Comparison of different classification techniques to determine a person's intention of whether an image is manipulated or not based on eye gaze tracking of a person.

Saswat Panda
College of Engineering and Computer Science
The Australian National Univeristy

Abstract. The aim of this paper is to determine a participant's intention, which is determined by the predicting what will be the participant's vote in determining whether an image is manipulated or unmanipulated image. This is done using the eye-gaze data of the participant. To obtain the result three major classification techniques, i.e., Decision tree classification, Deep Learning classification and Maximum Likelihood classification were trained and tested on the dataset, then the models were compared using different parameters. Finally, it was found that Deep Learning outperformed the other two models by achieving maximum train accuracy of 97.6, and average test accuracy of 100%.

Keywords: Eye Gaze; Decision Trees; Neural Networks; Maximum Likelihood; Classification

1 Introduction

The mind's reflector might be thought of as the eyes. As a result, data on eye movement has been used in a variety of applications. It has been used in teleoperation systems, or robotics grasping, as an indirect input to control a robot, for example, to determine the user's mental overload when performing sensitive activities, in shared autonomy systems to perform cooperative tasks more efficiently, and in teleoperation systems, or robotics grasping [1]. In this research I predict what the person will vote, and thereby predicting the person's intention of what he is going to vote.

The reason why Decision Tress are used for this research is that decision tree building requires no domain knowledge or parameter configuration, it is ideal for exploratory knowledge discovery, also, Multidimensional data can be handled via decision trees. Maximum Likelihood classification is used because it is simple and quick to forecast the test data set's class, and it works much better when the assumption of independence is maintained. Deep Learning is used in this research since it involves Time Series data, Trainable filters are used to capture spatial and temporal patterns, and trainable weights are used to assign emphasis to these patterns. The use of above techniques for classification were inspired from Milne et al (1995) [2].

In this research the vote of the participant is predicted, which can be either manipulated, unmanipulated or confused. The order followed in this research is, firstly data is pre-processed, then the important parameters are used to predict vote of the participant using Decision Tree, Maximum Likelihood and finally Deep Learning. At last, comparison on which model is the best for classification of manipulated and un-manipulated images is done.

2 Dataset description and preprocessing

This research uses two datasets taken from Caldwell et al (2015) [3]. The first dataset is the dataset containing eye data and the second one is a time-series based dataset. The description of the dataset is clearly presented in Table 1.

Table 1. Description of important the columns of the dataset.

Name of column	Description
X Pos	X position of the eye gaze of the participant, for a particular fixation.
Y Pos	Y position of the eye gaze of the participant, for a particular fixation.
Start Time	It is the start time when participant start looking at the image, for a fixation.
Stop Time	It is the stop time when participant stops looking at the image, for a fixation.
Duration	It is the duration of the fixation, basically difference between Start and Stop Time.
Samples in Fixation	Number of samples present for a particular fixation.
num_fixs	Total Number of Fixations made by the participant for a given image.
fixs_duration	Total amount of time the participant spent looking at a given image (in seconds).
num_man_fixs	The participant's total number of fixations when staring within the target region, for a given image.
man_fixs_dur	Total time spent looking within the target region (in seconds) by the participant, for a given image.

Image Manipulated	Whether the image is actually manipulated or not. 0 is unmanipulated, whereas 1 is unmanipulated.
Vote	Participant's opinion on whether image is manipulated or not. 0 is unmanipulated, 1 is unmanipulated, and 2 is confused.

In this research, first it was ensured that there were no null values present in both the datasets. The next step was to merge the two datasets. This was done by keeping 'Image Id' and 'Participant id' as primary key, and treating the two datasets as two tables, and finally joining both of them. Since, for participant_Id = 78 and Image_Id = 14 data whether the image is manipulated or not is not present, so those rows were dropped. The next step was to find the correlation between all the parameters with vote, and plotting a heat map. The participant id column, though it shows comparatively better correlation, was dropped because it does not give any information about the participant like his/her age, profession which might have been useful, rather it is just a number. Including participant id will result in better accuracy of models, but it will cause overfitting. For the same reason, other id columns were also dropped.

