**Logistic regression or Convolutional neural network for classification**

Jerrin Joe Varghese

Department of Data Science

Mercyhurst University

[jvargh81@lakers.mercyhurst.edu](mailto:jvargh81@lakers.mercyhurst.edu)

**Abstract**

Machine learning algorithms are becoming popular and are widely used for giving machines the ability to learn for them self without human intervention. Hence, these algorithms are used for object detection, image classification, stock prediction etc. Some machine learning algorithms are complex and requires more memory and processing power. This paper proposes the use of logistic regression to overcome the problem of memory and processing power, if the data can be turned into a binary classification problem. In order to test this hypothesis, the paper goes through the problem of driver distraction and uses both convolutional neural network (CNN) [8] and logistic regression [10] to analyze the performance of both models on different machines with different memory and processing power.

**1. Introduction**

Machine learning is a type of artificial intelligence technique that learns to identify new pattern in data. This technique is now widely used in several industries for various tasks. There are different types of machine learning algorithms such as supervised learning, unsupervised learning, and reinforcement learning. So, it’s important to know which type of machine learning algorithm is best suited for a machine learning problem. Supervised learning is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances. In other words, the goal of supervised learning is to build a concise model of the distribution of class labels in terms of predictive features [1]. We provide the machine with data that is already labeled and it’s not learning own its own, we call this type of technique as supervised learning. Supervised learning can be further divided into two: classification, where the output is in the form of categories and regression, where the output is in the form of real values [3]. Examples of supervised learning is Linear Regression, Logistic Regression, CART, Naive Bayes, KNN etc. Unsupervised on the other hand, studies how systems can learn to represent input patterns in a way that reflects the statistical structure of the overall collection of input patterns [2]. In this type of machine learning, the machine simply receives as data, but obtains neither supervised target outputs nor any rewards from its environment [4]. The machine learns patterns that it feels are present in the data. Unsupervised learning can be further subdivided into three sub categories such as clustering, association and dimensionality reduction. Examples of unsupervised learning are K-means, PCA etc. Finally, reinforcement [6] learning is a type of machine learning that helps the machine learn by helping it decide the next action by rewarding it. This methodology is typically used in robotics where the machine learns by trial and error technique. One of the most recent use of reinforcement learning is its use in Deep mind’s StarCraft project [5]. Where rewards are given for learning based on the score obtained from the StarCraft II engine against the built-in computer opponent. This paper focuses on supervised learning and whether we can use logistic regression to solve machine learning problems that could be converted to binary classification. The paper is looking into the binary classification problem because it relies on the premise: if any problem can be converted to a binary classifier and gives us a result that is close to deep learning or convolution neural networks with Logistic regression, then we could save time and solve the machine learning problems faster and utilize computers with low memory and computer processing power. Hence, driver distraction [7] is the problem used in the paper as it can be converted to a binary classification, where the classes are divided into distracted and undistracted drivers. This machine learning problem is solved using convolutional neural network and logistic regression [10]; therefore, developed three models, one with binary classification using logistic regression, the second multiple models of logistic regression and multi-class using convolutional neural network [8] and multi-class logistic regression [9]. All the models are available on Kaggle for reference [11,12,13].

**2. Dataset**

The dataset is collected from Kaggle’s State Farm dataset [14]. Figure 1 shows two examples of the dataset. The dataset consists of 10 classes such as safe driving, texting–right, talking on the phone–right, texting–left, talking on the phone–left, operating the radio, drinking, reaching behind, hair and makeup, and talking to passenger [14]. The dataset consists of 22424 labeled data for training and 79726 data for validation. The shape of each image is of 480 x 640.

A person driving a car

Description automatically generated A picture containing car, person, car seat, transport

Description automatically generated (a) (b)

Figure 1. Example of a driver distraction dataset: (a) safe driving (b) dangerous driving- texting.

**3. Relevant Work**

Convolutional neural network is traditionally used for image processing as they extract features by convolving the images and extracting useful information. Logistic regression is another machine learning algorithm which is widely used for binary classification. The paper goes through both algorithms to understand whether a binary classification problem or the one that can be converted to a binary classification problem needs CNN as it requires more memory and processing power.

