International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

An Improved Stacked Denoise Autoencoder with Elu Activation Function for Traffic Data Imputation

S. Narmadha, V. Vijayakumar

Abstract*: *Traffic data plays a major role in transport related applications. The problem of missing data has greatly impact the performance of Intelligent transportation systems(ITS). In this work impute the missing traffic data with spatio-temporal exploitation for high precision result under various missing rates. Deep learning based stacked denoise autoencoder is proposed with efficient Elu activation function to remove noise and impute the missing value.This imputed value will be used in analyses and prediction of vehicle traffic. Results are discussed that the proposed method outperforms well in state of the art approaches.

Keywords: Spatio-Temporal, Deep Learning, Elu Activation, Missing value, Auto Encoder.

I. INTRODUCTION

Advanced Traveller Information System (ATIS), Advanced Traffic Management System (ATMS), Intelligent Transportation System (ITS) solely depends on traffic data for congestion management, incident detection and other traffic related applications [11]. Traffic data (Total flow, Speed, Occupancy) collected from Loop detectors, GPS probes, Sensors and Cameras. However the data is often missing or incomplete due to hardware and software failures, malfunctions and also transmission errors [11]. The quality of transport service rely on accurate and complete source of data. Based on literature [14] broadly can categorize the imputation methods into three types : prediction, interpolation and statistical learning. Prediction based methods are Auto Regressive Integrated Moving Average(ARIMA)[9], Regression [14] etc. It predict the missing data based on historical data. Interpolation methods (k-NN,local least square (LLS))[14] fill the missing data by weighted average of neighboring days from the same detector. Statistical learning methods [6] learn the statistical features from observed data and find the unobserved value. Deep learning based stacked denoise autoencoder is a statistical learning methods [7,12,13] used to efficiently impute the traffic data. Activation functions performs a major role in terms of firing neurons to learn the input value and produce the exact output. Exponential linear unit (Elu) is an altenative activation function of ReLu and it produce more accurate results with extra alpha constant [4]. In this work Elu activation function is used with stacked denoise autoencoder to enhance the imputation accuracy. The organization of this paper is as follows. In section II presented the related works of traffic data imputation. Section III describes the methodology of stacked denoise autoencoder with Elu activation. Section IV contains experiments and discussion. Conclusion of this paper is in section V.

II. LITERATURE STUDY

Research towards traffic data imputation has more attention in recent years as ITS needed a complete data for further utilization. Historical imputation methods, Spline regression, ARIMA are proposed and called as vector based methods [6], which are performs well when the missing data is few because it supports only limited spatio temporal correlation [6]. Bayesian Principal Component Analysis(BPCA) and Probabilistic Principal Cmponent Analysis (PPCA) are matrix based methods [5] used to fill the missing value which incorporates more spatio temporal information and it provides more accuracy than vector based methods. Tensor based methods proposed [3,6] to impute the value with mutidimensional tensor pattern. İt provides high accuracy. Least Square Support Vector Machine technique (LS-SVM) proposed to predict the missing traffic data in the arterial road with spatio temporal information [10,15]. Neural network based algorithms [1] have been used for traffic data imputation. Existing methods separates the observed and unobserved data to deal with missing values. Deep learning based approach was proposed [12] which takes the imputation procedure as a recovery of data and it contains both observed and missed value. Imputation is performed in both single station data with all temporal factors (simple structure) and multiple station data with multiple temporal factors (complex structure). However still training time and complexity is high in deep models. So efficient activation function is chosen with deep stacked denoise autoencoder to enhance the result.

Manuscript published on 30 September 2019.
***Correspondence Author(s)**
S.Narmadha*, Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, India. Email: narmadhas17@gmail.com
Dr V.Vijayakumar, Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, India. Email: veluvijay20@gmail.com

III. METHODOLY

A. Denoise Autoencoder

Denoise autoencoder is a variation of autoencoder **(fig.1)**, it is a process of removing noise and finding the missing values from corrupted version of input [2]. It contains two parts (encoder, decoder) and three layers (Input layer, hidden layer, output layer) [1]. Algorithm is given and explained in **Table.1**.