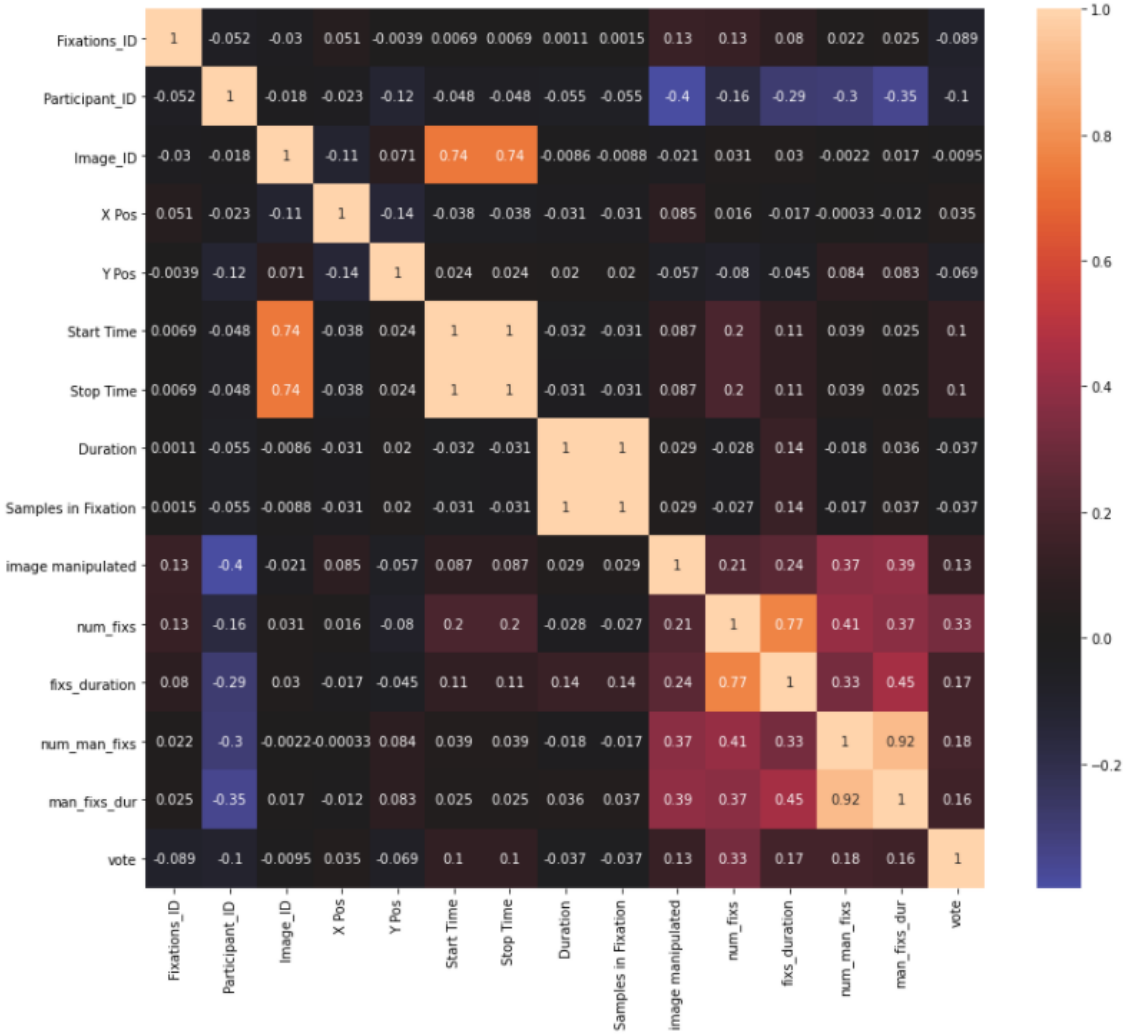

Fig. 1. Heat map showing the correlations between different parameters.

The data was then divided into input features and output label. The input features contained the important parameter having a correlation value of at least 0.06 with the output label(vote). So, the input features found are *Y Pos, Start Time, Stop Time, num_fixs, fixs_duration, num_man_fixs, man_fixs_dur,* and *Image Manipulated*. Though Start and Stop Time have a correlation of 1, including one without other will not be meaningful. So, it can be noted that apart from the eye-gaze data, the data of whether the image has been actually manipulated or not is used to predict the output label. Then, it was split into train and test, with the *test set containing 20%* of the dataset (total rows of the dataset is 30,794). Before applying the classification techniques, Min-Max Normalization was performed on the dataset. Basically, it is applied to all the features, and in each feature minimum value is converted to a 0, the maximum value is converted to a 1, and all other values are converted to a decimal between 0 and 1. It is done in order to convert the values of numeric columns in a dataset to a similar scale without distorting the ranges of values.

3 Decision Tree Classification

One of the predictive modelling methods used in analytics, data mining, and machine learning is decision tree learning. It goes from assumptions about an item to conclusions about the item's target value using a decision tree. Classification trees are tree models in which the target variable may take a distinct collection of values. In these tree structures, leaves represent class labels and branches represent feature combinations that lead to certain class labels. In this research sckit library has been used to implement decision tree classification. After successful computation the mean avg. precision was found to be 1.00, avg. recall 1.00 and the accuracy was found to be 1.00. A 100% accuracy indicates that the model is overfitting and hence, not suitable for this classification problem.

4 Maximum Likelihood Classification

It is based on Maximum Likelihood Estimation (MLE) which is method of optimizing a likelihood function to estimate the parameters of a probability distribution such that the observed data is most likely under the assumed statistical model. MLE is a special case of maximal a posteriori estimation from the perspective of Bayesian inference. Therefore, in order to implement it Naïve Baeyes Classifier from scikit has been used, which uses maximum likelihood for classification. After successful computation the mean precision was found to be 0.675, mean recall 0.64 and the test accuracy was found to be 0.61. Considering the dataset size and correlation values the accuracy can be better than this. So, this model can be considered as an average model for this classification problem.

