iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** II **Month of publication:** February 2016

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Spatiotemporal Traffic Prediction using Semantic Traffic Analytics and Reasoning(STAR) With Big Data Environment

L. Pravin Kumar[1], Ramesh Krishnan[2]
Internship, BMW India Pvt. Ltd., Chennai, India.
Research Scholar, Department of Electronics and Communication Engineering, Easwari Engineering College, Anna University, Chennai, India-600 089.

Abstract: In Urban Mobility Report, delays due to heavy traffic costing Americans $78 billion in the form of 4.2 billion lost hours and 2.9 billion gallons of wasted fuel. In addition, 2/3 of traffic delays are caused not by recurring congestion but by point-based spontaneous congestion due to traffic incidences. STAR-CITY, which integrates (human and machine-based) sensor data using variety of formats, velocities and volumes, has been designed to provide insight on historical and real-time traffic conditions, all supporting efficient urban planning. The real-time traffic situation to the most effective predictor constructed using historical data, thereby self-adapting to the dynamically changing traffic situations. Also includes in proposed with the distributed scenarios with the global traffic prediction.
Keywords : big data,Traffic prediction, spatiotemporal, context-aware, online learning

I. INTRODUCTION

The two most important commodities of the 21st century are time and energy; traffic congestion wastes both. Several disciplines, such as in transportation science, civil engineering, policy planning, and operations research have studied the traffic congestion problem through mathematical models, simulation studies and field surveys. However, due to the recent sensor instrumentations of road networks in major cities as well as the vast availability of auxiliary commodity sensors form which traffic information can be derived (e.g., CCTV cameras, GPS devices), for the first time a large volume of real-time traffic data at very high spatial and temporal resolutions has become available.While this is a gold mine of data, the most popular utilization of this data is to simply visualize and utilize the current real-time traffic congestion on online maps, car navigation systems, sig-alerts, or mobile applications. However, the most useful application of this data is to predict the traffic ahead of you during the course of a commute. This predictive information can be either used by a driver directly to avoid potential gridlocks or consumed by a smart route-planning algorithm (e.g., [6]) to ensure a driver picks the best route from the start. According to a study by McKinsey Global Institute [1], using traffic information that avoids congestion can save consumers $600 billion annually by 2020.Working with real-world data, we have identified certain characteristics of traffic data, such as temporal patterns of rush hours or the spatial impacts of accidents, which can be incorporated into a data-mining technique to make it much more accurate. For example, for generic time-series, the observations made in the immediate past are usually a good indication of the short-term future. However, for traffic time series, this is not true at the edges of the rush hours. In that case, the historical observations (perhaps for that same day, time, and location) are better predictors of future. Hence, an auto-regression algorithm such as ARIMA [3], which by itself cannot capture sudden changes at the temporal boundaries of rush hours, can be enhanced by incorporating historical patterns. While predicting short-term future has many applications, for example in fixing the errors of sig-alerts during rush hours, it is not useful for smart path-planning where some- times we need to know the traffic of a road-segment ahead of us by 30 minutes in advance. Again, historical data can improve long-term predictions because most probably the traffic behavior in 30 minutes at the desired location is similar to (say) yesterday's traffic at the same time and location. The main challenge is how to properly incorporate all the knowledge from historical and real-time data into an appropriate time-series mining technique. This is exactly what we accomplished in this paper by enhancing ARIMA.

II. EXISTING SYSTEM

Our experimental results with real-world LA data show that our enhanced ARIMA outperforms ARIMA by 78% when there is no unexpected events, and over 91% in the presence of events. In addition, we compared our enhanced approach with other competitor

©IJRASET 2013: All Rights are Reserved

Volume 4 Issue II, February 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

techniques for traffic prediction (e.g., [7] and [6]) and showed the superiority of our approach.

