Name: Jason Zhou
Mentor: Dr. Dongjin Song
Status Report #: 5
Time Spent on Research This Week: 5.5
Cumulative Time Spent on Research: 21.5
Miles Traveled to/from Mentor This Week: 0
Cumulative Miles Traveled to/from Mentor: 0
==

Wednesday, October 6th: (0.5 Hours)

I spent this time reviewing concepts I had previously learned. For this day, I focused on reviewing tensors, which are essentially just values that hold data. For example, it can be a number, array, or 2D array. These are all computer science terms; however, the mathematical equivalent would be scalars, vectors, and matrixes, respectively (Different names, same meaning).

Additionally, I also reviewed ideas like ranks and shapes. A rank is just the dimension of the tensor (1D, 2D, 3D, etc). Also, a rank is capable of going above the 3rd dimension (4D, 5D, 6D, etc). A shape is how the tensor is visualized. For example, I can have two 2D arrays; however, their shapes can be different. One can have 2 rows and 3 columns (2 by 3), while the other could have 5 rows and 2 columns (5 by 2).

Friday, October 8th: (1.5 Hours)

Similar to Wednesday, I spent my first 30 minutes reviewing concepts I had already learned. In particular, I focused on classification and the code I would need to know to replicate the model.

After this, I spent time practicing what I had reviewed. While watching the Tensorflow video course, I was exposed to some practice; however, it was not enough to make me feel comfortable with it. Essentially, I put more time into creating dense and convolutional neural network models. I also worked with reshaping tensors. For example, I might take a 2x3 array and restructure it to be a 6x1, so not only did I change the shape, but I also changed the dimension (2D to 1D). As practical skills will be important later on, I will probably put more time into practicing machine learning later on.

Additionally, my mentor had a personal issue to attend to, so we had to reschedule our meeting for Monday, October 11th.

Monday, October 11th: (0.5 Hours)

On this day, I met with my mentor to discuss the research project I would be conducting. Although my mentor and I had talked about my research project, I was a little confused about what exactly I would be doing. Essentially, I will be analyzing the different methods of anomalous sound detection and determining which one is the best method. He also said that if I have more time, I could try to create a new algorithm to create a new, more efficient way of processing audio data and making predictions. By talking with him, I was able to clarify this so that I would be able to do my research gathering with a good idea of what I would be looking for.

Beyond clarifying what the focus of my research project was, he also let me know about some methods of processing audio that already exist such as K-means, gaussian mixture models, DBSCANs, local outlier factors, and many more. I already know what K-means is; however, I do not exactly know what the other methods do, which makes me excited to learn them!

Additionally, he also gave me some tips for initial steps to start the experiments for my research. First, I should load the data and then scan the data. This will allow me to extract the important features of the audio, which will allow the machine learning model to classify the audio.

There was a lot of information exchanged in this Zoom meeting, so I ended up taking notes on it for future reference.

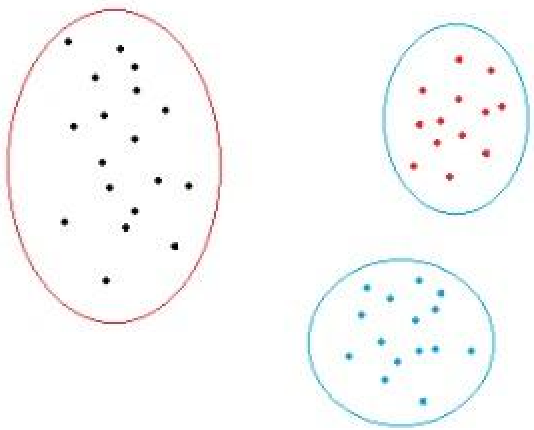
Thursday, October 14th: (1 Hour)

