Interactive Visual Explorer for Error Detection in Machine Learning Models

Author: Yujia Zhang | Faculty Mentor: Dr. Enrico Bertini | Other Mentor: Jun Yuan
Department of Computer Science and Engineering

Background

With today's machine learning models, machine learning is capable for making relatively accurate decisions. However, in order to help domain experts to validate model decisions and end users to trust and use the model, it becomes more important for model developers to further understand the model's behaviors and then improve the model. A statistical overview is usually a part of the model diagnosis[1]. By checking the prediction score distribution, projection and the confusion matrix, model developers can spot potential problems of the dataset. Comparing different model's statistical overviews can also help with further interpretations of the models' predictions.

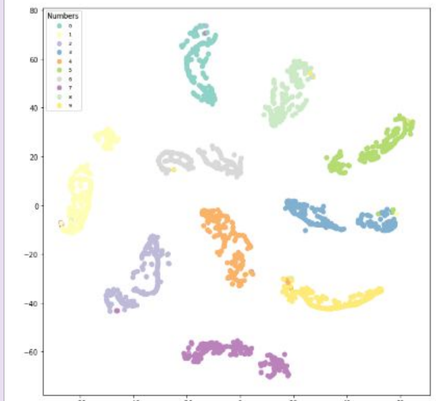


Figure 1: t-sne projection of pretrained model using CNN with 3 Conv2D layers

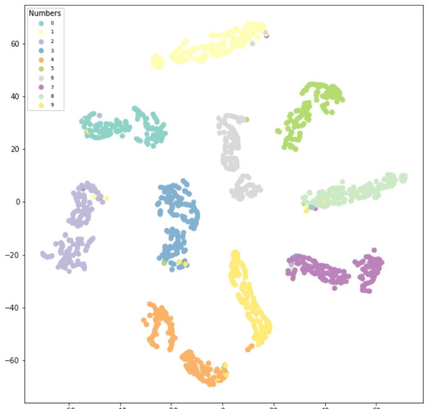


Figure 2: t-sne projection of pretrained model using CNN with 1 Conv2D layer

The figure 1 model has a prediction score of 99.2% and the figure 2 model has a prediction score of 95.6%. From the projection, there are less misclassified dots in figure 1 than figure 2

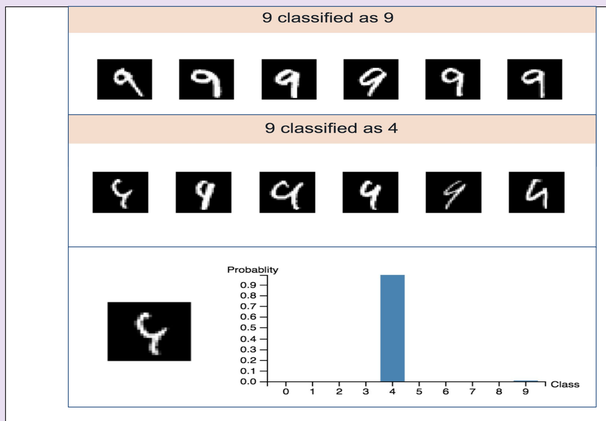


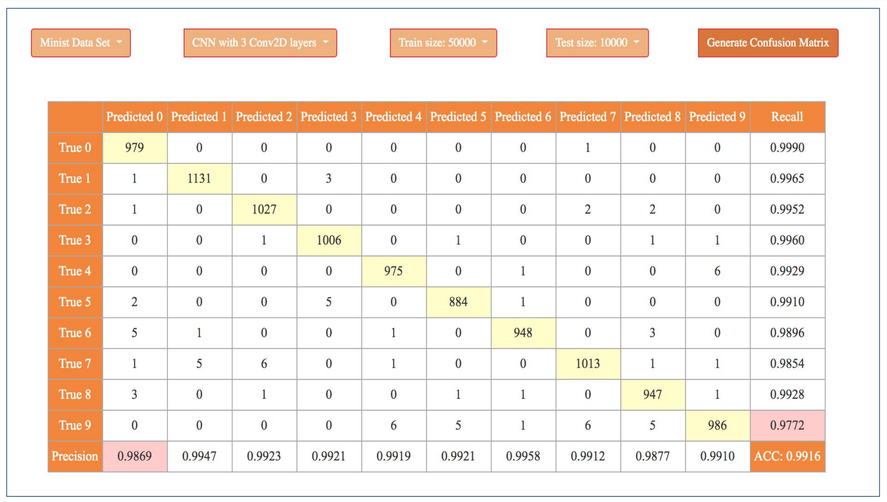
Figure 4: Selected data view from Figure 3 's confusion matrix

Clicking the corresponding block in figure 3 will show a closer view of correctly predicted data and wrongly predicted data. A click on individual data can show a histogram of probability of prediction. (In this case, the wrongly predicted 9's probability of being 4 is 99.8% while being 9 is 0.02%.) By observation, this particular data is misleading and can be considered to be removed from the dataset.

Abstract

The project is intended to build an interactive web application for users to compare prediction results from different models and detect potential model errors by visualizing the results. The visualization includes t-SNE, PCA projection, confusion matrix and some further detailed view on selected class or item.

The project started with MNIST dataset and pretrained models based on convolutional neural network with different architectures to visualize different model's accuracy and performance on image classification.



	Predicted 0	Predicted 1	Predicted 2	Predicted 3	Predicted 4	Predicted 5	Predicted 6	Predicted 7	Predicted 8	Predicted 9	Recall
True 0	979	0	0	0	0	0	0	1	0	0	0.9990
True 1	1	1131	0	3	0	0	0	0	0	0	0.9965
True 2	1	0	1027	0	0	0	0	2	2	0	0.9952
True 3	0	0	1	1006	0	1	0	0	1	1	0.9960
True 4	0	0	0	0	975	0	1	0	0	6	0.9929
True 5	2	0	0	5	0	884	1	0	0	0	0.9910
True 6	5	1	0	0	1	0	948	0	3	0	0.9896
True 7	1	5	6	0	1	0	0	1013	1	1	0.9854
True 8	3	0	1	0	0	1	1	0	947	1	0.9928
True 9	0	0	0	0	6	5	1	6	5	986	0.9772
Precision	0.9869	0.9947	0.9923	0.9921	0.9919	0.9921	0.9958	0.9912	0.9877	0.9910	ACC: 0.9916

Figure 3: Confusion matrix for figure 1 model generated by data explorer

Yellow blocks highlight True Positive. Pink blocks highlight the class with the lowest precision / recall.

Future work

Further work on the project includes expanding it to other datasets and including more visualizations on the dataset 's features. Developing an algorithm to compute and display visualizations faster is also a part to work on.

Reference

[1] Josua Krause, Aritra Dasgupta, Jordan Swartz, Yindalon Aphinyanaphongs,Enrico Bertin. 2017. A Workflow for Visual Diagnostics of Binary Classifiers using Instance-Level Explanations at the IEEE Conference on Visual Analytics Science and Technology (IEEE VAST 2017). 1-5

[2] Josua Krause, Adam Perer, Kenney Ng. 2016. Interacting with Predictions: Visual Inspection of Black-box Machine Learning Models in ACM CHI 2016 1-5

Acknowledgement

The authors thank NYU Tandon School of Engineering's Office of Undergraduate Academics for generous funding of the project. Thanks can go to others (Thompson-Bartlett, for instance).