International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

Normalization of Behavioral Features in Autism Spectrum Disorder With Influencing Attributes

K. Vijayalakshmi, M. Vinayakamurthy, Anuradha

Abstract- The brain developmental disorder that affects both behavior and communication is autism spectrum disorder (ASD) considered and recognized as a major medical issue affects the increasing population approximately 0.5%– 0.6%. It is a highly heterogeneous neuro-developmental condition that has severe symptoms with various comorbid disorders. Applied Behavior Analysis involves various methods to understand the change in behavior through therapy. The goal is to identify and increase relevant behaviors which can help to diagnose attributes that affect learning. Data mining as the technology handles such medical grounds to predict by analyzing patterns in huge data sets. The outline of the proposed work is to find the relevant attributes from the dataset by normalizing and ranking the attributes. The CFS subset evaluator using various search methods like best first, greedy stepwise and exhaustive search are used to filter relevant feature from the dataset. The ultimate objective of this paper work is to examine the ASD applied behaviors with subject to normalization and ranking. Applying these to the feature selection methods would help for better understanding on various currently wide spread complex medical condition.

Index Terms: Autism Spectrum Disorder (ASD), Applied Behavior Analysis (ABA), Data Mining, Feature Selection.

I. INTRODUCTION

With the emergence of technology and science, fast and efficient automated processes are being possible for health care industries to diagnosis complicated health issues. But the major challenge is to provide such diagnosis in time as even the slight delay can make the situation of the patient more worst for recovery. Nowadays, the data mining, Big data and machine learning are the most extensively used technologies to fasten the prediction process in the field of healthcare for the proper medical diagnosis and treatment. ASD affects child or adult affects how a person reacts and interacts with other people, communicates, difficult to understand others feelings, poor eye contact, does not using facial expressions, speaks with abnormal voice or odd rhythm. Males are mostly affected by ASD compare with females. It is also known as spectrum disorder because they have not a same range of symptoms as it differs from each ASD affected people. Most of the children either underreact or overreact to sensory stimuli. The variety of diagnosing experts includes child psychologists & psychiatrists, speech Therapists, Physiotherapists, neurologists & audiologists for children, specially trained teachers etc. The various tests that are required to assess the behavioral issues of ASD are mentioned below: interview with parent, medical examination, audio-visual test, observation & language screening and evaluation, reasoning, verbal testing etc., Figure.1 shows that autism spectrum disorder exploration. In the below figure clearly mentioned the group of developmental brain disorder range present in ASD.

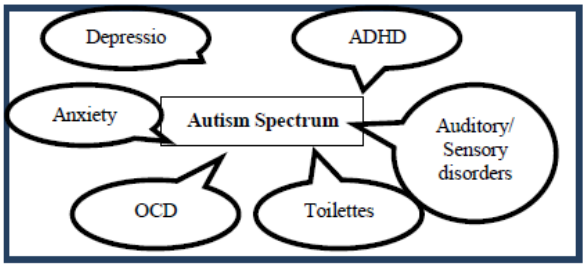

Figure.1 Exploring Autistic spectrum disorder

Data Preprocessing does removal of errors, maintains the consistency and accountability of missing values in most of the real world data which are incomplete with aggregate and missing values, noisy data with errors, outliers and duplication records. The quality of the mining depends on quality data for the quality decision making. Hence this paper work addresses the importance of pre-processing task through normalizing the attributes in the given dataset. The feature selection is an essential process that mostly applied in pattern recognition, machine learning and statistics. The major benefits of feature selection methods are

i. reducing computation time,
ii. improving classification performance and
iii. a better understanding of the data in various mining applications.

The main objective of the proposed work is to identify the useful characteristics of the dataset using normalization and ranking the attributes. The CFS subset evaluator with various search methods like best first, greedy stepwise and exhaustive search analyzed in order to eliminate the attributes that are not relevant from the dataset. The analysis of applied behaviors of ASD is carried on by applying the data preprocessing to feature selection methods for better understanding the various critical and complicated medical issues.

Manuscript published on 30 August 2019.
*Correspondence Author(s)
K. Vijayalakshmi, Research Scholar, REVA University, Bangalore, India.
Dr. M. Vinayakamurthy, Professor, REVA University, Bangalore, India.
Dr. Anuradha, PG Department, STC College, Pollachi, Tamilnadu. India.

