January - February 2020
ISSN: 0193 - 4120 Page No. 9773 - 9789

Deep Learning Techniques for Traffic Flow Prediction in Intelligent Transportation System: A Survey

Shiju George[1], Ajit Kumar Santra[2]
School of Information Technology and Engineering
Vellore Institute of Technology, Vellore, Tamil Nadu, India

Article Info
Volume 82
Page Number: 9773 - 9789
Publication Issue:
January-February 2020

Article History
***Article Received**: 18 May 2019*
***Revised**: 14 July 2019*
***Accepted**: 22 December 2019*
***Publication**: 14 February 2020*

Abstract

Intelligent Transportation System (ITS) intends to provide progressive services linking different modes of transportation and related service system. Road transportation is one of the major and complex entity of the traffic management system. Traffic flow prediction (TFP) contributessignificantrole inpredicting various parameters for road transportation thatgenerates stochastic and nonlineardata through sensors. Traffic flow is demarcated as the average number of vehicles present in a specific region given the historical flow data.Accurate forecasting of macroscopic parameters such as volume, density, speed and flowof traffic can improve the efficiency of traffic management system.In order to predict traffic flow, spatial and temporal traffic features act as a raw data input for the predicting models. Most of the shallow models are incapable to reveal bothspatiotemporal information in big data. Deep neural networks (DNNs) have recently highlighted the potentiality of capturing andextracting important features for various application frommassive dataset.Our work depictsthe current state-of-the-art deep learning techniques (DL) and itsinspirations in TFP incorporating various contextual factorssuch as construction zones, weather conditions, special events, traffic incidents, weekdays and holidays apart from spatiotemporalfeatures for predicting the flow.Finally, provides open challengesyet to be exploredfurther for enhancing deep learning techniques and approachesin forecasting accurate flow.

Keywords; *Intelligent Transportation System (ITS), Deep Learning (DL), Traffic flow Prediction (TFP), Machine Learning (ML).*

I. Introduction

Different modes of transportation and servicesacquire an important role inday-to-day activity of humanlife.Similarly,ITSprovide smart and intelligent organisation of information to support transportation systems for the economic growth and stability of a country.Various factor affects transportation system. One such parameter is traffic flow.Traffic flow has major challenges due to the non-linearity caused by various contextual factors. For the development of cities, ITS has been paying more attentionto develop an accurate forecasting model for the challenges faced by theroad transportationsystem.In the

January - February 2020
ISSN: 0193 - 4120 Page No. 9773 - 9789

traditional forecasting methods, timely and precise traffic flow forecasting is a tough processdue to the effort in constructing and solving the utilized mathematical model [1]. Subsequently, researchers have utilised the capability of classical ML algorithmsto build models that utilizes real-world dataset to make prediction incapacitating the problem of developing complex computational mathematicalmodels. Most of the existing computational methodsare likely to be limitedby multiple parameters forchaotic time seriesdata and uses shallow methods to predict the traffic flow [2]. Primarily, these ML approaches utilizes supervised, semi-supervised, unsupervised and reinforcement-learning techniques that predicts traffic conditions based on the context of the domain problem.

Standard data records with sufficient features essentially support the learning models to fulfil the objective of TFP. Therefore, every small movement observed and recorded spatially and temporallyon road are stored as raw data,andpre-processedfor input to the prediction algorithm. These data in the dataset are current and historical in nature to support the prediction process. With the development,dissemination and enhancement of current technology, the amount of data collected and associated with the traffic, using various sensory devices on road has become very huge.Emergence of thisBig Data era has broughtvital opportunities to explore newer techniques for the improvement ofprediction accuracywith sufficient information [3].Due to the accessibility of hugetrafficdata from diverse& disseminate sensors placed in ITS, data-driven traffic forecasting has confirmed extensive importance over the model-driven methods.This high dimensional non-linear data encompass useful traffic features for predicting various important information to support Intelligent Transportation Systems. Adoption of large volume of traffic data during the training process will avoid many failures caused by assumptions. The information and correlations hidden in thedatacan improve learning accuracy of traffic prediction [4][32].Conventional ML model rely on features developed by human being manually to capture the underlying traffic characteristics. These manually selected features fails todescribe the traffic characteristics and inadequate to comprehensively abstract the granular characteristicsdue to the non-linearity in traffic data. Non-linearity occurs due to the contextual factors such as construction zones, special events, weather conditions, traffic incidents, weekdays, accidents and seasons that alters the pattern of traffic flow invariably leading to inefficient traffic flow and thus cannot achieve accurate predictions.Because of theincreasingneed of accurate real time traffic flow information, a good number of forecasting algorithms have been developed. To an extend researchers have used multifaceted techniques including computational intelligence (CI) algorithms, ML algorithms and hybrid methods to resolve various issues. Results shows that the inclusion of CI and bio-inspired algorithms with ML outperforms many traditional approaches but the real problem exists due to the unprecedented data availability from various sources and incapability to extract features automatically. It is even difficult to state or compare the dominance of one method with the other since most of the existing methods uses insignificant amount of traffic data for evaluation and prediction of traffic flow. In addition, the veracity of traffic flow methods is dependent on both spatiotemporal features enclosed in the traffic data.The theory of DL drawn ever-increasing research attention for the capability to remarkably extracts features from raw unstructured and non-linear data through generalised learning procedure and transform it to suitable representation or feature vector. Multiple layers in DL between input and output performs feature identification, extraction and processing that enables complex functions with non-linear modules to transform from lower to higher

January - February 2020
ISSN: 0193 - 4120 Page No. 9773 - 9789

abstract level. DL extract useful information required for the traffic prediction considering all relevant features through the hidden layers. It provide solution to the problem by overtaking the job models built.

DLalso works well in the noisy environment to filter and extract information hidden within noise. It recognises the pattern and information embedded in the noisy datathrough visual analytics display.As TFP is an intricate task, prior knowledge represented through the hidden layers of the deeplearning algorithms greatlyimprove the forecasting accuracy [4].Through this ability, DL outperforms many other techniques.

In this work, we aim to present:

(i) ExistingDL techniques and models.
(ii) DL in traffic predictionfor both time series and location based historical data.
(iii) Open challenges in DL for TFP.

II CURRENT DEEP LEARNING TECHNIQUES& MOTIVATION

DL has attracted researchers from various domains[29]-[31]. To avoid complex mathematical approaches and limitations of classical ML models, numerousproblems started to adopt the benefits of DL techniques. Traffic data are complex, non-linear and nonstationary in nature. Manual designing of features in huge traffic dataset are often over-specified, incomplete and take a long time to design and validate by the traditional machine learning algorithms.DL attempt to transfer the input data from the traffic dataset into more abstract patterns or features using a hierarchy of multiple layers and predict the traffic flow by combining those features in the output layer.Supervised, semi-supervised and unsupervised techniques are used in DL to automatically learn and excerptoutputfor TFP.Fig.1 shows DL process with various input data sources, feature extraction steps and output parameters.

