Hierarchical Multi-label Classification for Activity Recognition

Nina Rešcic, Mitja Luštrek
Jozef Stefan Institute
International Postgraduate School Jozef Stefan
Jamova cesta 39
1000 Ljubljana
nina.rescic@ijs.si, mitja.lustrek@ijs.si

ABSTRACT: *Activity recognition using wearable sensors is very important in many domains of health monitoring and is therefore well researched. Most commonly classification considers all activities to be 'equal' (we will use term at classification). However, intuition suggests better results could be achieved using a hierarchical approach for classification. In this paper we compare three different approaches to classify activities: (i) Flat classification - classes are equal and we build one model to classify all of them; (ii) Multi-model hierarchical classification - classes are arranged in trees, we build different models to classify activities on different levels. We apply two different approaches; (iii) Hierarchical classification using CLUS software*[1].

Keywords: Activity Recognition, Hierarchical Multi-label Classification, Wearable Sensors

Received: 8 September 2018, Revised 21 October 2018, Accepted 31 October 2018

1. Introduction

Activity recognition (AR) using wearable sensors has been addressed many times, some of the most important application being personalized health systems. Many of developed methods for recognizing different activities used triaxial accelerometers worn on different body parts. With development of wrist-worn devices in past several years and with their growing popularity in everyday life, methods for recognizing sports activities [2], daily activities [3] and handspecific activities [4] using just wrist-worn sensors were proposed. Although the performance gets better with adding additional body sensors, as Attal and al. [1] proved in 2015 by reviewing the research done by then, we decided to focus our research on wrist-worn sensors due to before mentioned accessibility and popularity.

Vens and al. [6] defined hierarchical multi-label classification (HMC) as a variant of classification, that differs from normal classification in two ways: (1) a single example may belong to multiple classes simultaneously; and (2) the classes are organized in a hierarchy:an example that belongs to some class automatically belongs to all its super classes, the

[1] https://dtai.cs.kuleuven.be/clus/index.html

so called hierarchy constraint.

Although hierarchical approach might seem quite intuitive for AR, as certain activities are pretty obvious grouped together, the usage of hierarchical classification for AR has only been addressed a few times. None of the cases was specifically directed towards usage of wrist-worn device for recognizing different hierarchical activities (physical, daily, hand-movement activities). Khan and al. [8] proposed a hierarchical recognizer for recognition of limited amount of physical activities (static, transitions, dynamic) using a chestworn sensor device. Zheng [9] explored human activity based on the hierarchical feature selection and classification framework. He explored 2D and 3D motion (jumping, running, walking forward/left/right, upstairs/downstairs, static activities).

2. Dataset

The dataset we are working with consists of data from seven people involved in different activities (sport, rest, handwork, eating chores...). We organize the activities in hierarchy as presented in Table 1. First we tried to create structure tree by using Orange[2] software for hierarchical clustering. We calculated features as will be explained later in paper and put them into Orange software. We were looking for some indications of the hierarchy for different groups of activity. However, there was no clear or extremely obvious structure visible. The final structure was designed using knowledge achieved from previous research on the same dataset where at classification (for instance in research made by Cvetkovic et al. [4]) has been used for recognition of activities.

Group	Activity
Daily activities	chores eating handwork washing
Exercise	nordic running walking
Static	lying sitting standing

Table 1. Activity grouping

3. Methods

In this paper we are comparing three different approaches for activity recognition. First we addressed at classification, which is commonly used in previous research. Next, we implemented two multi-model hierarchical algorithms, based on approach proposed by Paes et al.[11]. We use the term multi-model as different models were used for different levels of hierarchy. Finally we used Clus software, which has algorithms for hierarchical multi-classidfication (HMC) already implemented and is mostly used in the field of functional genomics and text classication as shown by Vens et al.[6].

The users were wearing a wearable device (wristband or smartwatch) on their non-dominant hand. For the purpose of this paper we only considered triaxial accelerometer data, however for further research other measurements are available as well (heart rate, galvanic skin response..)

From raw measurements we crated instances using 2 second sliding window and computed set of various features from accelerometer data that were shown to perform well in similar setting (mean, average, skewness, kurtosis, peak counts) [4]. Additionally

[2] https://orange.biolab.si/

we computed the Euler angles pitch and roll and calculated some extra features from them as well - for instance, pitch and roll manipulation, amount of roll motion, regularity of roll motion... Altogether we computed 105 features. Afterwards feature selection was applied and the best of them were used to build models.