The paper work flows as follows: section II contain Literature review on autism spectrum disorders and feature selection from different papers, section III on proposed work, section IV have results of the proposed work and section V have the conclusion of the paper work.

II. RELATED WORKS

The variable selection is for analysis of diet related chronic diseases prediction using the clustering approach [1]. Behavior of autism spectrum disorders based on self-organizing approach to find the relationship between crisis duration and frequency of self-organized critically using statistical analysis. It has three-level classification such as mild, disruptive but low risk, moderate, minor injury and severe [2]. An evaluation of autism spectrum disorders such as statistical analysis of mental disorders and individuals with disabilities for improvement in education [3]. The core features, diagnosis and DSM-5 and DSM-IV classification of behavioral manifestations of autism spectrum disorder [4].

Feature selection algorithms are categorized into filter [7], wrapper [5] and hybrid [6]. The filtering model provides subsets of feature without using any mining algorithm, in turn the wrapper makes use of predetermined mining algorithm performance as its evaluation metric. Due to that, wrapper model is computationally more expensive than the filter model [10]. Most of the earlier research contributions majorly worked on attribute selection under classification with label, in which class information is available[8](supervised feature selection). But whereas, the latest work can be applied to feature selection on clustering techniques that comes without label(or unsupervised feature selection)[9]. The behavioral issues of learning and feeding skills of autistic clusters by highlighting the importance of early identification and intervention of ASD characteristics. The parents can identify the safe food that are preservatives, gluten and casein and sugar free which can reduce the hyper behavioral attributes of autistic children[12].

An intelligent feature selection should have an integrated system that recommends for the most suitable algorithms automatically to the user[13]. The Boruta feature selection algorithm used to solve the feature selection problem for data coming from evaluation sheets of subjects with Autism Spectrum Disorders (ASD)[14]. The fusion of Conditional Mutual Information Maximization and the SVM Recursive Feature Elimination give raise to a hybrid feature selection method using optimal SNP subset for higher accuracy [15]. A feature weighting method using a multiple kernel learning (MKL) algorithm using acoustic features for ASD children to classify[16].

III. PROPOSED WORK

In the proposed work ASD dataset is referred from the UCI Machine learning repository for feature relevance analysis. The ASD dataset[11] consists of 21 attributes (20 - input attributes,1-target attribute) and 704 instances. In this dataset, 10 behavioral and 10 individual features proved to be the most effective attributes to detect the ASD cases regarding behavior issues. The exploration of data should be completed by using attribute selection procedure which are more relevant for the process which in turn help to classify the ASD with better accuracy. Feature selection is nothing but a subset of features task should be selected that explain the hypothesis at least as well as the original set. The characteristics of Feature selection algorithms are given below:

1. 1. Starting point:

a. A. forward selection - start with no attributes to proceed forward then add attributes based on search space

b. B. backward selection – begins with all attributes and remove them successively by proceeding backward through the search space.

2. 2. Search organization: Heuristic search strategies has been proved as feasible when compared with exhaustive ones due to the better results.

3. Evaluation strategy:

a. The irrelevant attributes gets filtered from the dataset before the learning begins in the filter as it is completely independent of any learning algorithm. These algorithms use heuristics considering the generic features of the dataset in order to evaluate the merit of feature subsets.

b. An induction algorithm will be used along with a sampling techniques in the wrapper model in order to estimate the accuracy of feature subsets.

4. Stopping criterion: A feature selector based on the evaluation carried, features will be either added or removed until none of the alternatives provides the better improvement on the current feature subset merit.

The best subset of feature should be chosen, which is describes in hypothesis given as : S'⊂ S , where S' is the output by the feature selector with Q features and S is the original set with p features.