**3.1 Convolution Neural Network (CNN)**

Convolution Neural Network [8], consist of an input layer, hidden layer, and an output layer. Some of these layers in the network are: Convolution [15], Activation [16], Pooling [17], Dropout [18], Dense, and SoftMax [19].

The Convolution layer consists a set of filters, where each filter can be considered as a small square that extends through the full depth of the input volume. During each pass, the filter convolves across the width and height of the input, which results in a 2-d activation map that gives the responses of that filter at every spatial position. To avoid over-fitting, pooling layers are used to apply non-linear down sampling on the activation maps. It means that, this layer is aggressive at discarding information, but can be useful if used appropriately. Dropout layers also help to reduce over-fitting by randomly ignoring certain activation functions, while dense layers are fully connected layers and often come at the end of the Neural Network. The output of the layers of the neural network are processed using an activation function, which is a node that is added to the hidden layers and output layers. You’ll often find that the RELU activation [16] function is used in hidden layers, while the final layer typically consists of a SoftMax activation function. The idea is that by stacking layers of linear and non-linear functions, we can detect a large range of patterns and accurately predict a label for a given image. SoftMax is often found in the final layer which acts as basically a normalizer and produces a discrete probability distribution vector. Because of these benefits, CNN is most widely used in image classification or problems related to images.

**3.1.1 Pooling**

The pooling layer reduces the spatial dimensions of the input and the computational complexity of our model. Pooling also helps in controlling the overfitting problem, as it operates on every slice independently. There are different functions such as Max pooling, average pooling or L2-norm pooling. Max pooling is the most used type of pooling that takes the most important part from each slice of the input data.

**3.1.2 Rectified Linear Unit (Relu)**

Relu is an activation function that simply outputs 0 when x < 0, and conversely, it outputs a linear function when x ≥ 0 [16].

f (x) = max (0, x)

**3.1.3 Dropout**

Dropout is one of the most effective regularization that is used in a neural network. Using dropout helps us to randomly keep only a neuron active with some probability ‘p’. This helps it to force the network to be accurate even if some information is not present, which in turn helps the network not to be dependent on any one neuron.

**3.1.4 Fully Connected Layer**

In a fully connected layer, every neuron in one layer is connected to every neuron in the other layer. The last fully connected layer is the SoftMax activation function that classifies based on the generated features from the trained data.

**3.2 Logistic Regression**

Logistic regression is a binary classification statistical machine learning model. The logistic regression is a sigmoid function, which takes any real input and outputs a value between 1 and 0 [21]. The sigmoid function is given by the formula:

Sigmoid(x) = 1/ (1 + )

**4.** **Proposed Solution**

This paper uses three model to understand whether it is possible to use logistic regression for machine learning problems that can be converted to binary classification and get the same results.

**4.1 Pre-processing**

This paper also focuses on different algorithms and how they perform with the driver distraction problem. The problem tackled here is to understand how each algorithm is different in their prediction and will this analysis help in understanding, if even logistic regression can play vital role in problems like driver distraction. The image data from the dataset is split into training and validation set, where the training set consists of images of size 240 x 240 and the image class number. The training set is then split into features and labels. The features are then converted to a 4d array using NumPy array. The data is further used by the models for training.

**4.2 Convolutional neural net model**

The convolutional neural network uses Keras’s sequential model [11] and is divided into three convolutional groups. Each group consists of two convolutional layer of filter size 32,64 and 128 and kernel of 3x3. The convolutional layer also uses zero padding and Relu as the activation layer. The convolutional layer is followed by batch normalization to normalize the data and each group of convolutional layers is added with a max pooling layer and dropout layer at the end.

The convolutional layers are flattened and added to the fully connected layer. The fully connected layer consists of three dense layers with 512,128 and 10 neurons respectively. The loss function used is categorical cross entropy and optimizer as Adam.

