Analysis of Eye Gaze Pattern of Infants at Risk of Autism Spectrum Disorder using Markov Models

David Alie[1], Mohammad H. Mahoor[1], Whitney I. Mattson[2], Daniel R. Anderson[1], and Daniel S. Messinger[2]

[1]University of Denver, Department of Electrical and Computer Engineering, Denver, Co 80208,
[2]University of Miami, Department of Psychology, Coral Gables, FL 33146.

Emails: {dalie, mmahoor, danderso}@du.edu, w.mattson@umiami.edu, and dmessinger@miami.edu

Abstract

This paper presents the possibility of using pattern recognition algorithms of infant gaze patterns at six months of age among children at high risk for an autism spectrum disorder (ASD). ASDs, which must be diagnosed by 3 years of age, are characterized by communication and interaction impairments which frequently involve disturbances of visual attention and gaze patterning. We used video cameras to record the face-to-face interactions of 32 infant subjects with their parents. The video was manually coded to determine the eye gaze pattern of infants by marking where the infant was looking in each frame (either at their parent's face or away from their parent's face). In order to identify infants ASD diagnosis at three years, we analyzed infant eye gaze patterns at six months. Variable-order Markov Models (VMM) were used to create models for typically developing comparison children as well as children with an ASD. The models correctly classified infants who did and did not develop an ASD diagnosis with an accuracy rate of 93.75 percent. Employing an assessment tool at a very young age offers the hope of early intervention, potentially mitigating the effects of the disorder throughout the rest of the child's life.

1. Introduction

Recent advances in computer vision and pattern recognition have yielded techniques for investigating questions in social development relevant to psychologists. Facial expressions and eye gaze direction are important markers that describe the early emotion-related behaviors of children in face-to-face interaction (e.g., mother and infant interaction) [1, 2]. Computer vision and pattern recognition techniques can be used either for automated measurements of emotional expressions and eye gaze attention or for analysis, modeling, and even the prediction of the diagnoses of children at risk for autism. Children with autism spectrum disorders (ASDs) often show impairments in nonverbal social interaction, emotional expressions, and visual gaze attention [3, 4]. The increasing number of children diagnosed with autism in the United States (about one in every 110 children) [8], makes the utilization of innovative technologies such as computer vision and pattern recognition for early identification and potential treatment essential.

Among the several well defined communicative deficiencies exhibited by children with ASDs, impairment in visually disengaging attention in a variety of contexts is exhibited at a particularly early age [2]. In this paper we turn our attention to the patterns of visual disengagement exhibited by infants with an ASD and those without to see if gaze pattern analysis can help shed light on this aspect of the disorder. This paper postulates that the gaze patterns, accessible at much younger ages, of children with an ASD are different from those of comparison (typically developing or non-ASD) children. The purpose of this research is to take advantage of this information in order to identify high risk children at a much younger age, using only three minutes of face-to-face interaction.

In this interaction, parents were instructed to play with their infant while both parent and subject were videotaped for three minutes using three video cameras (two cameras recording the subject, the third recorded the parent). A fourth camera recorded both the parent and infant in the same frame. All four cameras were synchronized in time. The videos were then examined frame by frame by a trained gaze coder and the infant was identified as either looking at their parent or looking away at each time stamp as shown in Figure 1.

A substantial amount of work has been done by Shic et al. [7] regarding the pattern of visual engagement to different parts of the human face by subjects with an ASD.

978-1-4244-9497-2/10/$26.00 ©2010 IEEE

They focused on how the facial scanning strategies of children with an ASD changed over time and how these strategies differed from those of comparison children. Using fixed order Markov processes, they concluded that scanning strategies among children with an ASD do not evolve over time in the same way as comparison children. We did not focus on these changes in gaze patterns over time, but rather focused on differences between ASD gaze patterns at a single age. Our work uses a binary visual process (looking towards the face vs. looking away from the face) rather than distinguishing between gaze to various facial features.

