Automatic Autism Spectrum Disorder Detection Thanks to Eye-Tracking and Neural Network-Based Approach

Romuald Carette[1(✉)], Federica Cilia[2], Gilles Dequen[1], Jerome Bosche[1], Jean-Luc Guerin[1], and Luc Vandromme[2]

[1] Laboratoire Modélisation, Information et Systèmes, Université de Picardie Jules Verne, Amiens, Picardie, France
{romuald.carette,gilles.dequen,jerome.bosche, jean-luc.guerin}@u-picardie.fr

[2] Centre de Recherche en Psychologie: Cognition, Psychisme et Organisations, Université de Picardie Jules Verne, Amiens, Picardie, France
{federica.cilia,luc.vandromme}@u-picardie.fr

Abstract. Autism spectrum disorder (ASD) is a neurodevelopmental disorder quite wide and its numerous variations render diagnosis hard. Some works have proven that children suffering from autism have trouble keeping their attention and tend to have a less focused sight. On top of that, eye-tracking systems enable the recording of precise eye focus on a screen. This paper deals with automatic detection of autism spectrum disorder thanks to eye-tracked data and an original Machine Learning approach. Focusing on data that describes the saccades of the patient's sight, we distinguish, out of our six test patients, young autistic individuals from those with no problems in 83% (five) of tested patients, with a results confidence up to 95%.

Keywords: Neural network · Long Short-Term Memory (LSTM)
Data processing · Eye-tracking · Autism spectrum disorder · eHealth

1 Introduction

The notion of autism spectrum disorder implies different profiles in this disorder, shared by 1% of the world's population. Even though it is possible to detect at an early age, the methods used for this detection are, to date, mainly manual like parents interview or observations. As a decision support, a system called *eye-tracker* allows the reading of patients sight on a screen. It displays a set of information including the pixel each eye is aiming at, the gaze state, the position of patient's eyes, the pupil size and position, etc. As raw data, the system considers a set of three states: fixation, saccade and blink. While "fixation" and "saccade" represent sequential data with eye-focus staying into an area

The work presented in this paper is supported by Evolucare Technologies grant.

M. U. Ahmed et al. (Eds.): HealthyIoT 2017, LNICST 225, pp. 75–81, 2018.
https://doi.org/10.1007/978-3-319-76213-5_11

for the first, continuously moving for the latter, "blink" is quite different [1]. This state indicates the lost track of the eye, without knowing whether the patient actually blinked or just turned his/her head. It has been found that the dispersion of the patient's eye focus helps to indicate whether he/she is affected by an autism spectrum disorder. Research points towards this idea to find applicable solutions in diagnosis. These types of solutions remain manual, requiring data reading and expertise to be sure of a patient's situation. We propose an automatic approach based on LSTM neural networks and focusing on the saccade part of the readings. Our resulting solution ends up validating the state of 83% of the patients. Section 2 focuses on the autism spectrum disorder and the eye-tracker system. Section 3 provides a description of Neural Networks. In Sect. 4, we describe our approach and Sect. 5 provides some experimental results. We finally conclude and provide some future works in Sect. 6.

2 Autism Spectrum Disorder and Eye-Tracker System

In this section, we will state what autism spectrum disorder is, present the complexity of its diagnosis, and the eye-tracker system considered in our situation.

