Evaluating the robustness of machine learning algorithm on Human Activity Recognition

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Roberto Costa Fernandes (rcf6@cin.ufpe.br)
Orientador: George Darmiton da Cunha Cavalcanti (gdcc@cin.ufpe.br)
Área: Aprendizagem de Máquina

Março/2019
Abstract

Physical inactivity can cause several diseases, like diabetes, obesity, hypertension, cardiovascular diseases, depression, osteoporosis, and many other physical problems. It is needed to monitor the movements performed by someone and classify them as one of the Activities of Daily Living, like walking, laying down, standing, and other activities to diagnose physical inactivity and avoid this kind of problems. In this context, signals from smartphone sensors can be used as input to machine learning models to perform such classification. However, some of these activities, as sitting and laying, produce almost the same, resulting in noise among data labels thus making the training process harder. This work aims to analyze the impact of data with label noise defining the most robust machine learning technique for that kind of data.
**Resumo**

A falta de atividade física pode causar diversas doenças, como diabetes, obesidade, hipertensão, doenças cardiovasculares, depressão, osteoporoses e diversos de outros problemas físicos. Para diagnosticar a inatividade e evitar tais problemas é necessário monitorar as atividades feitas por uma pessoa e classificá-las em atividades do dia-a-dia, como andar, deitar, ficar em pé e outros. Nesse contexto, sinais de sensores presentes nos celulares podem ser utilizadas como entrada em modelos de aprendizagem de máquina para classificar tais atividades. Porém algumas dessas atividades podem gerar sinais parecidos e resultando em ruído, o que pode dificultar o processo de treinamento. Assim, o presente trabalho visa estudar o impacto de dados com ruído em suas classes e definir qual técnica de aprendizagem é mais robusta a esse tipo de problema.
Introduction

The World Health Organization (WHO) classifies physical inactivity as the fourth most significant risk factor for global mortality [1]. This risk emerges because physical inactivity can cause a couple of health problems [2], like cardiovascular diseases, diabetes, cancer, hypertension, obesity, depression, and osteoporosis. One of the most important actions to treat this sedentary behavior is to identify if the amount of activity performed by a person reach the time of active time recommended by the WHO [1].

Human Activity Recognition (HAR) is the research field that intends to identify the activities performed by someone, and it is the main task is to classify the Activity of Daily Living (ADL). The ADL can be defined as the activities performed by one person during a day, such as walking, sitting, standing, laying down, walking upstairs or walking downstairs. The HAR process can be divided into two parts, first, it must acquire data from the person, to secondly be able to analyze and classify the ADL.

There are two different ways to execute the first step of HAR. The first one is to use cameras and other sensors to identify activities on a well-defined environment. Although this approach has the downside of being used only in a controlled area, can detect a much wide range of ADL [3]. The second approach is to place sensors in different positions in the body [4] [5] [6], like wrist, waist, and chest, also known as body-worn sensors. This method solves the problem of the first approach, but it cannot be used as a long-term solution because it is not always comfortable to wear a sensor daily — the needed for recalibrating the sensors every time after dressing is another problem.

The rising popularity of smartphones in recent years is creating a new method to acquire data to use on HAR. This demand is occurring because these devices already have a triaxial accelerometer and gyroscope built in, making this approach very similar with body-worn sensors, with the advantage of almost everyone, is carry one smartphone on their ADL. Some works are emerging proposing using that methodology to obtain data [7] [8] [9], and this direction is showing to be very promising.

These data obtained in the first part of HAR are used to identify and classify the ADL. The first approach to classify ADL was using Digital Signal Processing (DSP) [4] [5]. One of the approaches [4] using DSP combines accelerometer data with sound data to recognize workshop activity, and achieves an accuracy of 84.4%. Another technique using DSP [5] is based on a waist-mounted triaxial accelerometer, aiming to identify 12 activities and reaches an accuracy of 90.8%.

Although the results using only DSP can be considered good, recent works propose the combination of DSP to extract features and use those features in machine learning models to classify the ADL. One of these approaches [8] extracts features using a histogram of gradient and Fourier descriptor and then uses a SVM to classify achieving an accuracy of 97.12%. Other techniques extract only simple features of the
signal from a sample window and use a multiclass SVM [9], achieving 96% of accuracy or an ensemble of classifiers achieving an accuracy of 99.22% [10].