**4.3 Logistic Regression model**

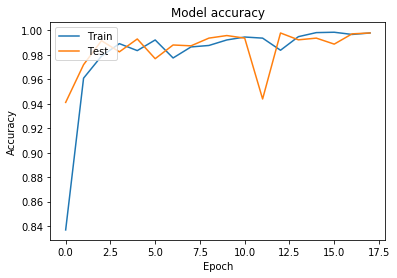
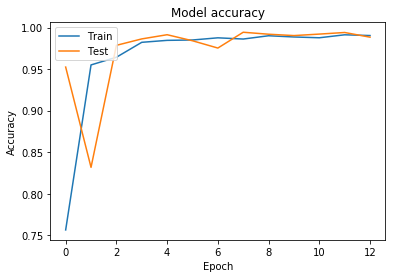
In the paper Logistic regression uses Keras’s sequential model [13] and it uses one batch normalization and one dense layer with cross entropy as the loss function and Adam as optimizer. The model also uses early stopping to avoid overfitting.The model splits the data into nine different groups, where each group has a good driver and bad driver (for each distraction) combination. The data is then trained to predict if the driver is good or bad. This information can be further processed and amalgamated to get similar output as the convolution neural network.

The second logistic regression model [12] splits the data into two classes i.e. good driver and bad driver. Where the bad driver data is a combination of all the distracted classes matching the class size of the good driver. The model uses the Keras’s sequential model and configuration like the previous logistic model. The model is then trained to predict different images such as good or bad driver.

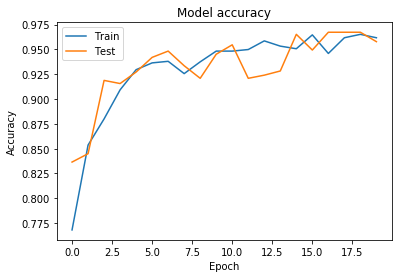
**5. Results**

This paper compares the r squared results, confusion matrix and accuracy to understand whether logistic regression can be used for problems that can be converted to binary classification problem instead of CNN. The result also focuses on the time required by the models to train, as the computers with less CPU power and memory are likely to take more time compared to powerful machines. The r squared value for the CNN model and the two logistic models are shown in table 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model Name | GPU Used | R square | Accuracy | Time consumed |
| CNN | Yes | 1.0 | 0.99 | 1.5 hours |
| CNN | No | 1.0 | 0.99 | 40 hours |
| Individual Logistic model (average of all models) | No | 0.97 | 0.99 | 20 mins |
| Logistic model | No | 0.837 | 0.89 | 10 mins |

Table 1 : R squared values of each model.

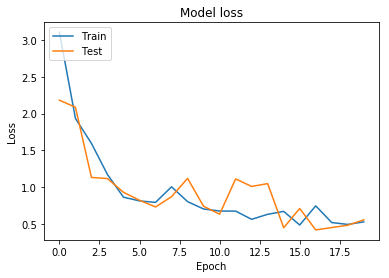
(a) (b)



**(c)**

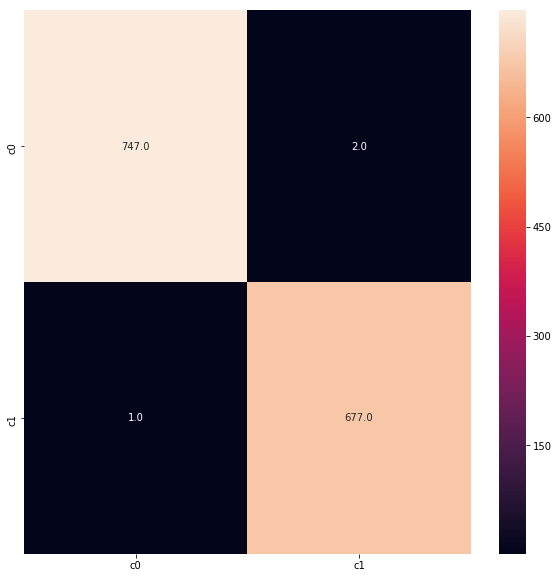
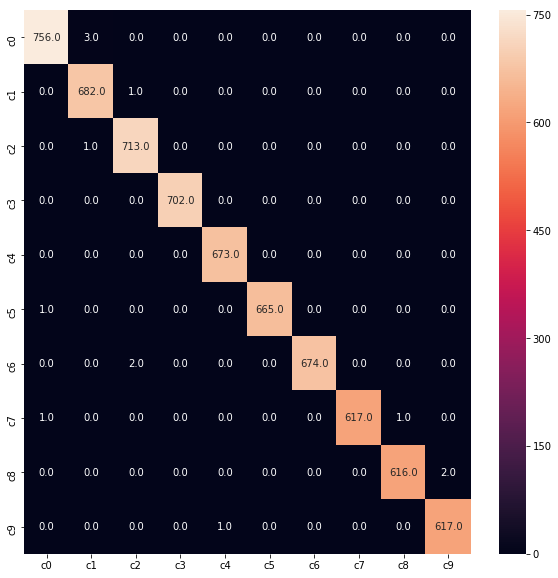
Figure 2. Accuracy graph for all models: (a) Convolutional neural network (b) Individual Logistic regression model (c) Logistic regression model.

