International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Class Association Rules based Feature Selection for Diagnosis of Dravet Syndrome

B. Dwarakanath[1], K Ramesh Kumar[2]

[1]Hindustan University, Department of Information Technology, Chennai, Tamil Nadu, India

Abstract: ***Medical science industry has huge amount of data, but most of this data is not mined to find out hidden information in data. Data mining techniques can be used to discover hidden patterns that can be effectively applied to disease diagnosis which will help the physicians to take effective decision. Due to the significance of data mining techniques, association rule mining which is one of the data mining techniques to find the hidden rules from the data is applied to disease diagnosis by various researchers. In this paper, we propose a class association rules based feature selection for the diagnosis of dravet syndrome. The proposed technique for the diagnosis of Dravet syndrome contain two important steps, namely preprocessing, feature selection through association rules, and the performance is compared between them.***

Keywords: Association rules, feature selection, support, confidence and dravet syndrome.

1. Introduction

Information retrieval from huge amount of data available at different health centers is quiet challenging. Therefore it is quite impossible to pull out expert knowledge from the world of medical dataset; also there was lot of problems because of imperfection of knowledge. Recently it brings attention for computer scientists, especially in the area of knowledge mining and artificial intelligence [4]. It is because of the computers and informatics, there is improvement in overall health delivery system [5].

Data gathering as well as consequent storage, retrieval and handling of the data are enhanced by efficient computerization through database in static fashion in medical practice. Such decision-making through computers is fruitful and improves accuracy [6]. Using knowledge gained from previous known data, data classification process is mostly applied in problems of medicine, social science management and engineering. Different types of problems such as disease diagnosis, medical image recognition, and credit evaluation using classification techniques [3].

One of the disease diagnoses achievable in the medical field is Epilepsy which is clinically and genetically heterogeneous group of disorders in human. 'Severe myoclonic epilepsy of infancy' (SMEI) or 'Dravet syndrome' is a fatal infantile-onset epilepsy affecting about 1 in 20 000 to 40 000 children with a two-fold predominance in males [7].

Mostly, mutations in genes encoding ion channels in brain neurons have been analyzed in various epilepsy syndromes [8]. It delivers extraordinary advancement in terms of diagnostic molecular pathology and reduces over-reliance on traditional clinical classification [7]. An initial diagnosis of Dravet syndrome is significant for treatment and genetic counseling [9]. But the detection of Dravet syndrome among children was a challenging task. However, the Information on the pervasiveness and attributes of early, vaccination-related seizures in Dravet syndrome make better recognition of Dravet syndrome among children presenting with seizures following vaccinations. The knowledge that vaccination, although possibly the trigger for the first seizure, is not the reason for genetic epilepsy syndrome of their child, could support immunization of other children in the family [10].

Statistical programming styles are efficient and effective approach, in medical data classification. In this paper, we propose a class association rules based feature selection and regression model for the diagnosis of dravet syndrome. Here, the given dataset is divided into training dataset and testing dataset. The training dataset is then converted into .arff format to apply the class association rule. The rules are generated using class association rule based on the constraints given. The weight value for each attribute in the training dataset is calculated using the rules generated. The attribute that has least weight value is removed from the training dataset as well as from the testing dataset.

2. Related Work

In this section, some classification method related with disease diagnosis based on association rules and regression analysis is presented. An automated identification system for breast cancer recognization had been designed by Murat Karabataka and M.Cevdet Ince, [3]; based on association rules (AR) and neutral network (NN). Moreover, feature extraction is the significant key for pattern recognition and classification. Here credit goes to Neutral Network, since the dimension of breast cancer dataset was reduced by association rule mining and intelligence classification, which is based on Neutral Network. The performance of designed AR+NN system has been compared with NN model.

