www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 5 Issue 08 August 2016 Page No.17799-17804

Significance of Data Mining Techniques in Classifying Dyscalculia

Mrs. Sampada Margaj[1], Dr. Seema Purohit[2]

[1]Department of Computer Science, Mumbai University,
Kalina, Santacruz East, Mumbai, Maharashtra 400098
sampada.malhar@gmail.com
[2]Department of Mathematics, Kirti M. Doongursee College,
Off. Veer Savarkar Marg, Dadar West, Mumbai, Maharashtra 400028
supurohit@gmail.com

Abstract:

Detection of learning difficulties among school children is a topic of research as predicting human behavior is a challenge for entire research community by taking this concern the study aims at predicting and classifying Dyscalculia among primary school children. Basically by following the approach of positive psychology towards learning detecting and designing intervention program for learner is a crucial aspect which actually requires targeted investigation so that right child should get the right help. In the present study we have used three classifiers rule-PART, J48 and SMO to predict the different classes of Dyscalculia.

Keywords: Dyscalculia, J48 classifier, rule-part classifier, Sequential Minimal Optimization, SMO

1. Introduction

In our country there has not been much research and work in the field of learning disabilities. Also there isn't enough awareness or rehabilitation measures available for children with learning problems. Learning disability is broadly classified as Dyslexia, Dysgraphia and Dyscalculia. In this work we have focused on dyscalculia which is nothing but the inability of children for performing mathematical operations. Under dyscalculia, study specifically focuses on Developmental Dyscalculia. Developmental Dyscalculia is further divided into several classes, so we have tried to predict those classes using data mining techniques. Since international research reports that 5 - 8% of school-age children experience difficulties that interfere with their acquisition of mathematical concepts or procedures [1][2] and that an average of 3.6 – 6.5% have severe difficulties with acquiring numeracy and mathematics [3] an increasing interest has been shown towards the subject by international professionals.

Thus this study aims at targeting the children who are under tremendous pressure due to their bad performance in the school examinations. We have considered schools in and around Mumbai, which do not have computer facilities and other amenities unlike the private schools, where implementation and detection of dyscalculia is much easier. This paper studies the three classifiers namely rule-PART, J48 and Sequential Minimal optimization. These classifiers are evaluated based on Sensitivity, Specificity, Accuracy, Error Rate, True Positive Rate and False Positive Rate. Among these Sequential Minimal Optimization classifier is turned out to be best for the prediction of several classes of Dyscalculia. All the above work focuses the data stored in single table. Using these classifiers, we have tried to predict different classes of Dyscalculia. Author [4] has compared the Naïve Bayes, Logistic function, RBF Network, Decision Table, SMO function algorithms, and the results showed that the Naïve Bayes algorithm outperforms with 86% accuracy in 0 second, this comparative study was applied to determine the most effective techniques that are capable for the detection of heart valve disease with a high accuracy. PART works by building a rule and removing its cover, as in the separate-and-conquer technique, repeatedly until all the instances are covered. The rule construction stage differs from standard separate-and-conquer methods because a partial pruned decision tree is built for a set of instances, the leaf with the largest coverage is made into a rule, and the tree is discarded [4]. A comparative study of corporate credit rating analysis using support vector machines and back propagation neural network were analyzed [5]. In this study the performances of three models were compared with the Naive Bayes classifier, tree augmented Naive Bayes, the SVM, C4.5 and the nearest neighbor classifier and the obtained results demonstrated that the proposed models could significantly improve the performance of the naive Bayes classifier, yet at the same time maintain its simple structure[6].Julie M., determined the relevance of various data pre-processing methods in classification it was done along with dimensionality reduction for the long list of attributes. The results obtained from this study was illustrated that the data pre- processing method gives good results in prediction system and can be used to improve the performance of classifiers [7].

2. Significance of Study

In our country there has not been much research and work in the field of dyscalculia. As it is always being confused with mathematical difficulty. Also there isn't enough awareness or rehabilitation measures available for children with learning problems. Thus this study aims at targeting the children who are under tremendous pressure due to their bad performance in the school examinations. We have considered schools in and around Mumbai, which do not have computer facilities and other amenities unlike the private schools.

