Indian Journal of Science and Technology, *Vol 8(28), DOI: 10.17485/ijst/2015/v8i28/73756, October 2015*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

The Minimum Normalized Dissimilarity between Objects based Rough Set Technique for Elucidating Learning Styles in E-learning

K. S. Bhuvaneshwari[1*], D. Bhanu[2] and S. Sophia[3]

[1]Department of CSE, Karpagam College of Engineering, Coimbatore - 641032, Tamil Nadu, India; bhuvana.me@gmail.com
[2]Department of CSE, Karpagam Institute of Technology, Coimbatore - 641105, Tamil Nadu, India; bhanu.saran@gmail.com
[3]Department of ECE, Sri Krishna College of Engineering and Technology, Coimbatore - 641008, Tamil Nadu, India; sofia_sudhir@yahoo.com

Abstract

Objectives: The objective of this research work is to analyse the learning styles of individual users in an e-learning system and to formulate a mathematical model to determine it. **Methods:** This research work proposes MNDBO, a rough set based clustering technique for elucidating learning styles by finding minimum normalized dissimilarity between objects in e-learning. The proposed clustering technique uses a normalized score value for estimating the deviation between data's through the equivalence property of rough set theory. **Findings:** Further, the result predicts that the clusters produced by MNDBO algorithm perform better than MADO by 11%, 14% than SDR and 16% than MMR in terms of cohesion. Furthermore, MNDBO algorithm also produces better results than MADO by 15%, 23% than SDR and 27% than MMR in terms of coupling. In addition MNDBO algorithm maximizes the cohesion and simultaneously reduces the coupling rate based on varying number of cluster size on an average 15% and 19% respectively. **Applications/Improvements:** If this Rough set based clustering technique is used means we can able to discover successfully relations with inconsistent or incomplete data.

Keywords: Clustering, Dissimilarity between Objects, E-learning, Learning Styles, Rough Set Theory, Standard Deviation

1. Introduction

E-learning is one of the emerging technologies incorporated by the worldwide educational organization for the purpose of enabling services like:

- Providing virtual learning facilities.
- Creating content for various domains for learners.
- Creating the virtual class environment by means of online admission, online attendance and online conduction of classes[1].

In order to give successful online administrative services in an e-learning system, the information about the learners and their interested domains of learning must be known. This type of learners' information assumes a vital role for the effective usage of e-learning framework[2]. However, the problem associated with this implementation methodology is growth of learners' information exponentially towards the time factor[3]. Then, the analysis of learning factors in a large amount of learners' information becomes a challenging issue. Hence, there is a need arises to incorporate a set of adaptable rules to analyze the learners' information for designing an effective and efficient e-learning system.

This paper contributes a rough set theory based data analysis model for mining relevant and significant information from the large amount of learners' data of the e-learning system. This model incorporates the principle

*Author for correspondence

of reducts in rough set theory for extracting knowledge from the learners' information[4]. For an effective e-learning system we need to analyze the learning style of an individual in the collected learners' information. Hence, in this paper, we incorporate rough set theory based data analytics model for mining rules and analyzing about the learners' learning styles in order to facilitate efficient learning. The main advantage of using rough set theory for this data analytics model due its potential towards:

- Extraction of relevant information.
- Decision friendly.
- High users understand ability.

In the recent past, a number of rough set theory based clustering mechanisms have been contributed by the researchers. Some of the existing rough set theory based approaches are enumerated below. Initially in 1980, the rough theory was proposed by Zdzislaw Pawlak[6], to analyze the information present in the data tables for deriving relationships among the given data. Further, this theory is also used to reduce the size of the data, deriving hidden patterns of the data and extraction of rules from the data[7]. It can also be well applicable for refining improper or incomplete information given. Researchers of the past decade have proved that rough set theory can be implemented for wide range of problems such as:

- Correlated and uncorrelated analysis.
- Rule extractions for expert systems.
- Learning from examples and switching circuits design.

