Using Behavioral Patterns in Treating the Autistic

Sudarsun. S, *Member, IEEE*, Varun Kant Vashishtha, and Avijit Nayak

***Abstract*—The development of behavioral therapy regimens for autistic patients is relatively challenging as these patients may not be able to express feedback to the applied treatment. The response to a treatment course is mostly estimated qualitatively and with little systematic feedback between therapy and response. Collecting and analyzing data about a patient's daily activities could yield patterns linking these activities, thereby providing therapists with some foreknowledge of likely possible behavioral outcomes related to their therapies. We propose a method for anomaly detection system, which can monitor behavior patterns of the patient based on the data collected on a daily basis. The knowledge gathered about the patient could prove suggestive of the patient's feedback to the applied therapy. Upon mining the behavioral patterns, the system could predict the response of a patient to a stimulus, given a list of recently displayed behaviors and/or completed activities. The knowledge thus gathered could also be used to treat other patients of similar disability.**

***Index Term*—Anomaly Detection, Autistic, Behavioral Patterns, Data Mining, Sequence Analysis, Statistical Modeling.**

I. INTRODUCTION

MANY individuals with autism lead lives that are significantly restricted because of problem behaviors such as self-injury and aggression. The autism is characterized by a wide variety of possible symptoms such as extreme withdrawal, lack of social behavior, repetitive behaviors, violent behavior, and etc. Because of wide variety of symptoms and intensity, treatment needs to be individualized for every person. Therapeutic interventions that leverage information technology are still in their infancy in this area. Therapist and researchers alike point to the need for more research and applications to help autistic persons achieve a higher quality of life.

Research is needed to find the cause and optimal matching of autistic persons to treatment. Data mining is optimal for this. Two approaches are possible. The first is the use of data mining to find biological associations, such as specific genes related to autism [8]. This type of data mining research is similar to data mining in molecular biology and genetics. Data mining can also help fine-tune diagnosing and find underlying causes for drug related treatments. However, our approach, mining the actual behaviors, may provide more specific insights into when specific behavior occurs, when specific treatment is successful. Our work focuses on the use of technology to study the patterns of behavior. We will use data mining techniques to examine the sequences of behavior of autistic people. Our goal is to improve the quality of life of those with autism by helping their therapists to understand more systematically and more immediately what effect the therapy is having on the patient, particularly in cases where changes in condition may not be so visible and may take a long time to manifest. Data mining techniques can be grouped into several categories depending on the goal of the application (description, classification, prediction, or optimization). Not all approaches are suitable for our problem since our goal is to prevent any unwanted situation and to help therapist evaluate his treatment process. The goal is to predict on the fly when a patient will behave appropriately or not.

We focus on the increase in appropriate behavior and the decrease of inappropriate behavior of the autistic patient. For this we need to collect a large amount of behavioral data on the patient and categorize the data broadly into sets of appropriate and inappropriate behaviors and then into sub-classifications of specific behaviors. Two general mining characteristics can contribute better insight into the treatment's effect. For example, if we present a set of rules of conditions that lead to mostly inappropriate behavior, therapist get more detailed information, which may lead to additional focus and caregivers may get a better idea of when to expect inappropriate behavior. Moreover comparing such rules over different stages during treatment is informative.

II. BACKGROUND

A number of frameworks [1] have been proposed or implemented for research and development to facilitate computer-aided learning for those with autism. These frameworks argue that it would be most beneficial if computer-aided learning software would focus on three main impairments: social and interpersonal skills, communication difficulties, and rigidity of thinking. Most of the current research and development can be classified into these three groups. Social and interpersonal skill training refers to helping people with autism understand why other people behave as they do. Some research is going on to building tools [2] that will help patients to better understand social situations. Others focus on the second impairment: communication. The use of information technology varies from approaches that use low complexity technology, from the therapist's point of view, to those making use of high complexity technology. For example, some researchers simply used PowerPoint slides [3] with video segments to teach socio-dramatic play. Others trained autistic people to point to pictures to aid communication [4]. This approach has been taken a step further by others [5] who developed a communication system for use with personal digital assistants. Even more technology intensive is simulations used to teach verbal communication [6] or social robots for diagnosis and treatment [7]. Finally, there are also attempts to address the third major impairment: rigidity of thinking. The most popular types of interventions use information technology to interact directly with autistic patients. However, compared to fields such as biology or commerce, the use of information technology within this field is still in its nascence, particularly with regard to the systematic analysis and prediction of behavior. Large-scale systematic studies could help specify characteristics of the syndrome and explain the wide variety of these characteristics; it could also help provide profiles for autistic patients and identify the therapies that are most suitable for them. To our knowledge, few researchers have used advanced data mining to study systematic differences, or the lack thereof, in therapy outcome for such patients.