1. **(b)**

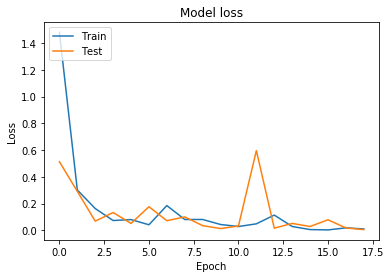
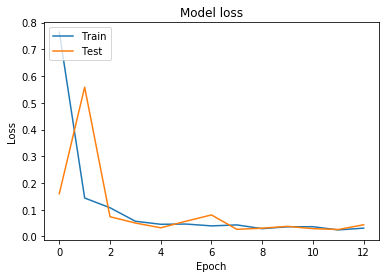


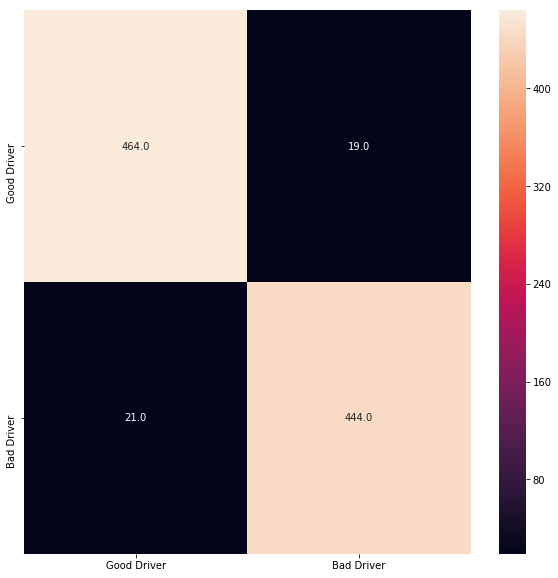
**(c)**

Figure 3. Loss graph for all models: (a) Convolutional neural network (b) Individual Logistic regression model (c) Logistic regression model.



1. **(b)**





**(c)**

Figure 4. Confusion matrix graph for all models : (a) Convolutional neural network (b) Individual Logistic regression model (c) Logistic regression model.

The figures 2,3 and 4 shows the accuracy, loss and confusion matrix for each model. The table 1 also shows the time taken for training each model. The results show that the values are approximately equal, and the time required by logistic regression is less compared to CNN. Hence, we can also use logistic regression when the data can be used as binary classification machine learning and when the memory and CPU power is less.

**6. Conclusion and Future Work**

Memory and processing power have been major issues for machine learning models. The paper compares logistic regression models and convolutional neural network in order to understand whether it is possible to replace convolutional neural network. This is because Convolutional neural network needs more processing power when compared to logistic regression. Our result shows that the logistic regression also gives us similar results to convolutional neural network, which is promising. The future work will include to use more complex machine learning problems, where the data is more complicated with more diverse images to see if the results are the same.