Analyzing learning style of a learner plays a key role in designing an e-learning system. Each and every learner has different style of learning. Studying and analyzing learning styles based on various classification methods have been proposed by the researchers in the past decade[8]. Many of them were focused on learning style scales, some of them focused on learning style inventory[9,10]. Few researchers have given a survey on learning style analysis, learner preference checklists, preferable questionnaire to assess the learning style and ability of the learners[11–13]. Hence, this review concludes that, there is lack of mathematical model for determining the learning style of an individual in order to design an effective e-learning system.

In addition, the first benchmark system is the Min-Min-Roughness (MMR) technique proposed in[14] that utilizes the minimum of mean roughness by considering only a single attribute into account for cluster analysis. The maximum value of mean roughness is considered for estimating the partitioning attribute. Further, Tripathy and Gosh[15] presented an algorithm that clusters categorical data together based on the property of standard deviation based score for calculating estimated roughness (SDR). The technique incorporates a characteristic attribute with minimum SDR value for choosing the partitioning attribute. Prabha D and Ilango[16] proposed the Minimum Average Dissimilarity between Objects (MADO) that utilizes the elements of rough set theory for clustering data through the estimation of dissimilarity between objects. Finally, the learner characteristics like active, reflective, group or solo learning qualities called 'learner portfolios' are analyzed by a collaborative agent in an e-learning environment[20]. This methodology determines e-learner characteristics from respective user profiles and interacts with any adaptive e-learning system in an asynchronous mode. This research work also introduces a collaborative agent based model for correlating learner characteristics.

From the literature review carried out with the various clustering techniques[17–19] that involves rough set theory, it is found to have following limitations:

- A rough theory based normalized score technique that incorporates an equivalence property in clustering has not been explored to the best of our knowledge.
- A rough set theory based clustering mechanism that could identify different learning styles of e-learning has not been much explored.

Hence, a Minimum Normalized Dissimilarity between Objects (MNDBO) techniques that maximizes cohesion and minimizes coupling has been proposed. The remaining part of the paper is organized as follows. Section 2 presents the proposed work Minimum Normalized Dissimilarity between Objects (MNDBO) techniques with the associated algorithm. Section 3 presents the experimental comparison carried out with MNDBO with the considered benchmark systems. Finally, Section 4 ends with the conclusion.

2. Proposed Work

2.1 The Minimum Normalized Dissimilarity between Objects (MNDBO)

In an e-learning environment, the manipulation of dependencies and roughness between the attributes that

determine the learning capability of students depends on factors like interest, psychology, graphics content and audio content. But, it is highly difficult due to dynamic learning capabilities of target audience. However, Minimum Normalized Dissimilarity between Objects (MNDBO) techniques overcomes this limitation by incorporating a significant property of rough set theory called equivalence property. The clustering attribute is determined based on the deviation of scores estimated between the objects of each equivalent class.

Let 'S_1' and 'S_2' be the sets that contain the attributes in the data and the data's whose value is equal for a specific attribute. In each and every manipulation, the set 'S_2', elucidates the data's of each and every equivalence class. The deviation score (Dev_{Score}) estimated between data's within the set 'S2', is determined through equation (1) as:

$$Dev_{Score} = \frac{\sigma(x_a x_b)}{M_{(x_a x_b)}} \times 100 \quad (1)$$

Where, the standard deviation and expected mean of the datum of the attributes are derived through equation (2) and (3) as:

$$\sigma_{(x_a x_b)} = \sqrt{\frac{\sum_{i=1}^{n}(x - \bar{x}_i)^z}{n}} \quad (2)$$

$$\sigma_{(x_a x_b)} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (3)$$

From the deviation score manipulated from equation (1), the normalized_score that depicts the actual deviation between data of each attribute from each of the sets of data is given by equation (4) as:

$$Normalized\_score = \frac{Dev\_score(i) - Dev\_MinScore(i)}{Dev\_Max\,score(i) - Dev\_MinScore(i)} \quad (4)$$

This normalized_score is calculated based on the ratio of difference of deviation between individual cluster score and the minimum individual cluster score to the difference of deviation between maximum individual cluster score and the minimum individual cluster score. This normalized_score factor is considered as the significant factor utilized for optimal identification of clusters in order to extract knowledge from the datasets to interpret the learning styles of students during the e-learning process.