3. Data Representation and Pre-processing

In this study we used 237 real world datasets from schools in and around Mumbai. The population is from primary section and all sample falls under same age group. The final sheet is arranged with names and the scores of all the subtests which are our attributes in this case. Below Table1 has the list of attributes with their descriptions based on which we are going to predict the student is affecting with dyscalculia or not. For final counseling the result based on these attributes, child’s earlier history and general observations by parents and teachers are also considered.

Table 1: Set of Attributes

Sr.	Attribute	Description
1	DSR	Difficulty with Shape Recognition
2	DSD	Difficulty with Size Discrimination
3	DNA	Difficulty with Number Arrangements
4	DGS	Difficulty with Grouping Sets
5	DPV	Difficulty with Place Values
6	DNC	Difficulty with Numeric Calculations
7	DVA	Difficulty with Verbal Analysis
8	DCC	Difficulty with Counting Concepts

4. Methodology Used

In this paper, we studied J48, Rules-PART and Sequential Minimal Optimization (SMO) classifiers which help for the prediction of dyscalculia.

Following evaluation measures are used to assess each of the classifiers:

- Accuracy: Overall how often the classifier gives the correct results
- Error rate: Overall how often the classifier predicts wrong results.
- Sensitivity: When it is actually yes how often the classifier predicts yes
- Specificity: When its actually No how often classifier predicts No
- Precision: When the classifier predicts Yes how often does it correct
- Prevalence: How often the Yes condition actually occurs in our sample.

5. Results and Discussion

5.1 J48

The J48 Decision tree classifier follows the following simple algorithm. In order to predict classes, it first creates a decision tree based on the attribute values from the available training data. Whenever it encounters a training set item it identifies the attribute which categorizes the various instances clearly. This feature that is able to tell us most about the data instances so that we can classify them the best is said to have the highest information gain. Then among the possible values of this feature, if there is any value for which there is no ambiguity have the same value for the target variable, then they terminate that branch and assign to it the target value that we have obtained [4].

Internal node in tree structure denotes a test on an attribute branch which represents an outcome of the test and leaf nodes represent class labels. Decision tree generation consists of two phases Tree construction, initially all the training examples are at the root. Partition examples recursively based on selected attributes tree pruning identify and remove branches that reflect noise or outliers. Decision tree is mainly used to classify an unknown sample test the attribute values of the sample against the decision tree [8] [9].

In this study, J48 algorithm is used, which is a greedy approach in which decision trees are constructed in a top-down recursive divide and conquer manner [10].

In decision tree, each arc is labeled with predicate, which can be applied to the attribute at the parent node. The basic steps in the decision tree are building the tree by using the training data sets and applying the tree to the new data sets.

The classification results are generated by J48 on training set:

Correctly Classified Instances 154 66.67 %

Incorrectly Classified Instances 77 33.33 %

Table 2: Confusion Matrix for J48

Actual value / Predicted value	Graphical	Practognostic	Numeric	Ideaognostic	Lexical	Operational	Verbal	Sequential	Actual Total
Graphical	23	2	0	0	0	0	1	0	26
Practognostic	2	15	0	0	0	0	2	0	19
Numeric	2	6	8	0	0	0	14	0	30
Ideaognostic	1	7	0	0	0	0	24	0	32
Lexical	1	0	0	0	30	0	0	0	31
Operational	0	0	0	0	2	23	0	6	31
Verbal	0	0	2	0	0	0	30	0	32
Sequential	4	0	0	0	0	1	0	25	30
Predicted Total	26	17	36	15	31	30	45	31	231

Table 2 denotes the confusion matrix for J48 classifier on training set. It shows the predicted value by the classifier against actual value. From the above mentioned matrix following values are calculated:

Table 3: Accuracy of Decision Tree for J48

Class	TP Rate	FP Rate	Precision	Recall	F-measure
Graphical	0.885	0.049	0.697	0.885	0.780
Practognostic	0.789	0.069	0.500	0.789	0.612
Numeric	0.267	0.010	0.800	0.267	0.400
Ideognostic	0.000	0.000	0.000	0.000	0.000
Lexical	0.968	0.010	0.938	0.968	0.952
Operational	0.742	0.005	0.958	0.742	0.836
Verbal	0.938	0.166	0.423	0.938	0.583
Sequential	0.833	0.030	0.806	0.833	0.820

