

Evaluation and Deployment of Models for Activity Recognition

Rita Pucci

University of Pisa, Department of Computer Science, pucci@di.unipi.it

Advisors: Alessio Micheli and Stefano Chessa at University of Pisa, Department of Computer Science.

Abstract. There is a growing need to monitor humans and animals in order to observe their behaviour. To understand animals in their natural environment, and to monitor children and the elderly, scientists rely on remote assessments. Automatic monitoring systems are a possibility to obtain direct observation of a subject. An automatic monitoring system can support biological and medical studies providing identification of the activity, without the intrusiveness of a human observer. AR is an emerging field but will soon be providing innovative solutions to many problems. In literature many models have been presented as a core part of an automatic monitoring system to recognise the activity of a subject. During my PhD, Machine Learning models were developed to detect physical activities of subjects using Activity Recognition (AR) techniques. Machine Learning models were programmed to autonomously identify activity patterns in accelerometer data. The versatility of Machine Learning models make them useful when managing monitoring activity where direct observation would not otherwise be possible.

Keywords: Artificial Neural Network, Machine Learning algorithm, Human activity recognition, Biologging, Sensors, Accelerometers

1 Introduction

AR deals with a broad range of applications in many areas connected to computer science. I focused on three areas where AR is requested to supply the need of monitoring. Ambient Assisted Living (AAL), HealthCare (HC), and Biologging have recently increased their interest in automatic monitoring systems. I started my PhD by familiarising myself with the field of AR. I then developed and applied different Machine Learning (ML) models to different case studies. This provided a comparison between models over the same datasets and vice versa. Using knowledge gained in the research phase, I choose libraries and tools useful in developing and analysing AR systems: Theano (Python library), Shark (C++ library), and NNTool (MatLab library). These libraries were used to verify results obtained over the same dataset and the same model. My research then moved towards managing the lack of common formats in raw accelerometer

datasets. This has prevented comparison among ML models. The evaluation of models over uniform datasets highlight the analysis of aspects such as accuracy, latency and processing resources required. The results obtained over the same dataset allow to analyse the trade-off between performance and obtrusiveness. This trade-off plays a crucial role in AR systems shift the focus of the research from the performance to the applicability of the system. In particular, I investigated and developed models providing a functional application for animal AR with the project Tortoise@ and compared the results of different models to objective of the Tortoise@ project. Lastly, I used the datasets to obtain a comparison among models and cases of studies.

2 Background

Literature shows that there have been several models designed to recognise activity in raw sensor data. The development of microelectronics described in [1] (hereafter a device) and computer systems made it possible for sensors and mobile devices to interact with people in their daily activities. Due to this, AR became a more attainable way of addressing problems in these fields. Research in AAL is grounded in Ambient Intelligence. Ambient Intelligence technologies allow people to overcome physical limits, and monitor daily personal and self-care activities ([3],[7]). This technology aims to improve a persons capabilities using a digital environment that is sensitive, adaptive, and responsive. A similar development was evident in the HC area where the research was increased and pushed by the still continuous request from the medical field. In fact the HC is a particular medical area of AAL. The human AR for the health care of elderly is a very active research area, which is recently proposed researches to prevent physical and psychological deterioration of people. A prior lack of research in Biologging left a vacuum filled by AR projects addressing biodiversity and conservation [8] – [9]. These Biologging projects provide autonomous systems for monitoring animals in the wild, and in particular, endangered species. The versatility of AR is growing but we currently lack a unifying model. Authors compare presented works in surveys and provide a uniform qualitative comparison. Unfortunately, the comparison of AR approaches is hindered by the fact that each model uses a different data-set. The specialisation of software for a specific dataset lends no development to the versatility of AR methods.

3 Methodology

In this context the AR problem, as described in [6], is defined as a temporal partition in intervals of time-sequence, hereafter called pattern, which are labeled concerning the activity performed. The patterns must be consecutive and non-empty assuming that they have fixed length. This defines the AR problem as to find a mapping function that can be evaluated for each pattern to obtain a classification as similar as possible to the actual activity performed. To deal with AR problem, I used automatic classifier structure. It consists in a filter stage and

a classifier stage. The filter stage pre-processes data to increase the classification performance. Even if the filter stage is a thorny stage, the classifier stage is the main and most important part of the system. The classifier stage is developed by a ML model that classifies each time sequence with a label activity.

The set of ML models presented in literature consists mainly in IDNN, SVM, ESN, and, in recent time, CNN. The IDNN is a subclass of Time Delay Neural Network introduced for speech recognition and specifically designed to treat sequential data. The inputs consist of the outputs of earlier nodes (as in multilayer perceptron) but not only of the current time slot, but also a number of previous time slots (it is called time delay). The basic units of IDNN have a delay introduced on the inputs which allows the model to relate the current input to the past history of events. The IDNN scans the input window over time so that the units implement the property of translation invariance, [11]. This property is recommended to deal with the dynamic nature of the sensors' sequences, the activity has to be recognised independently from a precise moment in time. I evaluate different structures of the model and three training algorithms: the Backpropagation algorithm [2], the Resilient Propagation algorithm (RP) [19], and the Levenberg-Marquardt algorithm (LM) [18].