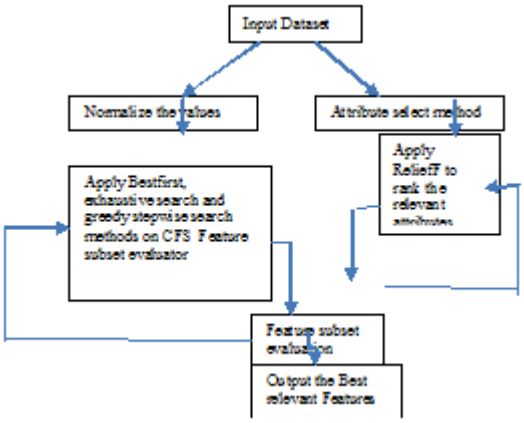

Figure.2 Architecture of Proposed work

The brief description about the attribute selection techniques used in this paper mentioned below:

CFS Subset Evaluation:

Correlation-based Feature Subset Selection (CFS) for Machine Learning evaluates any subset for its efficiency based on ability of predicting each attribute with its redundancy measures. It is beneficial to various machine learning due to its simplicity. This algorithm ranks the subsets of attributes using correlation based heuristic evaluation function.

All the irrelevant features must be removed or ignored as those attributes may have less correlation with the class. Similarly, the redundancy among features need to be identified as they may have high correlation among the remaining features. Any feature can be accepted, if it can predicts classes where the instance space is not predicted already by other features. To handle scalability on multidimensional databases, filters are the better choice as it executes faster than wrappers.

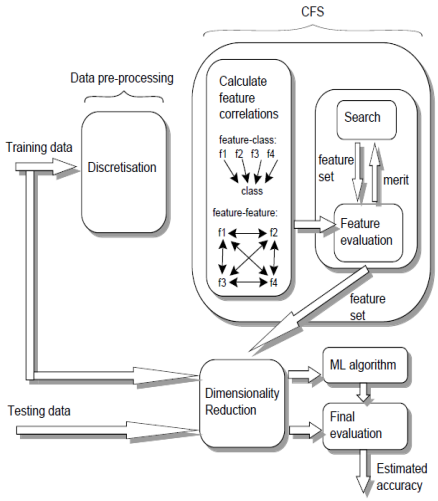

Figure. 3 CFS workflow

Search Method:

a. **Exhaustive search:** Every possible solution can be tried as an exhaustive way of searching which referred as an exhaustive search. It is also known as brute force with a requirements that has an ability to generate all candidate solutions and for efficient candidates. Exhaustive search is a simple and widely applicable algorithm for search-oriented problems

b. **Greedy Stepwise:** It is used to implement the forward and backward strategies. The forward method is proceed given below. To determine the variable that should be contributes really between the discrimination of groups. This process should be done in each and every step. The backward method is to implement the complete model of all descriptors.

c. **Bestfirst :** It is an AI search technique with backtracking capability. It migrates it searching by updating the changes to the subset obtained as like greedy hill climbing. It backtrack on its promising subsets, if the path is less promising. It explores the entire search space within the given sufficient time.

The purpose of this work is to investigate the behavioral analysis by identifying relevant characteristics of ASD through various feature selection algorithms.

IV. RESULTS & DISCUSSION

The feature selection algorithms such as CFS feature selection with various search methods like exhaustive search, greedy stepwise and bestfirst are used to interpret the dimensionality reduction of the dataset with/without normalization. The paper highlights the importance of applying data preprocessing on feature selection, may probably improve the performance of mining techniques.

The results are shown in the following tables and figures. Table.2 shows the comparison of feature selection algorithms with and without preprocessing.

S.No.	Search Method for CFS Feature subset Evaluator	Without preprocessing			With Preprocessing (Normalize)		
		No. of attributes before filtering	No. of attributes after filtering	Merit of best subset	No. of attributes before filtering	No. of attributes after filtering	Merit of best subset
1	Exhaustive Search	21	12	1	21	12	0.493
2	Greedy Stepwise	21	12	1	13	11	0.469
3	Bestfirst	21	12	1	21	12	0.493

Table. 1 Behavioral Attribute Normalization using feature selection

Figure.4 denotes graphically on CFS Feature subset Evaluator with and without normalization. It is observed from the following figure that the efficiency of mining techniques can be improved in a better way due to dimension reductionality.