In the next section, the proposed Minimum Normalized Dissimilarity between Objects (MNDBO) algorithm is presented, the MNDBO algorithm initially considers the set S_1 as a single cluster, then based on the deviation_score and normalized_score, the equivalent classes are derived from S_1. The equivalent classes enumerated from the S_1, is considered as S_2 that contains collection of sets of data based on the number of clusters 'k' considered for cluster analysis. The number of elements of each set grouped from set S_2 depends on the number of elements that are present in each of the individual cluster. Further, the partitioning element utilized for each clustering process depends upon the minimum normalized_score, which determines the point of datum that is considered as

Algorithm 1: Pseudo code for MNDBO clustering algorithm

Notations:

S1- data set considered for cluster analysis

S2- set of equivalent classes derived from S1 based on Normalized_Score

k – Required number of clusters

Procedure (DS, k)

1: begin

2: set –Initial number of clusters INC = 1

3: initialize 'k' as the number of clusters.

4: set Parent_Node = DS

5: do

6: for each ai from S1 (i = 1 to n, where 'n' denotes the number of attributes in each data set S1)

7: for j = 1 to m, where 'm' is the possible different values of each attribute in ai

Algorithm new_ Parent_Node (INC)

18. begin

19. for (i = 1 to INC)

20. size of cluster (i) = Number of elements of cluster (i)

21. next

22. estimate 8: in Parent_Node (DS), determine the group of equivalence classes for 'ai 'with attribute value'j' denoted through set S2

9: manipulate $Dev_{Score}(S_2)$ based on $\sigma_{(x_a x_b)}$ and $M_{(x_a x_b)}$.

10. next

11. next

12. set Normalized_Score = Min (Normalized_Score (S2)) for every set with constraint $|S_2| \geq 1$

13. estimate the partitioning attribute ai based on minimum Normalized_Score

14. INC = INC+1
15. parent_Node (DS) = New_ Parent_Node (INC)
16. while (INC < k)
17. End.
Maximum (Size of cluster (i))
23. return (Number of elements of cluster (i) equals to Maximum (Size of cluster (i)))
24.End.

the estimation point of knowledge used for cluster analysis through rough set theory.

Furthermore, the proposed algorithm is compared with existing benchmark techniques like MADO, SDR and MMR through parameters like cohesion and coupling for estimating the superior performance of the proposed algorithm.

The MNDBO clustering algorithm is given below:

3. Experimental Results

In the experimental analysis, real data sets are elucidated from the feedback form of three e-learning tutorial institutions are collected for segmentation and cluster analysis. The feedback form was collected for a period of three years. The data-set1 contains 5642 records, data set2 contains 4539 and data set3 contains 3403 records. From the feedback form, four attributes of e-learning viz., accessibility, and cost effectiveness, understanding and, time-saving were used to define the values of A, C, U and T values. Where, 'A' represents the e-content accessibility index, 'C' represents the cost incurred in accessing the e-content, 'U' represents the understanding quotient that differs from each and every student and 'T' represents the amount of time saved through e-learning rather than the traditional method. Hence, all the three datasets contains only four attributes viz., A, C, U and T for each of the student.

The values of A, C, U and T are normalized as follows:

- Arrange the data set in ascending of A, C, U and T.
- Partition the data set into five equal parts with 20% of the available records in each part.
- Assign classification index to each of the divided part into highly significant, significant, moderate, tolerable, least significant.

Initially, the MNDBO algorithm is applied into the normalized dataset for segmenting the learning styles of the students into various categories. Then, the benchmark techniques considered for study like MADO, SDR and MMR are the applied to the same three normalized dataset for studying the superior performance of the proposed MNDBO algorithm. Further, performance metrics like cohesion and coupling are considered for measuring the consistent quality of the cluster, in which cohesion defines the mean similarity among each elements of the cluster while coupling denotes the degree of similarity between each pair of elements of the cluster. Furthermore, in a dataset, the degree of cohesion must be greater than the degree of coupling.

From Table 1, it is evident that the dataset 1 clusters produced by MNDBO algorithm perform better than MADO by 11%, 14% than SDR and 16% than MMR in terms of maximizing cohesion. Further, on an average the proposed MNDBO algorithm enhances the degree of cohesion by 15%. Since, the proposed clustering technique utilizes a normalized score for estimating the degree of deviation between the each data of the equivalence class.

