have been included in the core architecture principles and implemented as a platform. It has been proven that standard application programming interfaces (APIs) and common data models with data marketplaces, which are the keys, increase interoperability and guarantee ecosystem extensibility.

Smart cities are everywhere these days. Like we have been hearing for the Internet of Things (IoT) from TV commercials, such as smart home services, even on load signs there are big sign boards promoting smart cities and their solutions.

Since smart city is not a technical terminology but rather a concept, there are a lot of definitions for a smart city from standard development organizations, research papers and also private corporates [1]. The one commonality of those definitions is that a smart city enhances citizen life quality by solving city problems. There should be many non-technical aspects to achieve them, but from the perspective of information and communication technologies (ICT), data can help in realizing the smart city mission. To do so, smart city systems, whether it currently existing or newly deployed later, should be interoperable in terms of data sharing and utilization.

"Driver authentication in VANETs based on intravehicular sensor data,"

Internet of Vehicles (IoV) has emerged as an advancement over the traditional Vehicular Ad-hoc Networks (VANETs) towards achieving a more efficient intelligent transportation system that is capable of providing various intelligent services and supporting different applications for the drivers and passengers on roads. In order for the IoV and VANETs environments to be able to offer such beneficial road services, huge amounts of data are generated and exchanged among the different communicated entities in these vehicular networks wirelessly via open channels, which could attract the adversaries and threaten the network with several possible types of security attacks. In this survey, we target the authentication part of the security system while highlighting the efficiency of blockchains in the IoV and VANETs environments. First, a detailed background on IoV and blockchain is provided, followed by a wide range of security requirements, challenges, and possible attacks in vehicular networks. Then, a more focused review is provided on the recent blockchain-based authentication schemes in IoV and VANETs with a detailed comparative study in terms of techniques used, network models, evaluation tools, and attacks counteracted.

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

Lastly, some future challenges for IoV security are discussed that are necessary to be addressed in the upcoming research.

With the huge increase in the number of vehicles on roads nowadays, more accidents and traffic congestion issues are encountered. This raises the need for serious arrangements to ensure roads' safety and traffic efficiency. Different technologies have been introduced towards maintaining safer and time-efficient driving on roads, such as, Vehicular Ad-hoc Networks (VANETs) in which the vehicles exchange data about their speed, location, etc., and other road-related information to raise their awareness about surrounding road conditions and help them making better and effective decisions. However, with the rapid advancement in today's technologies such as ubiquitous connectivity, wireless technologies, sensor devices, smart vehicles, and cloud computing platforms, the need for more powerful vehicular networks has increased. Hence, the IoV has appeared that can exploit and incorporate all these advanced technologies in order to provide more satisfying real-time services for vehicles' drivers and passengers.

IoV has emerged with great potential to support various services and offer several benefits to the transportation system such as cost effectiveness, time efficiency, road safety [1], traffic management [2,3,4], evolution of smart cities [5,6], autonomous driving [7,8,9,10] alarms and dynamic warning systems [11,12,13] as well as recording fatal occurrences [14]. However, in order for the IoV system to be able to secure such services, enormous amounts of data need to be generated and exchanged among the different IoV entities including vehicles, pedestrians, and roadside infrastructure. Since this information exchange takes place through an open-channel wireless network, the exchanged messages are vulnerable to various security attacks that could undermine the privacy of the communicating entities and the confidentiality of their data via eavesdropping or even affect the integrity of the transmitted messages by tampering them before reaching their target destination. Other types of security attacks that could be encountered in IoV environments are the attacks that target the authenticity of users. Here, a malicious entity masquerades a legitimate user and may commit malicious activities in the network. Therefore, efficient authentication is necessary to prevent such attacks in IoV.

EXISTING SYSTEM

Flood prediction is complex and challenging due to its dynamic nature and

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

reliance on climate conditions [26]. Changes in weather directly relate to flooding probability due to global warming and melting snow [27]. In literature, various techniques, models, frameworks and methods have been proposed to predict or somehow control vast disasters caused by floods.

Data-driven statistical methods used for flood prediction include autoregressive integrated moving average (ARIMA) [28], multiple linear regression (MLR) [29] Flood frequency analysis (FFA) [30] and regional flood frequency analysis (RFFA) [12] have been proposed. Later, empirical orthogonal function (EOF) [31], Bayesian forecasting models (BFM) [32], quantile regression techniques (QRT) [33], and climatology average method (CLIM) [34] have also been focused for flood prediction. Advanced flood forecasting systems for long-term and short-term prediction of floods are also significant for the generation of early flood warnings. The Hydraulic models of flow have been practiced forecasting rainfall, storms and tsunamis [35]. These models are also used to predict impact of climatic change [36], ocean waves [37] and floods [38].

Proposed System

In this article, Flood Forecasting Model (FFM) has been developed to forecast floods in multiple rivers and barrages of the selected region. The proposed model is composed of five layers as presented in the proposed system. Flood prediction involves a huge amount of multidimensional data. In physical layer, data collection centers are present at the edge, where sensors have been used to collect and transmit data to the local client station for local model training. These sensors include rain gauge sensors, water flow calculating sensors and water level sensors. In the proposed model hydrological and meteorological dataset of a region selected from Central Asia have been processed.

Advantages

1) We Proposed FFNN model which contained three layers of hidden nodes for transformation of input into output to make flood forecast.

2) We implemented Machine learning (ML) methods which is highly contributed in the advancement of prediction systems by providing better performance and cost-effective solutions.