A combination method for identification of disease like erythematosquamous, was designed by Murat Karabataka and M. Cevdetince, [2], based on Association Rule (AR) and Neutral Network (NN). For pattern recognition and classification, feature extraction was the key. Again, the best classifier will performs poorly, if the feature were not selected well.

R. Chaves *et al.,* [1] have designed AR-based Fs technique to overcome the negativity of size problem of a functional

database. Previously, it was designed for the early detection of Alzheimer's disease (AD) in order to design a CAD system. The ARs were acquired from activated blocks of controls at different minsuff and mincounf values. Also, the maxima values are selected from the most important voxels, rejecting the rest which in turn cent 100%. In order to verify the dependability of the designed AR FS-based method in terms of Acc, Spe, Sen converging to ideal operating point of 91.75%, 95.12%, 89.29% respectively for 12 PLS features (SPECT) and 90%, 90.67%, 89.33%(PET), which performed better than recently reported techniques by many try out conducted.

Claudia B. Catarino *et al.,* [11] have stated a string of 22 adult patients, together with three adult post-mortem cases with Dravet Syndrome. They checked the medical history, seizure types and frequency, antiepileptic drugs, cognitive, social and functional outcome and consequence of probes for all patients. They performed a methodical neuropathology study comparing the controls and a range of relevant paediatric tissue, from post-mortem materials from three adult cases with Dravet syndrome. New mutations were described for 11 adult patients. Multiple seizure types despite polytherapy, and age-dependent evolution in seizure semiology and electroencephalographic pattern Features included in Dravet syndrome in adulthood. Their study has offered an indication that Dravet syndrome was at least in part a epileptic encephalopathy. The effect of severe epilepsy on the developing child had been studied by Berten Ceulemous and Patrick CRAS, [12].

A severe epilepsy syndrome with mental retardation present with febrile seizures in infancy is developed by the children on later stage. Most of the patient's DNA-analysis is diagonised by Dravet Syndrome, a neuronal sodium channel SCN1A. In the past few years the treatment of severe myoclonic epilepsy in infancy has been improved. By avoiding anti-epileptic drug which block sodium channels, epilepsy was prevented, by combining two major anti-epileptic drugs. To improve the quality of life for these patients a significant acute seizure treatment is necessary. Continuing development was essential to evaluate if they improved the development possibilities for these children.

3. Motivation of the Work

In the research works given in [1]-[3], association rules are mined to find the important features to segregate the irrelevant and redundant attributes from a data set, thus the dimension of the data set can be reduced such a way the complexity is reduced and the performance can be increased. When analyzing all these works, the traditional apriori algorithm was applied with the interesting measure like, support and confidence. In order to bring more correlation among the attributes and class label, class association rules will be mined using apriori algorithm with a newly derived interesting measure which considers both frequency and correlation to find the rules. The overall steps of the proposed technique for diagnosis of Dravet syndrome contain two important steps, i) preprocessing, ii) feature selection through association rules, At first, the input data is transformed to suitable format to mine the association rules. Then, a newly derived interesting measure will be used to mine the class association rules that will be used to select the best features form the input data. Once the features are selected will be used for the diagnosis of Dravet syndrome diseases.

4. Proposed Class association rules-based feature selection

This section delineates our proposed class association rules-based feature selection and regression model for diagnosis of dravet syndrome. The Fig.1 shows the process of our proposed technique.