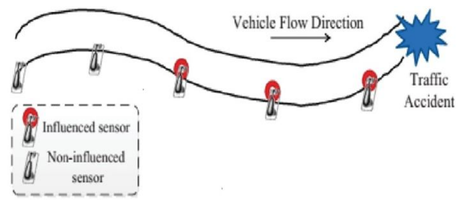


Fig.1 Influenced sensor and non-Influenced sensor

Rahul Mangharam, Oleg Sokolsky et al Vehicle traffic probing is a core issue in end-to-end travel time prediction as the network is dynamic and prediction estimates must be made within a short duration or they will be out-of-phase with the current network state. Travel-time prediction is fundamentally a Cyber-Physical Systems problem of massive scale as computations must consider real-time trajectory data from ≥1 million vehicles in a region along with spatio-temporal environmental variables such as weather, road conditions, sports events, school and factory hours, etc. to accurately estimate a canonical set of fastest paths for all drivers by a deadline. Effective congestion probing and prediction algorithms must consider both recurring traffic patterns and non-recurring spontaneous traffic events. The traffic infrastructure must support computation over a large amount of historic traffic data on one end and acquire timely updates from a large number of vehicles on the other end. Given the dynamical nature and scale of the problem, neither classical centralized nor distributed algorithms will perform well. Furthermore, unlike classical Real-Time Systems, where all tasks are assumed to be compliant, in the case of traffic prediction, the drivers in-the-loop exhibit a varying degrees of non-compliance. A new class of data-dependent real-time algorithms and system architectures are necessary to compute across large statistical variations and must continually deliver outcomes with the desired level of QoS within a deadline. If for example, street segments are treated as shared resources in time and space, then scheduling of fastest vehicle paths along all origin-destination pairs requires that drivers conform to a weighted estimate of both centrally and locally computed paths.

Vehicular traffic congestion is a significant and growing problem that affects the entire nation's population. Traffic congestion is a dynamical problem where both sporadic traffic incidences affect travel time delay in a disproportionately large manner and spatiotemporal environmental factors affect traffic behavior in the region and for the duration of the event. Finally, traffic congestion prediction is a data-dependent CPS-problem of massive scale where traffic assignment decisions have to be made by a deadline based on terabytes of semi-global historic traffic data on one hand and hundreds of thousands of real-time vehicle trajectory updates on the other.

III. RELATED WORK

Traffic prediction approaches can be grouped in two main categories: Simulation Models and Data Mining Techniques. Simulation Models: The traffic prediction techniques developed in the first category use surveys and/or simulation models. In [5], Clark proposes a non-parametric regression model to predict traffic based on the observed traffic data. In [7] and [2], authors use microscopic models upon trajectories of individual vehicles to simulate overall traffic data and further conduct prediction Yuan at el. estimate the traffic flow of a road segment by analyzing taxi trajectories. The major limitation of such studies is that they rely on sporadic observations and are often restricted to synthetic or simplified data for simulations.

A. Data Mining Techniques

The increase in the availability of real-time traffic allowed researchers to develop and apply data mining techniques to forecast traffic based on the real-world datasets. Since early 1980s, univariate time series models, mainly Box-Jenkins Auto-Regressive Integrated Moving Average (ARIMA) [3] and Holt-Winters Exponential Smoothing (ES) have been widely used in traffic prediction. In the last decade, Neural Network (NNet) models also has been extensively used in forecasting of various traffic parameters, including speed, travel time, and traffic flow. Nowadays, ARIMA, ES and NNet models are used as benchmarking methods for short-term traffic prediction [17], [16]. However, these approaches consider traffic flow as a simple time-series data and ignore phenomenons that particularly happen to traffic data. For example, for generic time-series, the observations made in the immediate past are usually a good indication of the short-term future. However, for traffic time-series, this is not true at the edges of the rush hours, due to sudden speed changes.

©IJRASET 2013: All Rights are Reserved

Volume 4 Issue II, February 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Traffic Event Analysis The effect of events on traffic prediction has also been studied in the fields of data mining and transportation engineering. The majority of these studies focused on real time event/outlier detection using probabilistic or rule-based approaches (e.g., [14], [9], [13]). There are also several studies that mainly concern the cause of the events, aiming at how to design the network or re-direct the traffic flows to avoid the delay of events (e.g., [4], [16]). However, none of these studies incorporate events into traffic prediction techniques, and hence fail to provide realistic estimations in the presence of events.

The focus of this paper, on the other hand, is to integrate the impact of various events into forecasting models. The most relevant work to our study is the model proposed by Kwon and Varajya[11]. Their model utilizes a nearest-neighbor technique to detect cumulative delays and impact regions caused by traffic incidents. The impact regions are defined with fixed thresholds. However, the impact of events on traffic congestion varies based on space and time.

