FROM BINARY TO MULTI-CLASS DIVISIONS: IMPROVEMENTS ON HIERARCHICAL DIVISIVE HUMAN ACTIVITY RECOGNITION

Tomás Vieira Caldas

Supervisor: *João Pedro Mendes Moreira*
Co-Supervisor: *João Manuel Paiva Cardoso*
Institution: *Faculty of Engineering of the University of Porto*
Course: *Integrated Master in Informatics and Computing Engineering*

Keywords Hierarchical Classifiction, Clustering, Machine Learning, Classifier, Pruning, Nodes.

CCS Categories Computing methodologies - Cluster analysis, Information systems - Hierarchical data models, Information systems - Clustering and classification, Computing methodologies -Supervised learning.

1. Abstract

In a world where there are more mobile devices than televisions, the amount and type of information a smartphone can provide is significantly increasing. Due to the use of sensors like the accelerometer and the gyroscope, a smartphone can retrieve specific data on-the-fly. Combining this with the factor that people take their device everywhere, and with the help of classification methods, the collected data can be converted into useful information about the user's daily activities and habits. People, often underestimate physically activity, and nowadays being physical inactive is almost considered a disease. With almost no progress made in order to reduce inactivity levels, friendly reminders can be introduced on your smartphone related to the lack of exercise.

In order to process the information needed to classify one's activity, we will resort to machine learning techniques, applying a divisive hierarchical approach. In this work, we will use a modified version of DIvise ANAlysis Clustering, a hierarchical clustering top-down technique. DIANA collects all the activities to be predicted in a big cluster and then starts to split them from a more generalized group to specific ones. Our version, will apply three classifiers (K-Nearest Neighbor, Decision Trees, Naive Bayes), instead of one, on each iteration and choose the one with higher accuracy. This way, instead of a unique way to classify an activity, we always have three (increased chance of correct classification). Once the technique is completed, a pruning phase will occur with the purpose of finding the optimum hierarchical structure to classify our data.

The goal of this project is to implement modifications to the already modified DIANA algorithm in order to have the best possible classification tree before moving to an online approach. The tree generated by the DIANA technique will have specific nodes replaced by their subtrees by turning binary into multi-class divisions. All these cuts will be analyzed and only the ones that provide a better classification accuracy will be accepted and executed. Additionally, since divisive methods have always been neglected by the scientific community, the results will be compared to a flat multi-class classification in order to see how well they behave.

With this work, we expect to provide the scientific community an improved version of the resulting DIANA hierarchical structure. That structure can later be exported to an Android mobile application to provide on-the-fly recognition on the physical activity the user is performing. From there all kinds of features can be implemented: target advertising, new eating habits or tracing of daily routines.