Versione PROVVISORIA del contributo presentato al Convegno Annuale

DISCLAIMER

Questa versione dell'abstract non è da considerarsi definitiva e viene pubblicata esclusivamente per facilitare la partecipazione del pubblico al convegno AIUCD 2021

Il Book of Abstract contenente le versioni definitive e dotato di ISBN sarà disponibile liberamente a partire dal 19 gennaio sul sito del convegno sotto licenza creative commons.

Analyzing the Sensor Data Stream for Monitoring and Visualization of Early Autism Signs (MoVEAS)

Mariasole Bondioli[1], Michele Zoncheddu[1], Stefano Chessa[1], Antonio Narzisi[2], Susanna Pelagatti[1],
[1] Dipartimento di Informatica, Università di Pisa, Largo Pontecorvo, 3 56127 Pisa (Italy)
[2] IRCCS Stella Maris, Calambrone (Pisa)
susanna.pelagatti(«»)unipi.it

ABSTRACT – Observing children affected by autistic spectrum disorders while they play is an important and widely used diagnostic method, that is conducted in a clinic under the supervision of a specialist. In this work we present a prototype of a sensorized toy that can classify its own movements (as they are applied by a children), to allow an observation of the play of the children even without the direct supervision of the specialist. This would allow to perform such tests even in environments more familiar to the children (at home, at school), in a non-invasive way.

KEYWORDS – e-health, internet of things, sensors, autistic spectrum disorders, smart toys

1. INTRODUCTION

MoVEAS (acronym of Monitoring and Visualization of Early Autism Signs) is a project that aims to detect, in a non-invasive way, early signs of autistic spectrum disorders (ASD), by monitoring play activities in young children through sensorized toys and classifying the activities through a neural network [1, 2]. Observing children while playing is a widespread technique for diagnosing ASD, especially in early age [3], and videotaped play session of groups of children with typical development, against groups of children with ASD, have been proven to be useful for the diagnosis [4, 3, 5].

Children with ASD use the toys in atypical and restricted ways (e.g. spinning), play with repetitive behaviors, and visually explore objects in unusual ways [6, 3]. Diagnosing ASD can be difficult, since there are no simple medical tests to diagnose the disorders, and the spectrum width makes it even harder.

This project's aim is to simplify the specialists' diagnosis process, by identifying ASD-related behaviors in non-clinical settings (home, school…) and allowing to indirectly monitor a play session. In particular, we present the MoVEAS project and discuss its activity recognition component based on machine learning that is aimed at classifying the movements applied to the toy by the children. We discuss the selection of the relevant data to classify the movements, the analysis of the consistency of the data obtained and the performance of the final model in a real scenario simulation.

2. THE MoVEAS PROJECT

The observation of children while they play is a commonly carried out on at the clinic with standard set of toys. This setting for the observation is not ideal in that the environment and toys used are totally new for the child. The idea of MoVEAS is to embed sensors in the toys used in standard evaluation sessions, in order to let the clinicians to collect data more freely, even at home or at school. In this way, they can monitor the play changes when toys become more familiar to the child.

In this perspective, the system reference scenario is based on the interaction steps between the toys moved by children, the server collecting and elaborating the data of the children's actions and the user interface, where the server gives the data back to the user (usually the clinician). The steps are the following: (1) The user accesses his/her private interface with a normal browser; (2) To start a new data collection session, the user selects a patient and the related toys and presses the "start" button; (3) The corresponding devices automatically acquire from the sensors (and send to the server) the data of the children's movement during the play session; (4) The *sensor detection module* detects the shift of the toy, elaborates the data and recognizes a specific movement; (5) The server communicates the information to the application web pages, where the data are visualized and analyzed for the diagnosis goals.

During this process, the server provides for a punctual data storage in its database, constantly receptive to the entry information from the software of the motion capture devices. The server also returns the processed information to the user through a web user interface. Through this interface, a user can also perform all management operations like create/remove a new patient in/from the database, connect a patient with a toy, start/end a new capture session etc.. This recruitment of children for this study was conducted according to the Helsinki Declaration.