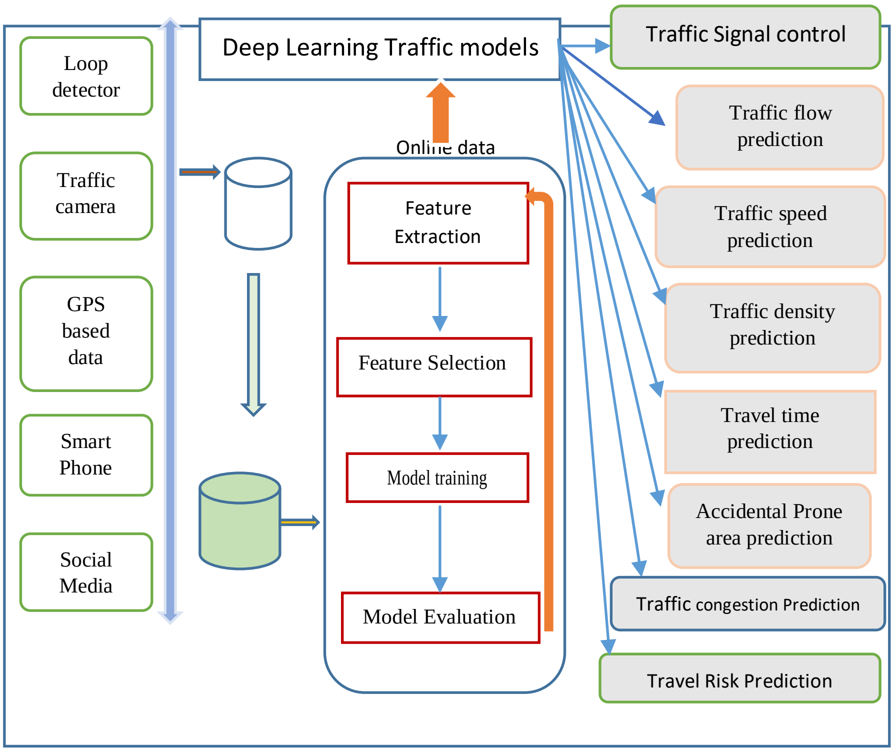

Fig 1. Deep Learning architecture for various macroscopic parametersof traffic

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 9773 - 9789

DL algorithms arecategorisedinto fivedifferent types, they are:

(a) Deep Neural Network (DNN) which is the basic model of DL
(b) Convolution Neural Network (CNN)
(c) Recurrent Neural Network (RNN)
(d) Generative Adversarial Networks (GAN)
(e) Deep Reinforcement Learning (DRL)

All these models are non-parametric in nature, supporting scalar or vector data model and widely used for traffic prediction through several software packages. Caffee, Lasagne, Torch,Keras, CuDNN, and Theano are some of the open source packages used for building DL models. TensorFlow deploys DNN, RNNand CNN.

2.1. Deep Neural Networks

DNNcan be explicitly categorised as Multilayer Perceptron (MLP), Deep BeliefNetwork (DBN) and Stacked Auto-Encoder (SAE).They differ mainly in the design of the hidden layers.An MLP is a feedforward artificial neural network (ANN). It consist ofone input layer, one hidden layer (computational engine) and one outputlayer. Every individual node is fully connectedto all other nodes in the subsequent layer. The weight factors in the hidden layer utilizes supervised learning with backpropagation (BP) for training dataset.MLP distinguish data that is not linearly separable using non-linear activation function by each layers except input nodes and gives promising results when anadequate amount of categorizeddata are present.

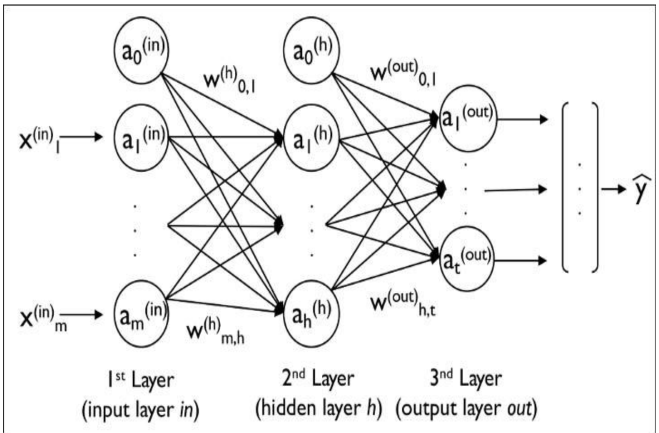

Fig 2.Multilayer Perceptron Network

The Deep Belief Network (DBN) is an unsupervised probabilistic algorithm with the capability to detect and learn features using stochastic latent variables. It deals with bothlabelled and un-labelled databy constructing models using supervised and unsupervised techniques respectively. DBN architecture is made up of input layer, hidden layers and output layer. The hidden layer can be viewed as composition of Restricted Boltzmann Machine (RBM)which captures the correlation present between the datapresent in each hidden layer sequentially and essentially learn the probabilistic distribution or Autoencoder that perform efficient data coding in an unsupervised manner respectively. Latent variables values in the hidden layers are learned by generative weights in the reverse direction using fine-tuning. Stacked Restricted Boltzmann Machines (SRBM) incorporate a huge amount of data in a parallel fashion with less processing time.

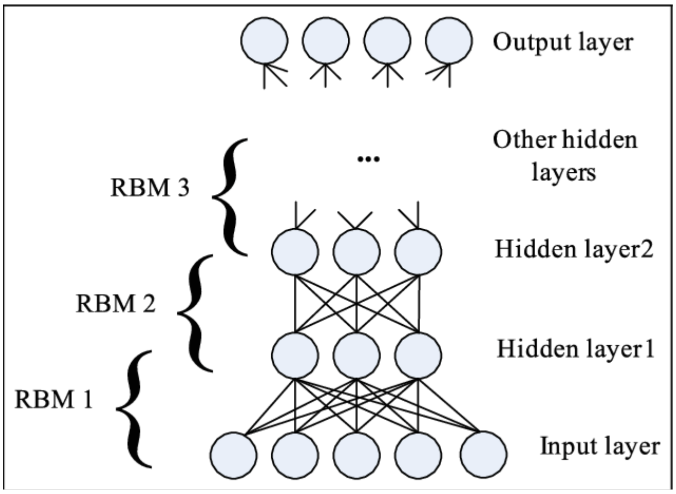

Fig 3. Structure of the Deep Belief Network.