Very little is known about autism, its causes, and its manifestations. There is substantive variation across cases. As a result, there is no fixed treatment process. The therapist has to decide on a particular treatment process based on his knowledge and understanding of the problem. But as patients are not socially or intellectually matured it is very difficult for a therapist to decide on the treatment. And, because of the painstaking pace of behavioral change regimens, the therapist may not have any feedback on the efficacy of his treatment process for a long time. This is further complicated by low levels of program adherence, particularly when a therapy's effects are not quickly apparent. To answer the problem faced by the therapist, we propose the following solution.

III. Problem Statement

Predictive analysis of human behaviors hinges upon the researcher's ability to collect a substantial history of the patient's behaviors and to undertake a detailed and careful classification exercise, coding each observation according to a systematic schema. Because human behaviors can be complex, subtle, and varied, the most efficient means of data collection will likely involve training caretakers to maintain an ongoing log, though ex post facto analysis of video recordings, while much less practical, may also be possible. Significant observation will first be required simply for the development of a comprehensive coding scheme against which the full range of a patient's possible behaviors may be represented. The next step is to record the specific sequences of behavior of a given patient and to create a database of such sequences. The accuracy of the prediction of patient's behavior will depend both on the breadth of different behaviors exhibited by the patient and by the size of the database. If the number of behaviors is large or if the number of observations is relatively small, the accuracy of prediction is likely to be low. For this reason, it may initially be most practical to apply these techniques for the analysis of severely autistic people whose range of behaviors may be less broad than for more socially functional patients. Predicting the next behavior of the patient will involve matching such patterns of behavioral activities. Suppose that A=>B=>C=>D=>E represents a sequence of activities (where A, B, C, D and E are specific activities) followed by the patient. From a search of that sequence in the database, we should be able to identify the frequency of various subsequent activities such that, from this specific sequence, we should be able to identify the probability of a certain behavior F as the ensuing activity.

The analysis of behavioral sequence is useful not only in predicting future behaviors but also in working backwards to identify those sequences of activity consistent, based on past observation, with the onset of severe outbursts such that remediation can be applied on an emergent basis.. Indeed, the data may in fact serve to identify possible remediation by suggesting relationships across those cases where a similar pattern of activity did not result in an outburst. These "good cases" can be used to derive information about which medication or reinforcement techniques can be used to divert the patient to follow the good pattern. This will help in improving the lives of both patients and caretakers.

The same kind of sequences of behavioral activities should be recorded after the implementation of the therapy. The measure of similarity or dissimilarity between the data sets before and after the implementation of the therapy can be utilized to evaluate the effectiveness of the therapy. For example, if the frequency of occurrence of behavioral patterns leading to inappropriate behavior is found to be decreasing or if the frequency with which outbursts follow a "known bad" pattern is found to be decreasing, then we can confidently say that the therapy is having good impact on the patient and vice versa.

IV. Proposed Solution

There are three basic problems we wish to address through these data mining techniques: a) how to estimate the effectiveness of a treatment from the observations on the patient; b) given a pattern of behavioral symptoms, how to predict the upcoming changes in the patient's behavior; and c) how to detect anomalies in a patient's behavior.

A. *Estimating the effectiveness of a treatment*

As previously noted, it is challenging to determine the effectiveness of a therapy because of the difficulty in obtaining direct, narrative feedback from many severely autistic patients. If we visualize this problem as a set of hidden events and a set of observed events, one possible solution may involve the use of Hidden Markov Models (HMM) [9].

HMM works by finding patterns that appear over a space of time. There are many examples of such temporal patterns, from the pattern of commands someone uses in instructing a computer to the sequences of words in sentences to the sequence of phonemes in spoken words. Our case clearly fits this description as we are not only examining behavioral observations but also their specific sequence.