2.1 Autism Spectrum Disorder

Autism spectrum disorders are expressed notably through limited capacities for communication and social interaction, linked to difficulty for making eye contact with others. These difficulties can be demonstrated, for example, by analyzing family films, which highlight an abnormal visual behavior towards others, in persons with autism spectrum disorders [2], as well as a preference for non social stimuli over social stimuli such as faces [3]. This preference may be explained by a specific deficit in processing faces [4], leading to recognition disorders [5]. However, deficits at a lower level, such as perceptive and/or attentional ones [6] put a question mark over the indispensable prerequisites in building these processing capacities in the child with autism spectrum disorders. In this regard, systematic studies of visual activity in persons with autism spectrum disorder, carried out using recordings of visual activity (eye-tracking), regularly highlight unadapted visual exploration strategies [7]. Nonetheless, heterogeneous or even contradictory results do not allow any conclusions to be reached concerning the quality of visual exploration of faces by persons with autism spectrum disorders. While specific patterns of visual activity depending on groups are sometimes demonstrated, they are not systematically replicated: the results could be dependent on the age of the participants, on the severity of their disorder and/or on their cognitive ability, but also on the type of stimuli used (static versus dynamic) or on the task required (free exploration, recognizing faces or emotional expressions, etc.) For a review, see Falck-Ytter and von Hofsten [8]. Results of studies using dynamic social stimuli come out for the most part in favor of atypical visual exploration of participants with autism spectrum disorders compared to control participants, notably with a diminution in eye fixation time [9].

2.2 Eye-Tracker System

Oculometry (eye-tracking or gaze-tracking) combines a group of techniques allowing eye movements to be recorded. Combined with pupillometry, these techniques are currently used in a variety of fields such as psychology and psycholinguistics, as well as experiencing a considerable momentum in the field of neurosciences and physiology. In the field of autism, these non-invasive techniques are currently used to improve diagnosis and therapeutic strategies through a better understanding of the physiopathological and psychological mechanisms which underlay the observed symptoms. This technique allows the ocular activity of an individual to be analyzed, demonstrating where his eyes are fixing, and thus revealing what he is and what he is not looking at. It enriches the more traditional methods of observation by allowing assessment of what a person is looking at while carrying out a cognitive and/or relational activity. It is a tool among others allowing the assessment and understanding of human activity as well as the analysis of a percept.

Within this study, visual activity has been recorded with a sampling frequency of 60 Hz using the SMI RED250mobile IView XTM RED device.

3 About Neural Networks

Neural Networks are state-of-the-art tools from the Artificial Intelligence field. Mimicking the functioning of the human brain, it takes given data as input and results in the form needed by its creator as output, transitioning the information though nodes and weighted connections (like brain cells and synapses). After a step of training (or learning), the network is ready to interpret any set of additional data to give the expected answer. That answer can be of many forms: a prediction, a classification or just a correlated value.

Many different implementations of Neural Networks exist. From Artificial Neural Network (ANN, Fig. 1a), Recurrent Neural Networks (RNNs, Fig. 1b) have been created. Due to RNNs limitations, LSTMs have been proposed. More robust than their past implementation, LSTMs are also harder to train [10]. A

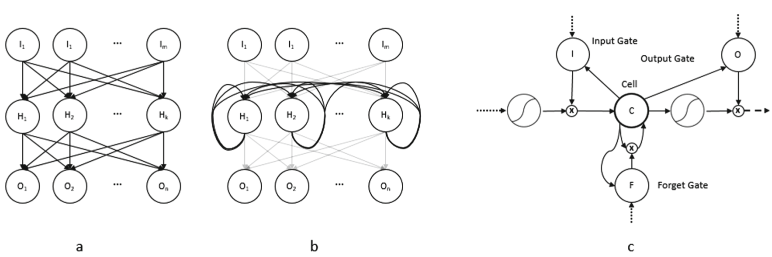

Fig. 1. (a) A ANN with m input nodes, n output nodes and a single hidden layer of k nodes ; (b) Same as (a), but as a RNN, recurrent connections are bold ; (c) A LSTM node, the inputs are dotted, the output is long-dotted

LSTM follows the implementation of a classic RNN, except from some specific nodes. Said nodes are now called cells as they are more complex: a cell is preceded and followed by other nodes (fed respectively by an input gate and an output gate), one more node is used as memory eraser (fed by a forget gate). The gates, as well as the entrance of the cell, are all given the same set of data. One single output then gets out of the ensemble. The cell is connected to every gate. In comparison with a classic node, a cell is itself constituted of more connections (see Fig. 1c).