InStacked Autoencoder,multiple layers of Autoencoder (AE) are utilised as the hidden layer. The main goal istraining the network to obtain output value same as the input vectors, to minimize the reconstruction error. The hidden layer will encode the input and then decode it by theoutput layer.In a trained network, the number of units in the input is equal to the number of unit

Published by: The Mattingley Publishing Co., Inc.

resulted at the output layersduring the encoding and decoding processes (Fig 4).

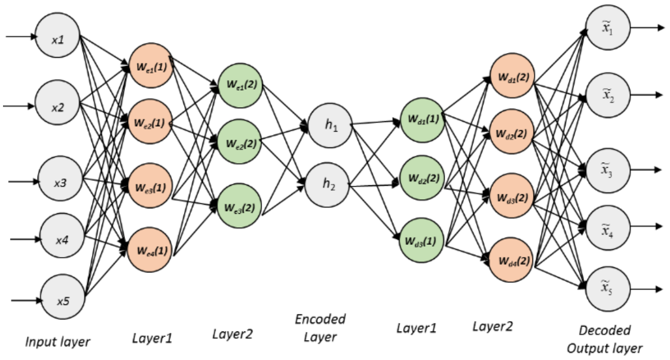

Fig 4. Stacked Autoencoder

2.2. Convolutional Neural Networks

A Convolutional Neural Network (CNN) consist of multiple layers for feature mapping and classification. A feature map is created by CNN to predict the probabilities of class for each feature applying a filter, whichscans the whole input image. The pooling layer also known as downsampling layer maintains the most essential information by scaling down the amount of information generated in the CNN layer as an input to it.Architecture uses uniform sparse matrix and attainan optimal network topology with the highly correlated units that combines to form input to the next layers [16]. Fully connected layer applies weights over the input generated by the feature analysis to predict an accurate label in the output layer.Concept of weight sharing in CNN substantially reduces the training needed by the number of parameters, resulting inimproved generalization. Various extended architectures can be developed and implemented usingCNN such as LeNet Architecture for image recognition task, AlexNet Architecture (modified variant of LeNet) and GoogleNet (inception V1) for higher performances.

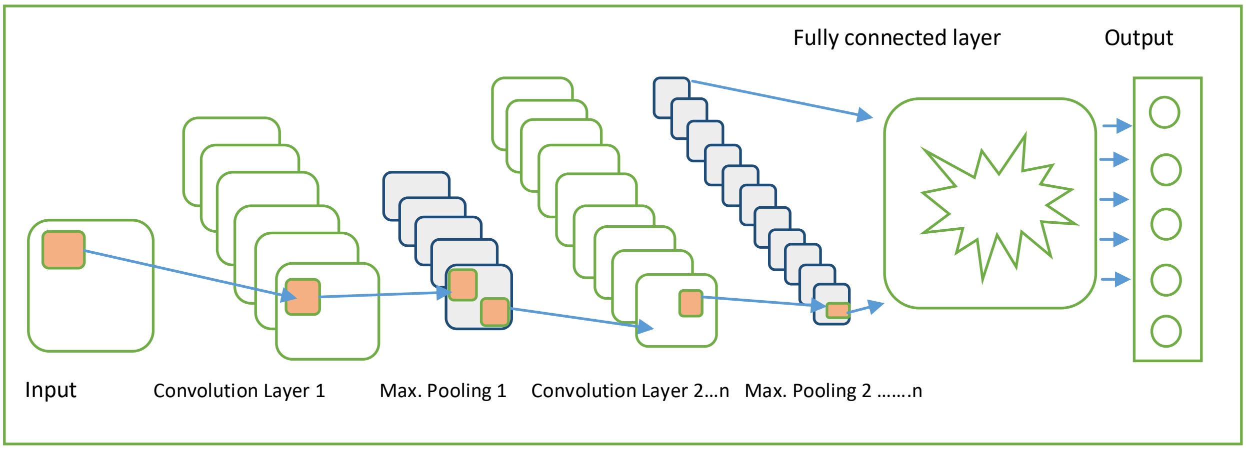

Fig. 5 CNN Architecture

2.3 Recurrent Neural Network

RNN recursively uses previous steps or information in the neural network as an input to the current step. RNNs can use their intrinsic state for processing and retaining successive inputs. RNN provides same weight factors and biases to all the layers in order to change the independent activations into dependent activations. Thereby, reduces the intricacy of increasing feature parameters and learning previous outputs by giving each output as input to the next hidden layer. An RNN is useful for time series application because it remembers therecent information throughout the training process.Recurrent neural network provides an effective pixel neighbourhood with CNN while dealing with images. The large number of gradient error multiplication quickly lead to gradient vanishing and exploding problems in RNN. Modified version of RNN includes Long

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 9773 - 9789

Short Term Memory (LSTM) and Gated Recurrent Units (GRU), which employs appropriate initialisation, batch normalization, and activation functions, tosolve gradient vanishing and exploding problems.

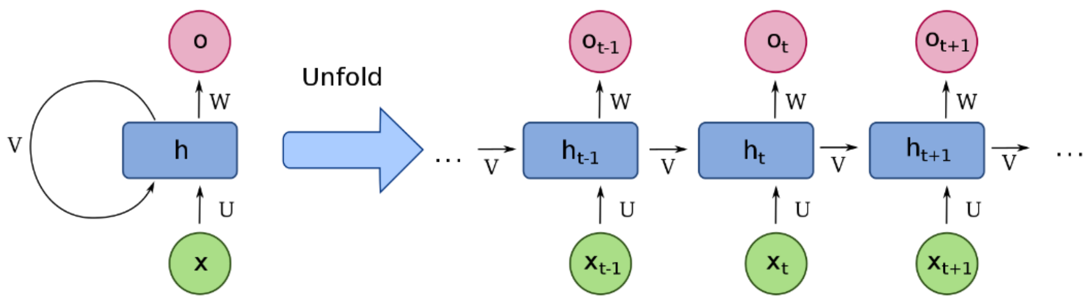

Fig 6. Recurrent Neural Network

2.4 Generative Adversarial Networks

Generative Adversarial Networks (GAN) belong to the set of generative models. GAN has two parts. Generative network that takes random points and decodes it into synthetic image. Secondly, discriminator network that takes synthetic or real images as input and predicts whether it came from training set or created by generator. Generator is trained to fool the Discriminator by evolving to create realistic images.Discriminator adapts to the increasing capability of Generator and sets high bar for realism.Finally, after training, Generator will be able to transform any random points to believable images. For GANs,The optimization process is not seeking a minimum, but an equilibrium.