3.1 Feature Selection

Feature selection was used only in the cases of at classification and MM-HMC. For feature selection, we first ranked the features by gain ratio. After that, we used a wrapper approach. We started with an empty feature set and added features in the order of their rank. After each feature was added, we evaluated its contribution by building random forest classifiers and internally cross-validating them on the training set. The feature was kept only and only if it increased the overall average accuracy. The ranking by gain ratio and the random forest algorithm were implemented in the Weka machine-learning suite and run with default parameter values.

3.2 Flat Classification

The most common approach for AR is the so-called at classification. All classes are considered equal, hierarchy is not taken into account. Algorithms were implemented in java, using Weka[3] library.

3.3 Multi-Model Hierarchical Classification

We implemented two different approaches for hierarchical classification. The first one, traditional hierarchical strategy Per Parent Top Down (PPTD - Figure 1), based on “local per parent node”model, and the second one, named Sum of weighted Votes (SWV - Figure 2), “local per level” model, proposed by Paes et al. in [10]. On the upper level we built a model to distinguish between three groups - daily activities, exercise and static. This was done the same for both approaches. From here on, the approaches differ.

1. PPTD For this approach, we split instances into three different subsets regarding to the classified group. We then run feature selection for each of the subsets separately and built three different models - on for each group of activities. Features were dierent for each group.

2. SVW After the first level, the classified group has been added to instances as an additional feature. Feature selection has been done again - this time for the whole level, and one model has been built to distinguish between activities.

Same authors have explored feature selection for both approaches in [11], where they have shown that the best results are obtained when using the lazy approach - this approach executes feature selection at the classification time of each instance. We have decided to use the eager approach, where feature selection is done prior to classification.

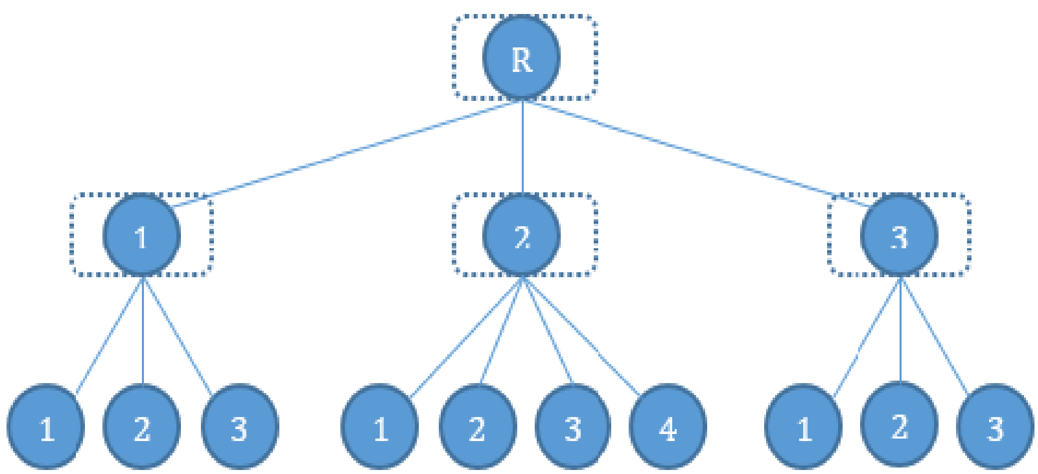

Figure 1. PPTD - local per parent node approach

[3]https://www.cs.waikato.ac.nz/ml/weka/

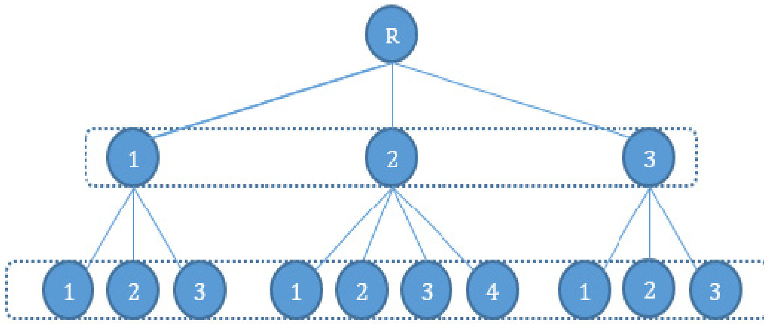

Figure 2. SVW - local per level approach

3.4 CLUS-Classification

CLUS is a decision tree and rule learning system that works in the predictive clustering framework. One of its important functionalists is the CLUS-HMC algorithm for hierarchical multi-label classification. The software has been shown to work very well in the field of functional genomics [6], so the idea to use it in hierarchical classification for activity recognition seems reasonable. Clus-HMC algorithm is a variant of standard greedy top-down algorithm for decision tree induction. To achieve the task of predicting a set of classes instead of a single class, additional changes to the learning procedure are needed, as shown in [12].

```
[Hierarchical]
Type = Tree
HSeparator = /
WType = ExpMinParentWeight
WParam = 0.5
OptimizeErrorMeasure = AverageAUPRC
ClassificationThreshold = [0.5, 0.7, 0.9, 0.95]
MEstimate = Yes
SingleLabel = Yes

[Tree]
PruningMethod = M5
M5PruningMult = 2.0
FTest = 0.1

[Constraints]
MaxDepth = 20

[Ensemble]
EnsembleMethod = RForest
FeatureRanking = Genie3
PrintAllModelInfo = Yes
```

