| | | |
|---|---|---|
| Course Code | : | AIT201 |
| Course Name | : | Applied Machine Learning |
| Lecturer | : | Goh Sim Kuan |
| Academic Session | : | 2021/09 |
| Assessment Title | : | Implementation of Machine Learning Classification in Human Activity: an Attempt on Maximizing Accuracy |
| Submission Due Date | : | 25/12/2021 |

Prepared by :

| Student ID | Student Name |
|---|---|
| AIT2009357 | Bi Xiaoyang |
| AIT2009376 | Wang Qipeng |
| AIT2009374 | Shi Ziyue |
| AIT2009369 | Huang Yiyang |
| | |
| | |
| | |
| | |
| | |
| | |

| | | |
|---|---|---|
| Date Received | : | |

Feedback from Lecturer:

## Own Work Declaration

I/We hereby understand my/our work would be checked for plagiarism or other misconduct, and the softcopy would be saved for future comparison(s).

I/We hereby confirm that all the references or sources of citations have been correctly listed or presented and I/we clearly understand the serious consequence caused by any intentional or unintentional misconduct.

This work is not made on any work of other students (past or present), and it has not been submitted to any other courses or institutions before.

Signature:

Bi Xiao Yang AIT 2009357 Shi Ziyue Wang Qipeng 黄逸扬 Huang YiYang

Date: 25/12/2021

# Implementation of Machine Learning Classification in Human Activity: an Attempt on Maximizing Accuracy

## 1. Introduction

Machine Learning (ML) is one of the fastest-growing areas in Artificial Intelligence and has shown its significance and potential in previous development. There are several data science problems in ML, the most remarkable of which is classification. Classification is the process of finding models (or functions) that describes and distinguishes classes or concepts of data, to be able to use the model to predict classes of unknown objects. In our project, we intend to use this technology in solving a real-world problem, Human Activity Recognition.

In this project, we focused on using all our knowledge to adjust the algorithms and parameters in different stages to get the solution with the highest accuracy. To give a comprehensive overview of our conception and implement the project, we organize the report in 4 parts:

• Introduction;

• Research Approach, in which we would present how we formulate the problem, and our data preparation and processing pipeline;

• Research Process, where we would make a detailed elaboration on ML models with the highest accuracy and the result we have gotten based on our experiment;

• Discussion to summarize our results and reflect on them.

## 2. Research Approach

In this section, we will focus on abstracting our problems to machine learning models, and our effort in fetching, pre-processing the data. As for the data processing part, we will present in the next section along with the ML models being applied.

2.1 ML Problem Formulation

Task T: Building a model that can recognize and classify human activities recorded by smartphones.

Performance Measure P: the overall accuracy and F1 score that the classification system presents.

Training Experience E: the accelerometer and gyroscope's recording of 30 study participants performing six activities (WALKING, WALKINGUPSTAIRS, WALKINGDOWNSTAIRS, SITTING, STANDING, LAYING).

## 2.2 Data Preparation

Our project intends to recognize and classify human gestures and our dataset based on the research of the recording of 30 study participants performing activities of daily living. With the guidance of our lecturer, the ideal dataset was obtained in previous relating research. An experiment has been carried out with a group of 30 volunteers within an age bracket of 19-48 years, they all performed six activities (WALKING, WALKINGUPSTAIRS, WALKINGDOWNSTAIRS, SITTING, STANDING, LAYING) and these activities were labeled manually.

The dataset has been randomly partitioned into two sets, where 70% of the volunteers were selected for generating the training data and 30% for the test data. After browsing the .csv files, there is an interesting finding that the "subject" column (ranging 1-30) perfectly corresponds to the 30 smartphones used in the experiment, and being split 7:3, the same as the ratio of train and test dataset.

Then we invoked corresponding functions to check whether there is a null or duplicated value and did feature extraction and analyzed the relationship between data and visualization tools to reveal them.