In our case, we have three sequences: a) the medication event sequences, b) the patient feedback event sequence, and c) the recorded patient's behavior observation sequence. The medication event sequence is the controller input that needs to be manipulated based on the patient feedback sequence. But, the patient feedback sequence is not really available readymade to us, but present as a hidden sequence. We still have the recorded observations on the patient's behavior. We know that the recorded observations about the patient are dependent on the hidden patient feedback sequence. As a result, HMM becomes a direct solution to our problem in modeling the patient's response to therapy.

There are three canonical problems associated with HMM: 1) Given the parameters of the model, how can we compute the probability of a particular output sequence? This problem is solved through the use of the Forward-Backward algorithm [9]. 2) Given the parameters of the model, how can we determine the most likely sequence of hidden states that could have generated a given output sequence? This problem can be solved through the use of the Viterbi algorithm [10]. 3) Given an output sequence or a set of such sequences, find the most likely set of state transition and output probabilities. In other words, discover the parameters of the HMM given a dataset of sequences. This problem can be solved through the use of the Baum-Welch algorithm [9].

We are particularly interested in the canonical problem "b", which could generate the sequence of a patient's feedback events, given the patient's behavior observation sequence. We will build the HMM model using the patient's feedback events as validated by the patient's behavior observation sequence as recorded.

B. *Prediction of upcoming behavioral changes*

When we have recorded all the activities of the patient, we should be able to discern patterns of relevance connecting the various activities. Any given activity is, of course, dependent on some other past activity, whether on the same day or before. When these activities are represented as observation states sequences, with the mental state change as the hidden state sequence, and the treatment sequence as the external instigator sequence, we may use an HMM to predict the next hidden mental state transition by using canonical problem "b" as in above.

If we constrain our problem to a limited set of mental states, when we have a model in place, we may be able to query the model for transition probability to any mental state given the observation sequence. The length of the observation sequence may be varied from few minutes to hours to days together. Using a first order HMM may not be appropriate to this problem, as we have to allow the next state change to be dependent on more than one previous state.

Conditional Random Fields (CRF) [11] would be an ideal alternative to n^{th} order HMM. Conditional random fields are a probabilistic framework for labeling and segmenting structured data, such as sequences, trees and lattices. The underlying idea is that of defining a conditional probability distribution over label sequences given a particular observation sequence, rather than a joint distribution over both label and observation sequences. The primary advantage of CRF's over hidden Markov models is their conditional nature, resulting in the relaxation of the independence assumptions required by HMM in order to ensure tractable inference. Additionally, CRF's avoid the label bias problem, a weakness exhibited by Maximum Entropy Markov models (MEMMs) and other conditional Markov models based on directed graphical models.

To train a CRF model, we will also have to generate a feature list which could describe explicit or implicit binding of 1) patient behavior observation and the mental state, 2) previous mental state and current state, 3) previous behavior observation and current observation and so on.

C. *Anomaly detection on behavioral patterns*

Once a recording system is in place, we will have accumulated a considerable volume of behavior sequence data specific to a particular patient. The simplest anomaly detection system should have a manually classified set of normal and abnormal behavior patterns. When a new pattern arrives, we may compare them with known normal and abnormal patterns and raise a flag accordingly. During the learning or knowledge base building phase, all the patterns are marked as abnormal. An expert is asked to review the pattern to change the marking from abnormal to normal if the pattern is indeed a normal pattern.

Fingerprinting is a viable technique for sequence comparison. Fingerprinting is a process of generating a unique hash code(s) for the given sequence of data. Fingerprinting algorithms are popular in the media industry because of their ability to deliver faster searches across huge media databases. Fingerprinting techniques involve two basic operations: 1) alignment of the sequence data, where the sequences are aligned to maximum overlap; and 2) computation of similarity scores.

V. LIMITATIONS

While data mining tools can be very powerful, they are not self-sufficient applications. To be successful, data mining requires the involvement of skilled technical and analytical specialists who can structure the analysis and interpret the output that is created. Consequently, the limitations of data mining are primarily human, whether due to weaknesses in collecting or coding observations or in analyzing or modeling that data once it has been collected.

Although data mining can help reveal patterns and relationships, it does not tell the user the value or significance of these patterns. The user must make these types of determinations. Similarly, the validity of the patterns discovered depends heavily on their fidelity to "real world" circumstances.

