Optimizing a Deep Learning Model in Order to Have a Robust Neural Network Topology

Riaz Ullah Khan, Rajesh Kumar, Nawsher Khan, Xiaosong Zhang, and Ijaz Ahad

Abstract—**In this study, a method based on different feature engineering / feature extraction / feature derivation is proposed for improving air passenger forecasting by machine learning existing libraries. In this kind of formulation, we kept focus on creating different kinds of datasets that differ one from another by methodology so we extracted new features and compared new feature space with original feature space in terms of variable importance. We conducted experiments to improve the variance by aggregating all the features in final feature space. Finally, we optimized a deep learning model to have a Robust Neural Network Topology.**

Index Terms—**Deep learning, logistic regression, feature extraction, neural network.**

I. Introduction

As an important aspect in air passenger prediction or forecasting research is the prediction of airport usage. Original forecasting models [1]−[3] don't consider the influencing factors comprehensively, especially at the time of dealing with changeable state or complex airport usage states. This paper recognizes different attributes as classification vectors and designs new time series algorithms through calculating representatives in all kinds of clusters between observing time series and representatives. We investigate a new approach of feature selection, and demonstrate the features which are useful for the prediction of the number of passengers traveling in different seasons. The Deep Learning and GA-based feature extraction is guided by higher probabilities of survival for fitter combinations of features, where the features extracted by Feature-Space = *{'Year', 'Month', 'Destination Airport', 'Destination Airport Name', 'Destination City Name', 'Destination Region Name', 'Destination Airport Name'}* in figure 1. We studied the effects of different parameters of different techniques through classification accuracy and compared the results to those obtained with step-wise feature addition. The importance of correlations among different features is also discussed with better understanding the correlation among different features [3]−[5]. In this study, Attention is given to features by experimenting on new features obtained from optimized neural network hidden layers. We used the strength of existing machine learning models to visualize the importance of feature selection. After choosing the set of features, adding new features in different steps and visualizing the correlation among all the features, we chose the methodology of incorporating deep neural network. This methodology ensures that all best practices of modelling are maintained. Many scholars give a strong argument against machine learning techniques that the process of variable selection is not well defined, therefore scholars are focusing on deep neural network to achieve accurate forecasting results [6], [7]. The figure segment needs to deliver forecasts for each admission level and flight date. The ticket request at a charge level (e.g. per take-off date) can be displayed as a period arrangement.

Shockingly, practically speaking just couple of strategies have been found to deliver satisfactory figures due to the structure and nature of the current information. For our application the world is changing so rapidly that when all is said in done just few chronicled information are accessible and much of the time various pertinent esteems are absent. Numerous investigations on this theme have demonstrated that for our information the straightforward and vigorous time arrangement determining models like basic normal, diverse renditions of exponential smoothing or relapse models are altogether superior to anything various surely understood more refined techniques. The purpose behind this lays in the straightforward strategies' capacity to influence satisfactory estimates to even on few chronicled information and to have the capacity to adjust all the new circumstances.

In this paper, we also kept focus on the above-mentioned methodology and observed good results which are shown in the following sections of the paper.

A. Main Contribution of this Paper

- To propose an optimized method of Robust Neural Network technique with existing machine learning models.
- To extract new features from original feature space.
- To observe the importance of new features on existing features through different experiments i.e. Deep Neural Network

B. Structure of the Paper

This paper is organized as follows.

Section I discusses a brief introduction and main contribution of the paper. Section II describes the Experiments, Dataset and parameter setting. Section III explains creating features using Robust Neural Network Topology. Finally, Section IV concludes the paper and

Manuscript received December 5, 2017; revised January 5. 2018.

Riaz Ullah Khan, Rajesh Kumar, and Xiaosong Zhang are with school of Computer Science, University of Electronic Science and Technology of China, China (e-mail: rerukhan@ gmail.com, rajakumarlohano@gmail.com, johnsonzxs@ uestc.edu.cn).

Nawsher khan is with Abdul Wali Khan University Mardan, Pakistan and King Khalid University Abha, Kingdom of Saudi Arabia (e-mail: nawsherkhan@ gmail.com).

Ijaz Ahad is with School of Information and Software Engineering, University of Electronic Science and Technology of China, China (e-mail: ijazahad1@ gmail.com).

DOI: 10.7763/IJMO.2018.V8.639

enlightens the future work.

II. Experiments

A. Dataset and Parameter Setting

Dataset along with values and properties is shown in Table I. To evaluate the accuracy and performance of different methods, we use K-fold cross-validation in our experiments. For each experiment, the dataset was split 75-25 ratio, each split into 2, one for training, and the other for testing.