## 2.3 Data Pre-processing

After realizing the identification of required data and the definition of training and testing set, we designed experiments in testing which pre-processing method may get the result the highest accuracy. Firstly, we tried to drop the column of "subject" and make a comparison between the results. The accuracy dropping the column would get slightly higher. Then we invoked label encoder to target activities and tested 3 pre-processing methods on the rest of the features subsequently: Standard Scaler, Minmax Scaler, and Principal Component Analysis (PCA). The test followed with 3 basic algorithms (Random Forest, KNN, and Decision Tree) and the result indicates that Standard Scaler's and PCA's accuracy is relatively better but the differences are subtle and unexplainable according to our limited knowledge. In case the combination of different ML models and pre-processing methods would result in unexpectedly high accuracy, we would test all the pre-processing methods in each ML model.

## 3. Research Process

### 3.1 Parameter Tuning

In order to get the proper parameter and design the most efficient algorithm for real-world applications, we applied two methods: Grid Search and Random Search, to find the best parameters for each model. As both methods are based on exhaustive enumeration and cross-validation, the accuracy of these two methods is basically the same while the time consumption of random search is consistently better. Such experimental results are also consistent with the efficiency of the two methods in real-life applications. Therefore, the random search would be applied in the following research.

### 3.2 ML models

10 algorithms were applied in search of the ML model that could yield the highest accuracy within our limited knowledge base. The code of these algorithms will be presented in the jupyter notebook, so we simply list their names here and select the ones with the highest accuracy for a detailed description. The ML models we applied are listed below: Random Forest, Decision Tree, KNN, SVM, Gaussian Naive Bayes, Logistic Regression, Linear Support Vector Machine, Multiple Layer Perceptron, XG Boost, and ANN.

### 3.3 Summary of the experiment results

#### 3.3.1 Result Overview

Since we did multiple groups of control trials, the amount of data derived was huge. To intuitively present the experimental results and make better comparisons, we made a statistical table of accuracy and F1_score as follows:

| | Standard Scalar's accuracy | Standard Scalar F1_Score | Minmax Scalar's accuracy | Minmax Scalar F1_Score | PCA's accuracy | PCA F1_Score |
|---|---|---|---|---|---|---|
| Random Forest(dropping subject column) | 0.92399 | 0.923825 | 0.745504 | 0.726173 | 0.908381 | 0.907462 |
| Random Forest(remaining subject column) | 0.917543 | 0.917246 | 0.785884 | 0.780656 | 0.905667 | 0.904784 |
| Decision Tree(dropping subject column) | 0.810655 | 0.80673 | 0.686121 | 0.644368 | 0.796403 | 0.796793 |
| Decision Tree(remaining subject column) | 0.783848 | 0.776547 | 0.675942 | 0.643778 | 0.794706 | 0.794718 |
| K-Nearest Neighbors(dropping subject column) | 0.891754 | 0.891353 | 0.879199 | 0.880551 | 0.890736 | 0.889945 |
| K-Nearest Neighbors(remaining subject column) | 0.891076 | 0.890635 | 0.884289 | 0.885413 | 0.798778 | 0.798801 |
| Support Vector Machine(dropping subject column) | 0.956227 | 0.956154 | 0.863929 | 0.862855 | 0.93926 | 0.939177 |
| Support Vector Machine(remaining subject column) | 0.956227 | 0.956154 | 0.863929 | 0.862855 | 0.906006 | 0.90555 |
| Gaussian Naiye Bayes(dropping subject column) | 0.571089 | 0.4879 | 0.655243 | 0.635235 | 0.872413 | 0.870661 |
| Gaussian Naiye Bayes(remaining subject column) | 0.571089 | 0.4879 | 0.655243 | 0.635235 | 0.879539 | 0.877995 |
| Logistic Regression(dropping subject column) | 0.961317 | 0.961308 | 0.85375 | 0.849226 | 0.939939 | 0.939674 |
| Logistic Regression(remaining subject column) | 0.961317 | 0.961308 | 0.85375 | 0.849226 | 0.921955 | 0.921414 |
| Linear Support Vector Machine(dropping subject column) | 0.958602 | 0.958576 | 0.857482 | 0.854459 | 0.937564 | 0.937224 |
| Linear Support Vector Machine(remaining subject column) | 0.958602 | 0.959576 | 0.857482 | 0.854459 | 0.923651 | 0.923215 |
| Multiple Layer Perceptron(dropping subject column) | 0.954191 | 0.954377 | 0.865965 | 0.854459 | 0.938242 | 0.938307 |
| Multiple Layer Perceptron(remaining subject column) | 0.954191 | 0.954377 | 0.865965 | 0.8603 | 0.91415 | 0.914251 |
| XG Boost(dropping subject column) | 0.927723 | 0.927686 | 0.785545 | 0.766834 | 0.901595 | 0.900546 |
| XG Boost(remaining subject column) | 0.927723 | 0.927686 | 0.785545 | 0.766834 | 0.893112 | 0.891372 |
| ANN (dropping subject column): 0.9620 | | | | | | |
| ANN(remaining subject column): 0.9586 | | | | | | |