Figure 4. CLUS settings fisle example

In our experiment we worked with random forest (to make it comparable with other two approaches), and we allowed the decision tree to go up to depth 20. We have shown experimentally that performance increases sharply up to decided depth, while afterwards the contribution has become negligible. The error we used for optimization was the average AUPRC (area under the precision-recall curve). We have tested the performance by changing the threshold determining when the probability output by the model is considered to predict a class. All of the above mentioned parameters are set in the settings le as seen in Figure 4.

4. Experimental Setup and Results

In our case the hierarchy is very simple, reduced to two levels. For HMC problems Clus returns several error values. To get fair results for each person included in the dataset, leave-one-person-out approach has been used, as mentioned before. For evaluation of the results we decided to choose standard measurements - precision, recall and F-score. However, when it comes to the evaluation of highly skewed class distributions, similar as with our dataset where for instance daily activities have a much higher frequency than rest, precision-recall curves are the most suitable evaluation tool [7], so this was also added. Vens el al. [6] have addressed the problem of most eligible evaluation tools for hierarchical classification. From the proposed evaluation tools we used the area under the precision-recall curve.

To evaluate predictive models independently from the threshold, two types of evaluation are suitable: ROC analysis and analysis of precision-recall curves (PRC). ROC analysis is better known in machine learning, however for hierarchical multi-label classification PR is more suitable. [?] PR curve plots the precision of a model as a function of its recall, and although it helps understanding the predictions, single value is more appropriate for comparing quality of different models. A score often used to represent this is the so-called "area under the PR curve" (AUPRC). The closer the AUPRC is to 1.0, the better the model.

$$\overline{AUPRC_w} = \Sigma_i w_i \; AUPRC_i$$

If all the weights are set to $w_i = 1/|C|$, where C is the set of classes, score is called average AUPRC, and is denoted as $\overline{\text{AUPRC}}$. If the weigths are set to $w_i = v_i / \Sigma_j v_j$ where v_i is the frequency of class c_i in data, we call this weighted AUPRC and denote it as $\overline{AUPRC_w}$. We have compared the performance of the proposed methods by comparing the precision, recall, F-score and $\overline{\text{AUPRC}}$ score by activity. Validation has been done using "leave-one-person-out" approach. We computed all of the mentioned measures for each person and averaged them to get the performance accuracy by method. Methods that we compared are at classification, multi-model classification using SVW (local per level) approach and CLUS-classification using same approach. We decided to leave out the comparison of PPTD algorithm due to lack of data. Classes for static group were poorly represented from the beginning and after classification on the first level some were left with only few examples. To avoid losing data we propose additional approach, which is roughly explained in the conclusion.

Using the same dataset Cvetkovic and al. [4] have reported on 70% accuracy for five different classes (sports, eating, chores, handwork, washing). We expected high confusion in group of daily activities (handwork, chores, eating, washing) and some confusion between other groups and within them as hand movements can be very similar in this group. Table 2 and Table 3 show the results of the experiments. We could not compare the $\overline{\text{AUPRC}}$ of at classication when classifying groups, as we only get the values for classied activities on lower level. However, we could compare at classification to other two approaches using other measures. As shown in Table 2 MM-HMC performs the best for AR on the upper level, but not much better than at. On the lower level the results from at classification and from MM-HMC were quite similar, with one approach performing better in some cases and worse in others. From the fact that direct classification on the upper level (MM-HMC) is not much better from the indirect, it is safe to conclude that this is the reason, that for similar results between the mentioned two approaches on the lower level. The achieved average accuracy for at classification has been 70:5% and very similar for MM-HMC. Each works better in some cases. Results using CLUS are not the most promising. However, there are many possible combinations of settings available and the performance could be improved by choosing different set of parameters and their values. We tried many possible combinations and the presented results are the best so far.