TABLE I: The Dataset

Name	Values
Dataset Characteristics	Multivariate
Number of Attributes	13
Number of Instances	3300
Number of Web Hits	533922
Area	Air Industry
Associated Tasks	Regression, Deep Neural Network

B. Aggregating Data and Deriving New Features

We started the research with the initial dataset which played the main role as a starter file. This file contained the feature space with respect to the target variable;
Feature-Space = *{'Year', 'Month', 'Destination Airport', 'Destination Airport Name', 'Destination City Name', 'Destination Region Name', 'Destination Airport Name'}*

First step of feature selection with aggregating data and deriving new feature produced figure 1 as an output. Starting from this original feature space, the goal was to train a simple model on this feature space in-order to have some starting point from where we could observe variable importance and some performance measure. In this first approach, the main interest was to improve the "variance explained" measure. Fitting a simple decision tree with R random forest library, we obtained the variance and feature importance on the original data file in Fig. 1.

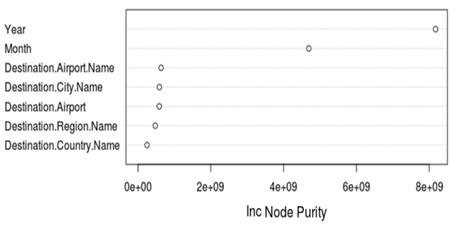

Fig. 1. Variable importance on original features.

Continuing with the first step of the research, even that the original feature space is not looking very promising, two new features were created:

- From the Month feature by doing cuts in order to see whether season corresponding to its particular months will be influential or not? We used regression technique by using Algorithm 1 to observe this. So, a new feature was created based on month in Fig. 2.
- From another source of information, where for each month is listed corresponding holidays, another feature is created in the following way:
  a) Zero will be appended for a particular month if there are no holidays.
  b) Total number of holidays for a particular month assuming that this will bring some kind of magnitude of all holidays and this will have a special influence regarding the number of passengers for that particular month.

Algorithm 1: Seasons Per Year

```
function YEAR (Month m)
    if m = 'Jan' || m = 'Feb' || m= 'March' then
        season1
    end if
    if m = 'Apr' || m = 'May' || m= 'June' then
        season2
    end if
    if m = 'July' || m = 'Aug' || m= 'Sep' then
        season3
    end if
    if m = 'OCT' || m = 'Nov' || m= 'Dec' then
        season4
    end if
end function
```

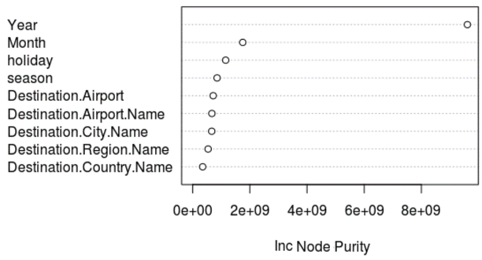

Fig. 2. Variable importance with additional features, seasons and holidays.

The variance increased from 31.28% to 48% in Figure 2because of creating these additional features that are also captured by decision tree feature importance.

Next step is to involve the initial research by aggregating other features like:

- GDP for the period, starting from 1998 to 2016
- Jet fuel prices
- Total population for period, starting from 1998 to 2016
- Interest rate for period, starting from 1998 to 2016
- Average base fair
- Distances from origin to destination

Aggregation all these features, another experiment was conducted to see if these features are influential or not and whether the explained variance is increasing or not. The results are presented explained: 66.68% in Figure 3.

After aggregating new features, creating another two features (holiday and season) we can conclude that they are influential by visual inspection of tree model feature importance. We also managed to increase the variance from 31.28% in Fig. 1 to 48% in Fig. 2 and then 66.68% Fig. 3. In Fig. 3 We added Jet Fuel price as a new feature to observe the influence on number of passengers i.e. If the Jet Fuel price increased so the air fare will be increased, and it may affect the number of passengers. In our case, we observe that Jet Fuel price is not strongly influential on the buying power of customers. We observed the influence of different features e.g. GDP, Average base fair, Population, and Distance.

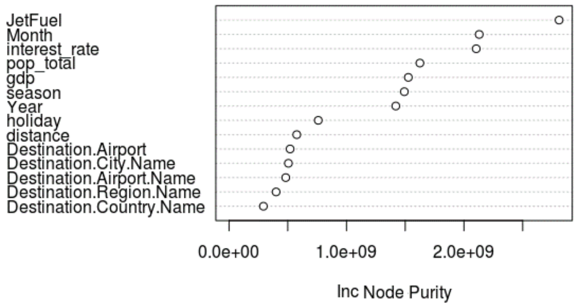

Fig. 3. Variable importance with all features.