3. THE ACTIVITY RECOGNITION COMPONENT

The activity recognition component operates in two steps. The first step is a data fusion algorithm (Fig. 1) that extracts from the 3-axis accelerometer, gyroscope and magnetometer time series two information for the user, which are the estimation of the orientation of the and the estimate of the intensity and direction of the movements of the toy. The two information in output are produced for each new tuple of data samples in input and hence form two time series. These time series, along with the raw data time series are in turn input to the second step based on neural networks that classify the time series as recognized movement of the toy.

For what concerns the data fusion algorithm, note that a correct estimation of the orientation and movement is a rather complex task. Each sensor taken individually (either accelerometer, gyroscope or magnetometer) is insufficient to compute either the orientation or the movement of the toy, for several reasons. The 3-axis accelerometer measurements have limitations in the detection of position in dynamic conditions and it is also impossible to estimate the orientation of the horizontal plane of the azimuth with reference to Earth's gravity (to which this sensor is subject). The estimate of the orientation can be improved by using the gyroscope, which, however, suffers of drifts in its measurements. To overcome the limits of the gyroscope we thus also the use a magnetometer (whose measurements refer to a global reference system given by the Earth’s magnetic field). The estimate of the toy orientation (in quaternion representation) is performed by fusing all these data by means of the Madgwick algorithm [7]. The estimation of the movement is obtained from the accelerometer data, from which is detracted of the contribution of the Earth’s gravity along each axis. In turn, this contribution is estimated on the bases of the orientation of the toy computed by the Madgwick algorithm. To make this estimation more stable, the accelerometer data is preprocessed with a standard Kalman filter. Both data of orientation and movement estimation are stored in a JSON record structured in two parts. The first contains the quaternion representation of the orientation , the raw acceleration along the three axes and the estimation of the power of the acceleration signal (this is used in the 3D real-time visualization), and the second part contains the orientation of the toy expressed in the Euclidean angles of roll, pitch and yaw, and the movement estimation in terms of magnitude and angles of direction (this is used for the storage and for the visualization of the movement diagrams).

The second step of classification of the movements of the toy analyses the time series of raw and pre-processed data by finding specific patterns. Not all patterns need the same recording time: for example, the duration of the movement of the toy pushed forward is much smaller than the one where the toy is thrown against a pillow – where the toy is thrown by the hand first, then flies in the air and finally hits the pillow. Our goal is to be able to recognize many different variable length sequences. This accomplished using a neural network.

In our work, we analyzed two solutions. The first one is a time delay neural network (TDNN), chosen for its simplicity (the training sequences are short, 14 samples but a complex model might easily go to overfitting); the second one is a recurrent neural network (RNN), chosen for its ability to recognize short sequences. To our purposes, both networks have been trained with the same dataset of fixed-length sequences. With TDNN, the classification is simply made by a window of size 7 that slides over the entire session’s data. For model training and validation, the dataset was composed of: 100 patterns with the toy going forward and 100 with the toy going backward; 200 patterns of simulated flight; 200 patterns with the still toy; 200 patterns of the toy carried while walking; 200 patterns of the toy thrown against a pillow. The training and validation portions were respectively the 60% and the 40% of the total, and the entire dataset is shuffled before every training. The test set was composed of 240 samples, disjoint from the training and validation ones. For what concerns the overfitting detection, this was based on the accuracy on the validation set: when the increment of accuracy on training did not match an increment on validation, the model was discarded. For both the TDNN and RNN has been chosen the model with the highest validation accuracy: more precisely, the model selection for the TDNN has been performed with a grid search on the number of filters and size of the kernel, while for the RNN the main hyperparameter was the number of units of the single recurrent layer.

In TDNN, model selection started from the simplest topology: an input layer with data normalization, a convolutional hidden layer with a ReLU activation function, that performed better than tanh, a max-pooling layer, that showed better results than the average-pooling one, and finally an output layer for the pattern classification, with a softmax function, to obtain values between 0 and 1, representing the probability of the classified pattern to belong to each possible class.

Only six features were considered relevant for the training: the gravity-free accelerations values in the three axes and the gyroscope raw data – the gyroscope data is used to discriminate some cases that are not well distinguished only by the acceleration. Some tests showed quickly that adding more features, like orientation, velocity data, or raw sensors data, made the network generally harder to train, and most often gave no advantage at all in the overall accuracy.