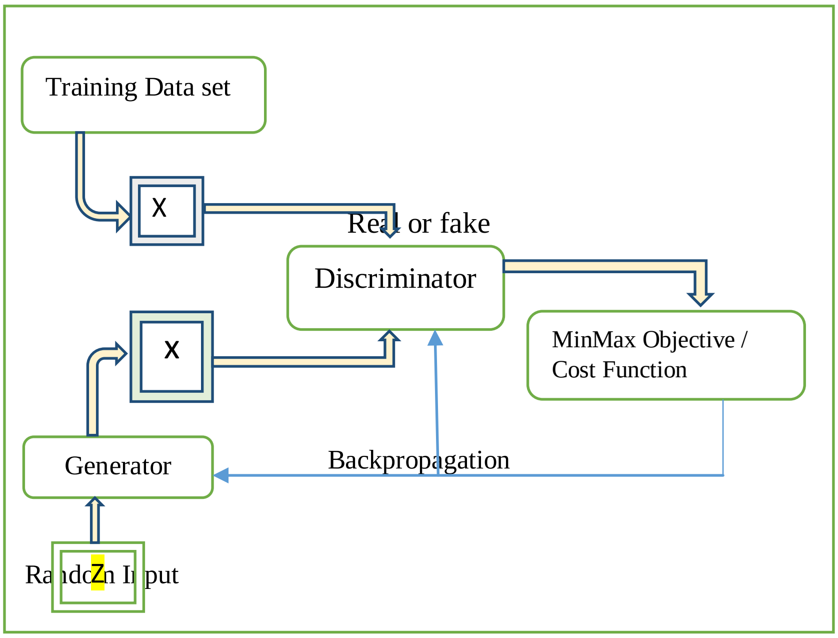

Fig 7. Generative Adversarial Networks

January - February 2020
ISSN: 0193 - 4120 Page No. 9773 - 9789

2.5 Deep Reinforcement Learning

Deep reinforcement learning is the composition of two methods that utilizes the advantages of both reinforcement learning (RL) and DL. RLhas various components including agents, environments, states, policy, actions and rewards.

Agent: An agent percept information from the environment and takes actions.

Environment: The environment represent the world or space where agent moves, act based on the percept and receives reward as an output.

State (s): A state defines the agent initial, current, intermediate and final situations.

Policy (π): Policy is the tactic applied by an agent based on the current state.to decide the next action. It maps states (s) to actions (a), the actions that promise the highest reward.

Action (Q) : Qπ(s, a) denotes return of an action under policy π from the current states. Q represents state-action pairs to rewards.

Reward: A reward is the responseto an action by the agent in a particular state to measure the success or failure.

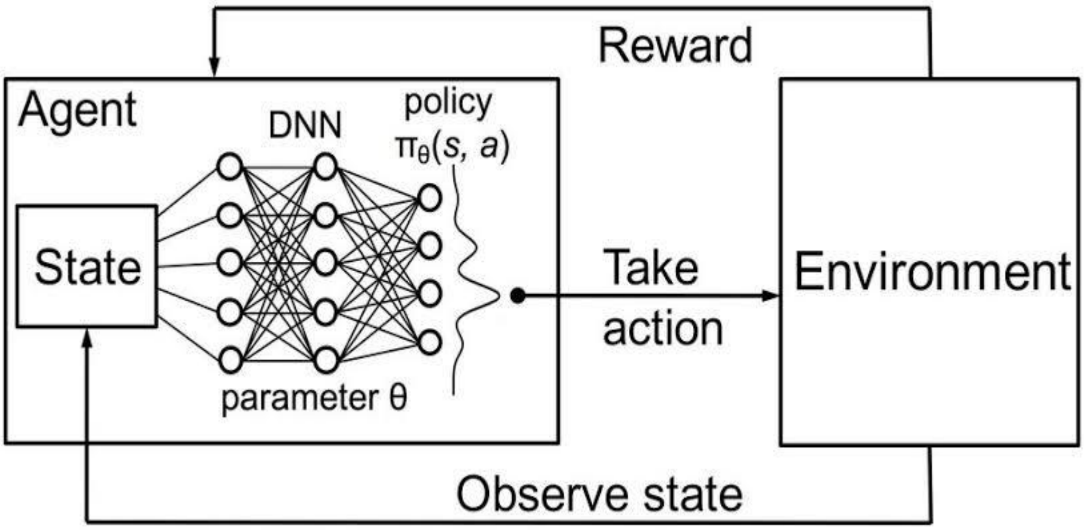

Fig 8. Deep Reinforcement Learning

III. DEEP LEARNING TECHNIQUES FORTFP

For urban areas, Traffic prediction is always a critical issue due to the complex infrastructure. Researchers have used multifaceted DL techniques to deal with various traffic situations. Even hybrid techniques resolved issues of capturing nonlinear spatiotemporaleffects.

Strenuous review for the current developments in TFP through DL will help researcher to understand and identify further flaws in accruing flow accuracy.

Li Deng et. al. (2012) presented a Deep Stacking Network that directs parallel learning for deep architecture. DSN is antiphonal to conduct parallel processing, thus ascendable over the possiblymassive size of the training data. DSN has superior classification accuracy than the Deep Neural Network (DNN) [3].It also has better local generalization abilities and quicker convergence speed when equated with other Neural Network methods.

J.Z.Zhu et. al. (2014) used radial basis function NN (RBFNN) model to overwork the spatiotemporal features from the volume of close intersection road for better prediction [4]. The RBFNN is a three-layer feedforward neural network with one hidden radial basis layer, which can evenly estimate any neighbouring function with a potential accuracy. Wenhao Huang et. al. (2014) developed multitask regression layer above the unsupervised Deep Belief Network

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 9773 - 9789

(DBN) for TFP [5]. For spatial data road, locations and link the multitask learning (MTL) are proven to perform 5% improvement over the existing methods.

Xiaolei Ma et. al. (2015) used Long Short-Term Neural Network (LSTM) neural network for traffic speed prediction [6]. Back-propagated error decay issue and determining the optimal time lags by default, due to the memory blocks present in the LSTM network. The model exhibits dominant capability with long temporal dependency for the time series TFP. Suitable for spatial information too.

Junbo Zhang et. al. (2015) dealt with various problems related to flow prediction influenced by manyintricateparameters [7]. The spatial correlation between different locations, temporalcorrelation among different time intervals, and external contextual parameters are the important issues solved by the multitasking model to predict the node and edge flow in the spatiotemporal network. Convolutional networkpredicts node and edge flow separately. The possibledelegacy of hidden layers iscoupled and trained together with gating fusion mechanism. It outperforms eleven baseline methods.