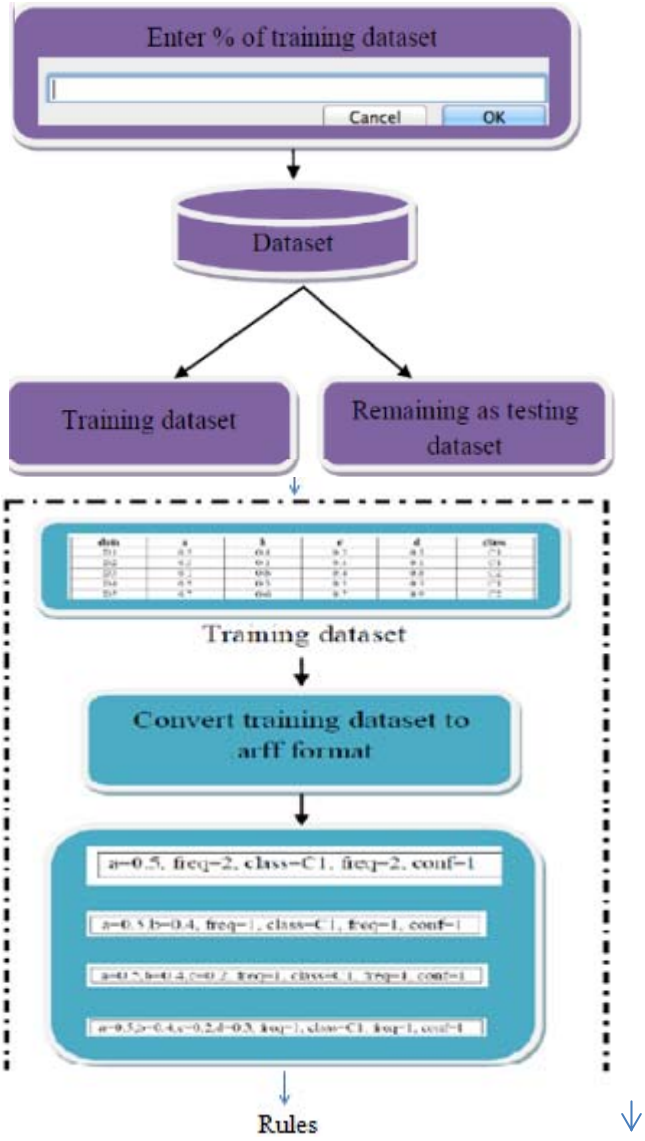

Figure 1: Process of our proposed technique

The Fig.1 explains as follows: Initially the dataset is split into two as training dataset and testing dataset. The training and testing datasets are split based on the percentage value. The remaining data except the percentage of data taken for training is chosen as testing data. The training data are then converted to .arff format to apply class association rule. The rules for the training dataset are generated using the class association rule.

```
1. y=0 58 ==> class=0 56    conf:(0.97)
2. v=0 y=0 58 ==> class=0 56    conf:(0.97)
3. w=0 y=0 58 ==> class=0 56    conf:(0.97)
4. x=0 y=0 58 ==> class=0 56    conf:(0.97)
5. y=0 z=0 58 ==> class=0 56    conf:(0.97)
6. y=0 aa=0 58 ==> class=0 56    conf:(0.97)
7. y=0 ac=0 58 ==> class=0 56    conf:(0.97)
8. y=0 ae=0 58 ==> class=0 56    conf:(0.97)
9. y=0 af=0 58 ==> class=0 56    conf:(0.97)
10. y=0 ag=0 58 ==> class=0 56    conf:(0.97)
11. y=0 ah=0 58 ==> class=0 56    conf:(0.97)
12. y=0 ai=0 58 ==> class=0 56    conf:(0.97)
13. y=0 aj=0 58 ==> class=0 56    conf:(0.97)
14. y=0 al=0 58 ==> class=0 56    conf:(0.97)
15. y=0 am=0 58 ==> class=0 56    conf:(0.97)
```

Figure 2: Sample association rules generated

The Fig.2 explains the diagnosis of Dravet Syndrome disease based on the association rule generated with minimum support and confidence. It is shown by an equation below:

$$conf = \frac{n(s)}{\max(n(s) in\ a\ class)}$$

The Algorithm for the analysis and diagnosis of Dravet Syndrome is Algorithm for Proposed Technique

Input: Dataset
Output: Classified data

1. Start the process by giving the dataset as input
2. Give the percentage level to separate the training dataset from the dataset given as input
3. The remaining data from the dataset are taken as testing dataset
4. Concert training dataset into .arff format to apply class association rule

5. Result and Discussion

This section explains the results obtained for our proposed technique. The results are compared using two regression techniques based on the metrics we set in class association rule. The metrics are minimum support and confidence. The result simply that after diagnosis of the dravet syndrome the severity for majority of the patients is 56 months. The results are compared for accuracy and execution time.