III. Optimizing a Deep Learning Model in Order to Have a Robust Neural Network

After finishing with sub-steps of first part of the experiment, the next stage involved finding a good neural network topology. We chose h2o library for implementation that runs under R environment. The main reason for choosing this implementation in presence of other neural networks libraries such as pure Python implementations or tensor-flow was the fact that h2o [8] has straightforward hyper-parameter tuning and is also fast, having Java back-end and option to run on some number of clusters k where k represents the number of cores on the processor [9]. After performing hyperparameter optimization using grid search [10], [11], the best neural network topology is tuned as Algorithm 2.

Algorithm 2: Tuned Model of Deep Net

```
function TUNEDMODEL(H20)
    x = predictors
    y = target
    training-Frame = train
    hidden = c (128,63,32)
    epochs = 100
    fold-Assignment = 'Module'
    l1 = 5.6e-05
    l2 = 7.4e-05
    input-Dropout-Ratio = 0.05
end function
```

This topology means we are dealing with a neural network with 128 neurons in first hidden layer, 63 neurons in second hidden layer and 32 neurons on third hidden layer. Regarding the number of epochs, the neural network should iterate, the best number from the possible states of the greed was 100 epochs. A dropout of 25% of the data from original train file was performed in order to have a 5-fold cross validation with 75% data for train and 25% data for evaluation to have some initial intuition about how the tuned deep learning model is performing by plotting predicted values on the un-seen test set against ground reality values. This is presented in the Figure 4 followed by plots of the original distribution in Figure 5 and the predicted distribution in Figure 6.

As of now demonstrated, there are two foremost ways to deal with adaptation to changes in our application: to adapt the verifiable information or to forecast themselves. Our approach is utilizing the two adaptations. The adaptation to attractivity changes lying later on or to short term influences is done as an undertaking of adaptation of the fundamental forecast. It doesn't influence the verifiable information and the generation of the fundamental forecasts. The adaptation to authentic attractivity changes is performed on the chronicled information.

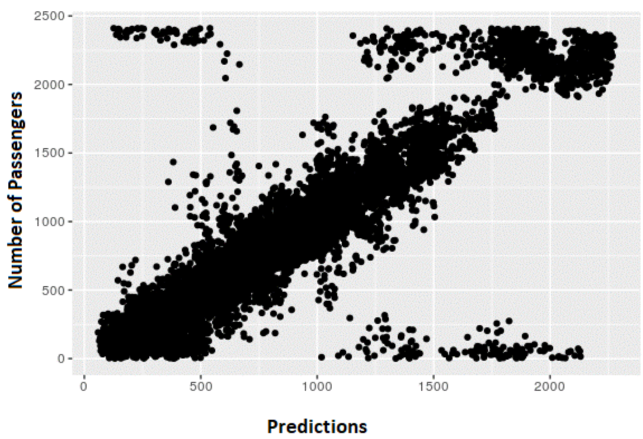

Fig. 4. Predicted values on the un-seen test set against ground truth values.

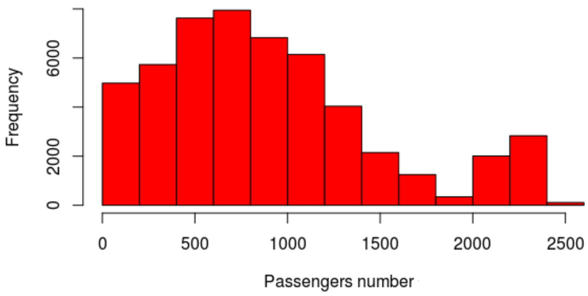

Fig. 5. Original distribution of target.

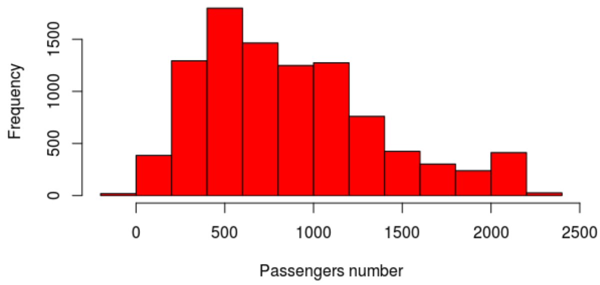

Fig. 6. Distribution of predicted values.