Yisheng Lv et al. (2015) used Stack Autoencoder model (SAE) for considering the spatial and temporal correlations integrally for TFP [8]. Multiple Autoencoder are stacked hierarchically to form deep net and logistics regression predictor layer is supplementedabove the SAE network for supervised TFP. Outperformed the random walk, Support vector machine and the radial basis function neural network model.

Yanjie Duan et. al. (2016) proposed a DL model, for traffic data imputation, which retrieves datasets through feature extractionand statistical dependency learning. Clustering method is applied to distinguish weekdays, weekends and holidays. Experimental results show that this deep leaning model outperforms ARIMAand Back Propagation neural network model.

DL method with an Error feedback Recurrent Convolutional Neural Network structure (eRCNN)is applied by Jingyuan Wang et. al. (2016),to predict traffic speed by adding the spatiotemporal traffic speeds of neighbouring roads as an input matrix [10].Transfer learning is performed for weight pre-training.It perceives prediction of errors originating from the suddendeviations of traffic speeds. eRCNN performance is found to be more accurate when compared with Auto Regression Integrated Moving Average, Support Vector Regression, Stacked Auto Encoders and 1D Convolutional Neural.

Arief Koesdwiady et. al. (2016), presented DBN combining the effects of weather on traffic flow [11]. It includes data fusion scheme at decision level to improve prediction accuracy using weather condition. Investigate cross-correlation between traffic condition parameters and weather conditions at severalcoarsenessto identify the weather influencing variables. The model could handle huge set of data but the time and space complexity is higher compare to ANNand Autoregressive Integrated Moving Average model. Social network rich information could support in future for more accuracy.

Hao-Fan Yang et. al. (2016) developedStacked Autoencoder Levenberg–Marquardt model (SAE-LM) [12]. One of the challenges with SAE-LM isthat by employing more Autoencoder the time consumed in the training and predictionphasesincrementsintensely.Performance also degrades if the determined traffic data has a largeflat dispersion, but for uneven traffic data, it has greater performance in TFP.

Hongsuk Yi, HeeJin Jung et. al (2017) developed the DNN model using TensorFlow and implemented using TFLearn [13]. In order to predict the flow conditions TPL was used to make a distinctionof clogged traffic conditions from non-clogged traffic conditions. Only limitation observed is the memory capacity that can hold onlyone percent oftraffic data for a particular day.

Published by: The Mattingley Publishing Co., Inc.

Improvement required for increasing the accuracy.

Xiaochuan Sun et. al. (2017), implemented a model known as Deep belief echo-state network (DBEN) to solve inheritedshortcomings of the back- propagation method including slow convergence and local optimum faced during time series prediction [14]. A progressive regression layer with an embedded echo-state learning function is madeabove DBN for supervised prediction. Echo-state network provides exceptional performance in modeling dynamical data.This network has an obstinate problem on how to find out largenumber ofparameters such as depth, neuron number, learning rate and batch-size, reservoir size, spectral radius and connectivity.

Yuankai Wu et. al. (2018) developedhybrid prediction model DNN based TFP model (DNN-BTF) [15]. Thismodel overworks the rewards of three major techniques of DL algorithms namely fully connected neural networks, recurrent neural networks, and convolutional neural networks to enhanceprediction accuracy. Apart from this, the model applies the attention model to traffic flow modeling thatascertain the significance of inputs of large-scalespatio-temporal points by default.DNN-BTF outperforms other methods traditional and individual DL methods.

Sakshi Indoliaa et. al. (2018)provided the theoretical understanding of CNN with Gradient Descent and Adaptive Moment Estimation (ADAM) Optimization [16].CNN has vast power to learn contextual characteristics to defeat the problems entangled in pixel wise classification. CNN has been used for various classification problems such as sensing object remotely,recognise ocean front, provide high-resolution data etc.

Weihong Chen et. al. (2018), has depicted, a novel fuzzy deep-learning approach called FDCN, for predicting urban traffic flow that reduces the forceful consequences of data precariousness in exploring the spatial and temporal correlation of traffic flow [18].FDCN learns the actual data for designing the membership parameters and fuzzy rules. This work was able to getbetter prediction performance when equated with the ARIMA,DeepST, CNN and FCNN methods.Further study canoptimize theorganization of the model and the power of contextualparameter on predictive performance.

ZongtaoDuani et. al. (2018), implemented ahybrid deep neural network model, an assortment of the convolutional neural network (CNN) andlong short-term memory (LSTM) method which extracts the spatial features andseizes the temporal information respectively to forecast the traffic flow [19]. A greedy policy consumesshort time to enhanceveracity when a network goes in depth considering spatial and temporal information. When compared to linear model, CNN model and general CNN-LSTM model, this work outperformed the others.Weather conditions and holiday data are not considered for the study.

Zibin Zheng et. al. (2019),developed an indepth and embedding learning approach (DELA) which can help to precisely learn important information fromfine-grained traffic data, route structure, and weatherconditions [20]. DELA consists of an engrafting feature, which captures the categorical and identify the correlated features, where a convolutional neural network (CNN) component learns 2-D traffic flow data and a long short-term memory (LSTM) component. The engrafting feature can captivate the categorical feature information and identify correlated features, utilises the gains of upholding a long-term memory of historical data. The model outperformed when compared with ARIMA, BPNN, RNN, LSTM, CNN, embedding & LSTM, embedding & CNN. The integration of the three models embeddingwith CNN andLSTM outperform the rest in terms of the prediction performance and accuracy. Limitations of the model are the poor model interpretability and

January - February 2020
ISSN: 0193 - 4120 Page No. 9773 - 9789

limited learning capability of embedding component.

Jiawei Wang et. al. (2019),depicted a mechanism for each critical paththrough the bidirectional long short-term memory neural network (Bi-LSTM NN) [21]. Multiple Bi-LSTM layers are arranged to explore spatial-temporal features that are given to a fully connected layer.The output of each path isensemble for network-wise traffic speed prediction. Selection of more criterion is an open option. Model is compared with KNN, ANN, LSTM NN, and CNN and outperform in terms of accuracy speed.

Licheng Qu et. al. (2019) developed aTFP mechanism for single day using a deep neural network based on historical traffic flow data and contextual factors [22]. A multi-layer supervised learning algorithm was used to train a predictor in order to find potential relationship between traffic flow data and contextual factors affecting the flow. Frequently used conventional method was compared against the proposed batch learning method.The prediction results this new method are found similar to that of the true traffic flow data.Contextual factors like accidents and construction activities can be taken for future studies.

Linchao Li et. al. (2019), implemented anenhanced multi-objective particle swarm optimization algorithm (MOPSO) that optimizesfewfeatures in deep belief networks to achieve TFP [23]. The bias-variance framework ensured increased stability and accuracy for MOPSO.It was observed that the efficiency of the hybrid model decreases when the layers are increasing more above four; it is an open challenge for researchers to still resolve.