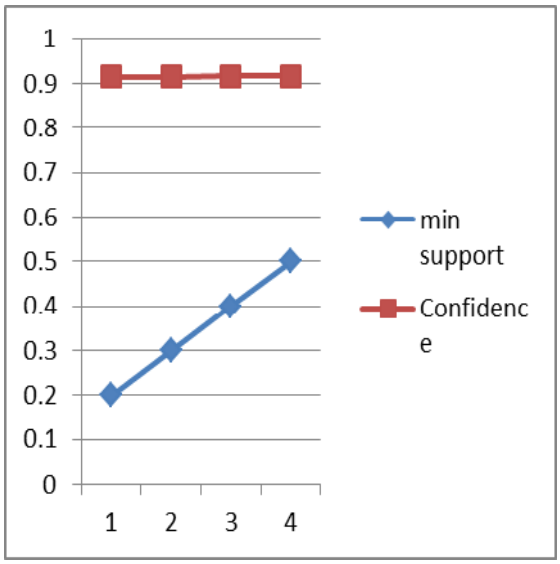

Figure 4: Accuracy based on Min. Support and Confidence

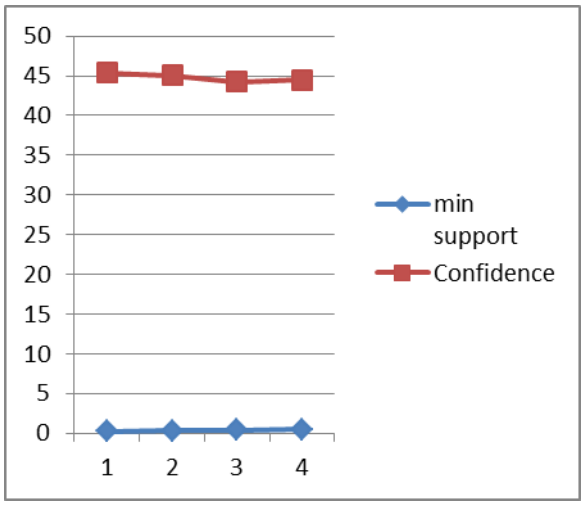

Figure 5: Time based on Min. Support and Confidence

5.1 Dataset

The dataset used to experiment the proposed technique is 'Dravet Syndrome GT' which is taken in real time. This dataset contains the details about ten patients who were affected by dravet syndrome. The attributes in the dataset are as follows: patient id, patient name, gender, age in years, age in months, number of months the patient fed mother's milk, first episode seizure occurrence month, number of visits to hospitals with respect to age in months, seizure occurrence, general reasons for seizure occurrence, general symptoms, tests, seizure type, vitamin drugs, seizure drugs, seizure status, activities and therapy.

5.2 Experimentation

In our experimentation, fifty percentages of data is taken for training and the remaining fifty percentages for testing. The experimentation is done the proposed technique based on minimum support and confidence for the analysis and diagnosis of dravet syndrome, and is compared in terms of accuracy and execution time.

6. Conclusion

In this paper we proposed a class association rule based feature selection for the diagnosis of dravet syndrome. The dataset is initially split into training dataset and testing dataset. The training dataset is converted into .arff format, the testing data is use for the analysis and the class association rule generated rules from the input dravet syndrome dataset. In future, measures can be used for analysis of dravet syndrome dataset missing values and regression technique may be applied to predict the dravet syndrome diagnosis.