The historical information is adjusted to the present circumstance to speak to a steady reason for the essential conjectures. Short term impacts are expelled from the recorded information. The upside of this approach is that the adjusted recorded information speaks to dependably a history coordinating to the latest takeoff and that it would then be able to be utilized to deliver every single essential forecast with no other requirement for adaptation. Another preferred standpoint is that the rehashed short-term impacts are segregated from the chronicled information and can in this way be likewise learnt for future adaptations. Close to the adaptation to attractivity changes and short-term influences, there is a plausibility to utilize the latest appointments for a future takeoff for the estimation of the flight particular conduct, which is translated as clamor in the time series models. There isn't just a solid seasonal conduct in the booking information, unique seasons like fairs, holidays, social or political occasions or on the other hand strikes are impacting the information, as well. There are various known and archived ways to deal with handle regular conduct in time series. These methodologies could be utilized and create better conjectures contrasted with the situation when no seasonal model is utilized. Be that as it may, in our application not just the appointments at a flight are impacted, yet in addition the advancements of the booking bends ought to be adjusted to the seasonal conduct.

IV. Conclusion

Accurate prediction of the air passengers forms an important part of an airport management/airline team management processes in the airline industry. In this paper, we have presented an overview of different features which effect the forecasting of air passengers. The experiments show that for accurate forecasting of air passengers, it is very important to select features which are closely related to each other's. We have noticed from our experiments that adding new features to the outputs of old features clearly changed the variance results in new outputs. We significantly improved the forecasting results by adding new features. Optimizing Deep Neural Network model also plays a vital role in improving the results. It is observed in general that Deep Neural Network model gives accurate results than individual machine learning model, even whatever method we use to combine different machine learning models [12]-[14]. The Deep Neural Network method is the new method in the predictive modeling world. The parameterization of the prediction is currently based on feature importance on the original features space. The basic prediction produced by different time series models are combined by a fixed rule depending on the booking level. The weighting parameters to adopt the forecasts to the season and holiday are calculated by fixed rules too, depending on the number of months to destination and the seasonality and holidays information, and though all these rules have been optimized with respect to performance, the intention is to go away from the fixed rules and to adopt a dynamic approach to combination. The combination should adjust the used basic methods as well as the adaptation to the new features.

A. Future Work

In this study, a method based on different feature engineering / feature extraction / feature derivation is proposed for improving air passenger forecasting by machine learning existing libraries. In this kind of formulation, we kept focus on creating different kinds of datasets that differ one from another by the methodology, so we extracted new features. In future work, we will extend this paper and particular attention will focus on features obtained by a modified Principal Component Analysis (PCA) using a different correlation matrix obtained from kinetic / information energy properties of the features represented by random vectors and also by experimenting on new features obtained from an optimized neural network hidden layers. We will introduce a method for improving prediction performance metric by ensembling predictions made on different engineered feature spaces with attention on replacing outliers of incorrect predictions. The main objective besides improving the performance metric is to obtain a distribution on some hold-out dataset that resembles the original distribution of the training dataset.

References

[1] A. Gosavii, N. Bandla, and T. K. Das, "A reinforcement learning approach to a single leg airline revenue management problem with multiple fare classes and overbooking," *IIE transactions*, vol. 34, no. 9, pp. 729–742, 2002.

[2] M. N. Laik, M. Choy, and P/ Sen, "Predicting airline passenger load: A case study," in *Proc. 2014 IEEE 16th Conference on Business Informatics (CBI),* vol. 1, pp. 33–38. IEEE, 2014.

[3] S. Riedel and B. Gabrys, "Adaptive mechanisms in an airline ticket demand forecasting system," in *Proc. EUNITE'2003 Conference: European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems*, Oulu, Finland.

[4] C. Udriste and O. Calin, *Geometric Modeling in Probability and Statistics*, London, UK: Springer publisher, 2014.

[5] D. Rizescu and V. Avram, "Using Onicescu's informational energy to approximate social entropy," *Procedia-Social and Behavioral Sciences*, 2014.

[6] A. Essa and H. Ayad, "Student success system: Risk analytics and data visualization using ensembles of predictive models," in *Proc. 2nd International Conference on Learning Analytics and Knowledge*, pp. 158–161. ACM, 2012.

[7] J. H. Friedman and B. E. Popescu, "Predictive learning via rule ensembles," *The Annals of Applied Statistics,* pp. 916–954, 2008.

[8] S. Landset and T. M. Khoshgoftaar, "A survey of open source tools for machine learning with big data in the Hadoop ecosystem," *Journal of Big Data*, 2015.