The search for the two most important hyperparameters, the kernel size and the number of filters, has been performed through a grid search. For each table's entry, the network has been tested 10 times, with shuffled data between training and validation set for each run. Outliers have been omitted from the average.
The values 9 and 7 were optimal respectively for the number of filters and the kernel size: they gave the best result in the validation set, and in particular:

- a number of filters up to 7 and a kernel size up to 5, made the network able to classify correctly only after a long and unstable training;
- from 7 filters upwards, the training curve was stable, and the overfitting occurred only from 15;
- kernel sizes from 9 upwards tended to overfit the network: the patterns were initially made of 22 samples, and the sliding window was be too big;
- with filters bigger than 20, the overfitting was mitigable only with very low kernel sizes, but in that case the network was hard to train, and very easy to underfit.

More complex topologies were tried, but adding hidden layers only led the neural network to the overfitting. The best training curve was obtained using 16-patterns long mini-batch.
The next step was to achieve the maximum accuracy with less possible data. The network was trained several times with different patterns' lengths, and the best choice for a stable and accurate training has proved to be 14 samples per pattern, that with the 22 Hertz sampling means 636 milliseconds long patterns. Up to 11-13 samples, the network was not able to achieve good performance in training, being hard to train and underfitting, while from 17 samples and more, the network fitted the noise, so overfitted. The convolutional layer's weights are initialized randomly in the uniform range [-0.000001, 0.000001]. To keep under control the network complexity, a weight decay approach has been used, with tested values in the range [0.001, 0.05]. The optimization algorithm chosen for the learning was Adam, because of its efficiency [20]. As done with the TDNN, the input passes through a normalization layer before going to the recurrent layer, and the activation function is ReLU for all layers, except for the last one, that uses the softmax function, particularly suitable for classification.
The best topology, that ensured a good fitting without instabilities during the training, is shown in the following schema: one recurrent layer and two densely connected layers, with two dropout layers before them. In particular, the number of nodes in the recurrent layer was crucial for the model selection:

- up to 5 units, the network was able neither to learn the training samples nor generalize;
- up to 10 units, the validation accuracy reached the 90%, but when the training continued, it overfitted;
- up to 12 units, the network was hard to train, and with 14 units the network was trained smoothly and the validation accuracy grew up to 93.5%;
- from 16 upwards, stricter regularization was needed, but the dropout layers made the training stable and avoided the overfitting. In the end, the best number of units was discovered to be 20.

The explored range of dropout for both layers was [0, 0.5]. The network was trained with a mini-batch size of 16 patterns.
Adding features besides the gravity-free acceleration and the gyroscope data gave most often no advantage at all, and sometimes made the training harder and unstable, also with more layers and more recurrent units. In this model, reducing the number of samples has not improved the validation accuracy as for the TDNN, but the performance significantly decreased only for less than 14 samples per pattern, accordingly to the results achieved with the TDNN.

4. EXPERIMENTAL RESULTS

The results obtained on the dataset are shown in Figures 2 and 3. We observe that TDNN provides a better validation accuracy and a smaller loss value. Moreover, in almost every test, showed smoother training and validation curves, and was easier to train effectively. The high accuracy obtained in validation and testing might not reflect the real network's performances in analyzing real play sessions: in the first two cases, the patterns have a well-defined start and end, while in a play session there is a continuous stream of data; for the classification, the network is given a fixed length subsection

Figure 2. Accuracy and F1 score for a) TDNN b) RNN

of that stream for each 22nd of second, and the subsection's bounds slide forward, like a sliding window, through the whole stream. When subsection contains only a single pattern, the networks behaves exactly like in the validation/testing scenario, but when there is an overlapping between two patterns (e.g. the toy stops flying and stays still), the classification is harder, because that's not a case that the network was trained to deal with.
For this reason, we performed a further experiment with data from a 46 seconds-long session of simulated play. The network was able to accurately recognize the still, forward, and backward patterns, while in flight and especially walking, the overall classification is correct, with some very short misclassifications. We plan to investigate this issue in detail with real play session recordings.

