Factored four way conditional restricted Boltzmann machines for activity recognition

Citation for published version (APA):

Document status and date:
Published: 06/11/2015

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 09. Mar. 2019
Factored Four Way Conditional Restricted Boltzmann Machines for Activity Recognition\(^1\)

Decebal Constantin Mocanu \(^a\)  Haitham Bou Ammar \(^b\)  Dietwig Lowet \(^c\)

Kurt Driesens \(^d\)  Antonio Liotta \(^a\)  Gerhard Weiss \(^d\)  Karl Tuyls \(^e\)

\(^a\) Eindhoven University of Technology, Dep. of Electrical Engineering, Netherlands  
\(^b\) Princeton University, Dep. of Operations Research and Financial Engineering, USA  
\(^c\) Philips Research, Dep. of Personal Health, Eindhoven, Netherlands  
\(^d\) Maastricht University, Dep. of Knowledge Engineering, Maastricht, Netherlands  
\(^e\) University of Liverpool, Dep. of Computer Science, Liverpool, UK

1 Introduction

Nowadays, there is an increasing need of autonomous agents (e.g., robots) capable to help elderly people in their daily life and to improve their life style by performing actions, such as monitoring and coaching. To fulfill these requirements, robotic support for people needs to be able to interact fully autonomously with humans and to understand properly what they are doing. This work focuses on the issues relating to automating the process of recognizing human activities.

2 The proposed method

Much research has been aimed at detecting human activities based on the output of a variety of low-power, low-bandwidth sensors, and power and pressure meters placed either around the home, or on-body \(^1\). The drawback of such an approach lies in the inability to capture sufficiently reliable data that allows to differentiate between subtly different activities. In principle, the most accurate and suited sensors for activity recognition would be video-cameras in combination with advanced computer vision algorithms to interpret the data, but this approach leads to significant privacy issues. As an alternative, we make use of motion capture data. More exactly, we use a Kinect\(^\text{®}\) sensor to generate a 3D point cloud and to extract the human skeleton joints from it. This approach yields relatively easy data to process and sufficient information to accurately recognize human activities.

Besides that, robotic agents should possess several cognitive capabilities, including: (1) the ability to recognize and classify human activities, (2) the ability to model and predict human motion, (3) the ability to realize when their predictions are wrong. Moreover, current machine learning techniques usually focus only on one of the previous three tasks. Aiming at a general unified framework that can simultaneously learn to recognize and classify human activities, auto-evaluate the classification results, as well as model and predict human motion, this paper proposes a novel machine learning model, namely Factored Four Way Conditional Restricted Boltzmann Machine (FFW-CRBM) \(^3\). FFW-CRBM extends state-of-the-art Conditional

\(^1\)The full paper has been published in *Pattern Recognition Letters*, Volume 66, 15 November 2015, Pages 100-108, ISSN 0167-8655, http://dx.doi.org/10.1016/j.patrec.2015.01.013.
Restricted Boltzmann Machine (CRBM) and Factored CRBM (FCRBm) [5] to a fourth order Restricted Boltzmann Machine consisting of: (1) a history, (2) a present or visible, (3) a hidden, and (4) a label layer. To enable classification and structured output prediction in one unified framework, the layers are connected using a factored four-way weight tensor, as depicted in Figure 1. Due to the complexity of the proposed model, the standard training methods for deep learning models are unsuited. As a second contribution, we introduce Sequential Markov chain Contrastive Divergence (SMcCD), an adaptation of Contrastive Divergence (CD) [2]. To illustrate the efficacy and effectiveness of the model, we performed two sets of experiments using real world data originating from (i) our smart companion robotic platform and (ii) a benchmark database for activity recognition [4]. For the sake of brevity, we present a snapshot with results from our prediction experiments on 6 physical exercise activities (i.e. EA1, ..., EA6) in Table 1 and 2. On the same data the classification accuracy of FFW-CRBm was 89.96 ± 5.38%, while Support Vector Machine (SVM) achieved just 74.93 ± 9.08% accuracy. In all the experiments performed, in terms of classification FFW-CRBMs outperformed widely used methods, such as SVM or K Nearest Neighbors, while in terms of human motion prediction, FFW-CRBMs performed better than standard models for multidimensional time series prediction, such as CRBMs or FCRBMs.

### Table 1: One step prediction errors.
The table shows the average RMSE values computed between the predicted human skeleton joints and the ground truth for physical exercise activities using FFW-CRBM, CRBM and FCRBM.

<table>
<thead>
<tr>
<th>Activities</th>
<th>CRBM</th>
<th>FCRBM</th>
<th>FFW-CRBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>EA1</td>
<td>0.110±0.005</td>
<td>0.054±0.002</td>
<td>0.028±0.008</td>
</tr>
<tr>
<td>EA2</td>
<td>0.138±0.003</td>
<td>0.036±0.006</td>
<td>0.018±0.012</td>
</tr>
<tr>
<td>EA3</td>
<td>0.106±0.012</td>
<td>0.044±0.021</td>
<td>0.023±0.011</td>
</tr>
<tr>
<td>EA4</td>
<td>0.126±0.011</td>
<td>0.094±0.008</td>
<td>0.027±0.014</td>
</tr>
<tr>
<td>EA5</td>
<td>0.125±0.004</td>
<td>0.068±0.011</td>
<td>0.026±0.009</td>
</tr>
<tr>
<td>EA6</td>
<td>0.123±0.026</td>
<td>0.093±0.007</td>
<td>0.048±0.019</td>
</tr>
</tbody>
</table>

### Table 2: Multi-step prediction errors.
The table shows the RMSE values computed between the predicted human skeleton joints and the ground truth and averaged over all physical exercise activities for various number of future steps.

<table>
<thead>
<tr>
<th>Future Steps</th>
<th>CRBM</th>
<th>FCRBM</th>
<th>FFW-CRBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 steps</td>
<td>0.116±0.007</td>
<td>0.048±0.016</td>
<td>0.038±0.021</td>
</tr>
<tr>
<td>20 steps</td>
<td>0.121±0.008</td>
<td>0.046±0.019</td>
<td>0.043±0.026</td>
</tr>
<tr>
<td>30 steps</td>
<td>0.121±0.011</td>
<td>0.049±0.021</td>
<td>0.047±0.017</td>
</tr>
<tr>
<td>40 steps</td>
<td>0.118±0.011</td>
<td>0.059±0.035</td>
<td>0.056±0.028</td>
</tr>
<tr>
<td>50 steps</td>
<td>0.119±0.012</td>
<td>0.086±0.102</td>
<td>0.059±0.027</td>
</tr>
</tbody>
</table>

### 3 Conclusion

This paper proposes a new machine learning technique for activity recognition and prediction. Thus, FFW-CRBMs together with an adapted training algorithm SMcCD are capable of: (1) classification, (2) prediction, and (3) self auto evaluation of their classification performance within one unified framework. The efficacy and performance of FFW-CRBm was demonstrated on real-world data acquired with our previously developed robotic platform for smart companions and on a benchmark dataset. Experiments showed that FFW-CRBMs are able to outperform current state-of-the-art machine learning algorithms in both, classification and regression.

### References


