Combining Neural Density Estimation and Computational State Space Models for Human Activity Recognition

Master Thesis

presented by
Benedikt Jakob Lasotta
Matriculation Number 1610104

submitted to the
Institute for Enterprise Systems
Dr. Stefan Lüdtke
University of Mannheim

September 2022

Contents

List of Algorithms

List of Figures

List of Tables

Chapter 1

Introduction

The task of recognizing human activities from sensor data is relevant for many real world applications. Its uses range from optimizing logistics in warehouses to providing healthcare and assistance systems for elderly and disabled people. Consequently, human activity recognition (HAR) has been an active topic of research in recent years. Methods to perform sensor-based HAR, broadly fall into either of two categories, each with their respective strengths and weaknesses: Symbolic methods as proposed in [3, 9] can employ prior domain knowledge and thus may need less training data, but often lack the capacity to learn from complex, time-series sensor data. One such approach are computational state space models (CSSMs) [3], which employ probabilistic inference on top of symbolic, rule based representations of possible actions within a domain to reason about observed activities. Contrary to this, modern, data-driven methods are capable of handling complex, multi-channel sensor data and have shown promising results in the past. An approach using a convolutional neural network (CNN) is described in [6]. The drawback of purely data-driven methods is, that they are oblivious to the causal structure of activities in the domain and thus may require more training data which is often hard to obtain. The goal of hybrid models is to combine both approaches in order to obtain state of the art performance while being able to incorporate prior knowledge to reduce the need for training data.

To benefit from the advantages of symbolic and data-driven methods, the main goal of this thesis is to develop a hybrid system for HAR by combining existing components in a novel way. The aim of this hybrid system is to achieve a greater sample efficiency than current state of the art systems. That is, with a fixed amount of limited training data, the hybrid system should outperform both data-driven and purely symbolic approaches. To build this hybrid model, an implementation of computational state space models (CSSM) [3] referred to as computational causal behavior models (CCBM) will be combined with an observation model that is obtained from sensor data via neural density estimation methods. Initial experiments will be conducted with a type of normalizing flows called masked autoregressive flow (MAF) [7] since it has shown promising performance on density estimation tasks, as shown in [2]. The system will be evaluated on a data set of raw sensor data from inertial measurement units (IMUs) of people performing the task of cooking a carrot soup, eating at, and cleaning up afterwards. It is available for download from the University of Rostock [**?**].

Additionally, an important research goal of this thesis is to examine which component contributes in what part to the final performance. To this end, an ablation study will be conducted. For this, the performance of the symbolic approach with a simple observation model, the density estimation component with a simple prediction, the full hybrid model, and a baseline of quadratic discriminant analysis (QDA) will be compared on the data set introduced above. The results of this ablation study can be found in chapter (insert ref).

1.1 Problem Statement

1.2 Contribution

1.3 Related Work

Intellectually closest to this thesis is the approach proposed by Rueda et al. introduced in [10]. They present the idea of using deep neural architectures as the observation model required by CSSMs. In particular the authors use a CNN to learn from time-series sensor data provided by inertial measurement units (IMUs), and predict the observed actions. In the line of this work, a similar approach will be taken. With the difference being, that the observation model is obtained by a neural density estimator like MAF. Which is subsequently combined with CSSMs to perform probabilistic inference. However, many other approaches to combine deep neural architectures with symbolic reasoning have been proposed in recent literature. DeepProbLog [5] is introduced as a framework which combines neural networks and probabilistic-logical modeling via the existing language ProbLog [8]. This is done by extending ProbLog with neural predicates. The evaluation of DeepProbLog shows that this framework is capable of performing symbolic reasoning as well as deep learning from examples in an end-to-end fashion. The downside of this approach is that its inference algorithms and language are ill-suited for dynamic systems, such as HAR. A recent approach that shows promising results for HAR is DUSTIN [1] which employs knowledge infusion on top of features extracted via a neural network. That is, the model uses a CNN to extract features from time series sensor data as well as the high-level context and concatenates to this representation a set of features which are obtained by a knowledge based reasoner. Just as for CSSMs, this has the added benefit that common-sense knowledge can be introduced to the model, which means that estimated action sequences are consistent with the user-context. The evaluation of DUSTIN shows that it is able to outperform other state-of-the-art neuro-symbolic approaches while having a high sample efficiency at the same time. Lastly, in [4] another hybrid architecture is proposed. This model combines high-level features extracted by a deep neural network with context information about process steps to obtain state-of-the-art HAR performance.