The second and third columns in the table denote True Positive Rate i.e. TP rate and False Positive Rate i.e. FP rate. TP Rate is the ratio of low weight cases predicted correctly cases to the total of positive cases. Using TP and FP rate we get the values of evaluation measures as follows:

Table 4: Derived values for evaluation measures

Evaluation Measures	Value
Accuracy	92%
Error Rate	8%
Sensitivity	67%
Specificity	95%
Precision	66%
Prevalence	12%

As shown in the Table 4 the accuracy for J48 is 92%, error rate is 8%, Sensitivity 67%, Specificity 95%, Precision 66% and Prevalence 12%. Confusion matrix shows the results of actual and predicted value.

5.2 Rules-part

The classification rules are constructed in two ways [11]

– Direct method
– Indirect method.

Advantages of rule-based classifications are,

– Easy to understand
– Easy to construct
– Highly expressive
– Can classify new instances quickly

It is an indirect technique for constructing classification rules. It employs partial decision tree to generate the individual rules and the tree is induced with C4.5 classifiers. After tree generation, rules are derived directly from the partial tree starting with the deepest leaf node in combination with every node along the path towards the root [12].

After performing this classifier on data set the following results we calculated which is based on the confusion matrix which is generated.

Correctly Classified Instances 169 73.1602 %
Incorrectly Classified Instances 62 26.8398 %

Table 5: Confusion Matrix for Rules-part

Actual value / Predicted value	Graphical	Practognostic	Numeric	Ideaognostic	Lexical	Operational	Verbal	Sequential	Actual Total
Graphical	24	1	0	0	0	0	1	0	26
Practognostic	4	13	0	0	0	0	2	0	19
Numeric	2	1	12	1	0	0	14	0	30
Ideaognostic	0	1	3	7	0	0	21	0	32
Lexical	0	0	0	0	29	0	0	2	31
Operational	0	0	0	0	1	23	0	7	31
Verbal	0	0	0	0	0	0	32	0	32
Sequential	1	0	0	0	0	0	0	29	30
Predicted Total	26	17	36	15	31	30	45	31	231

Table 5 denotes the confusion matrix for Rules-PART classifier on training set which shows the predicted value by the classifier against actual value.

Table 6: Accuracy of Decision tree for Rules-Part

Class	TP Rate	FP Rate	Precision	Recall	F-measure
Graphical	0.923	0.034	0.774	0.923	0.842
Practognostic	0.684	0.014	0.813	0.684	0.743
Numeric	0.400	0.015	0.800	0.400	0.533
Ideognostic	0.219	0.005	0.875	0.219	0.350
Lexical	0.935	0.005	0.967	0.935	0.951
Operational	0.742	0.000	1.000	0.742	0.852
Verbal	1.000	0.158	0.457	1.000	0.627
Sequential	0.967	0.045	0.763	0.967	0.853

The second and third columns in the table denote True Positive Rate i.e. TP rate and False Positive Rate i.e. FP rate. TP Rate is the ratio of low weight cases predicted correctly cases to the total of positive cases. Using TP and FP rate we get the values of evaluation measures as follows:

Table 7: Derived values for evaluation measures

Evaluation Measures	Value
Accuracy	93%
Error Rate	6%
Sensitivity	73%
Specificity	96%
Precision	73%
Prevalence	12%

As shown in the Table 7 the accuracy for rules-PART is 93%, error rate 6%, Sensitivity 73%, Specificity 96%, Precision 73% and Prevalence 12%. Confusion matrix shows the results of actual and predicted value. Table 5 denotes the confusion matrix for rules- PART classifier on training set.

5.3 Sequential Minimal Optimization

SMO is basically a new form of SVM (Support Vector Machine). Because SMO spends most of its time evaluating the decision function, rather than performing Quadratic Programming, it can exploit data sets which contain a substantial number of zero elements. SMO does particularly well for sparse data sets, with either binary or non-binary input data [13]. The data set pattern is matching with the same pattern as sparse data.
The classifier SMO gives following results:

Correctly Classified Instances 175 75.75 %

Incorrectly Classified Instances 56 24.24 %

Following Table 8 denotes the confusion matrix for SMO classifier on training set. It shows the predicted value by the classifier against actual value.