According to the table, we conducted a preliminary data filtering and listed the conditions in which the accuracy exceeded 0.9 and could even reach 0.95.

| | Standard Scaler's accuracy | Minmax Scaler's accuracy | PCA's accuracy |
|---|---|---|---|
| dropping subject column | More than 95%: | None of them is over 90% | More than 90% (But no more than 95%) |
| | [1]Support Vector Machine | | [1]Random Forest |
| | [2]Linear Support Vector Machine | | [2]Support Vector Machine |
| | [3]Logistic Regression:0.9613(highest) | | [3]Logistic Regression |
| | [4]Multiple Layer Perceptron | | [4]linear Support Vector Machine |
| | | | [5]Multiple Layer Perceptron |
| remaining subject column | [1]Support Vector Machine | None of them is over 90% | No more than 95% |
| | [2]Linear Support Vector Machine | | Logisitic Regression: 0.9399(highest) |
| | [3]Logistic Regression | | |
| | [4]Multiple Layer Perceptron | | |

Obviously, we can find that the models that perform rather outstandingly (whose accuracy is over 95%) are Support Vector Machine, Linear Support Vector Machine, Logistic Regression, Multiple Layer Perceptron when applying Standard Scaler. There also are models that result in a fair accuracy when applying Principal Component Analysis, and they are Random Forest, SVM, Logistic Regression, Linear Support Vector Machine and Multiple Layer Perceptron, their accuracy can be over 0.9 but not more than 0.95. As for the models applying Minmax Scaler, the result is not very ideal, which value never reaches 0.9.

### 3.3.2 ANN

Due to ANN's specificity (based on Deep Learning), we have rather high expectations of its performance, so we list it separately. According to the ANN model we built, the accuracy of remaining and dropping “subject” columns could respectively reach 0.9586 and 0.9620 at most, from which we can indicate that ANN is the best ML model in our project.

3.3.3 F1_score

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

In this report, we introduced another important measure for ML classification problems -- F1_score. It is the summed average of precision and recall, with a maximum of 1 and a minimum of 0. And from the experimental results we have had, the value of F1_score is extremely close to the accuracy value, just as expected. (The difference between them is as low as 0.01 for most models) This provided further evidence of the reliability of our experiments.

3.3.4 Summary

With the above elaboration, we here conclude this project:

1. Dropping the “subject” column could yield a result with slightly higher accuracy.

2. In pre-processing section, Standard Scaler has the best performance.

3. ANN, of all ML models, gets the highest accuracy, which is as high as 0.962.

4. The rest of the models that performed well are Support Vector Machine, Linear Support Vector Machine, Logistic Regression (can reach 0.9613 at most), Multiple Layer Perceptron. These 4 models’ accuracy could all be more than 0.95 when applying Standard Scaler Model.

## 4. Discussion

4.1 Assumption on the reason of receiving such result

After the statistics and analysis of the resulting data, we came to the conclusions mentioned in the previous part. We had a heated discussion and tried to use the knowledge we had acquired to explain the reasons for such results. In this part, we will try to explain the reason for accuracy when applying different pre-processing methods.

We proposed one assumption for the hypothesis that why PCA and Minmax could not reach as high accuracy as we wish.

We start our analysis with PCA. PCA is a useful statistical technique for the reduction of data dimensionality. By PCA, the sampling density can be increased by rounding off some information and reserving only the key information. It can effectively mitigate dimensional disasters, but the information reserved only focuses on the training set, which leads that the reserved information may not be important. Some seemingly useless information may be discarded, but this seemingly useless information happens to be important information that just does not have a great performance on the training set, so PCA may also aggravate the overfitting. This makes PCA extremely sensitive when applying to different ML models, well explaining why Gaussian Naive Bayes and PCA's combination has accuracy much higher than that with Standard Scaler and Minmax Scaler (0.879 vs 0.655 and 0.571) while most of the other models' performance applying PCA is not as good as those applying Standard Scaler.