None of these works does not consider the presence of noise data and outliers as one problem that can occur. Noise data can be easy collected once the smartphone can be not well placed or in a different pocket depending on the test case. In addition, some ADL, like standing and sitting, has features very similar and increase the number of misclassified examples [9], generating noisy labels.
Objectives

The main objective of this work is to analyze the impact of outliers and noisy labels on the classification of ADL using machine learning techniques. The techniques that will be analyzed will be Support Vector Machine (SVM), Random Forest, Multi-Layer Perceptron (MLP), Decision Tree and some ensemble techniques.

There are four specific objectives to accomplish to achieve the main one. They are:

- Train and test different machine learning models with data without noise to find the ground truth for these models.
- Define a method to add label noise in the dataset.
- Retrain the models with the new dataset.
- Compare the results of the different models and the impact of the noise in each model.
Methodology

The first challenge of this work could be to define a methodology to acquire the data needed to use in the following steps. However, instead of generating a new dataset, it will be used two public datasets available on the UCI Machine Learning Repository [11]. They will be: 1) Human Activity Recognition Using Smartphones Data Set [9], and 2) Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (AAL) Data Set. The author obtained both datasets using the same methodology. Therefore, they have the same 561 attributes and classify the same six ADL. The six ADL present in those datasets are; walking, walking upstairs, walking downstairs, sitting, standing, laying down. Also, these datasets are already divided into train set, 70% of instances, and test set, the remaining 30% of the examples.

Besides these advantages to use those datasets, there already have some results of classification accuracy using SVM [9] and precision, accuracy, recall and F-measure using Bagging, Adaboost, Rotation Forest, Ensemble of Nested Dichotomies and Random subspace [10]. The methodology used in these works consists of use 10-fold cross-validation on the train set to find the best parameters for the trained model and then use the test set to evaluate the model. It will be used the same methodology to be able to compare the results that will be obtained.

The use of a public dataset will help to avoid creating a new one, but it still needed to add noisy labels to the dataset. To do so, an approach that consists on adding label noise, given a probability, and changing the instance label to one of the other classes will be used. The first step is to test the models with no probability of changing the label to create a ground truth for each model and then testing with higher probabilities.

One of the methods that are showing good results to classify data with noisy labels and outliers is using an ensemble of classifiers based on instance hardness [12]. This method aims to remove outliers and noisy labels from the train set while trying to keep instances close to the border of classes. This method is based on the bagging algorithm, but differs when building the bootstrapped training sets, that originally choose an example with a uniform probability. In this method, an instance is chosen to the bootstrap with a given probability, defined as the inverse of an instance hardness. To define the hardness of one instance, it is used the k-Disagreeing Neighbors (kDN) method [13], that measures how much an instance differs from the instances close to it. The hardness of one example, given k neighbors, is defined as:

\[ kDN(x) = \frac{\{x'|x' \in kNN(x) \land label(x') \neq label(x)\}}{k} \]  

(1)

This approach aims to give a low probability to instances that have one label and are in the middle of the instances of other labels. Therefore, this method will avoid the
problem of removing examples close to class borders, when trying to filter the data before the training.

To evaluate this method and be able to compare with the results already presented to this dataset, accuracy (2), precision (3), recall (4) e F-measure (5) will be measured on the test set.

\[
\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)
\]

\[
\text{precision} = \frac{TP}{TP + FP} \quad (3)
\]

\[
\text{recall} = \frac{TP}{TP + FN} \quad (4)
\]

\[
F_1 = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (5)
\]
## Schedule

<table>
<thead>
<tr>
<th>Activity</th>
<th>Mach</th>
<th>April</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature Review</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Algorithm studies</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Algorithm implementation</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Experiments</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Evaluation</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Presentation Preparation</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

The table above outlines the schedule for activities across different periods from Mach to May.
References


Possíveis Avaliadores

Prof. Cleber Zanchettin
Prof. Ricardo Prudência
Prof. Leandro Almeida