Colton Proctor

Homework Week Eight

Dr. Aiman Darwiche

Grand Canyon University

In this paper the data from eight different participants was collected using three different accelerometers each. These devices measured a wide variety of motion data relating to the degrees of motion and their severity. This information was used to determine the activity that was being performed by each participant, such as sitting, walking, running, cycling, etc. To verify the accuracy of the information, the values were both normalized, and a subset of their overall values were curated. For the purposes of this experiment then a second subset of the overall features was chosen. The data set that was utilized for this experiment contained 18 features, as well as the activity label for which each feature vector corresponded to. The data set that I used was reduced from the overall data to remove missing values, and 10s of time was trimmed from both the beginning and the end. Activity labels were also added so that each activity had a label ranging from 1-8.

Three different Tree algorithms were used as baseline ensemble models, these were a bagging tree, a random forest, and an ADA boosted tree. There were also three non ensemble methods employed, these were an Artificial Neural Network, K-Nearest Neighbors Algorithm, and an SVM. Finally custom ensemble methods were used to improve the performance of the base models, all of which had similar performance. These three custom methods all used the same four base models, however differed in how they came to a decision on the class label. The first is weighted majority voting, the second was NB combination, and the third BKS combination.

To start the results of my testing began with the three base ensemble methods.
I used the mean f1 score of a 10-fold cross validation to determine the overall effectiveness of each of the different methods. The first to be tested was the bagging decision tree model. For this model I used the Bagging Classifier class with the base model being a CART decision tree. It

was created using 20 trees as the number of estimators to mirror the article. This was then trained using 10-fold cross validation, and the overall f1 score was 77.9. This was slightly lower than what was found in the paper, as theirs was 80.2.

Next to be tested was the Random Forest. I created a random forest having each tree use only the square root of the number of features in the input. This was done to mirror what was used in the paper. The random forest had an overall f1 score of 79.6 through my trials, while it was 81.4 in the paper. This continues the trend of the paper having slightly better results than what I was finding.

I continued the base ensemble methods by using the ADA boosting algorithm to build the final tree. The maximum number of iterations was kept to 100 to reduce the overall complexity of the tree, and the Discriminant weak learners were chosen. The mean f1 score for the 10-fold validation for this model was only 74.9. This was lower than the 80.3 reported in the paper.

Next, I tested the three different base models, K-Nearest Neighbors, Support Vector Machines, and Artificial Neural Networks. The simplest of the models is the KNN algorithm, so that was the one that I ran first. To mirror the paper the number of clusters was set to 7. The overall f1 score of this algorithm in my testing was 44.3, with runs as high as 63.8. This is significantly lower than the 78.9 reported in the paper. This leads me to believe that I fundamentally did something different than the experimenters in the paper did while building the base model.

After the KNN algorithm I built an SVM. To mirror the model built in the paper, multiclass SVM was used with the kernel set to linear. This resulted in an overall f1 score of 48.9, however on certain runs it would be as high as 83.2. In the paper they reported an average

of 83.22. My overall average was much lower than theirs, however I did see results that attained the f1 score that they reported in singular runs.

The ANN was the final method to be tested. I built a model that had an input layer that accepted 18 features, with a hidden layer that had 50 nodes. The output layer had 8 nodes to correlate to the 8 different classes that could be the result. My model however was very inaccurate, having only a 23.7 rating for the f1 score. In the paper they reported a f1 score of 81.4. This leads me to believe that I built my neural network incorrectly or am testing it incorrectly as the model should be performing better than what is being reported.

As of yet, I am unable to correctly implement the custom ensemble methods. I am having difficulties figuring out how to properly assign the probabilities for voting classification.

Works Cited

CHOWDHURY, ALOK KUMAR1; TJONDRONEGORO, DIAN1; CHANDRAN, VINOD1; TROST, STEWART G.2 Ensemble Methods for Classification of Physical Activities from Wrist Accelerometry, Medicine & Science in Sports & Exercise: September 2017 – Volume 49 - Issue 9 - p 1965-1973 doi: 10.1249/MSS.0000000000001291

Ensemble learning: Ensemble techniques. Analytics Vidhya. (2020, November 28). Retrieved December 24, 2021, from https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/