4 Our Approach

The training and test patients are from 8 to 10 years old. We have their data divided between two classes, namely autistic (class 1) and typical (class 2). We studied a total of 17 children from class 1 and 15 from class 2. The first step of work is to capture the data from the patients. To do that, each child is placed in front of a prerecorded video of a joint attention offer, staring a person who performs some actions: showing or speaking about a balloon, which may visible or not. Every child is shown the same order of actions. Each action should lead the child to look at some element of the video. If it is visible, the child should focus on it, else, he/she may search on the screen for the element. All tracking information are recorded by an eye-tracking system. The data we choose to focus on is the saccadic movement of the eye. This includes the amplitude, the acceleration (average and maximum), the deceleration (maximum), the velocity (average and maximum), the rate of maximum velocity for a given saccade and the duration of the saccade. Gathered data is used anonymously for privacy reasons. Instead we use a color as name. To be easily usable in our specific LSTM network, data is normalized so that each value is included in [0,1]. Data is divided between two sets, training and testing data (resp. about 75% and 25%). Test data consists in 4 patients from class 1 and 3 patients from class 2. We implement our solution within Pybrain module[1] framework, using the LSTM Layers it provides. Due to the sequenced data we use, we work with a network made of LSTM hidden layers. The hidden layers consists of 2 LSTM layers composed of 20 nodes each. These nodes amounts are high enough to provide us valid results, while being low enough to be easily trained by the limited hardware we have access to. Input and output layers are linear layers. Given the data we have at our disposal (see beginning of Sect. 4), 7 nodes are used as input, while we use 3 nodes as output: class 1, class 2 and undetermined. Undetermined is not used, except to express indecision of the network and seems to enhance the training phase. The network is trained using our training data set. Time needed to reach strong confidence of our neural network is about 2 days of sequential computation on Intel Core i7 3, 6 GHz. We save the state of the network at regular fitness milestones, from 1.0 and each time the fitness value decreases by 0.001 after that. A state includes the fitness milestone reached, the evaluations for the test data set, the validation rate, the average certainty of the

[1] http://pybrain.org/ version 0.3.

evaluations and the current values of the weights of the network (see 5 for said values, excepting weights). We end the training of the network at a 0.004 fitness value. Following the training on the initial patients data, additional patients are used for confirmation of the training. The same information is saved.

5 Experimental Results

We extract the network states with a fitness value from 0.008 to 0.004. We apply this network state to six additional patients to check its validity. Patients are classified into different learned classes: while DarkBlue/DarkCyan/CyanTRIS are TS (diagnosed autism), Cyan/CyanBIS/DarkRed are TC (typical children with the same chronological age without autistic history). The calculation of the TC or TS probabilities are computed as follows: for a given patient, our network returns a value for class 1 (TS), class 2 (TC), class 3 (undefined). Each of these three values are in [0,1]. Then, the probability of class i can be retrieved by dividing the output i by the sum of all outputs. For example, if results are 0.2, 0.55 and 0.05, the patient is recognized as 25% TS, 68.75% TC and 6.25% of pure uncertainty. To make the results more readable, only the highest probability is shown for each case in Table 1. Also, the last column indicates the average confidence on the valid result, thus it is not a mean of the listed confidences. Indeed, on mistakes, the valid result confidence can be very low.

Also, Table 2 shows the confusion matrix of the results from the test data.

We observe that, from fitness values 0.008 to 0.006, 5 of the six patients' diagnosis are confirmed. When the fitness value reaches 0.005, the network suffers from a over-fitting issue. Indeed, we observe that the test data for DarkCyan patient becomes uncertain at 0.005, then gets into the wrong class at 0.004. The confidence then falls to 67%. We notice the highest confidence of the network at over 90% on average, 98% at best (for 0.006 fitness value and 5 over 6 valid