Under this logic, we may reasonably extrapolate that Minmax Scaler's poorer performance for the same reason: Minmax Scaler is also a method of high sensitivity, the models we apply may not be appropriate. But as we have tested multiple models, the assumption may not be so valid and needs further testing.

## 4.2 Reflection

In this project, we have attempted to apply as many methods as possible by combining parameter tuning methods, data pre-processing methods, and ML models and the conclusion that we have drawn is fairly satisfying. We also tried to explain the reason that causes the unexpected result. But still, there remain unsolved problems, for example, (1) How to better explore the relationship between the data (columns) in the dataset (2) How to select suitable models and pre-processing methods for the predicted dataset more quickly. They need our further knowledge to understand more thoroughly. But overall, we have made our effort in designing our classification model, which is the first step in our academic career in machine learning.

**APPENDIX 1**

## MARKING RUBRICS

| Component Title | Group Project (Code, Report) | | | | Percentage (%) | 30 | |
|---|---|---|---|---|---|---|---|
| Criteria | Score and Descriptors | | | | | Weight (%) | Marks |
| | Excellent (15-13) | Good (12-10) | Average (9-7) | Need Improvement (6-4) | Poor (3-0) | | |
| Machine learning knowledge and understanding | Students demonstrate excellent technical & practical understanding of machine learning knowledge and data science pipeline. | Students demonstrate a good practical understanding of machine learning knowledge and the data science pipeline. | Students demonstrate an average understanding of machine learning knowledge and the data science pipeline. | Students demonstrate an insufficient understanding of machine learning knowledge and the data science pipeline. | Students fail to demonstrate an understanding of machine learning knowledge and the data science pipeline. | 15 | |
| The quality of Code & Report | Excellent programming practice is followed with very detailed documentation. The report provides a concise summary of what is accomplished with very detailed explanations and discussion. | Good programming practice is followed by documentation. The report provides a concise summary of what is accomplished with detailed explanations and discussion. | Good programming practice is followed by simple documentation. The report provides a summary of what is accomplished with explanations and discussion. | Programming practice is followed by simple documentation. The report provides a summary of what is accomplished with little explanation and discussion. | Programming practice is not followed with documentation. The report provides an unclear summary of what is accomplished with little explanation and discussion. | 15 | |
| | | | | | TOTAL | 30 | |

| Component Title | Group Project (Presentation) | | | | Percentage (%) | 20 | |
|---|---|---|---|---|---|---|---|
| Criteria | Score and Descriptors | | | | | Weight (%) | Marks |
| | Excellent (10-9) | Good (8-7) | Average (6-5) | Need Improvement (4-3) | Poor (2-0) | | |
| Content Delivery and Q&A | Project content was delivered clearly with strong findings. | Project content was delivered clearly with findings. Students can | Project content was delivered clearly with findings. Students can | Project content was delivered clearly with inferior findings. | Project content was not delivered clearly with inferior | 10 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Students can answer all questions by the lecturer and students.** | **answer questions by the lecturer and students.** | **answer most questions by the lecturer and students.** | **Students can answer most questions by the lecturer and students.** | **findings. Students cannot answer most questions by the lecturer and students.** | | |
| **Teamwork & Peer Assessment** | **All team worked well together to achieve objectives. Each member contributed in a valuable way to the project since the group was formed.** | **Almost all team worked well together to achieve objectives. Each member contributed in a valuable way to the project.** | **The team worked well together most of the time, with only a few occurrences of communication breakdown or failure to collaborate when appropriate. Members were mostly respectful of each other** | **Few of the team worked together, with only a few occurrences of communication breakdown or failure to collaborate when appropriate. Few members were mostly respectful of each other** | **The team did not collaborate or communicate well since the group formed. Some members would work independently, without regard to objectives or priorities. A lack of respect and regard was frequently noted.** | **10** | |
| | | | | | **TOTAL** | **20** | |

Note to students: Please print out and attach this appendix together with the submission of coursework