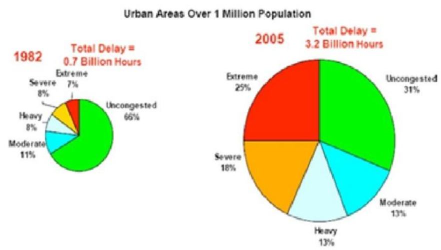


Fig.2 Growth of vehicle traffic congestion

For example, the impact region of an accident occuring during rush hour is usually more severe. Similarly, an accident at an inter-state street has a different impact region than that of a surface street. In this study, we consider such spatiotemporal characteristics of traffic events in training our models.

IV. PROPOSED SYSTEM

A. Spatio-Temporal Diagnosis of Traffic Status

How to identify the nature and cause of traffic congestion in real-time? How to capture diagnosis results on a spatial and (historical) temporal basis? How to understand the impact of city events on traffic conditions? These are general questions which cannot be answered by existing state-of-the-art traffic systems, but of really importance for city managers to better understand and plan her/his cities at any time. Such question remains open because (i) relevant data sets (e.g., road works, city events), (ii) their correlation (e.g., road works and city events connected to the same city area) and (iii) historical traffic conditions (e.g., road works and congestion) are not fully open and jointly exploited. STAR-CITY exploits the DL-based semantics of streams to tackle these challenges. Based on an analysis of stream behavior through change and inconsistency over DL axioms, we tackled change diagnosis by determining and constructing a comprehensive view on potential causes of changes [4]. Some extensions of the latter work have been achieved to support both scalable real-time and historical aggregation of diagnosis results.

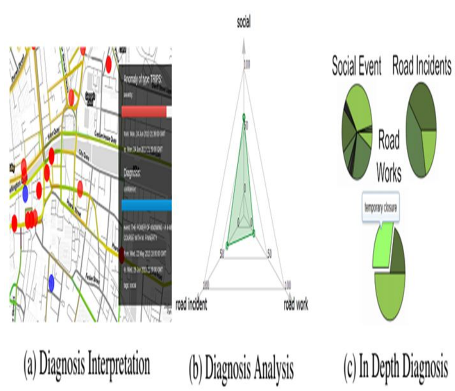


Fig. 3: Spatio-Temporal Historical and Real-Time Diagnosis in STAR-CITY (color print)

©IJRASET 2013: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Since the diagnosis reasoning of STAR-CITY strongly relies on classification of OWL 2 EL ontologies, we adopt a distributed classification [16] of OWL 2 EL journey times individuals to obtain a scalable diagnosis. The current implementation is limited to EL ++ expressivity for scalability reasons.

B. STAR Architecture and Technology

Semantic Representation: The model we consider to represent static background knowledge and semantics of data stream is provided by an ontology, encoded in OWL2 EL4. The selection of the W3C standard OWL 2 EL profile has been guided by (i) the expressivity which was required to model semantics of data in Table 1, (ii) the scalability of the underlying basic reasoning mechanisms we needed. Semantic Enrichment: All raw data streams in Table 1 are served as real-time OWL 2 EL ontology streams by using IBM InfoSphere Streams. Different mapping techniques are used depending on the data format. All the ontology streams have the same static background knowledge to capture time (W3C Time Ontology 5), space (W3C Geo Ontology 6) but differ only in some domain-related vocabularies e.g., traffic flow type, weather phenomenon, event type. These ontologies have been mainly used for enriching raw data, facilitating its integration, comparison, and matching. The DBpedia vocabulary has been used for cross-referencing entities.

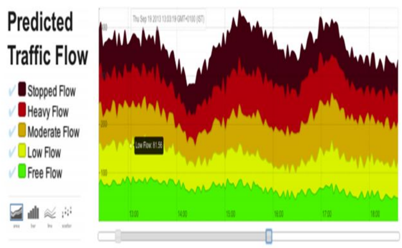


Fig. 4: Prediction of Traffic Status (color print).

Distributed Semantic Reasoning: Completion EL++ rules [11], used for classification, are distributed across various nodes based on their types. Each node is dedicated to at most one type of (normal form) axioms and runs the appropriate rule on axioms. Semantic Stream Reasoning: Real-time semantic comparison and matching of stream snapshots are operated. Such computing is required by predictive reasoning and realtime diagnosis for elaborating semantic context (events, weather, incidents) similarity and correlation over time, all in real-time. The stream correlation is established by comparing the number of changes i.e., new, obsolete, invariant ABox entailments between snapshots. The latter ensures context-aware diagnosis and prediction.