CONCLUSION

In this paper, we introduce DataFITS, an open-source data fusion framework that integrates diverse data by collecting, analyzing, and fusing it. We hypothesize

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

that heterogeneous data fusion increases data quantity and quality, thereby improving datasets for ITS applications. To verify this, we developed two ITS applications: one used polynomial regression to estimate traffic levels, while the other combined traffic incident data to classify events into accident, congestion, or non-incidents. Using real heterogeneous data from two German cities, we quantified the advantages of DataFITS by compiling a fused dataset. Our results indicate that DataFITS integrated data from multiple sources for 40% of all roads, thereby increasing the overall road coverage by 137%. In addition, the traffic estimation model, which uses polynomial regression, outperformed our previous approach based on descriptive statistics, achieving a high R^2 score of 0.91, low error metrics of 0.05, and provides accurate traffic estimations using the fused dataset. Compared to using a single sources dataset, the fused dataset estimation showed minor accuracy improvements but drastically improved the spatiotemporal coverage of the estimated areas. Our incident classification model relies on the fusion of traffic and incident data, achieving a 90% binary classification accuracy rate within our evaluation. Preprocessing the data, such as removing unclear traffic patterns, improved accuracy by an average of 29%. The classification of incidents into different categories resulted in a slightly lower accuracy of 86%, with unequal performance among classes indicated by F1 scores. To mitigate this problem, we oversampled the training dataset to create a more uniform representation of the data, resulting in an 80% accuracy for each class. Collecting more accident data can also solve this problem.

We plan to expand the DataFITS framework by collecting and fusing more data types, improving its performance and data quality, and expanding its data analysis. We focus on data types such as social media and images, which require methods such as Natural Language Processing (NLP) and image processing. For ITS applications, we aim to use automated machine learning to explore different models and hyper-parameters and compare them with our current models. We also plan to analyze the correlation between traffic and incidents and incorporate it into the traffic estimation models. In addition, we intend to explore the use of big data in military scenarios, combining information from the civilian and military fields to support strategic operations in urban warfare. To this end, our framework can be enhanced to collect and combine different types of information (image, text) to create common operational pictures and

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

verify/authenticate information, thereby avoiding misinformation that may influence political decisions.

REFERENCES

[1] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 383–398, Jan. 2019.

[2] Umweltbundesamt. (2022). *Verkehrsinfrastruktur und fahrzeugbestand*. Accessed: Dec. 12, 2022. [Online]. Available: https://www.umweltbundesamt.de/daten/verkehr/verkehrsinfrastrukturfahrzeugbestand

[3] German Federal Statistical Office (Destatis). (2022). *Passengers Carried in Germany*. Accessed: Jul. 12, 2022. [Online]. Available: https://www.destatis.de/EN/Themes/Economic-SectorsEnterprises/Transport/Passenger-Transport/Tables/passengerscarried.html

[4] G. Vítor, P. Rito, and S. Sargento, "Smart city data platform for real-time processing and data sharing," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Sep. 2021, pp. 1–7.

[5] A. B. Campolina, P. H. L. Rettore, M. Do Val Machado, and A. A. F. Loureiro, "On the design of vehicular virtual sensors," in *Proc. 13th Int. Conf. Distrib. Comput. Sensor Syst. (DCOSS)*, Jun. 2017, pp. 134–141.

[6] S. Jeong, S. Kim, and J. Kim, "City data hub: Implementation of standard-based smart city data platform for interoperability," *Sensors*, vol. 20, no. 23, p. 7000, Dec. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/23/7000

[7] L. Zhang, Y. Xie, L. Xidao, and X. Zhang, "Multi-source heterogeneous data fusion," in *Proc. Int. Conf. Artif. Intell. Big Data (ICAIBD)*, May 2018, pp. 47–51.

[8] P. H. L. Rettore, B. P. Santos, A. B. Campolina, L. A. Villas, and A. A. F. Loureiro, "Towards intra-vehicular sensor data fusion," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 126–131.

[9] P. H. L. Rettore, A. B. Campolina, L. A. Villas, and A. A. F. Loureiro, "A method of eco-driving based on intra-vehicular sensor data," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2017, pp. 1122–1127.

[10] P. H. L. Rettore, A. B. Campolina, A. Souza, G. Maia, L. A. Villas, and A. A. F. Loureiro, "Driver authentication in VANETs based on intravehicular sensor

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

data," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2018, pp. 00078–00083.

[11] [9] Karne, R. K. ., & Sreeja, T. K. . (2023). PMLC- Predictions of Mobility and Transmission in a Lane-Based Cluster VANET Validated on Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5s), 477–483. https://doi.org/10.17762/ijritcc.v11i5s.7109

[12] [10] Radha Krishna Karne and Dr. T. K. Sreeja (2022), A Novel Approach for Dynamic Stable Clustering in VANET Using Deep Learning (LSTM) Model. IJEER 10(4), 1092-1098. DOI: 10.37391/IJEER.100454.

[13] [11] Reddy, Kallem Niranjan, and Pappu Venkata Yasoda Jayasree. "Low Power Strain and Dimension Aware SRAM Cell Design Using a New Tunnel FET and Domino Independent Logic." International Journal of Intelligent Engineering & Systems 11, no. 4 (2018).

[14] [12] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Design of a Dual Doping Less Double Gate Tfet and Its Material Optimization Analysis on a 6t Sram Cells."

[15] [13] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Low power process, voltage, and temperature (PVT) variations aware improved tunnel FET on 6T SRAM cells." Sustainable Computing: Informatics and Systems 21 (2019): 143-153.

[16] [14] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Survey on improvement of PVT aware variations in tunnel FET on SRAM cells." In 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), pp. 703-705. IEEE, 2017