Table 1. Additional patients results

	TS children			TC children			
Fitness	CyanTRIS	DarkBlue	DarkCyan	Cyan	CyanBIS	DarkRed	Avg. Conf.
0.008	TS (88%)	TS (79%)	TS (94%)	TC (70%)	TS (92%)	TC (77%)	67.3%
0.007	TS (88%)	TS (86%)	TS (96%)	TC (73%)	TS (93%)	TC (87%)	66.3%
0.006	TS (89%)	TS (92%)	TS (98%)	TC (78%)	TS (94%)	TC (94%)	65.7%
0.005	TS (92%)	TS (97%)	TS (56%)	TC (79%)	TS (95%)	TC (90%)	60.1%
0.004	TS (95%)	TS (98%)	TC (89%)	TC (81%)	TS (95%)	TC (87%)	54.6%

Table 2. Confusion matrix for network between 0.008 and 0.005 fitness values

	TS Class	TC Class
TS patient	3	0
TC patient	1	2

results), while the mistake result (CyanBis) is strong (94%). On further analysis, we note, in the best case, a sensitivity value at 0.75 and a specificity value at 1.

6 Conclusion and Future Works

We have established that a neural network can distinguish the autism status of young patients thanks to the eye movement of the children. It might help to give a preemptive idea about further analysis to consider for a child who presents signs of this disorder. Such a system could be used with a desktop computer for testing kids at an early age, for example at school. It would be a solution to find more easily ASD suffering children, transmitting relevant data to a professional, including a more in-depth test if necessary. However, the system proposed here is not supposed to replace the work of professionals of the medical field. The help given only consists on an automatic way to focus as fast as possible on the needing children, and maybe point at false diagnoses. Our system works within its current state that has been defined arbitrarily. Still, further work should aim at optimizing the network topology, including the type of nodes used in it and/or the size of layers and their amount. A study on other machine learning solutions should also be conducted, to check the overall efficiency of LSTM-RNNs. Another issue consists in the very low amount of training data, for a neural network approach. Uncertainty and over-fitting might be reduced or removed if more data is gathered for training. Also, we must recall that various rates of autism spectrum disorder exist. Our present solution only gives an idea about the autism status: suffering from it or not. An analogous system might give an early direction about which autistic rate is actually from.

References

1. Cilia, F., Deschamps, L., Vandromme, L.: Investigation of interactive visual patterns during semi-structured joint attention sequences in children with Autism Spectrum Disorder. In: RIPSYDEVE Conference, Louvain-la-neuve (2016)
2. Adrien, J.L., Faure, M., Perrot, A., Hameury, L., Garreau, B., Barthelemy, C., Sauvage, D.: Autism and family home movies: preliminary findings. J. Autism Dev. Disord. **21**(1), 43–49 (1991)
3. Maestro, S., Muratori, F., Cesari, A., Pecini, C., Apicella, F., Stern, D.: A view to regressive autism through home movies. is early development really normal? Acta Psychiatrica Scandinavica **113**(1), 68–72 (2006)
4. Hobson, R.P., Ouston, J., Lee, A.: What's in a face? the case of autism. Br. J. Psychol. **79**(4), 441–453 (1988)
5. Trepagnier, C., Sebrechts, M.M., Peterson, R.: Atypical face gaze in autism. Cyberpsychol. Behav. **5**(3), 213–217 (2002)
6. Jemel, B., Mottron, L., Dawson, M.: Impaired face processing in autism: fact or artifact? J. Autism Dev. Disord. **36**(1), 91–106 (2006)
7. Pelphrey, K.A., Sasson, N.J., Reznick, J.S., Paul, G., Goldman, B.D., Piven, J.: Visual scanning of faces in autism. J. Autism Dev. Disord. **32**(4), 249–261 (2002)

8. Falck-Ytter, T., von Hofsten, C.: How special is social looking in ASD: a review. Prog. Brain Res. **189**, 209–222 (2011)
9. Speer, L.L., Cook, A.E., McMahon, W.M., Clark, E.: Face processing in children with autism: effects of stimulus contents and type. Autism **11**(3), 265–277 (2007)
10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural comput. **9**(8), 1735–1780 (1997)