Semantic Rule Association and Mining: Predictive reasoning is achieved following state-of-the-art principles i.e., rules association mining. The generation of association rules between streams (and their snapshots) is based on a DL extension of Apriori [12], aiming at supporting subsumption for determining association rules. Contrary to the initial version of Apriori, the association is achieved between any ABox elements together with their entailments (e.g., all congested roads, weather, works, incidents, city events). Rules are encoded in SWRL, and all consequents of each rule are validated though consistency checking. This ensures to obtain consistent, accurate prediction results. REST Interface : All functionalities of STAR-CITY are exposed through REST services, providing component-ization, evolve-ability via loose coupling and hypertext.4

Web User Interface : STAR-CITY strongly relies on HTML, CSS, Javascript (Dojo toolkit, D3, JQuery libraries) to produce an appealing user interface. Time-series, spider charts together with parallel charts are examples where Dojo and D3 components were combined with HTML and CSS.

Deployment : Our technology stack is based on (i) well-established commercial components from IBM e.g., IBM InfoSphere Streams for stream enrichment and processing, IBM WebSphere as the HTTP/Application Server, and (ii) state-of-the-art components such as pssh for parallel distribution of reasoning, Jena TDB 7 as RDF store.

V. CONCLUSION

In this current framework distributed scenarios where traffic data is gathered by distributed entities and thus, coordination among the Spatio-Temporal Diagnosis of Traffic Status updation, traffic prediction ,Semantic Rule Association and Mining. In future with the Programmable Car where functionality may be purchased and enabled on the fly via software services.

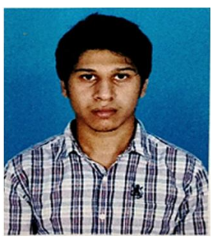
©IJRASET 2013: All Rights are Reserved

Volume 4 Issue II, February 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

REFERENCES

[1] http://www.nytimes.com/2011/05/13/technology/13data.html last visited june 14, 2012.
[2] M. Ben-akiva, M. Bierlaire, H. Koutsopoulos, and R. Mishalani. DynaMIT: a simulation-based system for traffic prediction. In DACCORD'98, Delft, The Netherlands.
[3] G. Box and G. Jenkins. Time series analysis: Forecasting and control. San Francisco: Holden-Day, 1970.
[4] M. M. Chong, A. Abraham, and M. Paprzycki. Traffic accident analysis using decision trees and neural networks. In IADIS'04, Portugal.
[5] S. Clark. Traffic prediction using multivariate nonparametric regression. In JTE'03, volume 129.
[6] B. Williams, P. Durvasula, and D. Brown. Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models. InTRR'98, volume 1644.
[7] H. M. Zhang. Recursive prediction of traffic conditions with neural network models. InJTE'00, volume 126
[8] M Khanjary, SM Hashemi, Route guidance systems: review and classification, in Proceedings of 6th Euro American Conference on Telematics and Information Systems (EATIS) (Valencia, 2012), pp. 1–7
[9] A. Ihler, J. Hutchins, and P. Smyth. Adaptive event detection with time-varying poisson processes. In ACM SIGKDD'06, NY, USA.
[10] S. Ishak and C. Alecsandru. Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings. In JTE'04, volume 130.
[11] J. Kwon, M. Mauch, and P. P. Varaiya. Components of congestion : delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. In TRR'06, pages 84-91.
[12] S. Lee and D. B. Fambro. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. In TRR'98.
[13] X. Li, Z. Li, J. Han, and J. Lee. Temporal outlier detection in vehicle traffic data. In ICDE '09.
[14] X. Li and F. Porikli. A hidden markov model framework for traffic event detection using video features. In ICIP '04, 2004.
[15] R. S. Marshment, R. C. Dauffenbach, and D. A. Penn. Short-range intercity traffic forecasting usingeconometric techniques. In ITE Journal, volume 66, 1996.
[16] N. L. Nihan and J. N. Zhu. Short-term forecasts of freeway traffic volumes and lane occupancies, phase 1. In TNW'93, volume 5.
[17] B. Park, C. J. Messer, and T. I. Urbanik. Short-term freeway traffic volume forecasting using radial basis function neural network. Number 1651.