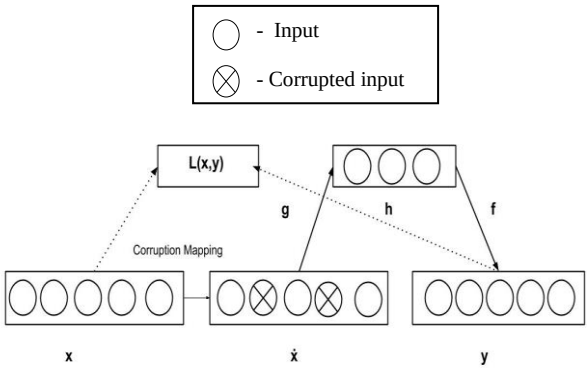

Fig. 1. Simple Denoise Autoencoder [7]

Initialize parameters (weight w, bias b), number of iterations. Input is passed into the hidden layer through mapping function $g = s(wx + b)$, then clean output is reconstructed from hidden layer by $f = s(w'g + b')$. Θ is a hyper parameter (w, b). Θ'(w',b') is a transpose of Θ.

Θ,Θ' $= \arg\min \frac{1}{n} \sum_{i=1}^{n} (x^{(i)} - y^{(i)})$ where L is the Loss function defined as L(x,y)=$\|x - y\|^2$

Table 1. Algorithm 1 - Autoencoder

Input: data X={$x_1 \ldots x_n$},Iterations I={1,2..n}
Initialize weights and bias randomly (w,b)
Pass X into hidden layer ;
Reconstruct the output from hidden layer through transpose T;
for j=1 to I do
  Perform forward propagation to compute y;
  Compute error e = x-y; // difference between input and output
  Perform backward propagation to compute ΔΘ and ΔΘ'
  Update the parameters to reduce e by Θ= Θ+ΔΘ;
  Θ'= Θ'+ΔΘ';
end for

B. Stacked Denoise Autoencoder (SDAE)

SDAE is structured with multiple autoencoders represented in **fig.2**. It comprises two processes: pre-training and fine tuning. Algorithm is summarized in **Table.2**. In Pre-training corrupted input is passed into AE_1 which appears in the bottom, it removes the noise and reconstruct the value through the mapping function $g = s(wx_1 + b_1)$ and transformation function $f = s(wx_1^T + b_1^T)$. For AE_2, AE_1 hidden layer parameters are extracted and passed as input and for AE_{ℓ}, $AE_{\ell-\ell}$ hidden layer parameters taken as input. Middle AE's learn the hidden representation of data and clean output is reconstructed by the final autoencoder. After train all autoencoder's fine tune the model to update all parameters of the SDAE.

C. Exponential Linear Unit

ELU is an activation function [4] used in SDAE's both input and output layer respectively. It is used to speed up the training of neural network and can alleviate the vanishing gradient problem. It has negative values which make unit activation closer to zero. So it reduces computational complexity and learning speed. ELU is defined by $\left\{f(r) = \begin{cases} r & r > 0 \\ \alpha.(e^2 - 1, & r \le 0 \end{cases}\right\}$ where r = input, **α** = constant, e= exponent.

