ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

ACCURACY ENHANCEMENT OF BRAIN EPILEPSY DETECTION BY USING OF MACHINE LEARNING ALGORITHMS

Rand Natiq AL-DAHHAN[1]

Altinbas University, Graduate School of Science and Engineering, Electrical and Computer Engineering, Istanbul. Randaldahan995@yahoo.com, ORCID No: 0000-0002-5218-7538

Osman N. UÇAN[2]

Altinbas University, Faculty of Engineering and Natural Sciences, Department of Electrical-Electronics Engineering, Istanbul. osman.ucan@altinbas.edu.tr ORCID No: 0000-0002-4100-0045

Abstract

Data has gained vital role in science and engineering applications; the proper data analysis has made it possible to boost the economical worthiness of those applications. Machine learning tools are used to classify the big data in order to discover the hidden patterns in them. That may lead to noteworthy advantages that related to future prediction of the data. The resultant information can be used to enhance the practical systems in such way only the profitable thing can be come on then. In other way, it helps to prevent any unpleasant occurrence that may harm the company or the organization. A brain epilepsy disease prediction system is implemented using four different algorithms namely: Naive Bayes algorithm, K-Nearest Neighbours algorithm, Random Forest algorithm and Long Short Term Memory Neural Network. The performance metrics are also initiate in order to evaluate the difference in prediction performance of the four tools. The accuracy of prediction the disease was recorded more likely 33.035, 95, 61.195 and 96.79 for the Naïve Bays, Random Forest, K-Nearest Neighbour and Long Short Term Neural Network.

Keywords: LSTM, FFNN, Random Forest, KNN, Naïve Bays.

BEYİN EPİLEPSİ TESPİTİNİ KULLANARAK DOĞRULUK GELİŞTİRME MAKİNA ÖĞRENME ALGORİTMALARI

Özet

Bilim ve mühendislik uygulamalarında veriler hayati bir rol oynamıştır; doğru veri analizi, bu uygulamaların ekonomik değerini artırır. Makine öğrenimi araçları büyük verileri sınıflandırmak için kullanılır ve veriler içindeki gizli kalıpların bulunmasını sağlar. Bu gelecek tahmini ile ilgili önemli avantajları sağlayabilir. Sonuçta elde edilen bilgiler pratik sistemleri sadece karlı olan şeyleri geliştirmek için de kullanılabilir. Başka bir şekilde bakıldığında, şirkete veya kuruluşa zarar verebilecek hoş olmayan olayların önlenmesine de yardımcı olur. Beyin epilepsi hastalığı tahmin sistemi dört farklı algoritma kullanılarak uygulanır: Naive Bayes algoritması, K-en yakın komşular algoritması, rastgele orman algoritması ve uzun kısa süreli bellek sinir ağı. Performans

ölçümleri de dört aracın tahmin performansındaki farkı değerlendirmek için başlatılır. Tahmin doğruluğu, bu dört yöntem için sırasıyla 33,035, 95, 61,195 ve 96,79 olarak kaydedildi..

   Anahtar Kelimeler: LSTM, FFNN, Rastgele Orman, KNN, Naive Bayes.

1. Introduction

Data gain paramount task in various sectors of human life in current days. Traditionally, data employed for recording business activates in markets, companies and banks, it rather than this used for prediction of future facts and circumstances of the business. The todays great utilization of data is making it acts as fundamental worth resources more likely mineral resources alike petrol and oil. Large applications and usage of data are folded under the so called data sciences which looks after the tools and technologies that can be used for mining the information and hidden patterns form the data (Wang et al., 2017)

Mining of data is gained extended interest by medical applications, finical applications, oil companies, health care and insurance organizations, scientific and engineering applications and many more. From the medical point of view, data can be gathered from the hospitals after the entry of patients arriving with their reports along with their case diagnosis on hospitals systems (Mirza and Cosan, 2018).

Data become paramount resource same likely the natural fortune of oil and minerals, hence, data science had established to deal with the facts underlaying in the said data. Data science field provide the algorithms that used to min the data to extract the hidden patterns herewith (Chen, Liu and Liu, 2017).

Due to the importance of the data in drawing the future strategies of business in large sectors in today's life; data collection techniques are also vital for efficient data mining. However, data is being collected using efficient systems such as accurate data entry and registry systems, sensors, digital devices for costumers' feedbacks entry and many more (Jithesh, Sagayaraj, and Srinivasa, 2017).

Medical data have been used for efficient diagnosis of diseases and helping to develop those systems that intern capable to predict the disease by only looking into the medical tests and examination reports. The accuracy of those systems is still under developments.

In this chapter, problem definition of the brain epilepsy disease detection in the large dataset is made along with the study objectives and dissertation report organization is demonstrated in the hereinafter (Yuhai, Shuo and Linfeng, 2018)

LSTM neural network algorithm is deployed for disease production by optimizing the number of forget gates inside the model structure. However, model is implemented with least possible forget gates which reduce the cost of computation and increase the performance of model accuracy.