Similarly, visual disengagement in the infant siblings of children with an ASD was studied by Ibanez et al. [6]. They analyzed infants with an older sibling with an ASD on the gaze pattern of infant subjects by measuring the length of time a subject would look at or away from an object. Their analysis focused on the average length of time elapsed between instances of gaze switching during several different interaction protocols.

Figure 1. Images captured during face-to-face interaction. In the image pair on the left the subject is looking at the parent, while in the pair on the right, the subject is looking away.

This research also differs from previous studies by focusing on the face-to-face protocols while using Variable-Order Markov models (VMM), rather than fixed order Markov models (as with Shic et al. [7]), to analyze the gaze data. The VMM enables us to better model human gaze behaviors by allowing us to tailor the model to place statistical emphasis on certain aspects of the gaze pattern (by allowing for a variety of context lengths) while simultaneously producing confident results from relatively smaller data sets.

Our experiment involved collecting and analyzing the gaze patterns of six-month old subjects using the face-to-face protocol [9]. ASD status of the subjects was determined by the presence or absence of an ASD diagnosis at 36 months. The subjects participated in the Autism Diagnostic Observation Schedule (ADOS) [10] at 30 months, and the Autism Diagnostic Interview-Revised (ADI-R) [4] was administered to their parents at 36 months. Clinical best estimate diagnoses were based on DSM-IV-TR criteria based on the ADOS, ADI-R, and clinical judgment of an experienced psychologist who was blind to subject status. Diagnostic classifications of either Autistic Disorder, Asperger's Disorder, Pervasive Developmental Disorder—NOS, or No Autism Spectrum Disorder were assigned. The first three groups were combined as these diagnoses fell under the autism spectrum (ASD). Reliability was calculated based on review of a second clinician on approximately 90% of the cases. Cohen's Kappa was .86 (93% agreement). The subjects were then grouped based on their three-year-old diagnostic outcome, and a VMM was used to generate a gaze pattern model for each group. The subjects were tested against these models using a leave-one-subject-out cross validation method and the program identified the subject as having an ASD diagnosis or not having an ASD diagnosis (comparison) at a rate of 93% for comparison subjects, and 100% for subjects with an ASD. The high degree of accuracy between models generated from data collected at six months of age and eventual diagnosis at three years of age suggests that gaze pattern, even at a very young age, is influenced by an emerging ASD.

The remainder of the paper is divided into three sections. Section 2 addresses the specific methods under which the data was analyzed. Particular attention is given to the mathematical processes employed in modeling and analyzing the gaze data. Section 3 of this paper outlines the data that was collected and the results of the analysis followed by a detailed discussion of experimental results and the conclusions that were drawn from these results. This section ties the formal mathematics of the earlier chapters with the implications to the psychological applications of this research. Conclusions and future work are finally given in Section 4.

2. Analytical Methods

Our aim was to test the gaze pattern of infants who later developed ASD against those who did not. Visual gaze data was gathered from 32 infant subjects, all with older siblings. Eighteen subjects had older siblings with ASD, the remaining 14 had typically developing older siblings. The 32 subjects returned at age three and were classified as either having an ASD (6 subjects) or not having an ASD (26 subjects). Infants with an older sibling with an ASD are at high risk for themselves developing an ASD; all six infants who subsequently were diagnosed with an ASD came from this group.

The subjects and their mothers were recorded on video and the visual behavior of the child in each frame was examined. Children were coded as either looking at their mother, indicated by a 1, or looking away, indicated by a 0, in each frame at a rate of 30 frames per second. The resulting data from each subject was arranged as a binary sequence.

