Factored four way conditional restricted Boltzmann machines for activity recognition

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

Factored Four Way Conditional Restricted Boltzmann Machines for Activity Recognition[1]

Decebal Constantin Mocanu [a]   Haitham Bou Ammar [b]   Dietwig Lowet [c]

Kurt Driessens [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® 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.

Figure 1: Schematic architecture of the proposed FFW-CRBM showing the four layer configuration of the machine as well as the factored fourth order weight tensor.

Besides that, robotic agents should posses 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-CRBMs extend state-of-the-art Conditional
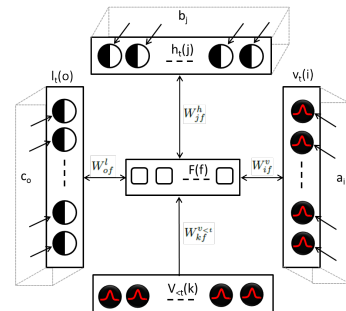
Activities	CRBM	FCRBM	FFW-CRBM
EA1	0.110±0.005	0.054±0.002	**0.028±0.008**
EA2	0.138±0.003	0.036±0.006	**0.018±0.012**
EA3	0.106±0.012	0.044±0.021	**0.023±0.011**
EA4	0.126±0.011	0.094±0.008	**0.027±0.014**
EA5	0.125±0.004	0.068±0.011	**0.026±0.009**
EA6	0.123±0.026	0.093±0.007	**0.048±0.019**

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.

Future Steps	CRBM	FCRBM	FFW-CRBM
10 steps	0.116±0.007	0.048±0.016	**0.038±0.021**
20 steps	0.121±0.008	0.046±0.019	**0.043±0.026**
30 steps	0.121±0.011	0.049±0.021	**0.047±0.017**
40 steps	0.118±0.011	0.059±0.035	**0.056±0.028**
50 steps	0.119±0.012	0.086±0.102	**0.059±0.027**

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 .

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 $89.96 \pm 5.38\%$, while Support Vector Machine (SVM) achieved just $74.93 \pm 9.08\%$ 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.

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

[1] Liming Chen, J. Hoey, C.D. Nugent, D.J. Cook, and Zhiwen Yu. Sensor-based activity recognition. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 42(6):790–808, Nov 2012.

[2] Geoffrey E. Hinton. Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, 14(8):1771–1800, August 2002.

[3] Decebal Constantin Mocanu, Haitham Bou Ammar, Dietwig Lowet, Kurt Driessens, Antonio Liotta, Gerhard Weiss, and Karl Tuyls. Factored four way conditional restricted boltzmann machines for activity recognition. *Pattern Recognition Letters*, 66:100 – 108, 2015.

[4] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy. Berkeley mhad: A comprehensive multimodal human action database. In *Proceedings of the IEEE Workshop on Applications on Computer Vision (WACV)*, 2013.

[5] Graham W. Taylor, Geoffrey E. Hinton, and Sam T. Roweis. Two distributed-state models for generating high-dimensional time series. *Journal of Machine Learning Research*, 2011.