Application of Enhanced CLARANS Algorithm for Faculty Performance Evaluation Rating

Ricardo Q. Camungao[1], Bartolome B. Tanguilig III[1], and Bobby D. Gerardo[2]

***Abstract*—**The Enhanced CLARANS (Clustering Large Applications based on Randomized Search) algorithm was developed through the integration of the Slim-tree method in its clustering phase. The middle level of the constructed tree was used as the sample objects for identifying initial cluster centers for clustering. The integration of the output of the slim-tree method provides lesser time for the Enhanced CLARANS algorithm in clustering objects in the dataset. In this paper, it addresses the task of applying the Enhanced CLRARANS algorithm for faculty performance evaluation rating by the students using the Enhanced CLARANS program interface. The application of the algorithm assumed that it provides a new k-medoids clustering algorithm for grouping objects in the dataset, alternative tool in identifying annual faculty performance and visual presentation of the clustered objects. The clustered output presented the strengths and weaknesses of the faculty from the five (5) assessment areas of the NBC 461 QCE: Student Evaluation form and used as bases for the administrators and top management in planning and decision - making.

***Keywords*—**Clustering, Enhanced CLARANS, Faculty performance evaluation rating, Slim-tree.

I. Introduction

THE birth of clustering techniques three decades ago provides new approaches for establishing relationships of the objects in the dataset. Briefly, clustering is the process of dividing the data into groups of similar objects according to a similar measure. The primary goal is that each cluster is composed of objects that are similar to each other and dissimilar to other objects of the other cluster [1].

CLARANS is one of the k-medoids clustering algorithms that use randomized strategies to identify best nodes in clustering objects [2]. The drawback of the k-medoids algorithms opens an opportunity for the enhancement of the CLARANS algorithm through integrating Slim-tree method. The integration of the Slim-method paved the way for the development of a new algorithm and named as Enhanced CLARANS algorithm. The Enhanced CLARANS algorithm employs the Slim-tree method to pre – cluster objects in the dataset [3]. The middle level of the Slim-tree that composed the actual cluster centers served as the starting nodes for clustering. With the integration of the Slim-tree method, the Enhanced CLARANS algorithm surpassed the clustering performance of the CLARANS algorithm in terms of time [3].

[1]Technological Institute of the Philippines
[2]Western Visayas State University.

This paper attempt to apply the Enhanced CLARANS Algorithm for Faculty Performance Evaluation Rating and it specifically aims to; (1) Use the Enhanced CLARANS algorithm's program interface for clustering and (2) Simulate Enhanced CLARANS algorithm using a real world dataset taken from the faculty performance evaluation rating by the students of the Isabela State University. The application of the Enhanced CLARANS algorithm believes that it provides a new approach to the interpretation of data and present visual presentation of the clustering output to easily grasp the strength and weaknesses of the faculty. The clustered output can be used for management planning and decision making.

The succeeding sections of this paper are arranged as follows. 2) Describes the Enhanced CLARANS algorithm and discuss basic concepts of faculty performance evaluation, 3) Presents the simulation of clustering the faculty performance evaluation rating by the students using the Enhanced CLARANS program interface and 4) provides the conclusions of this paper.

II. Related Literature

This section presents the Enhanced CLARANS algorithm and a brief description related to Faculty Performance Evaluation Rating by the students.

A. ***The Enhanced CLARANS** (Clustering Large Applications using Randomized Search)* **Algorithm**

The computational complexity of the CLARANS algorithm is $O(n^2)$ regarding the number of objects; it uses randomized search of cluster centers in clustering objects in the dataset. The reiteratively executed process is the identification of appropriate cluster centers. This process consumes time in accessing disk space until the best cluster centers are found. To reduce the time in grouping objects, a frequent approach is to identify sample objects from the original database as the starting objects use in the initial clustering process [2][3].