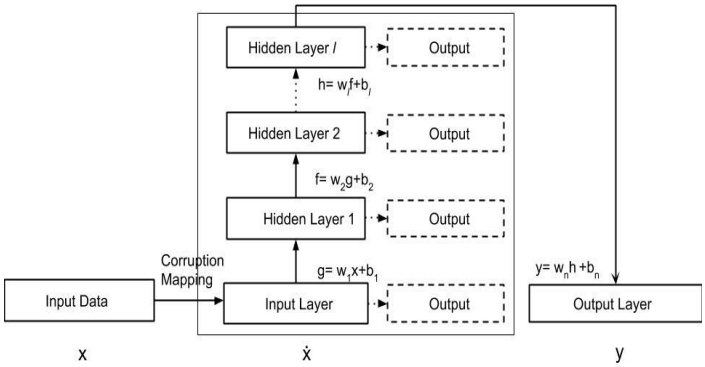

Fig. 2. Stacked Denoise Autoencoder

Table 2. Algorithm 2 - Stacked Denoise Autoencoder

Input : Traffic flow data X= {x_1, $x_2 \ldots x_n$} with missing values Ø And noise ¥;
Output : Reconstructed data without Ø and ¥ values;
Initialization : No of hidden layers H, No of hidden nodes h= {1,2...n}, pre-training iterations I, fine-tuning, iterations=T,epoch= Ω;
Step1 : Map the input data into noise $X = \bar{X}$;
Scale the value and pass into AE_1 (Train AE using algorithm1)
Extract the hidden layer weights(w ε h1),bias(b ε h1) and passed into the AE_2
Step 2 : **for i=2 to H do** //hidden layer calculations
Extract h_{i-1} and pass into i th AE (Train AE using algorithm1)
end for
Step 3 : **Fine tune** the model
Initialize weights and bias of all pre-trained auto encoders
for i=1 to T do
Perform forward propagation to compute y;
Perform backward propagation to compute x-y;
Using optimizer to reduce the error rate;
end for

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Data Description

Traffic flow data is collected from California based PeMS (Performance measurement system) [16] which contains traffic data of all freeways in the state. PeMS has 39000 individual detectors which spreads all over the state and it collects the data at every 30s and converted into 5 min interval.

The data collected from the year 2017 of district 5 (Los angeles). Total 243 vehicle detector stations(VDS) in the district. İn 365 days Feb 1st has some improper data except that have 364 days data and each day is represented as vector D. Each vector has 288 dimensions E as it is 5 minute data. Weekdays, non-weekdays(weekends and holidays), both weekdays and non-weekdays are taken as temporal factors. 115 non-weekdays and 250 weekdays in 2017.

Train and Test ratio is 80:20 for all experiments. To avoid overfitting early stop method is used. 20% to 40% of missing rate is represented as smoother [1], so here 30 % of missing rate is chosen for all experiments. Random Corruption ratio (RC) = $\frac{\sum_{i=1}^{D}\sum_{j=1}^{E} o_{ij}}{DE} * 100\%$ {where o =observed value [0 or 1]}. Tesla k80 12 GB GPU RAM machine is used to train and test the model. Three hidden layers are fixed with size 144, 72,144 respectively. Adam Optimizer is used to minimize the error. The following performance measures [13] are used to evaluate the proposed model such as

Root Mean Square Error (RMSE) = $\sqrt{\frac{\sum_{i=1}^{k}\sum_{j=1}^{l} o_{ij}\ (x_{ij}-y_{ij})^2}{\sum_{i=1}^{k}\sum_{j=1}^{l} o_{ij}}}$

Mean Absolute Error (MAE) = $\frac{\sum_{i=1}^{k}\sum_{j=1}^{l} o_{ij}\,|x_{ij}-y_{ij}|}{\sum_{i=1}^{k}\sum_{j=1}^{l} o_{ij}}$

Mean Relative Error (MRE) = $\frac{\sum_{i=1}^{k}\sum_{j=1}^{l} o_{ij}\frac{|x_{ij}-y_{ij}|}{x_{ij}}}{\sum_{i=1}^{k}\sum_{j=1}^{l} o_{ij}}$

Spatial and temporal factors are very important in imputation and prediction of traffic data. In this imputation process current VDS (single), current VDS with upstream and downstream (augmented) and all VDS are considered as spatial factors to impute single station. Week days (WK), Non-weekdays (N-WK) and both weekdays and Non-weekdays temporal data are evaluated with random corruption scenario.

B. Single VDS Imputation

VDS 500010092 is taken as current station for imputation and analysis of result. Find the missing values of single station based on the same station. Single Location and multiple periods of data are imputed and evaluate the result **(Table.3)**. Both weekdays and non-weekdays data imputes well on single station data.

Table 3. Comparison based on single VDS

Temporal Type	RMSE	MAE	MRE
Week days	10.32	9.16	0.44
Non week days	11.64	10.87	0.48
Both	**7.55**	**6.735**	**0.43**

C. Impact of Upstream (US) and Downstream(DS) on single VDS Imputation

Both US and DS locations are highly correlated with current station. Current station data is augmented with upstream and/or downstream. From the analysis single VDS with downstream gives better imputation result than others **(Table.4).**

Table 4. Single VDS with Upstream(US) and Downstream(DS)

Temporal Type	RMSE	MAE	MRE
Single VDS	10.98	9.56	0.46
Single VDS with US	10.55	9.10	0.44
Single VDS with DS	**9.82**	**8.51**	**0.41**
Single VDS with US & DS	9.91	8.71	0.43

D. Impact of all VDS on single VDS

SDAE with elu activation (specially using both weekdays and non weekdays) fill the missing values very effectively. In this part results (**Table 5**) compare with stacked autoencoder with sigmoid function. Elu activation perform well in reconstruct the observed value with minimal error rate and high performance.