5. Conclusion

In this work we compared three approaches to activity recognition. Our results show that for the purpose of activity recognition with 2 levels of activity (group and activity), at classification performs as well as both types of hierarchical classification - or even better. In some other uses of HMC, for instance functional genomics, fast performance and correct classification of higher levels is of greater importance than correct classification of lower levels. Unfortunately in the case of activity recognition fast performance was the only upside.

There are some possible improvements for future work. The dataset we were working on, was not really extensive. There were

	Flat	MM-HMC	CLUS
Fscore	82.05%	83.71%	74.36%
Precision	82.03%	83.73%	76.22%
Recall	82.12%	84.05%	73.10%
AUPRC		89.61%	81.09%

Table 2. Results upper level (group)

	Flat	MM-HMC	CLUS
Fscore	65:14%	66:79%	52:23%
Precision	68:29%	65:92%	58:31%
Recall	65:48%	67:69%	51:08%
AUPRC	68:63%	66:67%	54:76%

Table 3. Results lower level (activity)

many activities involved and not many instances of each. This could be solved with joining more similar datasets.

Some of the HMC-related papers mentioned different classifiers for classification. We used random forest, as it has performed the best in our previous research where we were only using at classication, however some other classifiers may perform better on the hierarchical problem. Better accuracy could as well be achieved by adding measurements from some other sensors (heart rate sensor), as maybe there are some more distinctive differences between subsets of the proposed hierarchy.

A possibility to improve the performance of MM-HMC is to add additional activities to each of the groups. For instance, we add exercise and static as two new activities in group of daily activities. Similar would be done for other two groups of activities. After building models for the lower level, we would then build additional models for all “new activities” classied to wrong group. We will try this approach in our future work.

References

[1] Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi. F., Oukhellou, L., Amirat, Y. (2015). Physical Human Activity Recognition Using Wearable Sensors. Sensors 15(12). Basel 31314-18.

[2] Siirtola, P., Lurinen, P., Haapalainen, E., Rnoning, J., Kinnunen, H. (2009). Clustering-based activity classification with a wrist-worn accelerometer using basic features. In: 2009 IEEE Symposium on Computational Intelligence and Data Mining. CIMD 2009 - Proceedings. 95-100.

[3] Chernbumroong, S., Atkins, A.S. (2011)Activity classification using a single wrist-worn accelerometer. In: 2011 5th International Conference on Software, Knowledge Information, Industrial Managrment and Applications (SKIMA) Proceedings. 1-6.

[4] Cvetkovic B., Drobnic V., Lustrek M. (2017) Recognizing Hand-Specic Activities with a Smartwatch Placed on Dominant or Non-dominant Wrist. *In:* Information Society. Ljubljana

[5] Nweke, H.F., Teh, Y.M., Al-gardi, M.A., Alo, U.R.(2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challanges. Expert Systems With Applications. 105. 233-261.

[6] Vens, C., Stryuf, J., Schietgat, L., Dzeroski, S., Blockeel H. Decision trees for hierarchical multi-label classsification. *In:* Machine Learning, 73 (2) 185-214.

[7] Davis, J., Goadrich, M. (2006). The relationship between precision-recall and ROC curves. *In:* Proceedings of the 23rd *Internation Conference on Machine Learning*. 233-240.

[8] Khan, A.M., Lee, Y.K., Lee, S.Y., Kim, T.S. (2010). A Triaxial Accelerometer-Based Physical-Activity recognition via Augmented-Signal Features and a Hierarchical Recognizer. *In:* IEEE Transactions on Information Technology in Biomedicine. 14 (5) 1166-72

[9] Zheng, Y. (2015). Human Activity Recognition Based on the Hierarchical Feature Selection and Classification Framework. *Journal of Electrical and Computer engineering*. doi:10.1155/2015/140820

[10] Paes, B.C., Plastino, A., Freitas, A.A. (2012). Improving Local Per Level Hierarchical Classification. *Journal of Information and Data Manegement* 3 (3) 394-409.

[11] Paes, B.C., Plastino, A., freita s, A.A. (2014). Exploring Attribute Selection in Hierarchical Classification. In: *Journal of information and Data manegment*- Vol. 5 (1) 124-133.

[12] Blockeel, H., Bruynooghe, M., D—zeroski, S., Ramon, J., Struyf; J. (2002). Hierarchical mulit-classification. *In*: Proceedings of the ACM SIGKDD 2002 Workshop on Mulit-Relational Data Mining (MRDM 2002). 21-35.

[13] Cesa-Bianchi, N., Gentile, C., Zaniboni, L. Incremental algorithms for hierarchical classification. *Journal of Machine Learning*. (2006) 31-54

[14] Clus Homepage (last accessed 26 Jun 2018) https://dtai.cs.kuleuven.be/clus/index.html.