Di Zang et. al. (2019), implemented a model to address theissue of long-term TFP [24]. This method consists of a generator and a discriminator; where the generator consists of multi-channel residual deconvolutional neural networks, and the discriminator consists of a convolutional neural network whose objective is get the most out of the hostile training process.The model uses two-dimensionalmatrix to hold back the original spatial andtemporal information of the data. When compared with CNN, ANN, ConvLSTM, MRDNN, ST-ResNet and RDBDGN_V, this model outperformed the performance of all.

Fanhui Kong et. al. (2019), used Restricted Boltzmann Machine (RBM) for traffic flow forecasting, taking into account of the traffic big data accumulation in Internet of Vehicles (IoVs), multimedia technologies to cater sufficientreal-world data for model training [25]. RBM deals the disorganized time series, using phase space reconstruction to organize the data.The traffic data is reconstructed using embedded dimension selected by virtual neighbour method, which transports on the phase space reconstruction to initial data. The model is equated with ANN, BP, SVR and ARMA indicates that RBM model can manifest the direction and quality in disorganized time series for traffic flow that is effective for real-time ioVs

Bailin Yang et. al. (2019) proposed an improved version of LSTM that connects thestrong -influence value of singularly long succession time steps to the current time step [26]. Attention mechanism is used to captivate thesestrong -influencetraffic flow values, followed by smoothening of data beyond the regular range to get better forecasting results. The proposed method LSTM + is compared with BPNN,SVM, RBFNN, SAE and LSTM. More exploration is still possible by making a combination of time and space parameters, taking in to account of the huge influence of historical data.

Shaojiang Deng et. al. (2019), proposed a stochastic subspace learning baseddeep CNN (RSCNN) for traffic flow forecasting which learns deep stratified feature representation with spatiotemporal connections over the traffic network [27]. In addition to deal with incomplete data cast other than the principals learning strategy via random subspace learning is

January - February 2020
ISSN: 0193 - 4120 Page No. 9773 - 9789

designed.RSCNN outperforms random walk (RW), Support Vector Regression (SVR) with polynomial kernel, NN, SAE and DBN.

Deep Learning Models	Enhanced Models	Referred works
DNN	Deep Stacking Network (DSN) [3] Radial Basis Function NN (RBFNN)[4] Tensorflowand Coded using TFLearn [13] DNN+ Batch Training [22]	Traffic flow prediction [3, 4, 13,22]
DBN	DBN-MTL[5] DBN + Data Fusion Scheme [11] Deep Belief Echo-State Network (DBEN) [14] DBN + MOPSO [23] RBM + IoVs [25]	Traffic flow prediction [5, 11, 14, 23,25]
SAE	SAE+ Logistic regression [8], Denoising Stacked Autoencoders (DSAE) [9] Stacked Autoencoder Levenberg–Marquardt [12]	Traffic data imputation[9], Traffic flow prediction [8, 12]
CNN	Fuzzy deep-learning CN [18] Multitasking CN [7] Random Subspace Learning Based Deep CNN (RSCNN) [27]	Traffic flow prediction [7,18,27]
RNN	Error Feedback Recurrent Convolutional Neural Network Structure (eRCNN) [10]	Traffic speed prediction [10]
LSTM	LSTM NN [6] Bidirectional LSTM [21] LSTM+ [26]	Traffic flow prediction [6,26] Traffic Speed [21]
HYBRID	NN+RNN+CNN [15] CNN+ LSTM [19] Embedded Component + CNN + LSTM [20]	Traffic flow prediction [15,19,20]
DGN	Residual Deconvolution Based Deep Generative Network (RDBDGN) [24]	Traffic flow prediction [24]

Fig 9. DL models for TFP

Published by: The Mattingley Publishing Co., Inc.

IV DISCUSSION AND CHALLENGES

The review has captured an in-depth understanding of DL methods for traffic flow forecasting. Traditional methods have handled numerous problem in traffic flow with mathematical and classical ML models, but fails to outperform when both spatiotemporal information is consider which is complex, uncertain and non-linear. DL has opened the door for researchers to solve these issues by featuring and extracting information at the granular level. With the vast research in the area of Neural Network techniques, the DL methods has become popular technique to deal with Big Data.

DL architecture needs improve data processing capabilities in order to deal with various traffic related problems such asshort-term andlong-termTFP using spatiotemporalinformation raw input data based on the input given and output desired. With the diversity of granular factors such as poor infrastructure, congestion, traffic incidents, constructions, weather, peak hours, weekdays, holidays, accidents, events and user behaviour the traffic flow patterns get affected in large scale. Apart from the contextual factors, information collected from various sources with corrupted, inconsistent, duplicate and missing data also affect accuracy offlow prediction.

Some of the major challenges related to TFP are given below.

a) Forecasting accuracy
Accurate forecasting of traffic flow in road transportation is a complex task due to the lack of unified model to solve all problems in the traffic.For a given traffic related problem, no two algorithms can be compared and claimed to provide universally optimal solution. Yet another in-depth study may provide sufficient knowledge for better solution.

b) Handling huge data

Managing huge data is a primary challenge with adequate support of effective organisation and storage. Learning from such massive data with traditional method is complex process due to difficulty in finding appropriate correlation. Even if the algorithm acquires massive data, the system needs sufficient processing capabilities to execute and extract useful information from the given data.

c) No better understanding

For the TFP, multifaceted algorithmsclaims to relate information present in the massive data but to understand the relevance of input and output depends on the derived links between the correlated input and output data. The hidden unit perform mysteriously extract the common pattern to fit the objective function. Exploring the interoperability of information in the middle layers of DL is a challenging task that may leads to newer findings.

d) Nonlinear data and large number of populations

Linear data provides an informalwayto fit the objectives functionby the traditional mathematical and ML models. In case of non-linear data, algorithms will abruptly underperform leading to bias-variance trade-off .and make undesirable prediction with huge loss function. Neural network with deep learning support to extract significant information because of its capability to correlate relevant information in the data. But based on the applications and complexity of data patterns due to large population of data, DLneeds to be exploited further.

e) Dealing Spatial and Temporal dependency

Most of the techniques used by the researcher are either spatial or temporal dependent. Spatial information locate a node, region, junction or area to map and predict the flow from the input traffic data. Time series data

Published by: The Mattingley Publishing Co., Inc.

consider short-term or long-term prediction based on the interval of time and analysis of current and historical flow data.For the accurate TFP, both spatial and temporal information plays a significant role. Literature review shows that most of the current techniques uses spatiotemporal information to predict traffic flow and need further improvement in large scale data.

f) Affect by the traffic condition
Non-linearity in traffic data leads to poor traffic conditions. Flow pattern changes abruptly due to the effect of contextual factors. Algorithms must be able to handle the outliers generated by the traffic conditions. DL occasionally train its model based on the limitation of available resources for processing, computing and evaluating. At the same time, infinite sources will also not help if the data is huge and corrupt. Sometimes early-stop method inDL helps to achieve desired output, but further continuation will lead to theoverlapping problem.

g) Sequences of random decisions

DL has a capability to extract useful information by initialising the input process with random values and acquire the desired value in the output layer. Sequence of random decision at the hidden layers and output layers depend on the weighting factor through back propagations. It is yet to explore the most appropriate way to address the vanishing and exploding problem occurs in DL due to randomisation.

h) Extrapolation of prediction with a statistical model.