AUTHORS



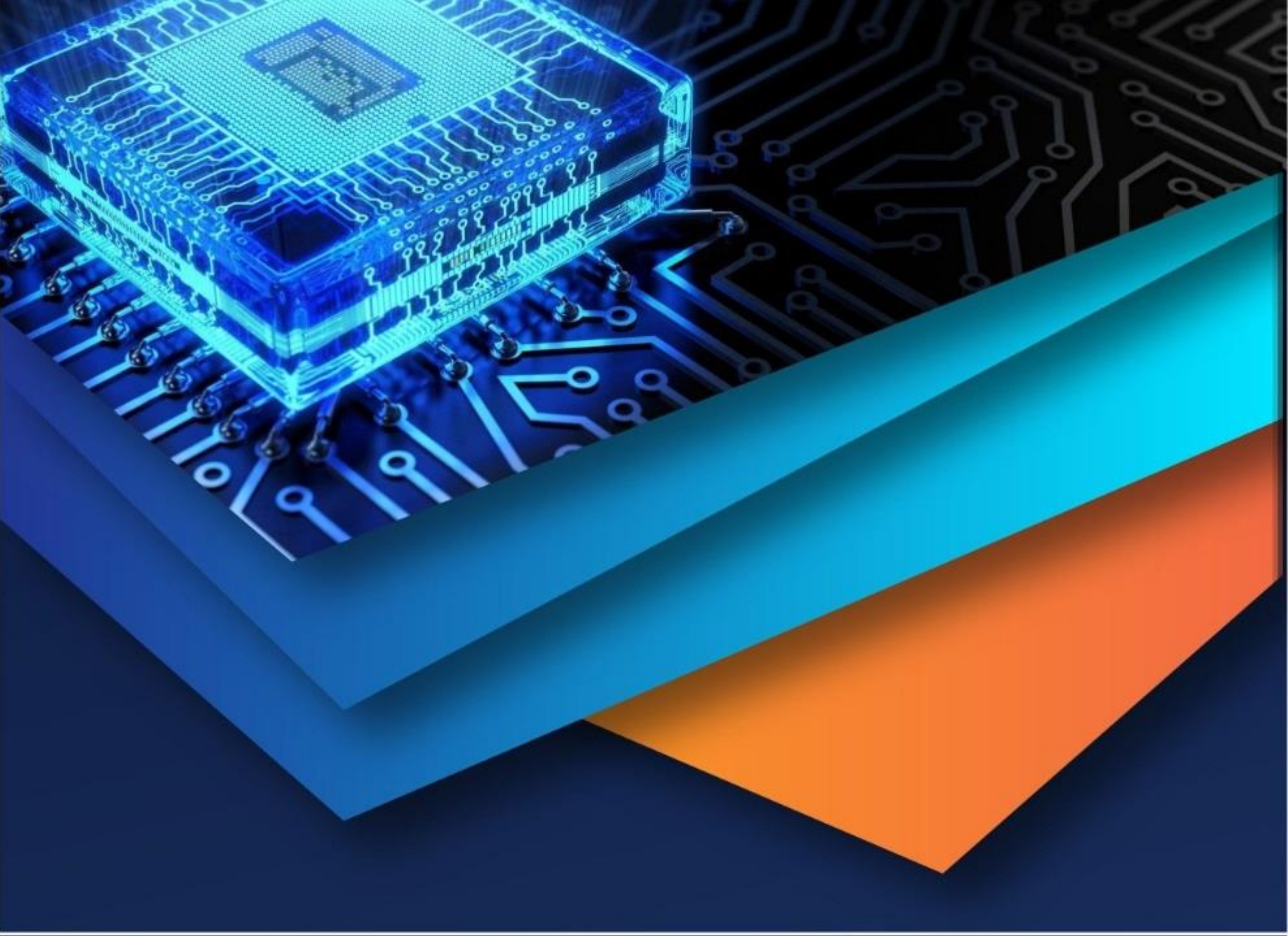
Pravin Kumar.L received B.E degree in computer science and engineering from Anna University,Tamil Nadu, India in 2015 respectively. Currently he is doing the full time internship in BMW India Pvt. Ltd. He has published one international journal.

His research interests focus on the area of BigData analytics, dynamic adaptation of MapReduce computations ,data mining and network security

Ramesh Krishnan received Bachelor of Engineering and Master of Engineering degrees from Anna University, India in 2007 and 2010 respectively. Currently he is pursuing the full time Ph.D Degree at Anna University. He has published papers in four international journals and three international conferences

His research interests focus on the area of low-power, high-performance VLSI architectures and circuit design for digital communications and digital signal processing.

©IJRASET 2013: All Rights Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)