The main idea of the SVM is to construct a hyperplane as the decision surface in order to have the maximum margin between positive and the negative patterns. Through the concept of margin maximisation the SVM represents an approximate implementation of the method of structural risk minimisation [10]. Our research emphasises the fact that the SVM learning algorithm is directly constructed using the kernel, and the inner product kernel (between support vectors). The kernel function allows the SVM to be applied either to linearly or nonlinearly separable patterns. Specifically we consider the Radial Basis Function (RBF), polynomial, and linear kernels [2].

The ESN model is investigated by Jaeger [12] and [13]. The ESN model is a Recurrent Neural Network (RNN) and is based on the Reservoir Computing (RC) paradigm. This paradigm provides the separation between the recurrent dynamical part, the reservoir, and the non-recurrent output part, the readout. Hence the ESN approach differs from the RNN in training phase where all weights are adapted. The fact of fixed recurrent connections among hidden units is the key feature of ESN, the encoding function which is implemented by the network is not adaptive. Because there are no cyclic dependencies between the trained readout connections, training ESN is a simple linear task. The effectiveness of Activity Recognition system based on the ESN has been validated on the EVAAL international competition [14].

The CNN is a biologically inspired architecture that can learn invariant features. The model is introduced by LeCun in [15]. The basic idea of CNN is to ensure a degree of shift and distortion invariance. These models benefit from good performance and low computational and memory requirements. Recently the CNN is proposed over accelerometer data.

The dataset is split at random in training and test subsets. The training subset is used to train each model, as well as to select the values for the hyperpa-

rameters, hereafter Model Selection (MS). Each model is also evaluated for the amount of memory resources needed, hereafter called Model Assessment (MA). MA focuses the attention over the applicability concerning the trade-off between performance obtained and the resources required. The automatic classifier structure it is evaluated both for the performance obtained and for the applicability. It is expected the identification of a trade off among hardware resources required, classification accuracy, and embedding design. The evaluation among settings for the two stages of the autonomous classifier shows that it can be possible to choose the setting concerning the interest of the research, higher accuracy or more generality of the model.

Concerning the deployment of activity recognition system, Tortoise@ is an autonomous system to identify the nest digging activity of tortoises using a device mounted atop the tortoises shell. An accelerometer, as well as temperature and light sensors, are embedded on a device MME called a MicaZ module. Accelerometer data was collected from devices of different tortoises during their two-month nesting. The device can discriminate between (nest) digging and non-digging activity (specifically walking and eating) over specific tortoise's movements by using an automatic system. The automatic system is modularly structured using an artificial neural network and an output filter. For the purpose of experiment and comparison, and with the aim of minimising the computational cost, the artificial neural network has been modelled according to three different architectures based on the input delay neural network (IDNN). All of them were developed in C and they are standard IDNN, IDNN with Local Receptive Fields (IDNN LRF), and the IDNN with Local Receptive Fields and Weight Sharing (IDNN LRF WS). The IDNN LRF takes the local receptive fields from CNN. Each hidden neuron scans the input using their local receptive field, units in a layer are connected in order to receive input from a set of units in a small neighbourhood of the previous layer. The IDNN LRF WS is still inspired by CNN with weight sharing (WS) among LRF hidden units. These further reduce the number of free parameters from the large amount of units sharing the same weight vector and obtains a certain level of shift invariance (the detection of features regardless of their position).

4 Discussion

During my Masters degree, and my PhD, I took part in the Tortoise@ project. Results of the three architecture of the IDNN are initially presented in [4] and extended in [5]. The evaluation of the proposed models, IDNN, IDNN LRF, and IDNN LRF WS, was performed to find a good trade-off between the performance and the applicability on a low power device. The performance measurements of the three models take into account the averages of errors computed on five different initializations of their weights for the ANN. The applicability it is evaluated considering the memory required for the storage of the weights of the IDNN. The highest performance, concerning the Tortoise@, is reached by the IDNN. IDNN provides an accuracy of 96.24% with a memory occupation of 1844 bytes.