Table 8: Confusion Matrix for SMO

Actual value / Predicted value	Graphical	Practognostic	Numeric	Ideaognostic	Lexical	Operational	Verbal	Sequential	Actual Total
Graphical	25	1	0	0	0	0	0	0	26
Practognostic	0	17	0	0	0	0	2	0	19
Numeric	0	6	8	4	0	0	12	0	30
Ideaognostic	0	2	1	10	0	0	19	0	32
Lexical	0	0	0	0	30	0	0	1	31
Operational	0	0	0	0	1	25	0	5	31
Verbal	0	0	1	1	0	0	30	0	32
Sequential	0	0	0	0	0	0	0	30	30
Predicted Total	25	26	10	15	31	25	63	36	231

From the above mentioned matrix following values are calculated:

Table 9: Accuracy of Decision tree for SMO

Class	TP Rate	FP Rate	Precision	Recall	F-measure
Graphical	0.962	0.000	1.000	0.962	0.980
Practognostic	0.895	0.042	0.654	0.895	0.756
Numeric	0.267	0.010	0.800	0.267	0.400
Ideognostic	0.313	0.024	0.667	0.313	0.426
Lexical	0.968	0.005	0.968	0.968	0.968
Operational	0.806	0.000	1.000	0.806	0.893
Verbal	0.938	0.141	0.476	0.938	0.632
Sequential	1.000	0.030	0.833	1.000	0.909

The second and third two columns in the table denote True Positive Rate i.e. TP rate and False Positive Rate i.e. FP rate. TP Rate is the ratio of low weight cases predicted correctly cases to the total of positive cases. Using TP and FP rate we get the values of evaluation measures as follows:

Table 10: Derived values for evaluation measures

Evaluation Measures	Value
Accuracy	94%

Error Rate	5%
Sensitivity	75%
Specificity	96%
Precision	75%
Prevalence	12%

As shown in the Table 7 the accuracy for SMO is 94%, error rate is 5%, Sensitivity 75%, Specificity 96%, Precision 75% and Prevalence 12%. Confusion matrix shows the results of actual and predicted value.

6. Conclusion

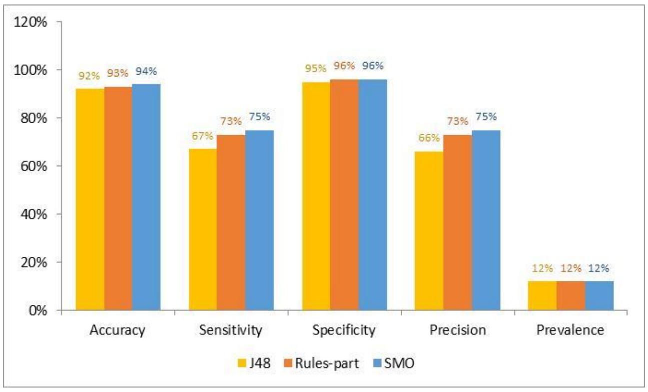

Figure 1

Figure 1 gives the overall summary of all three classifiers J48, Rules-PART and SMO. For accuracy SMO gives the higher value which is 94% compared to other two classifiers. For sensitivity and specificity also SMO gives the higher value 73% and 75% respectively. For precision, SMO is showing the better results than others. For prevalence all three classifiers are showing the same rate 12%. And all the three classifiers take the same time to build the model. So finally we can conclude saying SMO gives better results on classifying different subclasses of Dyscalculia compare to J48 and Rules-PART. There is still future work to try to find out the model which will give more accuracy.

References

[1] Geary, D. C. (2004). Mathematics and learning disabilities. Journal of Learning Disabilities, 37, 4-15.

[2] Fuchs, L.S., & Fuchs, D. (2002). Mathematical problem-solving profiles of students with mathematical disabilities with or without comorbid reading disabilities. Journal of Learning Disabilities, 35, 563-573.