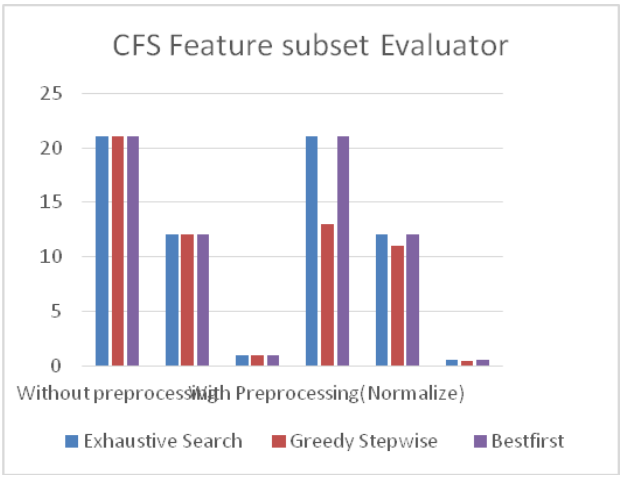

Figure. 4 CFS Feature subset Evaluator

RELIEF as an instance based learning, assigns a relevance weight to every attribute present in the dataset in order to rank those attributes. Those exceeded with a specified threshold value will form the final subset. Based on :

a. high weight - differentiates between instances from different classes and

b. same weight - instances of the same class.

It is observed from the table 2, the number of dimensions that are applied to filter is reduced due to preprocessing which optimize the time and maximise the computational efficiency.

Feature subset Evaluator	Search Method	Without preprocessing		With Preprocessing (Normalize)	
		No. of attributes before filtering	No. of attributes after filtering	No. of attributes before filtering	No. of attributes after filtering

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Relief Filtering	Ranking	21	20	13	12 (Attribute Selection Method)

Table. 2 Ranking Behavioral Attribute analysis using ReliefF with preprocessing

The following figures 5 & 6 represents the qualifying variables in the dataset through ranking and provided with and without preprocessing.

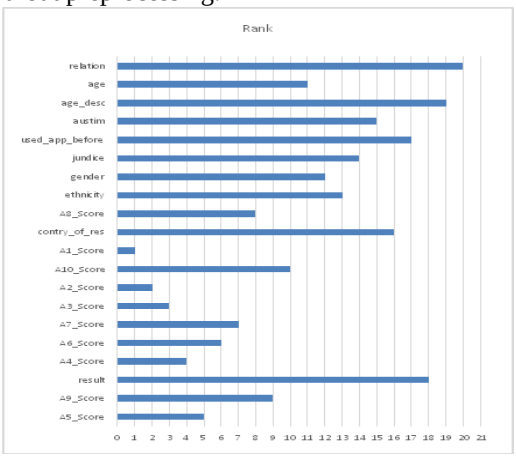

Figure. 5 Attribute Ranking without preprocessing

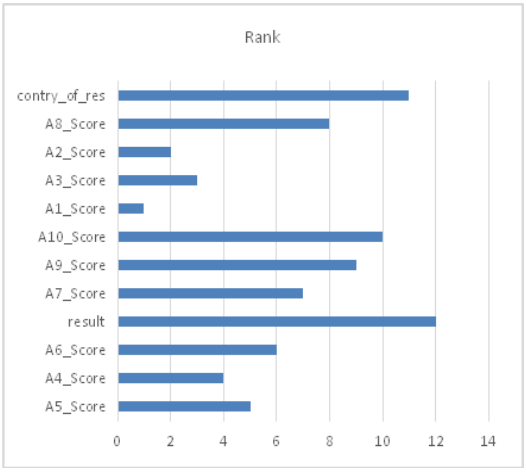

Figure. 6 Attribute Ranking with preprocessing

Based on the ranking measures, the capabilities of various mining process like classification or clustering can be improved and enhanced.

V. CONCLUSION

In this paper, the normalization and ranking of attributes of the given ASD dataset under feature selection is addressed. Using this study, the behavioral analysis of ASD attributes can be applied further to classify the instances for predicting the risk level of autism. Further in future work, the ranking feature selection method, ReliefF can be applied to evaluate on various machine learning classification algorithms in terms of prediction accuracy and precision for the best classifier.