[9] A. McGregor, M. Hall, P. Lorier, and J. Brunskill. "Flow clustering using machine learning techniques," *Passive and Active Network,* 2004.

[10] J Bergstra and Y Bengio. "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, 2012.

[11] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *Advances in neural Information Processing Systems 25, Curran Associates,* pp. 2951-2959, 2012

[12] D. Brzezinski and J. Stefanowski, "Reacting to different types of concept drift: The accuracy updated ensemble algorithm," *IEEE Transactions on Neural*, 2014.

[13] D. Jiménez. "Dynamically weighted ensemble neural networks for classification," *Neural Networks Proceedings*, 1998.

[14] Q. Zhang, J. M. Hughes-Oliver, and R. T. Ng, "A model-based ensembling approach for developing QSARs," *Journal of Chemical*, 2009.

Riaz Khan was born in District Malakand, KPK Province, Pakistan in 1980. Currently, he is a PhD candidate at University of Electronic Science and Technology of China. He earned his M.S. Degree from University of Sunderland, United Kingdom in 2010, Diploma in Information Technology from Metro College of Management Studies in 2008, MSc Degree from Hazara University, Pakistan in 2006 and Bachelor's Degree from University of Peshawar, Pakistan in 2003. Since 2010, he has worked in various industrial and educational institutions. From 2010 to 2012, he was appointed as an IT analyst in a reputable organization in United Kingdom (IU consultant, UK). From 2012 to 2016, he was appointed as a lecturer in Computer Science at Abdul Wali Khan University Mardan, Pakistan, and currently Dr. Riaz Khan is working in the Lab of Cyber Security, School of Computer Science and Engineering. University of Electronic Science and Technology of China as a PhD fellowship. His areas of interest include Cyber Security, Big Data, Dynamic Program Analysis, Cloud Computing, Data Management, Distributed System, Blockchain and Internet of Things. In addition, he has published more than 15 articles in various International journals and conference proceedings.

Rajesh Kumar belongs to Sindh, Pakistan. Dr. Rajesh Kumar is currently a PhD student Research Fellow at the University of Electronic Science and Technology of China. He got Master Degree from University of Sindh, Jamshoro, Pakistan. He received his bachelor degree in computer science from University of Sindh, Jamshoro, Pakistan. In addition, he has published 9 articles in various International journals and conference proceedings.

Nawsher Khan belongs to Samarbagh, Dir Lower, Pakistan. Samarbagh is situated on 34° 57′ 11.39″ N, 71° 41′ 9.75″ E co-ordinates which has natural beauty. Dr. Nawsher Khan was a Post-Doctoral Research Fellow at the University of Malaya Kuala Lumpur, Malaysia in 2014. He got Doctoral Degree from University Malaysia Pahang (UMP) Malaysia in 2013. He received his bachelor degree in computer science from University of Peshawar, Pakistan and master degree in computer science from Hazara University, Pakistan. Since 1999, he has worked in various educational institutions. In 2005, he was appointed in NADRA (National Database and Registration Authority) under the Interior Ministry of Pakistan and in 2008 has worked in (NHA) National Highways Authority. Now he is working as an Assistant Professor at Abdul Wali Khan University Mardan, Pakistan, and currently Dr. Nawsher Khan is working in King Khalid University Abha, Saudi Arabia as Associate Professor. His areas of interest include Big Data, Cloud Computing, Data Management, Distributed System, Scheduling, Replication and Sensor Network. In addition, he has published more than 50 articles in various International journals and conference proceedings.

Xiaosong Zhang was born in Sichuan Province of China in June 1968. He received his M.S and PhD degrees from University of Electronic Science and Technology of China in 1999 and 2011, respectively. He has been a Professor in Information Security at University of Electronic Science and Technology of China since 2011. His areas of interest include cryptography, Dynamic Program Analysis, and Information Security. He earned best achievement awards in 2012, 2013, 2014, 2015, 2016 and 2017. As of 2017, he published more than 60 academic papers in the field of network and information security, including CCF Class A top-level journal articles such as IT, TSE and TIFS. He published "Network Security Protocol," "Malware Analysis and Testing," " Software testing "and other monographs, textbooks and translation, was authorized international and domestic invention patents 28. 10 of software were registered as copyright.

Ijaz Ahad belongs to KPK province of Pakistan. Mr. Ijaz Ahad is currently a Postgraduate student at the University of Electronic Science and Technology of China. He got MSc Degree from Hazara University, Pakistan. He received his bachelor degree in computer science from University of Malakand, Pakistan. In addition, he has published 3 articles in various International journals and conference proceedings.