Because Markov models establish a probabilistic outcome for a given context, a binary training sequence can be limiting as it only allows for two possible states at each discrete step. We solved this problem by establishing a "phrase alphabet" whose "letters" represent binary sequences of a certain length. In this representation scheme, we define the parameter T as that length such that each letter corresponds to one of the 2^T possible binary combinations. In a simplistic example where $T = 4$ (therefore using 16 letters), the sequence 0000 would be re-coded as "*A*", while 0101 would be re-coded as "*F*", etc. A data string whose first units were 001001010000 would be broken down into three sequences of four frames which would be translated to their corresponding letters. This process is illustrated by the following diagram:

001001010000 → 0010-0101-0000 → *B-F-A*

In this way, the data string for each subject was translated from its original form into a sequence of representative letters. The combination of these two methods made the data more representative of the subjects' behavior and created a wider, more even data distribution to assist in the training of the Models.

2.1 Markov Models

The data was analyzed using both hidden and variable order Markov models with various parameters. Markov models are well suited to this sort of analysis because of their strength in classifying or predicting the subsequent elements of a sequence. Two models (one for comparison subjects, one for ASD subjects) were first generated using data sequences from all but one of the subjects. The models used the data strings to create a probabilistic understanding of which letters follow each other most often and used that information to create an "average behavior" model for each group. The untested subject was analyzed in the same way and then compared to the two models to identify the model which the subject most closely resembles. The subject was then classified as either being from the comparison group or the ASD group.

2.1.1 Hidden Markov Model

Hidden Markov Models (HMMs) [14] is a technique for analyzing time-varying data with spatio-temporal variability. Its model structure can be summarized as a hidden Markov chain and a finite set of output probability distributions. The use of HMMs includes two stages: training and classification. In the training stage, the number of states of an HMM must be specified, and the corresponding state transformation and output probabilities are optimized in order for the generated symbols to correspond to the observed data features (e.g., eye gaze signal). HMM model is usually trained using the Expectation Maximization algorithm and used for the classifying an unknown sequence of data. In the matching stage, the probability that a particular HMM possibly generates the test symbol sequence corresponding to the observed data features is computed. HMMs are superior to time-series analysis and dynamic time warping in processing unsegmented successive data [11], and are extensively being applied to different classification problems such as speech recognition [14,15] and gesture recognition [11-13,16,17].

In this paper we trained two separate HMMs for assigning an unknown sequence of eye gaze data into two classes, either comparison group or ASD group.

2.1.2 Variable-Order Markov Model

Variable-order Markov Models (VMM) are very similar in concept and function to HMM. In contrast to HMM, however, VMM do not use a fixed order throughout the training process when calculating the probability values. As a result, where a HMM produces a final state machine, a VMM will produce a probability "tree" with the null-space as the root and the elements of a sequence as the nodes and leaves. Figure 2 is one such tree for a simplified example in which $T = 2$ (four possible letters). Each generation is labeled with an order number. This is to illustrate the differences between generations. The nodes of order 0 have no context and are the first letters in a sequence. Nodes of order 1 have a context of length 1 (their parent node of order 0), etc. This can be formally expressed by stating a node k of order n will have a context of length n where the context is the concatenation of all the parent nodes of k beginning with the root (ε).

The example tree in Figure 2 was created by the same VMM algorithm as in the experiment, using data generated by the comparison group. For the sake of clarity, the phrases, or contexts, mentioned in the last section were limited to two characters instead of the six used in the experiment. A counter is set at each node and a record is kept of each time a node is encountered during training. These tallies are then used to calculate the probabilities of seeing a phrase or letter given a certain context, much like in HMM. More precisely, the training algorithm seeks to calculate the probability

$$P(x_i \mid s_n)$$

where x_i is a given state and s is a given context of concatenated states of any length n. The above statistic is generated for each node on the tree throughout the training sequence. In the diagram, the number next to each node is the probability of encountering that letter given the context that precedes it.