REFERENCES

1. Abby Flynt, Madeleine I. G. Daepp, Diet- related chronic disease in the northeastern united states: a model based clustering approach, Springer, 2015.
2. LucioTonello, Luca Giacobbi, Alberto pettrnon, Alessandro Scuotto, Massimo Cocchi, Fabio Gabrielli, and Glenda cappelo, Crisis Behavior in Autism Spectrum Disorders: A Self-Organized critically Approach, Hindawi, 2018.
3. Malinda L. Pennington, Douglas Cullinan, and Louise B. Southern, Defining autism: variability in state education agency definitions of and evaluations for autism spectrum disorders, Hindawi, 2014.
4. So Hyun Kim, Catherine Lord, The Behavioral Manifestations of Autism Spectrum Disorders, Elsevier, 2014.
5. R. Caruana and D. Freitag, Greedy Attribute Selection, Proc. 11th Int'l Conf. Machine Learning, pp. 28-36, 1994.
6. S. Das, Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection, Proc. 18th Int'l Conf. Machine Learning, pp. 74-81, 2001.
7. M. Dash, K. Choi, P. Scheuermann, and H. Liu, Feature Selection for Clustering-a Filter Solution, Proc. Second Int'l Conf. Data Mining, pp. 115-122, 2002.
8. M. Dash and H. Liu, Feature Selection for Classification, Intelligent Data Analysis: An Int'l J., vol. 1, no. 3, pp. 131-156, 1997.
9. M. Dash and H. Liu, Handling Large Unsupervised Data via Dimensionality Reduction, Proc. 1999 SIGMOD Research Issues in Data Mining and Knowledge Discovery (DMKD-99) Workshop, 1999.
10. R. Kohavi and G.H. John, Wrappers for Feature Subset Selection, Artificial Intelligence, vol. 97, nos. 1-2, pp. 273-324, 1997.
11. Tabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. Proceedings of the 1st International Conference on Medical and Health Informatics 2017, pp.1-6. Taichung City, Taiwan, ACM.
12. K. Vijayalakshmi, Dr. Vinayakamurthy, Dr. V. Anuradha, Mining Behaviors on Learning and Feeding Issues of Autism Clusters, IJSTR, vol 1 , issue 1, 2017.
13. Huan Liu, Lei Yu, Toward Integrating Feature Selection Algorithms for Classification and Clustering, IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 4, April 2005.
14. Krzysztof Pancerz, Wiesław Paja, Jerzy Gomuła, Random Forest Feature Selection for Data Coming from Evaluation Sheets of Subjects with ASDs, Proceedings of the Federated Conference on Computer Science and Information Systems pp. 299–302, ACSIS, Vol. 8. ISSN 2300-5963, 2016.
15. Raid Alzubi, Naeem Ramzan, Hadeel Alzoubi, Hybrid Feature Selection Method for Autism Spectrum Disorder SNPs, IEEE, 2017
16. Yasuhiro Kakihara, Tetsuya Takiguchi, Yasuo Ariki, Yasushi Nakai and Satoshi Takada, Acoustic Feature Selection Utilizing Multiple Kernel Learning for Classification of Children with Autism Spectrum and Typically Developing Children, Proceedings of the 2013 IEEE/SICE International Symposium on System Integration, Kobe International Conference Center, Kobe, Japan, December 15-17.

AUTHORS PROFILE

K. Vijayalakshmi received M.C.A degree from University of Bharadhidasan, Trichy India, in 1999, and the M.Phil degree in Computer Science from Manonmaniam Sundranar University, Trinelveli, India, in 2002. Currently she is pursuing Ph.D programme in Computer Science and Engineering at REVA University, Bangalore, Karnataka, India. She has presented and published research papers in national conferences & International journals. She is a life member of ISTE (Indian Society for Technical Education)

Dr. M. Vinayakamurthy is a Professor in School of Computer Science and Applications, REVA Unviersity, Bangalore, Karnataka, India. He has more than 30 years of teaching and research experience. His areas of specialization include Data mining, Fuzzy logic and Optimization Techniques. He has over 50 publications in International Conferences and Journals to his credit. He has also published books in the field of Mathematics.

Dr. Anuradha is currently working as HOD, PG in a reputed institution, Tamilnadu, INDIA. She has 20 years of experience in teaching with research experience. She published and presented in various conferences and journals both at national and international.