Reconhecimento de Padrões
Pattern Recognition

2016/2017

Project Assignment

Human Activity Recognition Using Smartphones

1 Background

The need for intelligent assistance systems is growing with the population aging. A straightforward approach of monitoring is the physical observation, which is highly human resource demanding, and will become infeasible with the population aging. So, automatic human activity recognition (HAR) in Ambient Assisted Living (AAL) environments is of major importance. Besides the monitoring of the elderly population, HAR also applies to other disorders and situations, like: physically impaired people, people suffering from diabetes and obesity, or just for monitoring of personal fitness activities. The need of minimally intrusive, real-time and easy-to-wear systems are required and has been a subject of many recent research studies in HAR domain.

2 Objective

Your task is to develop classifiers for Human Activity Recognition using data collected from 30 subjects performing activities of daily living while carrying a waist-mounted smartphone with embedded inertial sensors. Consider two scenarios:

- **Scenario A (Binary Problem)**: where one where one want to discriminate if a given person is walking or not;
- **Scenario B (Multiclass Problem)**: where differentiation among six different activities are required (walking, walking upstairs, walking downstairs, sitting, standing and laying).

3 Practical Assignment

3.1 Dataset Description

Data were acquired from 30 human volunteers with age in the range 19-48 years. The participants were asked to perform six different activities:

- WALKING
- WALKING UPSTAIRS
- WALKING DOWNSTAIRS

- SITTING
- STANDING
- LAYING

The participants wear a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, that enable the acquisition of 3-axial linear acceleration and 3-axial angular velocity at a sample rate of 50Hz. Annotation was performed manually based on a video recorded in parallel with the cellphone data. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The signals collected from the accelerometer and gyroscope were pre-processed by applying noise filters and then segmented by applying a rectangular sliding window of 2.56 sec and 50% of overlap (128 readings/window). The gravitational and body motion components were obtained by filtering the acceleration signal with a Butterworth low-pass filter (cutoff frequency of 0.3Hz). From each sliding window, a vector of features was obtained by calculating measures from the time and frequency domain.
Download data and obtain more details at https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones#. Details of the **561** computed features are described in the file *features_info.txt* and *features.txt*.

3.2 Feature Selection and Reduction

Some of the supplied features may be useless, redundant or highly correlated with others. In this phase, you should consider the use of feature selection and dimensionality reduction techniques, and see how they affect the performance of the pattern recognition algorithms. Analyze the distribution of the values of your features and compute the correlation between them. Make sure you know your features! Do not forget to present your findings in the final report.

3.3 Experimental Analysis

You should be able to design experiences in order to run the pattern recognition algorithms in the given data and evaluate their results. Keep in mind that you should develop classifiers for the binary and for the multiclass problem. Define the appropriate performance metrics for the different classification problems. Justify your choices!

Run the experiments multiple times and to be able to present average results and standard deviations (of the metrics used) you should split the training set in parts and use cross-validation. At the end chose the best classifiers and test them in the test set provided.
Do not forget that manually inspecting the predictions of your algorithms can give you precious insights of where it is failing and why, and what you can do to improve it (e.g. what makes the algorithm fail in this particular case? what special characteristic does it have that makes it so hard? how can I make the algorithm deal better with those cases?). Go back and forward to the Pre-processing, Feature reduction and Feature Selection phases until you are satisfied with the results. It is a good idea to keep track of evolution of the performance of your algorithm during this process. Try to show these trends in your final report, to be able to fundament all the issues involved (choosing parameters, model fit, etc.)

3.4 Pattern Recognition Methods

You can write your own code in your language of choice or use the functions and methods available in Matlab and in the Statistical Pattern Recognition STPRTool used in the classes (since you are already familiarized with it). The methods used in your work should be described as well as discussion of the parameters used. Try out different pattern recognition algorithms. You should try to understand how they perform differently in your data.

3.5 Results and Discussion

Present and discuss final results obtained in your Project assignment. This problem was already studied by other authors. Compare your results with the results from other sources.

3.6 Code & Graphical User Interface (GUI)

You should deliver your software code in MATLAB, or in any other programming language you used during the project.

For your project you should write code for a graphical user interface. The GUI should improve the interaction of the user with the code by providing options for data-loading, feature selection/dimensionality reduction, classification, post-processing, validation and visualization.
Remember to comment your code. Write also a help section to your code that tells the purpose of the function, usage, and explanation of parameters.

4 Documentation

Write documentation (in Portuguese or in English) about your project. The documentation should include a cover page where course name, project title, date, names and student numbers of the authors are mentioned.

Describe the methods used for classification in such detail that reader would be able to implement the same kind of functions for feature extraction and classification just based on your documentation and some basic background in pattern recognition. Always justify your choices, even when their are based on intuition. Do not forget to verify your assumptions! Include classification results with the given data to your documentation. At the end of your documentation you should have a list of all references used.

4.1 Requirements

Practical assignment is meant to be done in groups of two persons. If someone wants to work alone, this is also possible. Larger groups are not allowed.

4.2 Project Submission & Deadlines

1. **Project First Milestone (Deadline: 21st April 2017!)**
   Deliverables:
   - Data Preprocessing (Scaling, Feature Reduction (PCA & LDA), Feature Selection, etc.);
   - Minimum Distance classifier, Fisher LDA;
   - Code + short report.
2. **Project Final Goal (Deadline: 26th May 2017!)**
   Deliverables:
   - Data Preprocessing (Scaling, Feature Reduction (PCA & LDA), Feature Selection, etc.);
   - Several classifiers;
   - Final Report
   - Matlab code + GUI.
3. **Presentation and Discussion (5th and 6th of June 2017!)**

Acknowledgments

Credits to the UCI Machine Learning Repository[2].

References

[1] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

[2] Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.