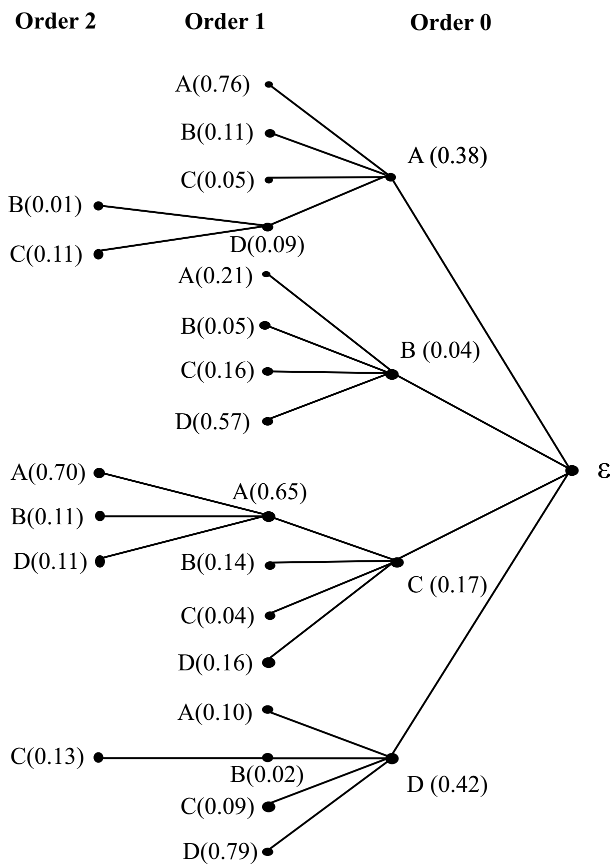

Fig. 2. This tree includes the probabilities for the first two generations of letters produced in a simplified training sequence. To calculate the likelihood of a sequence beginning with the letter *C,* we start at the root (ε) and move down the tree one step to the letter *C* where we find the probability to be 0.17. Taking this one step further to find $P(A \mid C)$ in this example, we begin at the root of the tree (the null set) and traverse the tree according to the context (which is *C* in this case). The probability of encountering an *A* given context *C*, we simply look to the letter *A* one generation below the context *C*: 0.65.

For our experiment, we used a type of VMM known as a PPM algorithm, which stands for Prediction by Partial Match [5]. This class of algorithms requires that a maximum bound on the context length, β, be set by the user. We used $\beta = T$. PPM models attempt to calculate $P(x_i \mid s)$ given some context s such that $|s_n| \leq \beta$ in much the same way as the general VMM case described above. PPM algorithms are unique in the case where there is no data in the training sequence upon which $P(x_i \mid s_n)$ can be calculated. If this is the case, the algorithm instead calculates

$$P(x_i \mid s_{n-1})$$

If this is still an empty data set, the algorithm calculates

$$P(x_i \mid s_{n-2})$$

and the process continues repeatedly until a match is found in the training sequence or the context becomes the null set. In the latter case, a fixed probability is assigned.

3. Experiment and Results

The experimental setup involved the infant subject sitting opposite their parent for the three minutes of face-to-face interaction. As we described in the introduction, the interaction was captured on video and each frame was coded, producing a binary sequence representing the subject's visual attention throughout the interaction.

Capturing data at a rate of 30 frames per second appeared to be overly rapid relative to the rate of gaze switching. This created artificially long strings of repeated values in the data. In order to offset the effect of these repetitions, the coded binary gaze data was smoothed by a low-pass filter (a moving average filter with size 18) and then down-sampled by a factor of 18 to create time units that more closely matched the observed rate of gaze switching, thereby optimizing the effectiveness of the program.

The sequences were analyzed by the different VMM programs using a leave-one-subject-out cross validation method. The program would identify one sequence to be tested then use the remaining sequences to train a model for the comparison group and a model for the ASD group. Classification of the test sequence was performed through a log-likelihood approach, in which the chosen sequence's probability of occurrence given a model is calculated. This yields a likelihood value for each model and the sequence is assigned to the model which corresponded to the higher value. The process repeats from the beginning until all sequences have been tested and assigned to one model or the other.