5. RELATED WORK

Previous research has established the connection between the way in which children interact with objects and the potential early identification of children with ASD. Those findings motivate our own work to develop "smart toys," objects embedded with wireless sensors which are safe and enjoyable for very small children, that allow detailed interaction data to be easily recorded. The ways in which infants and toddlers play with objects can be indicative of their developmental progress. Depending on their age, a child's object play activities can display simple physical milestones such as placing objects in their mouth to sophisticated cognitive tasks such as symbolically using a banana as a telephone receiver. In fact, observing the subtleties surrounding the way in which infants play with toys may highlight early indicators for developmental delays, such as autism. To better quantify how play may serve as an early indicator for autism, researchers have conducted studies examining the differences in object play behaviors among infants.
The work in [8] demonstrated that lateral glances toward objects were significantly more common in a group of children with autism (mean age 44 months) than in children with typical development. [9] found that children with autism (mean age 33 months) who spent a large proportion of time in restricted object use showed poorer joint attention, imitation, and social engagement abilities. Finally, [10] reported group differences in repetitive movements with objects (defined as 'at least 3 consecutive'). From a technological point of view, there has been a growing interest for using technology to ease several aspects of diagnosis and treatment of ASD people [11]. Automatic detection of stereotypical movements in children with autism using sensors has been investigated by Goodwin et al [12]. Their experiment involved six children and each child wore three wireless accelerometers, one on the left wrist, one on the right wrist, and one around the chest, with no restriction in movement. Session were taken at school or during therapy in situations that usually triggered stereotypical behavior. Each session was video recorded, and specialists annotated them both online and offline using an annotation software. A classifier algorithm was trained to automatically detect repetitive behaviors from the accelerometer readings with good results. Plotz et al. [13] used a commercially available device with a 3D accelerometer and a microcontroller combined with machine learning techniques to detect and classify anomalous behaviors. Other approaches [14, 15] used Kinect to recognize stereotypical movements. In [15], the Kinect camera was used to film 12 actors performing three separate stereotypical motor movements each. Then, a software recognizing these gestures was developed and instructed using machine learning approach. Manual grading was used to confirm the validity and reliability of the software.
Our approach differs from the above ones, since we aim at designing a device that is not invasive (does not requires wearable devices or cameras), and can be used even in the private houses, without the supervision of specialists. To this purpose our design exploits a state of the art IoT architecture, which is not present in the above works. In this sense we can highlight how several combinations of the three sensors introduced in our work have been already largely exploited in IoT literature regarding the Human Activity Recognition. In this field, in fact, we can find work as that of Najafi et al.s [16] in which the authors presents a new method of physical activity monitoring based on a 2D accelerometer and a gyroscope fused in a sensor systems to be put on the front of the chest, which is able to detect body postures (sitting, standing, and lying) and periods of walking in elderly persons using only one kinematic sensor attached to the chest.
Moreover, we highlight a wide number of human Activity recognition researches based on a single accelometer system, as those in which the researchers attempted to distinguish walking on level ground from walking on a stairway using waist
acceleration signals [17] and the studies focused on the design of an algorithm for analyzing and classifying human activity (as standing, sitting down with lowering subjects head, sitting down and leaning against, lying down straight, lying upside down, walking, going up/down stairs, running) using a body-fixed triaxial accelerometer on the back [18].
The accelerometer is in fact in literature defined as a sensor that shows promise in providing an inexpensive but effective means of long-term ambulatory monitoring (in most of the case treating elderly patients) [19].

6. CONCLUSIONS

We discussed the MoVEAS project focusing on its new component of activity recognition. In particular, we discussed the choice of the relevant data to classify the movements, on the consistency of that data over time, on the effective training and comparison between two neural network models, and on the performance of the deployed model in a real scenario simulation. Despite the good results achieved in this experimental phase, only working with children with ASD can reveal the actual project's usefulness. In this perspective, the next experimental phase is already planned, with a sample of already diagnosed ASD children and a control group of neurotypical children. This phase will also have the objective of devising and assessing a suitable diagnostic protocol that may make use of MoVEAS.

In the future, we will also address some current limitations of the software. For the movement classification, it might be useful to "flatten" the small spikes that affect the continuity of a classification, in order to obtain compact chunks to highlight with different colors in the session's scrollbar. Another path to be explored is to train the recurrent neural network with a variable-length patterns training set, and to find the best way to choose dynamically the length of the session's subsequence to feed the network with, in order to exploit the full network's capabilities.

7. REFERENCES

[1] Mariasole Bondioli et al. "Capturing Play Activities of Young Children to Detect Autism Red Flags". In: Jan. 2020, pp. 71–79. isbn: 978-3-030-24096-7. doi: 10.1007/978-3-030-24097-4_9.

