Name: Lidan TAN 谭荔丹

SID: 12232423

Date: 2023/1/2

PF Report

This report primarily discusses my contribution to the project, as well as an introduction to the techniques covered in the course and utilized for the project.

My Contribution to the Project

The project requires us to distinguish speech samples belonging to 250 individuals. Our group implemented our own data pre-processing procedure on flac files as well as a hierarchical network of experts to perform the classification. My key contributions to the project consist of the following: I was responsible for choosing the model architecture, outlining the division of work, developing the model, integrating the training and testing code for the model, executing the training and testing of the model, and composing the experimental report.

Choosing the Architecture

The first and foremost task for the project is to choose the model or architecture to use. The choice of this step directly affects the final outcome, and requires a great deal of background knowledge. I realized the challenge is too big for a single, straightforward machine learning model after completing an experiment that trains the 250 class classification job and produces an overall accuracy of 50.14%. I led the model selection process by researching and evaluating various options based on their performance. The "Network of Experts" architechtre (Ahmed, 2016) is finally chosen because it is relatively implementable and can effectively break down the problem, which is precisely what is required for this huge classification task. What's more, by breaking down the decision process, this architecture makes the decision process partially interpretable, and thus, debuggable and tunable. I then made a recommendation and presented my rationale to the team, who agreed on the chosen architecture.

Division of Work

I established the duties and responsibilities for each team member based on the selected architecture and produced a project schedule to guarantee the prompt and effective completion of the job. Specifically, I designed the workflow and regulated the input, output, and function for each funciton I asked my teammated to implement, so that they are readily integratable. I also regulated the data preprocessing workflow and the features needed for upcomming classification. The code we submitted has references to related files.

Developing the Model

As I have proposed the architecture, I implemented the base model. The base model is chosen to be a simple feed-forward neural network because I have found related projects on github and discovered such simple structure performs pretty well on speaker identification problems with a small number of classes.

Integrating Code and Training Model

After I received the building blocks my teammates provided, I integrated them. As for training, I first trained all the backbones and the specialties respectively. Then I integrated the models into a complete workflow and fine-tuned the parameters of the complete workflow. Finally, I used the complete workflow to make predictions on the test dataset.

Composing Report

Since I have the best knowledge of the workflow in our group, I completed the major part of our report, except for the parts concerning data pre-processing, which I left for my teammate who is responsible for data pre-processing.

To sum up, I contributed to the team's overall success by demonstrating strong technical skills, problem-solving ability, and leadership. I also actively participated in discussions, provided constructive feedback, and collaborated effectively with my teammates.

Techniques Utilized for the Project

The techniques covered in the course and utilized for the project include the k-means algorithm, feed-forward neural networks, and tree structures.

k-Means

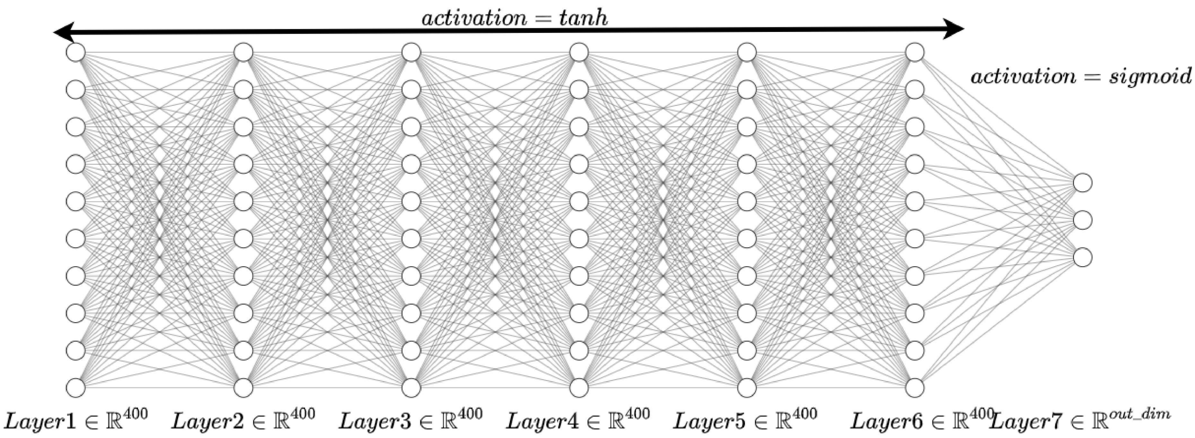
The k-Means algorithm is an clustering algorithm what is widely used in fields include data analyzing and processing, marketing, medical services, and risk management(Ikotun, 2023). It is an unsupervised clustering algorithm which is utilized in the project to divide the 250 classes of speakers of the training data into clusters named "specialties". These "specialties" are collections of classes that are similar in features. The "network of specialist" architecture divides the HUGE classification problem with 250 classes into about a dozen of small problems, each of them is a small classification problem with around a dozen of classes(Ahmed, 2016). The algorithm is shown in the pseudo code below(Sklearn, 2021):

```
function K_means(X, k)
  choose k samples from X as initial centroids
  old_centroids = empty list
  new_centroids = centroids

  while diff(old_centroids, new_centroids) > threshold
    for each sample in X
      assign sample to nearest centroid
    end for

    old_centroids = new_centroids
    new_centroids = update_centroids(X, old_centroids)
  end while

  return centroids
end function

function update_centroids(samples, old_centroids)
  new_centroids = empty list

  for each centroid in old_centroids
    samples_in_cluster = filter samples by samples assigned to centroid
```

```
    mean_of_samples = calculate mean of samples_in_cluster
    add mean_of_samples to new_centroids
  end for

  return new_centroids
end function
```

Feed-Forward Neural Network

A sort of machine learning model that derives from the biological brain system is the feed-forward neural network. It is composed of neurons that imitate the activation process of real neurons using activation functions and output values to the next layer of the neural network, taking the weighted sum of the preceding layer as input (Sazl, 2006). Back-propagation is used to train the feed-forward neural network, which employs the chain rule to iteratively calculate the gradient for each parameter in each layer (Sazl, 2006).