Several values for the parameter *T* were tested. Six was found to produce the optimal agreement percentage. This established a maximum context length of 6, and an alphabet containing 64 characters. Figure 3 contains the agreement percentages for the various values for *T*.

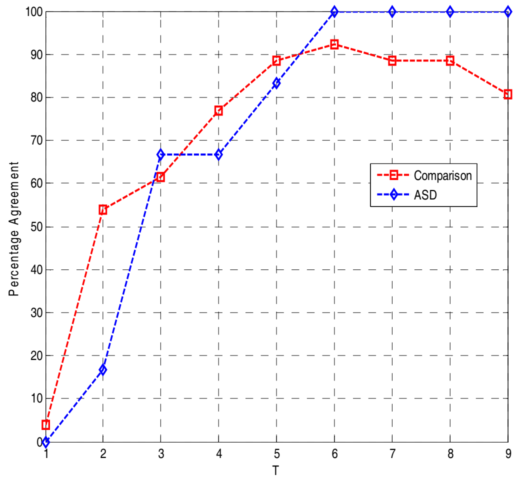

Figure 3. The percentage agreement results for the ASD (blue) and comparison (red) groups are plotted against various values for the parameter T. $T = 6$ was the optimum value with the highest percentage agreement rates for both groups.

The VMM algorithm with $T = 6$ correctly identified 92.30% of the comparison subjects and 100% of the ASD subjects with Cohen's Kappa of 0.82. The Cohen's kappa coefficient, which is a statistical measure of inter-rater agreement, ranges between 0 and 1 and is more robust than simple percentage agreement since it takes into account random agreement.

The HMM algorithm correctly identified 92.30% of the comparison subjects with Cohen's Kappa of 0.29, but only 33.33% of the ASD subjects (the best results were achieved using HMMs with 4 hidden states).. The results of the analysis are displayed in greater detail in Table 1.

3.1 Analysis and Discussion

The results from the VMM and HMM modeling are best understood in the context of the nature of Autism Spectrum Disorder, which, as its name suggests, affects people to varying degrees. This presents a challenge when trying to use a specific behavior (in this case gaze pattern) as the foundation for making generalized assertions (in this case a gaze pattern VMM) about a group of people with varied affectedness. Despite this hurdle, there exist consistent differences in the gaze patterns of ASD and comparison infants that lend themselves to analysis using machine learning techniques. Consequently we can use these techniques to examine an infant's gaze pattern and confidently say whether it more closely resembles those of comparison infants or infants who developed ASD.

VMM Algorithm		
	Comparison Group	ASD Group
Assigned to Comparison Group	24	0
Assigned to ASD Group	2	6
Percentage Agreement	**92.30% (specificity)**	**100.00% (sensitivity)**

(a)

HMM Algorithm		
	Comparison Group	ASD Group
Assigned to Comparison Group	24	4
Assigned to ASD Group	2	2
Percentage Agreement	**92.30% (specificity)**	**33.33% (sensitivity)**

(b)

Table 1. The comparison subjects and ASD subjects are separated by column with the rows denoting the subject distribution as assigned by the VMM algorithm with $T = 6$ (a) and by the HMM algorithm (b). The VMM algorithm correctly identified 24 of the 26 comparison subjects and all 6 of the ASD subjects, while the HMM algorithm identified 25 of the 26 comparison subjects, and 2 of the 6 ASD subjects.

One of the reasons the HMM algorithm may have been less effective is that these differences in the gaze pattern are less apparent when we limit ourselves to statistics of a fixed context length. Because of the spectrum nature of ASDs, we expect there to be greater uniformity in the comparison subjects than in the ASD subjects. This may help explicate why the HMM algorithm was less successful in identifying ASD subjects than was the VMM algorithm.