Chapter 2

Theoretical Framework

2.1 Preliminaries

2.2 Distribution Estimation

The task of estimating the joint probability distribution $p(x)$ from a set of examples $\{\boldsymbol{x}_n\}_{n=1}^{N}$ is a central topic for many machine learning applications. Let $\boldsymbol{x}$ be an example with D variables $x_1, x_2, ..., x_D$. The central fact is that any joint distribution of variables can be decomposed into a product of joint conditional probabilities, where variable x_d only depends on the prior variables $x_{1:d-1}$. This means that according to the chain rule of probability the joint distribution $p(x)$ can be expressed as

$$p(x) = \prod_{d=1}^{D} p(x_d | x_{1:d-1}) \tag{2.1}$$

2.3 Symbolic Reasoning

2.4 Differences and Similarities

Chapter 3

Algorithms

3.1 Computing the Good

3.2 Computing the Evil

3.3 Diagnosis

Chapter 4

Experimental Evaluation

4.1 Settings

4.2 Experiments

4.3 Results

Chapter 5

Conclusion

5.1 Summary

5.2 Future Work

Bibliography

[1] Luca Arrotta, Gabriele Civitarese, and Claudio Bettini. Knowledge Infusion for Context-Aware Sensor-Based Human Activity Recognition. In *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 1–8, June 2022. ISSN: 2693-8340.

[2] Ivan Kobyzev, Simon J.D. Prince, and Marcus A. Brubaker. Normalizing Flows: An Introduction and Review of Current Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3964–3979, November 2021. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[3] Frank Krüger, Martin Nyolt, Kristina Yordanova, Albert Hein, and Thomas Kirste. Computational State Space Models for Activity and Intention Recognition. A Feasibility Study. *PLOS ONE*, 9(11):e109381, November 2014. Publisher: Public Library of Science.

[4] Stefan Lüdtke, Fernando Moya Rueda, Waqas Ahmed, Gernot A. Fink, and Thomas Kirste. Human Activity Recognition using Attribute-Based Neural Networks and Context Information, October 2021. arXiv:2111.04564 [cs, eess].

[5] Robin Manhaeve, Sebastijan Dumancic, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. DeepProbLog: Neural Probabilistic Logic Programming. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[6] Fernando Moya Rueda, René Grzeszick, Gernot A. Fink, Sascha Feldhorst, and Michael Ten Hompel. Convolutional Neural Networks for Human Activity Recognition Using Body-Worn Sensors. *Informatics*, 5(2):26, June 2018. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[7] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked Autoregressive Flow for Density Estimation. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[8] Luc De Raedt. ProbLog: A Probabilistic Prolog and Its Application in Link Discovery. page 6.

[9] Miquel Ramirez and Hector Geffner. Goal Recognition over POMDPs: Inferring the Intention of a POMDP Agent. In *Twenty-second international joint conference on artificial intelligence*, page 6, 2011.

[10] Fernando Moya Rueda, Stefan Lüdtke, Max Schröder, Kristina Yordanova, Thomas Kirste, and Gernot A. Fink. Combining Symbolic Reasoning and Deep Learning for Human Activity Recognition. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 22–27, March 2019.

Appendix A

Program Code / Resources

The source code, a documentation, some usage examples, and additional test results are available at ...

They as well as a PDF version of this thesis is also contained on the CD-ROM attached to this thesis.

Appendix B

Further Experimental Results

In the following further experimental results are ...

Ehrenwörtliche Erklärung

Ich versichere, dass ich die beiliegende Master-/Bachelorarbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen. Ich bin mir bewusst, dass eine falsche Er- klärung rechtliche Folgen haben wird.

Mannheim, den 31.09.2022 Unterschrift