Temporal type	RMSE		MAE		MRE	
	SDAE (sigmoid)	**SDAE (Elu)**	SDAE (sigmoid)	**SDAE (ELu)**	SDAE (sigmoid)	**SDAE (Elu)**
Week Days (WD)	15.18	**10.98**	14.3	**10.14**	0.48	**0.46**
Non-Weekdays	20.51	**13.94**	18.24	**12.32**	0.57	**0.51**
Both	13.53	**8.50**	12.75	**7.24**	0.48	**0.43**

Table 5. All VDS on Single station

From all the observations both weekdays and non-weekdays gives more accurate value than considering single VDS and also Downstream and Upstream.

E. *Representation of sigmoid and Elu activation functions*

Activation function (called as transfer function) is used to compute the weighted sum of input and biases, which can be decide whether neuron can be fired or not. It can be divided into two types: linear activation and Non-linear activation [4].

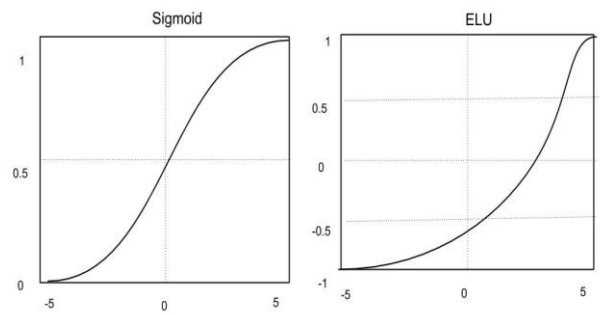

Fig. 3. Activation function representation (sigmoid & ELU)

Results are compared (**fig.4**) for single VDS with 30 % random corruption ratio. SDAE with Elu gives minimum error rate than sigmoid activation.

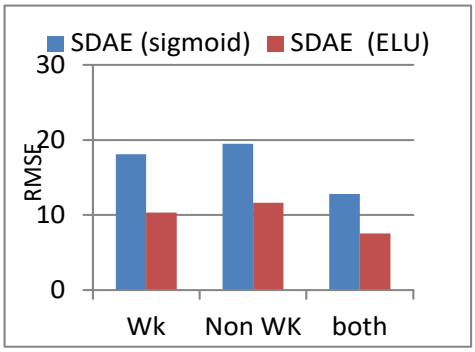

Fig. 4. Comparison of SDAE with Sigmoid & Elu

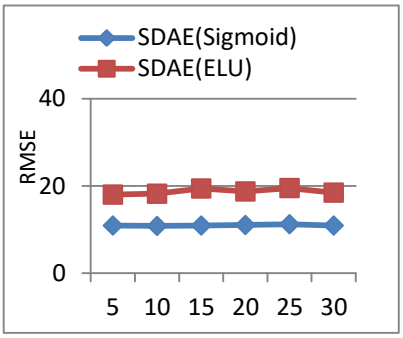

Fig. 5. Comparison from 5 – 30 % of corruption rate

Take the corruption ratio 5% to 30% and check the RMSE error rate for single VDS (both weekdays and non weekdays). As shown in **fig.5,** SDAE with Elu gives better accuracy than sigmoid in various error rates.

V. CONCLUSION

In this paper Stacked denoise autoencoder with Elu activation function is used for traffic data imputation. Activation functions play a key role to produce the accurate result in terms of reconstruct the output and find the missing values. Sigmoid is a linear activation function commonly used in all neural network models. Elu is an alternative of ReLu and it gives more accurate result, increase the training speed of network and reduces the computational complexity. Simple and complex structure of data (spatial and temporal) is evaluated with random corruption strategy. In future the model may evaluate based on continuous corruption with various missing rates.

ACKNOWLEDGEMENT

We thank Sri Ramakrishna College of arts and science for giving support for doing this research work.