Another limitation of data mining is that while it can identify connections between behaviors and/or variables, it does not necessarily identify a causal relationship.

VI. CONCLUSION

We have only touched the surface of what is possible with data mining of behavioral patterns. While these methods hold some promise for the analysis and extrapolation of patients' behavioral patterns, the development of effective therapy regimens will require a more comprehensive analysis combining physiological and psychological factors. It is also important to note that, absent experimental data, it will not be possible to estimate the expected levels of accuracy of these techniques; the quantum of behavioral scope and the difficulty in reliable behavioral coding will prove significant to the efficacy of these techniques. The validation of these concepts with experimental data is a logical future extension of this paper.

VII. REFERENCES

[1] D. J. Moore, P. McGrath, and J. Thorpe, "Computer Aided Learning for people with autism - a framework for research and development," Innovations in Education and Training International, vol. 37, pp. 218-228, 2000.

[2] J. Bishop, "The Internet for educating individuals with social impairments," *Journal of Computer Assisted Learning*, vol. 19, pp. 546-556, 2003.

[3] M. Dauphin, E. M. Kinney, and R. Stromer, "Using Video-Enhanced Activity Schedules and Matrix Training to Teach Socio-dramatic Play to a Child with Autism," *Journal of Positive Behavior Interventions*, vol. 6, pp. 238-250, 2004.

[4] M. L. Israel, L. Ruthel, M. Bates, and N. Smith, "Software to teach nonverbal persons with severe autism and retardation to communicate by pointing to pictures," presented at *Johns Hopkins National Search for Computing Applications to Assist Persons with Disabilities*, 1992.

[5] T. Miller, G. Leroy, J. Huang, S. Chuang, and M. H. Charlop-Christy, "Using a Digital Library of Images for Communication: Comparison of a Card-Based System to PDA Software," presented at *First International Conference on Design Science Research in Information Systems and Technology*, Claremont, CA, 2006.

[6] O. E. Hetzroni and J. Tannous, "Effects of a Computer-Based Intervention Program on the Communicative Functions of Children with Autism," *Journal of Autism and Developmental Disorders*, vol. 34, pp. 95-113, 2004.

[7] B. Scassellati, "Quantitative metrics of social response for autism diagnosis," presented at *IEEE International Workshop on Robot and Human Interactive communication (RO-MAN)*, Nashville, TN, USA, 2005.

[8] A. L. Yonan, A. A. Palmer, K. C. Smith, I. Feldman, H. K. Lee, J. M. Yonan, S. G. Fischer, P. Pavlidis, and T. C. Gilliam, "Bioinformatic analysis of autism positional candidate genes using biological databases and computational gene network prediction," *Genes, Brain and Behavior*, vol. 2, pp. 303-320, 2003.

[9] L Rabiner, B Jheong, "An Introduction to Hidden Morkov Model", *IEEE ASSP Mag*., vol. 3, no. 1, pp. 4-16, 1986.

[10] D. Froney, "The Viterbi Algorithm", *Proc of IEEE*., vol. 61, no. 3, pp. 268-273, 1973.

[11] J. Lafferty, A. McCallum, F. Pereira. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data"., *In Proceedings of the Eighteenth International Conference on Machine Learning (ICML-2001)*, 2001

[12] P. Cano, E. Batlle, E. Gmez, L. de C. T. Gomes, and M. Bonnet. "Audio fingerprinting: concepts and applications", *1st International Conference on Fuzzy Systems and Knowledge Discovery*, 2002

VIII. BIOGRAPHIES

Sudarsun S (M' 2002) is the Director – R & D at Checktronix India Pvt Ltd, Chennai. He holds an M.Tech Computer Science from IIT Madras. He received a BE degree in Electronics and Instrumentation Engineering from Madras University with a Gold Medal. His research experience includes Statistical Natural Language Processing, Machine Learning and Distributed Computing.

Varun Kant Vashistha is an Advanced Software Engineer – Trainee at Checktronix India Pvt Ltd, Chennai. He holds a B.Tech Chemical Engineering from IIT Kanpur. His research interest is Predictive Intelligence.

Avijit Nayak is an Advanced Software Engineer – Trainee at Checktronix India Pvt Ltd, Chennai. He holds a B.Tech Electrical Engineering from IIT Kanpur. His research interest is Pattern Analysis and Classification.