Traffic data creates deterministic and stochasticsprediction environment for the agents acting in the environment. The agents could be vehicles, vehicle types, traffic signals,sensors attached to the infrastructure and human behaviours as driver on road. The statistical data changes with time and location. An efficient outlook for the granular details by the algorithms will help to forecast better flow of traffic.

i) Affected by external or contextual factors

Weather, traffic incidents, constructions,accident-prone area, special events, road rallies, accidents and congestions are the major contextual factors that affects TFP. These factors bring non-linearity in data to be trained by the algorithms. DL though manage to handle such data to an extend but uncertainty may cause to increase the error rates.

j) Expensive computation cost

Review states that DL potentially manages and analyse Big Data using supervised or unsupervised learning in a short time but the amount of traffic data generated each day across the world is expanding at an exponential rate. Training DL methods on such a massive amount of data need enough clusters of computing resources. In general, for DLmethods the number of training data instances for an efficient learning should be 10X the number of parameters in deep.Even though many optimization techniques are appliedon massive data but unfortunately, it consumes largetime and requires high data processing capabilities.

k) Difficult theoretical analysisfor missing data

Other important transportation related problems are missing data imputation and traffic event detection. Transfer learning based on DL in a homogenous environment is an open means to resolve prediction in case of insufficient or missing data [9] [27].

l) Difficult to discovery the global optimal solution

The results of multifaceted techniques for TFP are delicate to the test instances and model features. The developmentof prediction models are concentrated to blind experiments and fine-tuning of the objective functions for traffic features with a minimal understanding of how and why they work.Major problem arises when the network is very large with multiple parameters that may encounter substantial divergence in error obtained in the training data set. Hyper-parameters play an important role in optimization for an algorithm, with initially,randomly set values as the offset of learning from training samples. Slight change in hyper parameter value will lead to substantial change in modelperformance and accuracy. Another challenge is Trial-and-Error learning by including a given number of hidden layers incorporated to learn the most efficient features and subsequently forward it to supervised layer to fine tune the deep network through back propagation. It performs well for the trained problem but perform poorly on any new task.Therefore, to find a global optimization for forecasting the traffic flow is yet a challenging task open for research.

V CONCLUSION

We summarize and analyse implementations that have accomplishedillustrious performance in TFP.DL has shown better performance overcoming the shortcoming of traditional mathematical model and ML algorithmsbased on the context of problem be solved given distinct domains and input data. Therefore, understanding of traffic theory using deep architecture and the usage of the same on road transportation networks are worth perusing. An attempt is made to analyse through literature reviewbased on a set of key challenges affecting the TFP handling big data, spatiotemporal information and contextual factors. The remarkable progress made by deep learning systems in ashort span of a decade proves that the science is still young to overcome number of challenges. Further exploration is required grounded on the capabilityof DL techniquesadapting to the changing necessitiesof ITS applications. The prolific future of traffic forecasting in ITS need to adapt a unified modeling, computing abilities, testing and explaining traffic phenomena. The characteristics of traffic flow is yet an open challenge for DL methods due to immense non-linearity and complex correlation of huge data exploding in transportation system.

REFERENCES

[1] Attila M. Nagy and Vilmos Simon,” Survey on traffic prediction in smart cities,”Pervasive

and Mobile Computing, Volume 50, October 2018, Pages 148-163

[2]Jürgen Schmidhuber (2014), Deep learning in neural networks: An overview,Neural

Networks 61 (2015) 85–117 http://dx.doi.org/10.1016/j.neunet.2014.09.003

[3] Li Deng, Dong Yu, and John Platt, “Scalable stacking and learning for building deep

architectures”, IEEE ICASSP 2012, Page 2133-2136.

[4] J. Z. Zhu, J. X. Cao, and Y. Zhu, “Traffic volume forecasting based on radial basis function

neural network with the consideration of traffic flows at the adjacent intersections,” Transp.

Res. C, Emerg. Technol., vol. 47, pp. 139–154, Oct. 2014.

[5] W. Huang, G. Song, H. Hong and K. Xie, "Deep Architecture for Traffic Flow Prediction: Deep

Belief Networks With Multitask Learning," in IEEE Transactions on Intelligent Transportation

Systems, vol. 15, no. 5, pp. 2191-2201, Oct. 2014.doi: 10.1109/TITS.2014.2311123

[6] Xiaolei Ma, Zhimin Tao , Yinhai Wang , Haiyang Yu , Yunpeng Wang, Long short-term

memory neural network for traffic speedprediction using remote microwave sensor data,

Transportation Research Part C 54 (2015) 187–197.

[7] J. Zhang, Y. Zheng, J. Sun and D. Qi, "Flow Prediction in Spatio-Temporal Networks Based

on Multitask Deep Learning," in IEEE Transactions on Knowledge and Data Engineering.

doi: 10.1109/TKDE.2019.2891537.

[8]Y. Lv, Y. Duan, W. Kang, Z. Li and F. Wang, "Traffic Flow Prediction With Big Data: A

Deep Learning Approach," in IEEE Transactions on Intelligent Transportation Systems, vol.

16, no. 2, pp. 865-873, April 2015.doi: 10.1109/TITS.2014.2345663.

[9] Yanjie Duan, Yisheng Lv , Yu-Liang Liu and Fei-Yue Wang, " An efficient realization of

deep learning for traffic data imputation," in Transportation Research Part C: Emerging

Technologies, ISSN: 0968-090X, Vol: 72, Page: 168-181, Year2016.

[10] Jingyuan Wang, Qian Gu, Junjie Wu, Guannan Liu and Zhang Xiong," Traffic Speed

Prediction and Congestion Source Exploration:A Deep Learning Method," 2016 IEEE 16th

International Conference on Data Mining," IEEE16th International Conference on Data

Mining pp 499-508, Year 2016.

