International Conference on Systems, Science, Control, Communication, Engineering and Technology 2015 [ICSSCCET 2015]

ISBN	978-81-929866-1-6
Website	icssccet.org
Received	10 - July - 2015
Article ID	ICSSCCET008

VOL	01
eMail	icssccet@asdf.res.in
Accepted	31- July - 2015
eAID	ICSSCCET.2015.008

An enhanced Rough Set Based Technique for Elucidating Learning styles in E-Learning System

K.S.Bhuvaneshwari[1], Dr.D. Bhanu[2], Dr. S. Sophia[3]
[1]Assistant Professor, Department of CSE, Karpagam College of Engineering, India
[2]Professor, Department of Computer Science and Engineering, Karpagam Institute of Technology, Coimbatore, India
[3]Professor, Department of ECE, Sri Krishna College Of Engineering and Technology, Coimbatore, India

Abstract: *Rough set theory is considered as the most essential strategy significantly suitable for illustrating distinctive sorts of learning styles controlled by the learners during e-learning process through feature information selection. The Rough set hypothesis is likewise utilized for effectively finding relations with conflicting or fragmented information which is incomplete in nature. Be that as it may, when harsh set hypothesis is consolidated, they are not sufficiently effective to evaluate ideal subsets. Hence, this paper provides a comparison of various rough set based techniques for adapting learning styles. The paper provides the analysis of rough set based clustering methods in terms of two parameters cohesion and coupling. In addition, the paper also proposes an enhanced methodology based on normalized score value for finding the deviation between data's through the equivalence property of rough set theory. The experimental results show that the proposed algorithm maximizes the stated metrics.*

Keywords: Rough set theory, e-learning, Learning styles, Normalization, 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 i) providing virtual learning facilities, ii) creating content for various domains for learners, iii) 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 of reducts in rough set theory for extracting knowledge from the learners' information [4].

The learning style of an individual is one of the imperative data to be derived from the learners' information. Since, the nature of the training and the adequacy of the information about the specific domain not just rely on the substance given in the e-learning framework

Cite this article as: K.S.Bhuvaneshwari, Dr. D. Bhanu, Dr. S. Sophia. "An enhanced Rough Set Based Technique for Elucidating Learning styles in E-Learning System." *International Conference on Systems, Science, Control, Communication, Engineering and Technology (2015)*: 27-32. Print.

additionally the presentation that has more effects on the learners [5]. 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, ii) decision friendly and iii) high user understandability.

The remaining part of the paper is organized as follows. Section 2 depicts the related work that portrays the role of rough set theory in elucidating relations for estimating the various learning styles in e-learning process. Section 3 presents the Minimum Normalized Dissimilarity between Objects (MNDBO) techniques with the associated algorithm. Section 4 presents the experimental comparison carried out with MNDBO with the considered benchmark systems. Finally, Section 5 concludes with the conclusion.

2 Related Work

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. Researcher of the past decade have proved that rough set theory can be implemented for wide range of problems such as i) correlated and uncorrelated analysis, ii) rule extractions for expert systems iii) 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] [12] [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. Finally, Prabha Dhadayudhan 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.

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.

3 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 very equivalence class. The deviation score (Dev_{Score}) estimated between data's within the set "S2"., is determined through equation (1) as

$$Dev_{Score} = \frac{\sigma_{(x_a,x_b)}}{M_{(x_a,x_b)}} \times 100 \qquad (1)$$

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

$$\sigma_{(x_a,x_b)} = \sqrt{\frac{\sum_{i=1}^{n}(x-\overline{x_i})^2}{n}} \qquad (2)$$

$$M_{(x_a,x_b)} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (3)$$

Cite this article as: K.S.Bhuvaneshwari, Dr. D. Bhanu, Dr. S. Sophia. "An enhanced Rough Set Based Technique for Elucidating Learning styles in E-Learning System." *International Conference on Systems, Science, Control, Communication, Engineering and Technology (2015)*: 27-32. Print.

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

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

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

The MNDBO clustering algorithm is given below:

Algorithm MNDBA_Cluster(Data_Set DS, Cluster Required C)

//*Input:* Data Set for Clustering, Number of Clusters needed

Begin
    Initialize the number of clusters IC = 1;
    Set C as the number of clusters;
    S_2- set of equivalent classes derived from S1 based on Normalized_Score
    Intialize the parent node to Data Set DS
    do
     for each attribute a_i from Data Set DS (i varies from 1 to n, where 'n' denotes the number of
        attributes in each data set DS
         for j varies from 1 to m, where 'm' is the possible different values of each attribute a_i
          in parent node(DS), determine the group of equivalence classes for a_i with attribute value 'j'
              denoted through set S2
          manipulate Dev_{score}(S2) based on $\sigma_{(x_a, x_b)}$ and $M_{(x_a, x_b)}$.
         next
         next
         set Normalized_Score = Min(Normlized_Score(S2)) for every set with constraints |S2|≥ 1
         estimate the partitioning attribute ai based on minimum Normalized_Score
        increment the number of cluster to 1 (ie)IC=IC+1
         set Parent_Node (DS) = New_ Parent_Node (IC)
     while (INC < k)
End
Algorithm New_Parent_Node(IC)
Begin
    for(i varies from 1 to IC)
        size of cluster (i) = number of cluster (i)
    next
    estimate Maximum(Size of cluster(i))
    return (Number of elements(i) equals to Maximum(Size of Cluster(i))
end

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.

