

Predict Parkinson Disease using wearable data

Aloi Alfredo, Riccio Francesco

December 29, 2021

Abstract

This is a project for the Deep Learning course held in the Master Degree in Computer Science of Università degli Studi della Calabria.

Contents

| | | |
|----------|--|----------|
| 1 | Introduction | 3 |
| 2 | Task 1: next value prediction | 4 |
| 2.1 | Data understanding and preparation | 4 |
| 2.2 | Modeling | 4 |
| 2.3 | Evaluation | 4 |
| 2.4 | Finding better parameters | 4 |
| 2.5 | Conclusions | 4 |
| 3 | Task 2: anomaly detection | 5 |
| 3.1 | Data understanding | 5 |
| 3.2 | Data preparation | 5 |
| 3.3 | Modeling | 5 |
| 3.4 | Evaluation | 5 |
| 3.5 | Conclusions | 5 |

1 Introduction

Parkinson disease, the second most neurological disorder that causes significant disability, reduces the quality of life and has no cure. Approximately, 90% affected people with Parkinson have speech disorders. Deep learning has the potential to give valuable information after processing, that can be discovered through deep analysis and efficient processing of data by decision makers. The main problem is the identification of OFF symptoms perceived by the study subject with reasonable accuracy from real-world data collected. OFF periods are times when Parkinson's disease medication, namely levodopa, is not working optimally. As a result, symptoms return. OFF periods can include both motor symptoms, such as tremor and rigidity, and non-motor symptoms, such as anxiety. The goal is to consolidate data collection by identifying the most important variables to solve the problem.

2 Task 1: next value prediction

We have 3 time series (X, Y, Z) recorded each 10 seconds. For the first sub-task we consider, for each time series, sequences of five minutes (`window_size` = 30) every one minute (`window_shift` = 6), predicting the next value in the series. The evaluation metric is the Mean Absolute Error, in particular we have to go below:

- 81.06 for X;
- 85.26 for Y;
- 79.94 for Z.

We then choose, for the second sub-task, the one time series with the worst value of the evaluation metric, looking for a better combination of window size/window shift.

2.1 Data understanding and preparation

We have two files, `train.csv` and `test.csv`, each one containing three columns, representing the three time series X, Y and Z. The `test.csv` file will be used for the evaluation, while the `train.csv` file, that contains 144910 rows, will be used for the training part.

CONTINUE CON LA PREPARATION

2.2 Modeling

Model description

2.3 Evaluation

Talking about evaluation and results

2.4 Finding better parameters

Talk about searching for best window size and shift

2.5 Conclusions

conclusions about this task

3 Task 2: anomaly detection

General description about the task

3.1 Data understanding

Description about the data...

3.2 Data preparation

All the steps: normalization etc.

3.3 Modeling

Model description

3.4 Evaluation

Talking about evaluation and results

3.5 Conclusions

conclusions about this task