[11] A. Koesdwiady, R. Soua and F. Karray, "Improving Traffic Flow Prediction With Weather

Information in Connected Cars: A Deep Learning Approach," in IEEE Transactions on

Vehicular Technology, vol. 65, no. 12, pp. 9508-9517, Dec. 2016.doi:

10.1109/TVT.2016.2585575.

[12] H. Yang, T. S. Dillon and Y. P. Chen, "Optimized Structure of the Traffic Flow Forecasting

Model With a Deep Learning Approach," in IEEE Transactions on Neural Networks and

Learning Systems, vol. 28, no. 10, pp. 2371-2381, Oct.2017.

doi:10.1109/TNNLS.2016.2574840

[13] Hongsuk Yi, HeeJin Jung and Sanghoon Bae, "Deep Neural Networks for traffic flow

prediction," 2017 IEEE International Conference on Big Data and Smart Computing

(BigComp), Jeju, 2017, pp. 328-331.doi: 10.1109/BIGCOMP.2017.7881687.

[14] Xiaochuan Sun, Tao Li, Qun Li, Yue Huang, andYingqi Li, " Deep belief echo-state

network and its application to time series prediction," in Knowledge-Based Systems 130

(2017) 17–29.doi.org/10.1016/j.knosys.2017.05.022

[15] Yuankai Wu, Huachun Tan,Lingqiao Qin, Bin Ran andZhuxi Jiang," A hybrid deep

learning based traffic flow prediction method and itsunderstanding," Transportation

Research Part C 90 (2018) 166–180.doi.org/10.1016/j.trc.2018.03.001.

[16] Sakshi Indolia, Anil Kumar Goswami, S. P. Mishra, Pooja Asopa, "Conceptual

Understanding of Convolutional Neural Network- A Deep Learning Approach,"

International Conference on Computational Intelligence and Data Science (ICCIDS 2018),

Procedia Computer Science 132 (2018) 679–688

[17] Ritika Wason (2018), Deep learning: Evolution and expansion, Cognitive Systems

Research,Volume 52, pp 701-708.

[18] Weihong Chen, Jiyao An, Renfa Li, Li Fu a, GuoqiXie, Md ZakirulAlam Bhuiyan ,Keqin

January - February 2020
ISSN: 0193 - 4120 Page No. 9773 - 9789

Li, “A novel fuzzy deep-learning approach to traffic flow prediction withuncertain spatial–

temporal data features,” in Future Generation Computer Systems 89 (2018) 78–88.

doi.org/10.1016/j.future.2018.06.021

[19] Z. Duan, Y. Yang, K. Zhang, Y. Ni and S. Bajgain, "Improved Deep Hybrid Networks for

Urban Traffic Flow Prediction Using Trajectory Data," in IEEE Access, vol. 6, pp. 31820-

31827, 2018.doi: 10.1109/ACCESS.2018.2845863.

[20] Z. Zheng, Y. Yang, J. Liu, H. Dai and Y. Zhang, "Deep and Embedded Learning Approach

for Traffic Flow Prediction in Urban Informatics," in IEEE Transactions on Intelligent

Transportation Systems, vol. 20, no.10, pp. 3927-3939, Oct. 2019.doi:

10.1109/TITS.2019.2909904

[21] Jiawei Wanga, Ruixiang Chen andZhaocheng He, Traffic speed prediction for urban

transportation network: A pathbased deep learning approach,” Transportation Research

Part C 100 (2019) 372–385.

[22] Licheng Qu, Wei Li, Wenjing Li, Dongfang Ma and Yinhai Wang, “Daily long-term traffic

flow forecasting based on a deep neural network,” Expert Systems With Applications 121

(2019) 304–312.doi.org/10.1016/j.eswa.2018.12.031.

[23] Linchao Li, Lingqiao Qin, Xu Qu, Jian Zhang, Yonggang Wang and Bin Ran,”Day-ahead

traffic flow forecasting based on a deep belief networkoptimized by the multi-objective

particle swarm algorithm,”Knowledge-Based Systems 172 (2019) 1–14.

doi.org/10.1016/j.knosys.2019.01.015.

[24] D. Zang, Y. Fang, Z. Wei, K. Tang and J. Cheng, "Traffic Flow Data Prediction Using

Residual Deconvolution Based Deep Generative Network," in IEEE Access, vol. 7, pp.

71311-71322, 2019.doi: 10.1109/ACCESS.2019.2919996.

[25] Fanhui Kong, Jian Li, Bin Jiang andHoubing Song, Short-term traffic flow prediction in

smart multimedia system forInternet of Vehicles based on deep belief network.Future

Generation Computer Systems 93 (2019) 460–472.

[26] Bailin Yang , Shulin Sun , Jianyuan Li , Xianxuan Lin and Yan Tian, “Traffic flow

prediction using LSTM with feature enhancement,” Neurocomputing 332 (2019) 320–327.

doi.org/10.1016/j.neucom.2018.12.016.

[27] Shaojiang Deng, Shuyuan Jia and Jing Chen, “Exploring spatial–temporal relations via deep

convolutional neuralnetworks for traffic flow prediction with incomplete data,” Applied

Soft Computing Journal 78 (2019) 712–721.doi.org/10.1016/j.asoc.2018.09.040.

[28] Lu Zhao, Yonghua Zhou, Huapu Lu andHamido Fujita, “Parallel computing method of

deep belief networks and itsapplication to traffic flow prediction,” Knowledge-Based

Systems 163 (2019) 972–987.doi.org/10.1016/j.knosys.2018.10.025

[2] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, Fuad E. Alsaadi, A

survey of deep neural network architectures and their applications, , Neurocomputing 234

(2017) 11–26. http:// dx.doi.org/10.1016/j.neucom.2016.12.038

[29] Dan Ciresan, Ueli Meier, and J¨urgenSchmidhuber. Multi-column deepneural networks for

image classification. In Computer Vision and PatternRecognition (CVPR), 2012 IEEE

Conference on, pp 3642–3649.

[30] Ramzi Abdelmoula. Noise robust keyword spotting using deep neuralnetworks for

embedded platforms. Master’s thesis, University ofWaterloo, 2016.

[31] LiqiangNie, Meng Wang, Luming Zhang, Shuicheng Yan, Bo Zhang,and Tat-Seng Chua.

Disease inference from health-related questionsvia sparse deep learning. Knowledge and

Data Engineering, IEEETransactions on, 27(8):2107–2119, 2015.

[32] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, “Data mining with big data,”IEEE Trans. Knowl.

Data Eng., vol. 26, no. 1, pp. 97–107, Jan. 2014.

[33] J. Schmidhuber, “Deep learning in neural networks: An overview,”Neural Netw., vol. 61,

pp. 85–117, Jan. 2015.