The base network, which is the shared neural network architecture between the backbone and the specialists, consists of feed-forward neural networks with seven fully connected layers and 400 nodes per layer overall, with the exception of the output layer, which has nodes corresponding to the target classification's number of classes. Except for the output layer, which employs softmax as its activation function, each layer's activation function is tanh. Depending on whether they are backbones or specialists, the loss of each base network is the cross entropy loss between the desired specialty label or class label and the projected specialty label or class label.

Tree Structure

A tree model is a process that uses hierarchical decisions to translate a set of characteristics to a decision. It consists of leaf nodes, which stand in for the decisions, and internal nodes, which represent decisions and perform data splitting depending on various criteria. Variance, entropy, the Gini index, classification error, information gain, gain ratio, and other common heuristical criteria can all be utilized to build a tree (Song, 2015). The tree can be trimmed when construction is complete to make it smaller and function better.

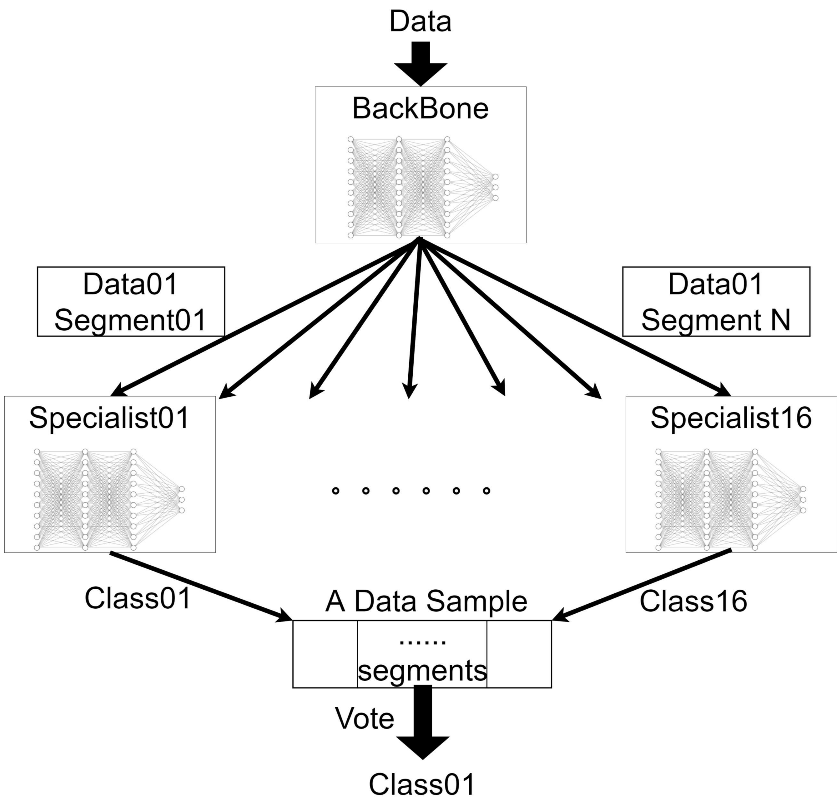


Figure 1

However, we don't expressly employ the tree model in this situation. We only employ the idea of "hierarchical decisions" instead. The project's architecture, as is shown in Figure 1, consists of two hierarchical decision phases. The first step is to divide samples into "specialties" where feature classes are more similar. This divides the 250-class classification challenge into smaller classification tasks. Specifically, $\lfloor\sqrt{250}\rfloor = 16$. It implies that the 250 class classification issue is split into about 16 problems, each of which is a 16 class classification problem. The accuracy of using the base model to classify 250 classes directly is 50.14%. However, as can be seen in the form below, scaling down the problem improves the accuracy of each classification task.

Specialty No.	**0**	**1**	**3**	**9**	**12**	**14**	**15**
Speaker Cont	26	62	7	1	1	151	2
Accuracy	70.72%	54.80%	80.57%	100.00%	100.00%	53.52%	100.00%

The second step is to classify, inside each specialties, each samples into each classes inside this specialty, completing the classification.

References

K. Ahmed, M. H. Baig, and L. Torresani, "Network of Experts for Large-Scale Image Categorization," in Computer Vision – ECCV 2016, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., vol. 9911, Springer, Cham, 2016, pp. 1-15, doi: 10.1007/978-3-319-46478-7_32.

A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," Information Sciences, vol. 622, pp. 178-210, 2023, doi: 10.1016/j.ins.2022.11.139.

Sklearn. (2021). Clustering - sklearn 0.23.2 documentation. [online] Available at: https://scikit-learn.org/stable/modules/clustering.html#k-means [Accessed 2 Jan. 2023].

M. H. Sazlı , "A brief review of feed-forward neural networks", *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, vol. 50, no. 01, pp. 0-0, Jan. 2006, doi:10.1501/commua1-2_0000000026

Song Y Y, Ying L U. Decision tree methods: applications for classification and prediction[J]. Shanghai archives of psychiatry, 2015, 27(2): 130.