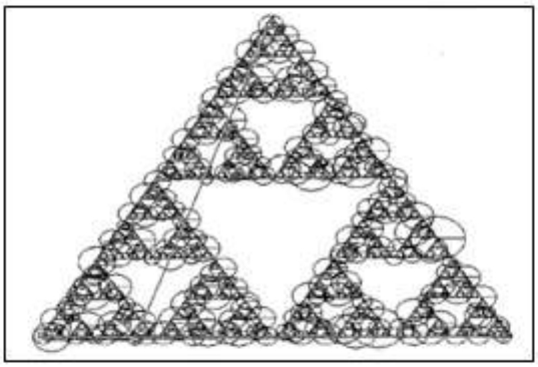
To speed up CLARANS algorithm using Slim-tree method, the strategy is to draw sample objects based on the features of the Slim-tree which hierarchically divides the metric space into groups or clusters from the number of objects and identify a cluster center for each set of objects. The assigned cluster center is the center object of a particular sub - tree. It is for this reason that only the objects of a chosen part of the tree is considered instead of considering all objects in the data set to evaluate the clustering algorithm [3].

To build a Slim-tree, it is necessary to have a policy employed when inserted a new object, and more than one node is qualified for the insertion of the new object (*ChooseSubtree* algorithm). The Slim-tree has three options for *ChooseSubtree* algorithm [5],

1. *Random:* randomly choose one of the qualifying nodes.
2. *MinDist:* choose the node that has the minimum distance from the new object and the center of the node.
3. *MinOccup*: choose the node that has the minimum occupancy among the qualifying nodes.

The default algorithms in the construction of the Slim-tree are given as *"minoccup"* option for the *ChooseSubtree* algorithm and the MST strategy for splits; the capacity node *C* is 60 for vector (L_2 distance) [7].

For the sampling strategy to be made possible, building the Slim-tree should be done first where each level of the constructed tree represents a data space division that grows from the root to the leaves, containing all the objects in the dataset. The question now is: which part of the tree has enough information regarding data distribution that can lead to a clustering with a lower computational cost? [2].



Fig, 1 A view of Slim-tree using Sierpinsky dataset [4]

According to [4] the middle level of the tree would be the best choice to get sample objects needed for the clustering for the reason that it contains enough information about the distribution of data that may yield to appropriate clustering output. The higher and lower part of the tree on the other hand, are not suited to use as the sample objects to start the clustering, because the upper levels of the tree contain less information about the data distribution where data are grouped by a smaller number of representatives while the lower levels contain too much information that may slow down the algorithm.

The sampling strategy using Slim-tree method is integrated into the CLARANS algorithm and named as Enhanced CLARANS Algorithm [3]. The algorithm is depicted as follows;

Slim-tree phase

1. Input parameters for building the tree: the page size and *ChooseSubTree* algorithm.
2. Build the Slim-tree.
3. Find the middle level of the tree. The number of cluster centers in this level will serve as the set of objects (*D*) the output of the sampling strategy.

Initialization phase

4. Input number of *k* cluster.
5. Initialize *mincost*to a large number, *numlocal* to 2 and determine the *maxneighbor*is (250, 1.25% of k(n-k)).

Clustering phase

Repeat

6. Set C an arbitrary node from the set of objects (D);
   (C = $[R_1, R_2, \ldots, R_k]$)
7. Set j =1;
8. **Repeat**
   Consider a random neighbor C* of C;
   Compute the TS_{ih} of C* and TS'_{ih} of C;
   If $TS_{ih} < TS'_{ih}$ **then**
   C = C*;
   Go to step 7;
   Else
   j = j + 1;
   go to step 8;
   Until j = maxneighbor;
9. **if** TS'_{ih}<mincost**then**
   mincost = TS'_{ih} ;
   BestSets = CurrentSets;

Else

increment i by 1;
Go back to Step 7;

10. **Until** i >numlocal;
    Bestnode = Currentnode;

End

The enhanced algorithm is composed of three phases namely Slim-tree phase, initialization phase, and clustering phase.