These results also suggest differences between the gaze patterns of the two groups are readily identifiable. It is important to note that the ASD group contains significantly fewer subjects that the comparison group. It is a concern that too few subjects in one group will generate a model that is not inclusive enough to yield confident results. Based on our trials with the varying parameter values, this appears not to be the case. The assignment results were consistently the best with a down-sampling rate of 18 and maximum word length of 6. This implies that the number of subjects in the ASD group does not impact the quality of the model that is generated. Put another way, the patterns from each group are significant enough to generate a model with the key characteristics using a relatively small number of subjects. Using many more subjects, as in the comparison group, will, we expect simply refine the model.

4. Conclusions and Future Work

This paper presented a pattern recognition approach for identifying infants at risk for ASD at 6 months of age using their eye gaze patterns. The results represent promising steps towards early detection of ASDs. It also suggests some of the future steps that may help achieve the eventual goal of a fully developed diagnostic tool. Further research should include expanding the study to include greater numbers of subjects. Incorporating details about what part of the parent's face the infant is looking at (as in Shic et al. [7]) or where the infants looked when they were looking away might increase the resolution of the possible conclusions.

References

[1] D. S. Messinger, M.H. Mahoor, S. Chow, J.F. Cohn. Automated measurement of facial expression in infant-mother interaction: A pilot study. *Infancy*, vol. in press, 2008.

[2] L. Ibanez, D. S. Messinger, L. Newell, M. Sheskin, and B. Lambert. Visual disengagement in the infant siblings of children with an Autism Spectrum Disorder (ASD). *Autism: International Journal of Research and Practice*, 12:523-535, 2008.

[3] American_Psychiatric_Association. *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision)*. Washington, DC: American Psychiatric Association, 2000.

[4] C. Lord, M. Rutter, and A. Le Couteur. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism Development Disorders*, 24:659-685, 1994.

[5] R. Begleiter, R. El-Yaniv, and G. Yona. On prediction using variable order Markov models. *Journal of Artificial Intelligence Research* 22:385-421, 2004.

[6] L. Ibanez, D. S. Messinger, L. Newell, B. Lambert, and M. Sheskin. Visual disengagement in the infant siblings of children with an Autism Spectrum Disorder (ASD). *Autism* 12:473-86, 2008.

[7] F. Shic, K. Chwarska, J. Bradshaw, and B. Scassellati. Autism, eye-tracking, entropy. *IEEE International Conference on Developmental Learning*: 73-78, 2008.

[8] Center for Disease Control. Prevalence of Autism Spectrum Disorders – Autism and Developmental Disabilities Monitoring Network, United States, 2006. *Morbidity and Mortality, Weekly Report,* 58, 2009.

[9] E. Tronick, H. Als, L. Adamson, S. Wise, & T. B. Brazelton. Infants response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child and Adolescent Psychiatry*, 17: 1-13, 1978.

[10] C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, M. Rutter. The Autism Diagnostic Observation Schedule-Generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders* 30:205-223, 2000.

[11] C. Myers, L. Rabinier, A. Rosenberg, Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 28(6):623-635, 1980.

[12] A. Bobick and A. Wilson. A state-based technique for the summarization and recognition of gesture. *Proceedings of the International Conference on Computer Vision,* 382-388, 1995.

[13] K. Takahashi, S. Seki et al. Recognition of dexterous manipulations from time varying images. *Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects,* 23-28:1994.

[14] A.B. Poritz. Hidden Markov Models: a guided tour. *Proceedings of the IEEE International Conference on Acoustic Speech and Signal Processing*, 7-13:1988.

[15] L. Rabinier, A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of IEEE*, 77(2):257-285, 1989.

[16] T. Starner and A. Pentland. Real-time American Sign Language recognition from video using hidden Markov models. *Proceedings of the International Symposium on Computer Vision,* 265- 270, 1995.

[17] C. Vogler and D. Metaxas. ASL recognition based on a coupling between HMMs and 3D motion analysis. *Proceedings of International Conference on Computer Vision*, 363-369, 1998.