REFERENCES

1. A F Costa, M S Santos, and J P Soares, "Missing Data Imputation via Denoising Autoencoders : The Untold Story", IDA 2018, pp. 87–98, Springer (2018)
2. Aurelien Geron, "Hands on Machine Learning with Scikit-Learn and TensorFlow" (2017)
3. B Ran, H Tan, J Feng, YLiu, and W Wang, "Traffic Speed Data Imputation Method Based on Tensor Completion", Computational Intelligence and Neuroscience, vol. 2015, Article ID 364089, 9 pages, Hindawi (2015)
4. C E Nwankpa, W Ijomah, A Gachagan and S Marshall, "Activation Functions: Comparison of Trends in Practice and Research for Deep Learning", arXiv:1811.03378v1 [cs.LG] (2018)
5. G Chang, T Ge, "Comparison of Missing Data Imputation Methods for Traffic flow", Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), December 16-18, Changchun, China, IEEE(2011)
6. H Tan, G Feng, J Feng, W Wang, Y J Zhang, and F Li, "A tensor-based method for missing traffic data completion", Transportation Research Part C: Emerging Technologies, vol. 28, pp. 15–27 (2013)
7. L Gondara, K Wang, "Recovering Loss to Followup Information Using Denoising Autoencoders", arXiv:1802.04664v1 [cs.LG] (2018)
8. L qu, L li, Y zhang, J hu, "PPCA-Based Missing Data Imputation for Traffic Flow Volume: A Systematical Approach", IEEE Transactions on Intelligent Transportation Systems, vol. 10, no. 3, September (2009)
9. M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques", no. 722, p. 116, Transportation Research Record (1979)
10. S Narmadha, V.Vijayakumar, "A Study on Imputation Methods for Vehicle Traffic Data", International Journal of Recent Technology and Engineering (IJRTE), Volume-7, Issue-5S (2019)
11. W C Ku, G R. Jagadeesh, A Prakash, T Srikanthan, "A Clustering-Based Approach for Data Driven Imputation of Missing Traffic Data", 2016 IEEE Forum on Integrated and Sustainable Transportation Systems (FISTS), Beijing, China, IEEE Xplore Digital library (2016)
12. Y Duan, Y Lv, W Kang, and Y Zhao, "A deep learning based approach for traffic data imputation", 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 912–917, IEEE (2014)
13. Y Duan, Y Lv, Y L Liu, FY Wang, "An efficient realization of deep learning for traffic data imputation", Transportation Research Part C (2016) 168–181, Elsevier (2016)
14. Y Li, Z Li, and L Li, "Missing traffic data : comparison of imputation methods", IET Intelligent Transport Systems, Volume 8, Issue1, pp. 51–57, IEEE Xplore(2014)
15. Z. Yang, "Missing Traffic Flow Data Prediction using Least Squares Support Vector Machines in Urban Arterial Streets", 2009 IEEE Symposium on Computational Intelligence and Data Mining, IEEE Xplore (2009)
16. http://pems.dot.ca.gov/

AUTHORS PROFILE

S.Narmadha, Research Scholar, Sri Ramakrishna College of arts and Science. Published 6 papers in International/ National Conferences and Journals. Research interest includes Machine learning and deep learning.

Dr V.Vijayakumar, Professor, Sri Ramakrishna College of Arts and Science, Coimbatore. He has twenty years of experience in Teaching, Research & Development. He awarded Research degree in the area Video Data mining. He has recognized as Research supervisor in Bharathiar University and guided 8 M.Phil scholars. Currently, he is guiding three Ph.D Scholars and one M.Phil Scholar. He completed AICTE Research Promotion Scheme grant to the tune of Rs.10.5 lakh and currently received Major Research Grant Rs. 10.5 Lakh form ICSSR-IMPRESS. He was appreciated by the Analog and Digital Systems. He was honoured with Best Academic Officer Award by EMC Corporation Bangalore. He has published twenty nine papers in International/ National Conferences / Journals including Springer and IEEE. He has delivered lectures in Big Data Analytics (Hadoop and Ecosystems), R-Programming, No-SQL (MongoDB), Internet of Things and SMACT. He is serving as Subject Matter Expert and Adjunct Faculty in Texila American University, Guyana, South America. He is a reviewer of various International Conferences and journals. He is serving as a member of various bodies. His research interest includes Machine Learning, Computer Vision, Data Science & Big Data Analytics and Internet of Things. He is a member in ISTE, IAENG, IACSIT and CSTA professional societies.