The memory space required for the IDNN is less than 30% if it is used the IDNN LRF, 398 bytes with an accuracy of 95.51%. It is worth to bear 1% in accuracy less for a reduction of 70% of memory space. With the IDNN LRF WS the advantage is less affecting, the accuracy obtained is of 94.34% with 196 bytes of memory space. It is worth noting that to this memory space it is to add the space required for the filter stage. Tortoise@ is a starting point for many different AR project for animals. In fact, a similar automatic system classifier was developed to be embedded on a mobile phone as a rudimentary prototype. Tortoise@ is one of the first projects about an automatic system classifier for the classification of tortoise movements. The advantage of the ESN, SVM, and CNN is due to the possibility of analyse long sequences without any pre-processing phase. In relation to the project to save biodiversity in the last year of my Ph.D., I spent six months at the University of Queensland in Brisbane, Australia. During this period I started (ongoing) collaboration with the University of Queensland and with Macquarie University. We analyse ML models to identify prey capture activity in Little Penguins and in seals. Results obtained among the architectures based on the IDNN were compared to results obtained with SVM, ESN, and CNN over the same dataset. The results are still under analysis and will be published in a paper in this year. Concerning the Human AR, I participated in Palumbo et Al. [3]. We compared the results obtained with ESN and IDNN over a dataset of Received Signal Strength (RSS) and accelerometer data to recognise seven different activities. We selected daily activities: Bending, Cycling, Lying, Sitting, Standing, and Walking. The performance of the activity recognition system is assessed on a purpose-specific collected real-world dataset. Our results show that the proposed system reaches a very high level of accuracy while maintaining a low deployment cost. The average test accuracy of 0.944 with ESN is comparable with the performance obtained with IDNN of 0.923. Single activities are well recognized with accuracy score from 0.825 (Standing) to 0.999 (Cycling). However, it is worth noting that the Sitting and Standing activities are hard distinguishable, due to the nature of the RSS input signals used. Specifically, if we include the phases of sitting down and standing up the ESN system obtains good results, as reported in [16],[17]. For each case of study it is possible to identify the better compromise between performances and applicability, the IDNN is a good compromise between performance and applicability, ESN and CNN are mainly adapt for higher performances and are flexible enough to further customized architecture that observe the applicability limits.

5 Future Work

The problem of AR with humans and animals is still open and is growing up looking for new solutions mainly in AAL, and HC. The Biologging area is emerging and it is even more necessary now that the biodiversity crisis that characterised last decades leads to the decline and extinction of many animal species worldwide. Future proposals will continue developing complete systems that has potential to be applied to many different case studies, providing a serious possi-

bility for monitoring and protection of endangered species and for helping and support the elderly and children.

References

1. Francis, L A and Iniewski, K (2013) *Novel Advances in Microsystems Technologies and Their Applications* RC Press
2. Haykin, S (2009) *Neural networks and learning machines* Pearson Upper Saddle River
3. Palumbo, F and Gallicchio, C and Pucci, R and Micheli, A (2016) Human activity recognition using multisensor data fusion based Reservoir Computing. *IOS Press* 8:87107
4. Barbuti, R and Chessa, S and Micheli, A and Pucci, R (2013) Identification of nesting phase in tortoise populations by neural networks. *Extended Abstract The 50th Anniversary Convention of the AISB, selected papers* 6265
5. Barbuti, R and Chessa, S and Micheli, A and Pucci, R (2016) Localizing Tortoise Nests by Neural Network. *PlosOne*
6. Lara, O D and Labrador, M A (2013) A survey on human activity recognition using wearable sensors. *IEEE* 15:11921209
7. LBocca, M and Kaltiokallio, O and Patwari, N (2012) Radio tomographic imaging for ambient assisted living. *Springer* 108130
8. Block, B A (2005) *Physiological Ecology in the 21st Century: Advancements in Biologging*. *Science Integrative and Comparative Biology* 45:305-320
9. Kooyman, G L (2004) *Genesis and evolution of biologging devices: 1963-2002* 58:522
10. Vapnik, V and Levin, E and Le Cun, Y (1994) Measuring the VC-dimension of a learning machine. *Neural Computation* 6:851876
11. Waibel, A (1989) Modular construction of time-delay neural networks for speech recognition. *Neural computation* 1:3946
12. Jaeger, H (2002) Adaptive nonlinear system identification with echo state networks. In *Advances in neural information processing systems* 593600
13. Ozturk, M and Xu, D and Principe, J C. (2007) Analysis and design of echo state networks. *Neural Computation*, 19:111138
14. Botía, J A and García, J A À and Fujinami, K and Barsocchi, P and Riedel, T (2013) Evaluating aal systems through competitive benchmarking.
15. LeCun, Y and Bengio, Y (1995) Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361:255258
16. Álvarez-García, Juan Antonio and Barsocchi, Paolo and Chessa, Stefano and Salvi, Dario (2013) Evaluation of localization and activity recognition systems for ambient assisted living: The experience of the 2012 EvAAL competition. *Journal of Ambient Intelligence and Smart Environments* 5:1 119132
17. Palumbo, Filippo and Barsocchi, Paolo and Gallicchio, Claudio and Chessa, Stefano and Micheli, Alessio (2013) Multisensor data fusion for activity recognition based on reservoir computing. *International Competition on Evaluating AAL Systems through Competitive Benchmarking* 2435
18. Moré, Jorge J (1978) The Levenberg-Marquardt algorithm: implementation and theory. *Numerical analysis* 105116
19. Riedmiller, Martin (1994) *Advanced supervised learning in multi-layer perceptrons from backpropagation to adaptive learning algorithms*. *Computer Standards & Interfaces* 16:3:265278