References

[1] R. Chaves, J. Ramírez, J.M. Górriz, C.G. Puntonet, "Association rule-based feature selection method for Alzheimer's disease diagnosis", Expert Systems with Applications, vol. 39, pp. 11766–11774, 2012.
[2] Murat Karabatak, M. Cevdet Ince, "A new feature selection method based on association rules for diagnosis of erythemato-squamous diseases", Expert Systems with Applications, vol. 36, pp. 12500–12505, 2009.
[3] Murat Karabatak, M. Cevdet Ince, "An expert system for detection of breast cancer based on association rules and neural network", Expert Systems with Applications, vol. 36, pp. 3465–3469, 2009.
[4] B. K. Tripathy, D. P. Acharjya and V. Cynthya, "A Framework For Intelligent Medical Diagnosis Using Rough Set With Formal Concept Analysis", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.2, No.2, 2011.
[5] R.D. Lele, Computer in Medicine. Tata McGraw Hill Publishing Company Ltd., New Delhi, 1988
[6] D.M. Shah & S.M. Patel, "IT in Healthcare", Presented at the Proceeding of the National Seminar on Computer Applications, 2001.
[7] Chloe M Mak, KY Chan, Eric KC Yau, Sammy PL Chen, WK Siu, CY Law, CW Lam and Albert YW Chan, "Genetic diagnosis of severe myoclonic epilepsy of infancy (Dravet syndrome) with SCN1A mutations in the Hong Kong Chinese patients, Hong Kong Medical Journal, Vol. 17, No. 6, 2011.
[8] Ingrid E. Scheffer, "Diagnosis and long-term course of Dravet syndrome", European journal of pediatric neurology, pp.1-4, 2012.
[9] Ceulemans B, Boel M, Claes L, Dom L, Willekens H, et al. "Severe myoclonic epilepsy in infancy", Toward an optimal treatment. J Child Neurol, Vol. 19, pp.516–521, 2004.
[10] Nienke E. Verbeek, Nicoline A. T. van der Maas, Floor E. Jansen, Marjan J. A. van Kempen, Dick Lindhout and Eva H. Brilstra," Prevalence ofSCN1A-Related Dravet Syndrome among Children Reported with Seizures following Vaccination: A Population-Based Ten-Year Cohort Study", PLOS ONE Journal, Vol. 8, No.6, 2013.
[11] Claudia B. Catarino, Joan Y.W. Liu, Ioannis Liagkouras, Vaneesha S. Gibbons, Robyn W. Labrum, Rachael Ellis,Cathy Woodward, Mary B. Davis, Shelagh J. Smith,J. Helen Cross, Richard E. Appleton, Simone C. Yendle, Jacinta M. McMahon, Susannah T. Bellows, Thomas S. Jacques, Sameer M. Zuberi, Matthias J. Koepp,Lillian Martinian, Ingrid E. Scheffer,Maria Thom and Sanjay M. Sisodiya, "Dravet syndrome as epileptic encephalopathy: evidence from long-term course and neuropathology", A Journal of Neurology, Vol. 134, pp.2982-3010, 2011.
[12] Berten CEULEMANS and Patrick CRAS, "Severe Myoclonic Epilepsy in Infancy" Relevance for the clinician of severe epilepsy starting in infancy", Acta neurol. belg., Vol. 104, pp. 95-99, 2004.2004].

Author Profile

B. Dwarakanath received the B.E. and M.Tech. degrees in Computer Science and Engineering from Bangalore University and Vellore Institute of Technology in 2000 and 2004, respectively. During 2001-2002, he stayed in Muthayammal Engineering College. His research interests are Data Mining, Image Processing and Network Programming. He is now associated with Hindustan University.

Dr. K. Ramesh Kumar received the Ph.D. degree in Computer Science from Alagappa University. During 2010-2011, he stayed in Karunya University. His areas of interest are Data Mining, Ubiquitous and Cloud Computing. Currently, He is now associated with Hindustan University.