This was the day that I began gathering research articles to begin the first step of my research proposal. I began looking for articles related to anomalous sound detection. I skimmed the articles, looking at the abstract, intro, and conclusions to determine if the subject was relevant to what I was looking for and took note of the various methods that were used to process the data. If the articles were related to my research proposal, I placed them in my library under the “ARM research” tag. Just from scrolling in Google Scholar, there seems to be a diverse number of methods; however, for this day, I decided to look at two articles in particular. One was an article about a method called AdaCos, which was based around a cosine loss function that helped to classify anomalous sounds within the audio data. Additionally, this article was significant because the method was used on the testing data from the computer science community challenge that my research project is based on. My second article used a method called ArcFace. Although ArceFace is mainly used in image recognition, the algorithm was tested with the same data set as the first article I mentioned. According to the authors, this method was also quite strong in sound recognition tasks and seemed to outperform the baseline method, which I believe is a synonym for the control of an experiment. Also, in this case, the control was a method called AUC, which I have to take the time to learn. This is significant because it seems that ArcFace will be a candidate to consider when trying to determine the most efficient and accurate processing method when conducting my own research.

Although I skimmed the articles, it took me quite a while to go through each one because of the complex language within the text. For now, I believe this was a good start, and I will make time to look in-depth at the articles I have collected at a later date.

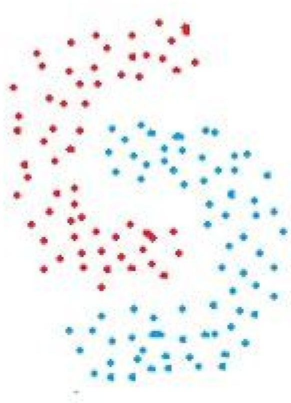
Beyond looking at scholarly articles, I also began looking into the processing methods that I was unfamiliar with, which I mentioned in my October 11th report. I decided to look into DBSCAN, which I later found out stands for density-based spatial clustering of applications with noise. As the name suggests, DBSCAN is a type of clustering that is similar to K-means.

Normally, for classifying data points that form groups or clusters, one would use K-means.



(This is a picture of a K-Means algorithm being applied to data. The circles mean that the algorithm recognizes that the data points within the shape can be classified or grouped together)

As one can see, the K-Mean algorithm is able to classify (or circle) data points based on the groups that they make. However, what if one had data that looked like this:



(A picture of data with points that can not easily be classified by the K-Means algorithm)

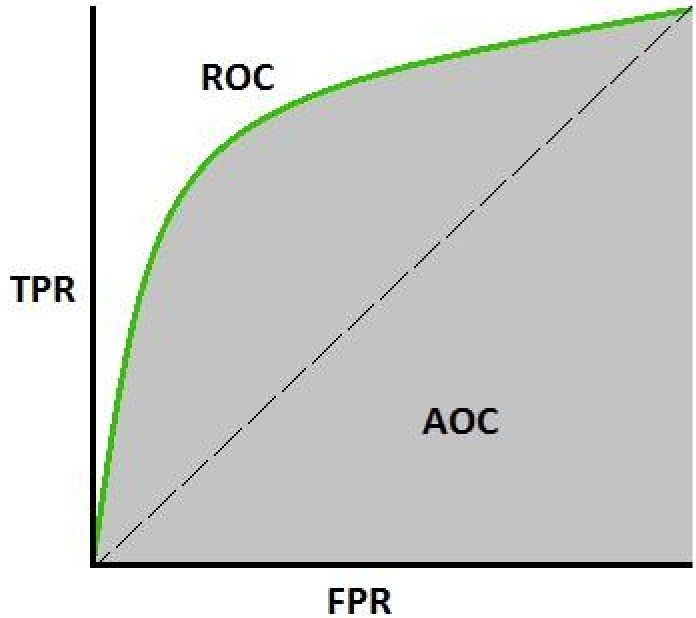
The shape that the red and blue points make it hard to draw a circle around one color without incorporating the other. As such, using K-Means to process and group this data would be inefficient and troublesome. However, a DBSCAN would be able to handle this data more efficiently because it does not rely on grouping through circles. Rather, this algorithm relies on grouping through density. Thus, it will detect areas of red that have the greatest density and classify them that way, which does not group them into a definitive shape as K-Means would.

Saturday, October 16th: (2 Hours)

On this day, I created a tag called “potential articles” in Google Scholar. I skimmed the internet for titles of articles that seemed to be related to anomalous sound detection and stored them in my library under the new tag I had created. Although I did not take a look at these articles, I will view them at a later date.