2. Data pre-processing

Data is gathered from large number of cases who are suffering from brain epilepsy issues. Those cases are ranging between two thousand to three thousand cases. The data pre-processing made as in the following points.

Data is collected for all the cases and tabled in excel sheet format where that consists of eleven columns and two-thousand, five hundred rows. The columns are representing the tests made to each case in order to diagnose the disease while the rows are representing the number of cases Martin Sundermeyer et.all,

Data cells are then observed individually by computer program (code) in order to indicated the missing values and to verify the values more likely identifying the minimum and maximum of the value along with value types (integer, float, logical or character) Dires Negash Fente and (Fente and Singh, 2018, Liu, Zhou and Li, 2018).

As the data range and type are uncovered now the missing values from the data must be identified for refilling process. However, missing values are been realized in most of big data due to various reasons such as: entry error from the source at the time of data entry; it might be due to bad sectors and damages of storage system which lead to data missing, this kind of causes can be tackled by recovering the data but always error is expected to accompanied with the recovered data and this error is seen as missing data N Maria Klara (Jędrzejewska, Zjawiński and Stasiak, 2018, Lu, Salem, 2017).

The data important step after identifying the missing data is refilling those data by the values which may not case further error while training. Amongst many proposed methods of tackling the drawbacks of missing data, the average method is the best over the others.

In order to refill the missing value of any column, the average of the particular columns is to be taken and all the missing values is to replace by the average value of the column. Figure 1 demonstrates the process of missing values replacement (Chen, Liu, and Liu, 2017, Liu, Zhou, and Li, 2018).
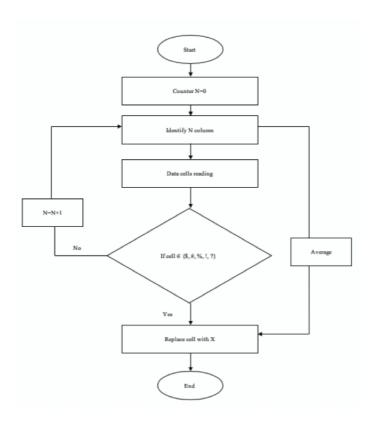
Start

Counter N=0

Identify N column

Data cells reading

N=N+1

No

If cell ∈ ($, #, %, !, ?)

Average

Yes

Replace cell with X

End

Figure 1: Process of missing values identification and refilling.

3. Predictor Designs

In order to predict the disease occurrence at big data, the record from every patient is studied after the data pre-processing. The values are analyzed carefully in order to discover the hidden patrons within values, The overall process that made in this regard can be summarized as the following:

1) data is firstly pre-processed in order to justify the missing values so as to the training quality is ensured.

2) from the other hand, values normalization is also made in order to minimize the variance between the values for optimizing the training performance.

3) each row in the data is realized in order to identify the patient data which will be used in both training and testing process. However, over two thousand and five hundred patients only two thousand of them will be used during the training stage while the rest five hundred will be used during the testing stage.

In order to ensure high performance of disease prediction, smart neural network classifier has been used called as long short term memory neural network. This model is consisted of two hidden layers and one input and one output layer.

This model is considered as smart version of artificial neural network, it used same architecture of recurrent neural network except some changes in the weights. It is capable to process large number of data values at one. It differs from classical feed forward neural network model by its large feedback loops between its layers.

The main terminology of the long short term neural network is word "gates" which is given to the layers. More likely, the input and output gates is representing the input and output layers in normal (classical) feed forward neural network. While the forget layer is representing the hidden layer which is popular in classical feed forward neural network N B. (Chandra and Sharma, 2017).

The structure of long short term memory neural network can be illustrated in the Figure below:
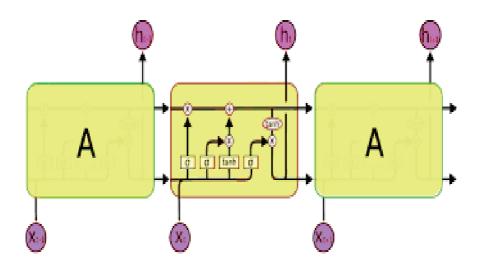

Figure 2: Long short term memory neural network structure N B. (Chandra and Sharma, 2017).

So to say, a two thousand objects (cases) of from the data is used for training the above model and the rest five hundred of the data is used to test the model. The model is supposed to predict whither the object is infected or not.

In order to evaluate the performance of this model over the other available algorithms, the performance of prediction s compared with the same while using, Random Forest and K nearest Neighbour and Naive Bayes algorithms.

4. Study Outcomes

After examination of all algorithms in prediction of disease, the performance metrics for all the algorithms are recorded and tabled in the following Tables.

287

Table 1: performance metrics of prediction measure for all the algorithm.