Table 1. Aggregate cohesion value for dataset-1

Aggregate Cohesion	Cluster Groups			
	4	5	6	7
MNDBO	1.24121	1.6323	2.1253	2.8121
MADO	1.23111	1.6010	2.0945	2.7122
SDR	1.22498	1.5813	2.0345	2.6012
MMR	1.22323	1.5345	2.0407	2.5119

From Table 2, it is evident that the dataset 1 clusters produced by MNDBO algorithm perform better than MADO by 13%, 18% than SDR and 20% than MMR in minimizing coupling. Further, on an average the proposed MNDBO algorithm minimizes the degree of coupling by 18%. Since, the proposed clustering technique estimates a normalized score based on standard deviation and mean that represents the central tendency of each equivalent class for estimating the degree of coupling between the each data of the equivalence class.

Table 2. Aggregate coupling value for dataset-1

Aggregate Coupling	Cluster Groups			
	4	5	6	7
MNDBO	0.38111	0.50121	0.71223	0.92151
MADO	0.41211	0.52212	0.72343	0.93121
SDR	0.42228	0.53223	0.73257	0.93862
MMR	0.43212	0.53455	0.74253	0.94819

From Table 3, it is evident that the dataset 1 clusters produced by MNDBO algorithm perform better than MADO by 11%, 14% than SDR and 16% than MMR in terms of maximizing cohesion. Further, on an average the proposed MNDBO algorithm enhances the degree of cohesion by 15%. Since, the proposed clustering technique utilizes a normalized score for estimating the degree of deviation between the each data of the equivalence class.

Table 3. Aggregate Cohesion value for dataset-2

Aggregate Cohesion	Cluster Groups			
	4	5	6	7
MNDBO	1.1771	1.8121	2.4151	2.9121
MADO	1.1621	1.8019	2.3232	2.9101
SDR	1.1611	1.8010	2.2112	2.8151
MMR	1.1522	1.7919	2.2001	2.8101

From Table 4, it is evident that the dataset 1 clusters produced by MNDBO algorithm perform better than MADO by 13%, 18% than SDR and 20% than MMR in minimizing coupling. Further, on an average the proposed MNDBO algorithm minimizes the degree of coupling by 18%. Since, the proposed clustering technique estimates a normalized score based on standard deviation and mean that represents the central tendency of each equivalent class for estimating the degree of coupling between the each data of the equivalence class.

Table 4. Aggregate Coupling value for dataset-2

Aggregate Coupling	Cluster Groups			
	4	5	6	7
MNDBO	0.54127	0.70121	0.90121	1.09122
MADO	0.55111	0.71131	0.92212	1.100120
SDR	0.55221	0.71291	0.93343	1.102241
MMR	0.55411	0.72483	0.93996	1.103112

From Table 5, it is evident that the dataset 1 clusters produced by MNDBO algorithm perform better than MADO by 11%, 14% than SDR and 16% than MMR in terms of maximizing cohesion. Further, on an average the proposed MNDBO algorithm enhances the degree of cohesion by 15%. Since, the proposed clustering technique utilizes a normalized score for estimating the degree of deviation between the each data of the equivalence class.

Table 5. Aggregate Cohesion value for dataset-3

Aggregate Cohesion	Cluster Groups			
	4	5	6	7
MNDBO	1.1771	1.8121	2.4151	2.9121
MADO	1.1621	1.8019	2.3232	2.9101
SDR	1.1611	1.8010	2.2112	2.8151
MMR	1.1522	1.7919	2.2001	2.8101

From Table 6, it is evident that the dataset 1 clusters produced by MNDBO algorithm perform better than MADO by 13%, 18% than SDR and 20% than MMR in minimizing coupling. Further, on an average the proposed MNDBO algorithm minimizes the degree of coupling by 18%. Since, the proposed clustering technique estimates a normalized score based on standard deviation and mean that represents the central tendency of each equivalent class for estimating the degree of coupling between the each data of the equivalence class.