Looking back to the notes I took on Monday, October 11th, I felt quite overwhelmed. I realized that there is an abundance of various methods used to process sounds, such as autoencoders, long short-term memory networks (LSTM), LSTM-Autoencoders, CovLSTM, ArcFace, AdaCos, etc. I have no doubt that there are more; however, these are just some I discovered while searching for research related to my topic. Although I feel that I have learned a lot, seeing all of these different machine learning models definitely put into perspective just how much I still need to understand. Additionally, as I mentioned in the report on Thursday, many of these articles use vocabulary that I am unfamiliar with. Therefore, I thought it would be a good idea to take a break from gathering research and take a moment to study the vocabulary and concepts that have come up repeatedly during my article search.

For example, one of the most frequent concepts that kept occurring was an AUC-ROC curve. AUC stands for area under curve, while ROC stands for receiver operating characteristics. Essentially, this is a graph that allows one to quantify the performance of a machine learning model, similar to a grade on a test.



(This is an example of an AUC-ROC curve. It is a graph with TPR graphed on the y-axis, and FPR graphed on the x-axis. TPR stands for True Positive Rate, while FPR stands for False Positive Rate)

The image above incorporates vocabulary such as TPR (True Positive Rate) and FPR (False Positive Rate). Simply put, a True Positive is when the machine learning model predicts that a value is positive and happens to be correct. In contrast, a False Positive is when the model predicts something to be positive but is incorrect. Just like there are True Positives and False Positives, there are also True Negatives and False Negatives. Keep in mind that the use of positive and negative only define values that are opposite of one another. For example, instead of positive and negative, one could think of it as 0 and 1 or yes and no (whatever happens to fit the situation).

In learning these terms, I was also introduced to concepts such as recall and precision. Recall is the number of true positives that the model was able to predict versus the total amount of total Trues. For example, this would be the number of patients a doctor diagnosed as having a sickness versus the total number of patients with a sickness (because it is possible that a doctor may not have identified all possible sick patients).

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

(This is the formula for calculating the recall of a machine learning model. Notice that False Negative is included in the denominator. A False Negative is for when a model guesses something is negative but has guessed wrong, meaning that the value is actually positive.)

Precision, on the other hand, is the number of True Positives that the model was able to identify versus the number of true and false positives. For instance, if a doctor diagnosed 10 people to have cancer, but only 6 people truly had cancer, then one would say that his precision was 6 (the number of people who actually have cancer)/ 10 (the number of people he believes to have cancer).

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)}$$

(This is the formula for calculating precision for a machine learning model)

Ultimately, these concepts are used to judge and evaluate the accuracy of a model, which allows computer scientists to make improvements later on. However, the articles I viewed used AUC as a baseline or control. As I am unsure how this was done, I will have to investigate further.

References

deeplizard. (2018, September 11). *Tensors explained - data structures of deep learning* [Video]. YouTube. https://www.youtube.com/watch?v=Csa5R12jYRg

Huilgol, P. (2020, September 4). *Precision vs. recall – an intuitive guide for every machine learning person*. Analytics Vidhya. Retrieved October 16, 2021, from https://www.analyticsvidhya.com/blog/2020/09/precision-recall-machine-learning/#:~:text=The%20recall%20is%20the%20measure,our%20model%2C%20Recall%20%3D%200.86

Narkhede, S. (2018, May 9). *Understanding confusion matrix*. Towards Data Science. Retrieved October 16, 2021, from https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Python Engineer. (2020, September 30). *TensorFlow tutorial 03 - first neural network (training, evaluation & prediction)* [Video]. YouTube. https://www.youtube.com/watch?v=Edhv7-4t0lc&t=1182s

Yildirim, S. (2020, April 22). *DBSCAN clustering — explained*. Towards Data Science. Retrieved October 14, 2021, from https://towardsdatascience.com/dbscan-clustering-explained-97556a2ad556

Zhou, Q. (2020). ArcFace based sound mobilenets for DCASE 2020 task 2.

Wilkinghoff, K. (2021, July). Sub-cluster AdaCos: Learning representations for anomalous sound detection. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.