- The **Slim-tree phase** is the building of the tree, which covers the inputting of parameters needed such as *ChooseSubTree* algorithm and node page size of the tree. As presented in section 2.2 the Slim-tree was built using different options of the *ChooseSubtree* algorithm. The number of nodes or cluster centers in the middle level of the tree will be used as the sample objects (*D*) for the initial identification of nodes in the clustering phase. As stated in section 2.2 that the representative objects stored in the index nodes of each level of the tree are in a sense approximate cluster centers that can be used as a convenient starting point to cluster the objects in the dataset [2].
- The **initialization phase,** this phase covers the initialization of the needed parameters for the clustering such as inputting the number of *k* cluster, assigning the value of *numLocal* to 2 and identifying the values of the *mincost* and *maxNeighbor(*250, 1.25% of k(n-k)).
- The **clustering phase** this phase deals with the clustering part of the CLARANS algorithm. The initial nodes are randomly chosen from the set of objects (*D*) identified in building the Slim-tree. The number of objects tried in this

phase is dependent on the number of *maxneighbor (*250, 1.25% of k(n-k)). After the number of *maxneighbor* attempts and no better solution is found the local optimum is reached. The procedure continues until *numlocal* is greater than 2 [1], at this point the best sets are identified having the lowest *mincost*. Finally, use the nodes of the best sets to cluster the objects in the data set.

B. Faculty Performance Evaluation Rating by the Students

In most organization evaluation of the employees' performance is considered essential. This can be used to determine how the individual employee contributes to attaining the goals of the organization. Feedback of these evaluation results may lead the employees to increase their productivity, improve their morale and motivation and allow collaboration to ensure the realization of the organization's goals and objectives [10].

Castetter, [9] contended that teacher/faculty performance appraisal could be an effective mean to identify the faculty/teachers' developmental needs. Information from evaluation process and inventory of needs are important inputs to the faculty development process and improvement of the teaching-learning situation. If the teachers' strengths and weaknesses are identified, they could be the basis for planning an appropriate faculty development programs that would ensure the attainment of the school's vision, mission, goals, and objectives.

In many educational institutions, students' evaluations of the teacher's performance are given importance. Student ratings are one of the several methods used by schools, universities and colleges to evaluate a faculty member's teaching performance. This is a practice continued at the University of San Jose-Recoletos (USJ-R). Every semester, a faculty member is evaluated by his/her students. Through the students' evaluations, his/her strengths and weaknesses are identified [10].

Following the evaluation process the NBC 461 QCE form: Student Evaluation was used for the students to rate their faculty. The teaching effectiveness of faculty was evaluated with the following assessment areas: (1) Commitment, (2) Knowledge of the Subject Matter, (3) Teaching for Independent Learning and (4) Management of Leaning and (5) Critical Factors.

III. Results and Discussions

Clustering the faculty performance evaluation rating by the students was simulated in the Enhanced CLARANS Algorithms program interface. The interface was designed and developed using the C# language and PhP scripting language as shown in Figure 2.