**References**

[1] S. B. Kotsiantis, *Supervised Machine Learning: A Review of Classification Techniques.*

[2] Appeared in Wilson, RA & Keil, F, *editors. The MIT Encyclopedia of the Cognitive Sciences.*

[3] Elnaz Barshana, Ali Ghodsib, Zohreh Azimifara, Mansoor Zolghadri Jahromi*, Supervised Principal Component Analysis: Visualization, Classification and Regression on Subspaces and Submanifolds.*

[4] Zoubin Ghahramani, *Unsupervised Learning : University College London, Kzoubin@gatsby.ucl.ac.ukhttp://www.gatsby.ucl.ac.uk/~zoubin.*

[5] Learning Oriol Vinyals Timo Ewalds Sergey Bartunov Petko Georgiev Alexander Sasha Vezhnevets Michelle Yeo Alireza Makhzani Heinrich K¨ uttler John Agapiou Julian Schrittwieser Stephen Gaffney Stig Petersen Karen Simonyan Tom Schaul Hado van Hasselt David Silver Timothy Lillicrap DeepMind Kevin Calderone Paul Keet Anthony Brunasso David Lawrence Anders Ekermo Jacob Repp Rodney Tsing Blizzard, *StarCraft II: A New Challenge for Reinforcement.*

[6] Marc G Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling, *The Arcade Learning Environment: An evaluation platform for general agents: J. Artif. Intell. Res.(JAIR), 47:253–279, 2013.*

[7] *Understanding the distracted brain why driving while using hands-freecell phones is risky behavior National Safety Council White Paper April 2012.*

[8] Wenling Shang, Kihyuk Sohn, Diogo Almeida, Honglak Lee*,* Understanding *and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units : University of Michigan, Ann Arbor;NEC Laboratories America; Enlitic; Oculus VR.*

[9] Peter Karsmakers, Kristiaan Packman’s, Johan A.K. Suykens Multi*-class kernel logistic regression: a fixed-size implementation.*

[10] *Applied logistic regression second edition David W. Hosmer university of Massachusetts Amherst, Massachusetts Stanley Lemeshow The ohio state University Columbus, Ohio.*

[11] <https://www.kaggle.com/jerrinv/driver-distraction> *Dated : May 01 2019*

[12] <https://www.kaggle.com/jerrinv/logistic-regression> *Dated : May 01 2019*

[13] [*https://www.kaggle.com/jerrinv/driver-distraction-using-logistic-regression*](https://www.kaggle.com/jerrinv/driver-distraction-using-logistic-regression)*. Dated : May 01 2019*

[14] [*https://www.kaggle.com/c/state-farm-distracted-driver-detection/data*](https://www.kaggle.com/c/state-farm-distracted-driver-detection/data)*. Dated : May 01 2019*

[15] Relation Classification via Convolutional Deep Neural Network Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou and Jun Zhao*, National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences 95 Zhongguancun East Road, Beijing 100190, China {djzeng,kliu,swlai,gyzhou,jzhao}@nlpr.ia.ac.cn*

[16] *Abien Fred M. Agarap, Deep Learning using Rectified Linear Units (ReLU)*

[17] Bonn, Germany scherer*, Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition Dominik Scherer, Andreas Muller, and Sven Behnke University of Bonn, Institute of Computer Science VI, Autonomous Intelligent Systems Group, Roberts. 164, 53117*

[18] *Dropout: A Simple Way to Prevent Neural Networks from Overtting Nitish Srivastava nitish@cs.toronto.edu Georey Hinton hinton@cs.toronto.edu Alex Krizhevsky kriz@cs.toronto.edu Ilya Sutskever ilya@cs.toronto.edu Ruslan Salakhutdinov rsalakhu@cs.toronto.edu Department of Computer Science University of Toronto 10 Kings College Road, Rm 3302 Toronto, Ontario, M5S 3G4, Canada*.

[19] *Multi-Category Classification by Soft-Max Combination of Binary Classifiers Kaibo Duan1, S. Sathiya Keerthi1, Wei Chu1, Shirish Krishnaj Shevade2, and Aun Neow Poo1 1 Control Division, Department of Mechanical Engineering National University of Singapore, Singapore 119260 fengp9286, mpessk, engp9354, mpepooang@nus.edu.sg 2 Department of Computer Science and Automation Indian Institute of Science, Bangalore 560*

[20] *Measures of Fit for Logistic Regression Paul D. Allison, Statistical Horizons LLC and the University of Pennsylvania*

[21] *wikipedia : https://en.wikipedia.org/wiki/Logistic\_regression#Definition\_of\_the\_logistic\_function*