[3] Lewis, C., Hitch, G., & Walker, P. (1994). The prevalence of specific arithmetic and specific reading difficulties in 9- and 10-year old boys and girls. Journal of Pupil Psychology and Psychiatry, 35, 283-292A. Bonnaccorsi, “On the Relationship between Firm Size and Export Intensity,” Journal of International Business Studies, XXIII (4), pp. 605-635, 1992. (journal style)

[4] Sushil Kumar Rameshpant Kalmegh,” Effective Evaluation of Classification of Indian News Using Various Data Mining Algorithm”, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 9, November 2014. www.ijiset.com ISSN 2348 – 7968.

[5] Zan Huang, Hsinchun Chena, Chia-Jung Hsu, Wun-Hwa Chen and Soushan Wu, “Credit rating analysis with support vector machines and neural networks: a market comparative study,” Decision Support Systems (Elsevier) 37 (2004) pp. 543– 558. M. Clerc, “The Swarm and the Queen: Towards a Deterministic and Adaptive Particle Swarm Optimization,” In Proceedings of the IEEE Congress on Evolutionary Computation (CEC), pp. 1951-1957, 1999. (conference style)

[6] Sona Taheri∗, Musa Mammadov∗,∗∗ ∗ Centre for Informatics and Applied Optimization School of Science, Information Technology and Engineering, University of Ballarat, Victoria 3353, Australia “LEARNING THE NAIVE BAYES CLASSIFIER WITH OPTIMIZATION MODELS”sonataheri@students.ballarat.edu.au, sona.taheri@unisa.edu.au Victoria Research Laboratory National ICT Australia, Victoria 3010, Australia e-mail: m.mammadov@ballarat.edu.au,musa.mammadov@nicta.com.au, Int. J. Appl. Math. Comput. Sci., 2013, Vol. 23, No. 4, 787–795 DOI: 10.2478/amcs-2013-0059.K. Deb, S. Agrawal, A. Pratab, T. Meyarivan, “A Fast Elitist Non-dominated Sorting Genetic Algorithms for Multi-objective Optimization: NSGA II,” KanGAL report 200001, Indian Institute of Technology, Kanpur, India, 2000. (technical report style)

[7] Julie M. David et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 , 2011, 829-835,”Prediction of Learning Disabilities in School Age Children using SVM and Decision Tree”.

[8] Ian H. Witten, Eibe Frank, Mark A. Hall., Data Mining Practical Machine Learning Tools and Techniques, Third IJISET - International

Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 9, November 2014. www.ijiset.com ISSN 2348 – 7968 552 Edition, Morgan Kaufmann Publishers is an imprint of Elsevier

[9] Anshul Goyal and Rajni Mehta., Performance Comparison of Naïve Bayes and J48 Classification Algorithms. International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11 (2012)

[10] Margaret Mary. T, Hanumathappa. M, "Diagnosis of Learning Disabilities in school going children using Data Mining Techniques: a survey", IJISET, International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 8, October 2014.ISSN 2348 – 7968

[11] M. Thangaraj, C. R. Vijayalakshmi "Performance Study on Rule-based Classification Techniques across Multiple Database Relations " , International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 5– No.4, March 2013

[12] Frank, E., Witten, I., J.1998. Generating Accurate Rule Sets without Global Optimization, Machine Learning: Proceedings of the Fifteenth International Conference, pp.144-151, Madison, Wisconsin, Morgan Kaufmann, San Francisco.

[13] John C. Platt, "Fast Training of Support Vector Machines using Sequential Minimal Optimization," 2000.Microsoft Research 1 Microsoft Way, Redmond, WA 98052, USA jplatt@microsoft.com

Author Profile

Sampada Margaj received the B.Sc. and M.Sc. degrees in Computer Science from University of Mumbai in 2002 and 2004, respectively. Further she did M.Phil. in Computer Science in 2008. She is carrying 12+ years of experience in teaching various subjects of Computer Science and Information Technology at graduate and postgraduate levels. Her expertise is in artificial intelligence and data mining.

Dr. Seema Purohit has a Post-graduate and Doctorate Degree in Computational Mathematics from University of Mumbai. With a rich experience of 25 years in the education field; she has an Administrative Experience of more than 14 years of heading the Department and an Institution. She has a unique combination of a Technological Expertise, Managerial and Operational Excellence by which understanding the needs of today's Education System, IT Industries and Business Organizations becomes easy. She has also delivered several key note addresses and guest lecturers on the topics related to Technology and Management in many national and international conferences.