4 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,

Cite this article as: K.S.Bhuvaneshwari, Dr. D. Bhanu, Dr. S. Sophia. "An enhanced Rough Set Based Technique for Elucidating Learning styles in E-Learning System." *International Conference on Systems, Science, Control, Communication, Engineering and Technology (2015)*: 27-32. Print.

'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:

1) Arrange the data set in ascending of A, C, U and T.
2) Partition the data set into five equal parts with 20% of the available records in each part.
3) 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 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.

4.1 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 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.

4.2 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 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.

4.3 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 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.

4.4 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 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.

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 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.

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

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.

5 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. Finally, the results also infers that the MNDBO algorithm is highly successful than the benchmark clustering algorithms like MADO, SDR and MMR considered for investigation.

References

Cite this article as: K.S.Bhuvaneshwari, Dr. D. Bhanu, Dr. S. Sophia. "An enhanced Rough Set Based Technique for Elucidating Learning styles in E-Learning System." *International Conference on Systems, Science, Control, Communication, Engineering and Technology (2015)*: 27-32. Print.

[1] C. Bean, and C. Kambhampati, “Autonomous clustering using rough set theory,” International Journal of Automation and Computing, 5(1), pp. 90-102, 2008.
[2] F.Cao, J. Liang, Li, D., L. Bai, and C. Dang, “A dissimilarity measure for the k-Modes clustering algorithm,” Knowledge Based Systems, 26, pp. 120-127, 2011.
[3] H. Chen, K. Chuang, and M.Chen, “On data labeling for clustering categorical data,” IEEE Transactions on Knowledge and Data Engineering, 20(11), pp. 1458-1471, 2008.
[4] C.Cheng, and Y.Chen, 2009, “Classifying the segmentation of customer value via RFM model and RS theory,” Expert Systems with Applications, 36(3), pp. 4176-4184, 2009.
[5] V.Ganti, J.Gehrke, and R.Ramakrishnan, “CACTUS-Clustering categorical data using summaries,” Proceedings of 5th ACM SIGKDD International conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, pp. 73-83, 1999.
[6] Z. Pawlak, “Rough sets: Theoretical aspects of reasoning about data,” Norwell, MA, USA: Kluwer Academic Publishers, 1992.
[7] Z. Pawlak, and A. Skowron, “Rudiments of rough sets. Information Sciences,” 177(1), pp.3 – 27, 2007.
[8] S. Guha, R. Rastogi, and K.Shim, “ROCK: A robust clustering algorithm for categorical attributes,” Information Systems, 25(5), pp. 345-366, 2000
[9] T. Herawan, M.M. Deris, and J.H. Abawajy, “A rough set approach for selecting clustering attribute,” Knowledge Based Systems, 23(3), pp. 220-231, 2010.
[10] T. Herawan, R. Ghazali, I.T.R. Yanto, and M.M. Deris, “Rough set approach for categorical data clustering,” International Journal of Database Theory and Application, 3(1), pp. 33-52, 2010.
[11] P. Kumar, and B.K.Tripathy., “MMeR: an algorithm for clustering heterogeneous data using rough set theory,” International Journal of Rapid Manufacturing, 1(2), pp. 189-207, 2009.
[12] T. Kunz, and J.P. Black, “Using automatic process clustering for design recovery and distributed debugging,” IEEE Transactions on Software Engineering, 21(6), pp. 515 -527, 1995.
[13] R. Ling, and D.C. Yen, “Customer relationship management: An analysis framework and implementation strategies,” Journal of Computer Information Systems, 41(3), pp. 82-97, 2001.
[14] B.K.Tripathy, and A. Ghosh, “SDR: An algorithm for clustering categorical data using rough set theory,” Proceedings of IEEE conference on Recent Advances in Intelligent Computational Systems, Trivandrum, India, pp. 867-872, 2011.
[15] B.K. Tripathy, and A. Ghosh, “SSDR: An algorithm for clustering categorical data using rough set theory,” Advances in Applied Science Research, 2(3), pp. 314-326, 2011.
[16] Prabha dhandayudhan and Ilango Krishnamurthi, MADO: an algorithm for clustering categorical data using rough set theory,” Data and Knowledge Engineering, 63(3), pp. 879-893, 2014.
[17] Z. Pawlak, “Rough set,” International Journal of Computer and Information Sciences, 11(5), pp. 341-356, 1982.
[18] Z. Pawlak, “Rough sets: Theoretical aspects of reasoning about data,” Norwell, MA, USA: Kluwer Academic Publishers, 1992.
[19] J.Y. Shyng, F.K. Wang, G.H.Tzeng, and Wu, K.S., “Rough set theory in analyzing the attributes of combination values for the insurance market,” Expert Systems with Applications, 32(1), pp. 56-64, 2007.

Cite this article as: K.S.Bhuvaneshwari, Dr. D. Bhanu, Dr. S. Sophia. “An enhanced Rough Set Based Technique for Elucidating Learning styles in E-Learning System.” *International Conference on Systems, Science, Control, Communication, Engineering and Technology (2015)*: 27-32. Print.