Tool	Accuracy	Time	MSE	MAE	RMSE
N. Bays	33.035	0.3905	7.604	10.644	2.757535
RF	95	62.915	0.065	0.065	0.254951
KNN	61.195	0.67	1.8764	1.2446	1.369818
LSTM	96.79	4	0.0726	0.0726	0.269393

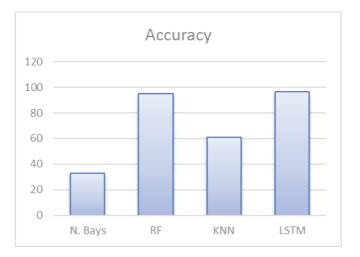The same can be graphically represented by the Figure hereinafter.

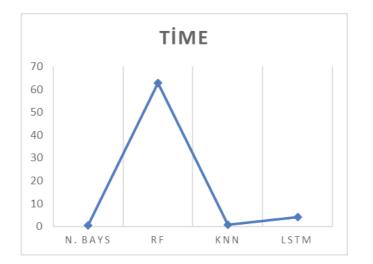**Figure 3:** Accuracy of prediction measure for all the algorithm.

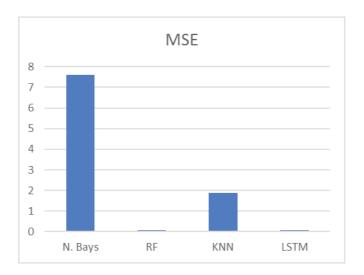**Figure 4:** Time of prediction measure for all the algorithm.

Figure 5: MSE of prediction measure for all the algorithm.

5. CONCLUSION

Machine learning approaches are gained extra importance in today's life technologies. It became deployed in many of applications in science and engineering. Machine learning and artificial intelligence is being used in medical applications to predict the disease occurrence after training the machine (model) with data.

Long short term memory neural network is can be used to analyze the big data with high efficiency and less number of error. This study involves using the brain epilepsy disease dataset to train the LSTM model and then the performance of the same is evaluated using the accuracy, time, MSE, MAE and RMSE.

The results are compared with the other machine learning tools such as K-nearest neighbours, Random Forest, Naive Bayes. Results are shown that Long short Term Memory neural network is outperformed over the others. An accuracy of prediction equal to ninety-six percent is observed in the results while using Long Short Term Memory Neural Network over the other algorithms.

Random Forest Algorithm could also achieve a good prediction accuracy but its process lasts for long time which considered as main degradation of the performance. The other algorithms also achieved different accuracy measures and all are lesser than our proposed model.

6. REFERENCES

Chandra, B., and R.K. Sharma. 2017. On improving recurrent neural network for image classification. In 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, 1904-1907

Chen, Z., Y. Liu, and S. Liu. 2017. Mechanical state prediction based on LSTM neural netwok. In 2017 36th Chinese Control Conference (CCC), Dalian, 3876-3881.

Chen, S., C. Peng, L. Cai, and L. Guo. 2018. A deep neural network model for target-based sentiment analysis. In 2018 international joint conference on neural networks (IJCNN), Rio de Janeiro, 1-7.

Fente, D.N., and D.K. Singh. 2018. April. Weather forecasting using artificial neural network. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, pp. 1757-1761.

Jędrzejewska, M.K., A. Zjawiński, and B. Stasiak. 2018. Generating Musical Expression of MIDI Music with LSTM Neural Network. In 2018 11th International Conference on Human System Interaction (HSI), Gdansk, 132-138.

Jithesh, V., M.J. Sagayaraj, and K.G. Srinivasa. 2017. LSTM recurrent neural networks for high resolution range profile based radar target classification. In 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, 1-6.

Liu, Y., Y. Zhou, and X. Li. 2018. Attitude estimation of unmanned aerial vehicle based on lstm neural network. In 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, 1-6.

Lu, Y., and F.M. Salem. 2017. Simplified gating in long short-term memory (lstm) recurrent neural networks. In 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, 1601-1604.

Mirza, A.H., and S. Cosan. 2018. Computer network intrusion detection using sequential LSTM neural networks autoencoders. In 2018 26th signal processing and communications applications conference (SIU), Izmir, 1-4.

Sundermeyer, M., H. Ney, and R. Schlüter. 2015. From feedforward to recurrent LSTM neural networks for language modeling. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(3), 517-529.

Xu, X., H. Ge, and S. Li. 2016. An improvement on recurrent neural network by combining convolution neural network and a simple initialization of the weights. In 2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS), Chongqing, 150-154.

Wang, Y., J. Zhou, K. Chen, Y. Wang, and L. Liu. 2017. Water quality prediction method based on LSTM neural network. In 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, 1-5.

Yuhai, G., L. Shuo, and H. Linfeng. 2018. Research on failure prediction using dbn and lstm neural network. In 2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Nara, 1705-1709.