Table 6. Aggregate Coupling value for dataset-3

Aggregate Coupling	Cluster Groups			
	4	5	6	7
MNDBO	0.54127	0.70121	0.90121	1.09122
MADO	0.55111	0.71131	0.92212	1.100120
SDR	0.55221	0.71291	0.93343	1.102241
MMR	0.55411	0.72483	0.93996	1.103112

4. Conclusion

In this paper, a Minimum Normalized Dissimilarity between Objects (MNDBO) algorithms is presented. This MNDBO algorithm estimates the degree of deviation between data's of the same equivalence class. This algorithm also estimates the quality of cluster for the three real data pertaining to student's learning styles during e-learning process. The experimental results also infers that the MNDBO algorithm generates clusters with high degree of cohesion and low degree of coupling when the cluster size is varied from 4 to 7 in increments of 1. The suitability of MNDBO algorithm is proved through the process of testing with synthetic data sets that contains high dimension.

5. References

1. Bean C, Kambhampati C. Autonomous clustering using rough set theory. International Journal of Automation and Computing. 2008; 5(1):90–102.

2. Caoa F, Lianga J, Li D, Bai L, Dang C. A dissimilarity measure for the k-Modes clustering algorithm. Knowl Base Syst. 2011; 26:120–7.
3. Chen H, Chuang K, Chen M. On data labeling for clustering categorical data. IEEE Trans Knowl Data Eng. 2008; 20(11):1458–71.
4. Cheng CH, Chen YS. Classifying the segmentation of customer value via RFM model and RS theory. Expert Syst Appl. 2009; 36(3):4176–84.
5. Ganti V, Gehrke J, Ramakrishnan R. CACTUS-Clustering categorical data using summaries. San Diego, CA, USA: Proceedings of 5th ACM SIGKDD International conference on Knowledge Discovery and Data Mining, KDD'99; 1999. p. 73–83.
6. Pawlak Z. Rough sets: Theoretical aspects of reasoning about data. Norwell, MA, USA: Kluwer Academic Publishers; 1992.
7. Pawlak Z, Skowron A. Rudiments of rough sets. Inform Sci. 2007; 177(1):3–27.
8. Guha S, Rastogi R, Shim K. ROCK: A robust clustering algorithm for categorical attributes. 15th International Conference on Data Engineering; 1999. p. 512–21.
9. Herawan T, Mat Deris M, Abawajy JH. A rough set approach for selecting clustering attribute. Knowl Base Syst. 2010; 23(3):220–31.
10. Herawan T, Ghazali R, Tri Riyadi Yanto I, Mat Deris M. Rough set approach for categorical data clustering. International Journal of Database Theory and Application. 2010; 3(1):33–52.
11. Kumar P, Tripathy BK. MMeR: An algorithm for clustering heterogeneous data using rough set theory. Int J Rapid Manuf. 2009; 1(2):189–207.
12. Kunz T, Black JP. Using automatic process clustering for design recovery and distributed debugging. IEEE Trans Software Eng. 1995; 21(6):515–27.
13. Ling R, Yen DC. Customer relationship management: An analysis framework and implementation strategies. J Comput Inform Syst. 2001; 41(3):82–97.
14. Tripathy BK, Ghosh A. SDR: An algorithm for clustering categorical data using rough set theory. Trivandrum, India: Proceedings of IEEE conference on Recent Advances in Intelligent Computational Systems; 2011. p. 867–72.
15. Tripathy BK, Ghosh A. SSDR: An algorithm for clustering categorical data using rough set theory. Advances in Applied Science Research. 2011; 2(3):314–26.
16. Prabha D, Ilango K. MADO: An algorithm for clustering categorical data using rough set theory. Data Knowl Eng. 2014; 63(3):879–93.
17. Pawlak Z. Rough set. Int J Comput Inform Sci. 1982; 11(5):341–56.
18. Pawlak Z. Rough sets: Theoretical aspects of reasoning about data. Norwell, MA, USA: Kluwer Academic Publishers; 1992.
19. Shyng JY, Wang FK, Tzeng GH, Wu KS. Rough set theory in analyzing the attributes of combination values for the insurance market. Expert Syst Appl. 2007; 32(1):56–64.
20. Jawahar S, Nirmala K. Quantification of Learner characteristics for collaborative agent based e-learning automation. Indian Journal of Science and Technology. 2015; 8(14).