5 Deep Learning Classification

A neural network with a single input layer, three hidden layer and one output layer has been used, since, the number of hidden layers is greater than three it is termed as deep learning. Three hidden layers has been used, since, on increasing or decreasing the number of layers, it was found that there was no significant change in the test accuracy, and three hidden layers gave the best test accuracy. By Experimenting with different activation function, like ReLU, PReLU[4], and Swish[5]. The dying out problem of ReLU when less features were used motivated to use PreLU and Swish. After comparing the activation functions, the best test accuracy was found for PreLu, without overfitting and using just 100 epochs. Dropout has been set to 0.1 to control overfitting, which it does by randomly removing some neurons. Adam optimizer has been used, especially to fix the initial bias. The learning rate used was 0.001, on increasing it the accuracy curve produced more noise, and decreasing it further no significant effect was found. 100 epochs were used to train the data with a batch size of 16. Since, in practice, it has been observed that when employing a larger batch, the model's quality, as evaluated by its ability to generalise, suffers significantly[6]. The maximum train accuracy found was 97.6% and the minimum loss was found to be 0.055. From the graphs below, it can be observed that there is a gradual and smooth increase in train accuracy, and finally flattening(Fig.3), whereas a gradual and smooth decrease in test accuracy, and finally flattening(Fig. 2). This suggests that there is no overfitting. When the model was evaluated the average test accuracy was found to be 100%, which is very good(Table 2) considering the fact that there is less variance in the dataset for certain features.

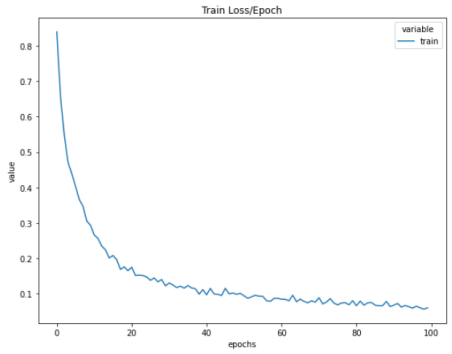

Fig. 2. Loss decreases with increase in epochs.

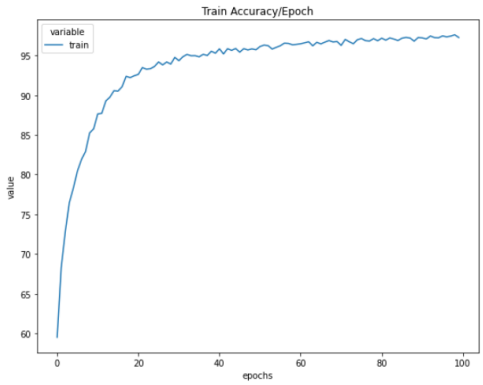

Fig. 3. Accuracy increases with increase in epochs.

Table 2. Comparison of average test scores for the three techniques of classification. Mean of Precision and Recall are taken.

Model	Precision	Recall	Accuracy
Decision Tree Classifier	100%	100%	100% (overfit)
Maximum Likelihood Classifier	51%	52%	61%
Deep Learning Classifier	99%	99%	100%

6 Conclusion and future work

Hence the purpose of the paper of predicting manipulated image using the three techniques of classification has been achieved. Deep Learning was successful in beating Decision Trees at not overfitting, and achieving very good test scores (Table 2). It has achieved an average test accuracy of 100% (maximum train accuracy 97.6%). Though the results are good overfitting should be removed properly. It can be done by introducing new more correlated features, which vary greatly with the fixation_id. Merging of the dataset resulted in less variance in data with respect to fixation_id. Those features should also be recorded in the Time series dataset.

References

[1] F. Koochaki and L. Najafizadeh, "Predicting Intention Through Eye Gaze Patterns," 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2018, pp. 1-4.

[2] L. &. G. T. &. S. A. Milne, "Classifying Dry Sclerophyll Forest from Augmented Satellite Data: Comparing Neural Network, Decision Tree & Maximum Likelihood.," in *Australian Conference on Neural Networks*, 1995.

[3] T. G. R. J. L. C. Sabrina Caldwell, "Imperfect Understandings: A Grounded Theory And Eye Gaze Investigation Of Human Perceptions Of Manipulated And Unmanipulated Digital Images," in *Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science (EECSS 2015)*, Barcelona, 2015.

[4] K. He, X. Zhang, S. Ren and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1026-1034.

[5] Ramachandran, Prajit & Zoph, Barret & Le, Quoc. (2017). “Swish: a Self-Gated Activation Function".

[6] N S Keskar, D Mudigere, J Nocedal, M Smelyanskiy & P T P Tang, “On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima,” 2016.