[2] Marco Lanini et al. "Sensorized Toys to Identify the Early 'Red Flags' of Autistic Spectrum Disorders in Preschoolers". In: Jan. 2019, pp. 190–198. isbn: 978-3-030-01745-3. doi: 10.1007/978-3-030-01746-0_22.

[3] Sally Ozonoff et al. "Atypical object exploration at 12 months of age is associated with autism in a prospective sample". In: Autism 12.5 (2008). PMID: 18805942, pp. 457–472. doi: 10.1177/1362361308096402.

[4] Grace T. Baranek et al. "Object Play in Infants With Autism: Methodological Issues in Retrospective Video Analysis". In: American Journal of Occupational Therapy 59.1 (Jan. 2005), pp. 20–30. issn: 0272-9490. doi:10.5014/ajot.59.1.20

[5] Amy Wetherby et al. "Early Indicators of Autism Spectrum Disorders in the Second Year of Life". In: Journal of autism and developmental disorders 34 (Oct. 2004), pp. 473–93. doi: 10.1007/s10803-004-2544-y.

[6] Cornelia Bruckner and Paul Yoder. "Restricted object use in young children with autism Definition and construct validity". In: Autism : the international journal of research and practice 11 (Apr. 2007), pp. 161–71. doi: 10.1177/1362361307075709

[7] Sebastian O.H. et Al., "Estimation of IMU and MARG orientation using a gradient descent algorithm", IEEE Int. Conf. on Rehabilitation Robotics, ETH Zurich (CH), 2011.

[8] P. Mitchell, S. Parsons, and A. Leonard. Using virtual environments for teaching social understanding to 6 adolescents with autistic spectrum disorders. Journal of autism and developmental disorders, 37(3):589–600, 2007.

[9] T. C. Pagani Britto and E. Brigante Pizzolato. Towards web accessibility guidelines of interaction and interface design for people with autism spectrum disorder. In Ninth International Conference on Advances in Computer Human Interactions - ACHI 2016, Venice, 2016. 65, 104, 106

[10] Z. E. Warren, Z. Zheng, A. R. Swanson, E. Bekele, L. Zhang, J. A. Crittendon, A. F. Weitlauf, and N. Sarkar. Can robotic interaction improve joint attention skills? Journal of autism and developmental disorders, 45(11):3726–3734, 2015. 42, 126, 127

[11] M. C. Buzzi, M. Buzzi, B. Rapisarda, and M. Tesconi. Teaching lowfunctioning autistic children: Abcd sw. EC-TEL, 2013. 51, 56, 127

[12] N. Goncalves, J. Rodrigues, S. Costa, and F. Soares. Automatic detection of stereotyped hand flapping movements: Two different approaches. Pages 392–397, 2012. 23, 84, 127

[13] C. Pilebro and Birgitta B. Backman. Teaching oral hygiene to children with autism. International Journal of Paediatric Dentistry, 15(1):1–9, 2005. 62

[14] N. Goncalves, S. Costa, J. Rodrigues, and F. Soares. Detection of stereotyped hand flapping movements in autistic children using the kinect sensor: A case study. In 2014 IEEE international conference on autonomous robot systems and competitions (ICARSC), pages 212–216. IEEE, 2014. 42

[15] N. Josman, H. M. Ben-Chaim, S. Friedrich, and P. L.Weiss. Effectiveness of virtual reality for teaching street-crossing skills to children and adolescents with autism. International Journal on Disability and Human Development, 7(1):49–56, 2008. 41, 127

[16] M.Sekine, T. Tamura, T. Togawa, and Y. Fukui. Classification of waist acceleration signals in a continuous walking record. Medical engineering & physics, 22(4):285–291, 2000. 85

[17] M. Monibi and G. R. Hayes. Mocotos: Mobile communications tools for children with special needs. pages 121–124, 2008. 60

[18] F. Albinali, M. S. Goodwin, and S. S. Intille. Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum. pages 71–80. ACM, 2009. 48

[19] A. Abran, A. Khelifi, W. Suryn, and A. Seffah. Usability meanings and interpretations in iso standards. Software quality journal, 11(4):325–338, 2003. 106

[20] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: International Conference on Learning Representations (Dec.2014).