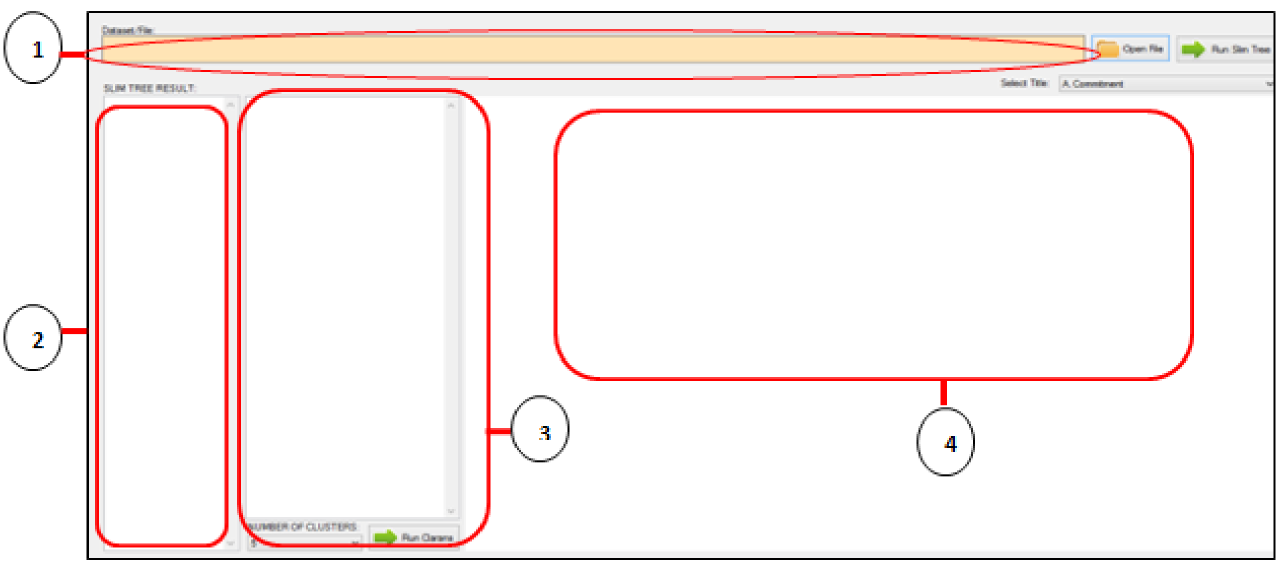


Fig. 2 E-CLARANS Program Interface

The program composes of functionalities intended for clustering faculty performance evaluation rating. The program's functionalities are as follows;

(1) Dataset loader provides the link or path name of the dataset.

(2) Slim-tree used for the pre-processing of objects to identify sample objects used for the initial identification of cluster centers.

(3) Enhanced CLARANS performs the clustering process that covers the identification of the *k* clusters centers, clustered objects, and objects of each cluster.

(4) Scatter Plot provides the visual presentation of the clustered output.

The dataset used was a real data taken from the faculty performance evaluation rating by the students from the first to second semesters of the school year 2015 – 2016. The faculty members were evaluated according to the assessment criteria of National Budget Circular (NBC) 461 Qualitative Contribution Evaluation (QCE) form: Student Evaluation and rated using the scale of 5 being the highest and 1 is the lowest. There were 652 faculty members from the ten (10) campuses of the Isabela State University were evaluated. As shown in Table 1 the summary of the faculty performance evaluation rating by the students from the school year 2015 – 2016.

TABLE 1.SUMMARY OF FACULTY PERFORMANCE EVALUATION RATING BY THE STUDENTS FOR THE S.Y. 2015 – 2016

NO	CAMPUS	COLLEGE	A		B		C		D		E		RATING	
			1ST	2ND	1^{ST}	2ND	1^{ST}	2^{ND}	1ST	2ND	1^{ST}	2ND	1^{ST}	2ND
1	Ilagan	AS	4.56	4.88	4.36	4.84	4.48	4.84	4.44	4.80	4.76	4.92	4.52	4.85
2	Cabagan	TE	4.68	4.28	4.64	4.24	4.72	4.16	4.68	4.16	4.78	4.31	4.70	4.23
3	Cauayan	IT	4.72	4.64	4.68	4.76	4.60	4.64	4.56	4.72	4.63	4.80	4.64	4.68
4	Cauayan	IT	3.92	4.16	4.00	4.40	3.96	4.40	3.68	4.36	3.48	3.94	3.81	4.25
5	Cabagan	Agriculture	4.84	4.92	4.88	4.88	4.84	4.92	4.84	4.88	4.63	4.90	4.81	4.91
6	Cauayan	BM	4.48	4.72	4.60	4.76	4.44	4.60	4.48	4.64	4.64	4.72	4.54	4.69
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
651	Cabagan	TE	4.64	3.72	4.68	3.80	4.60	3.80	4.68	3.68	4.69	3.65	4.67	3.73
652	Echague	CAS	4.20	4.20	4.20	4.20	4.24	4.24	4.32	4.32	4.36	4.36	4.27	4.27

Legend **(A)** Commitment, **(B)** Knowledge of the Subject Matter, **(C)** Teaching for Independent Learning, **(D)** Management of Leaning and **(D)** Critical Factors

A. Clustering the Faculty Performance Evaluation Rating by the Students using the Enhanced CLARANS Program Interface

The simulations were performed in the Enhanced CLARANS Algorithm program interface. Loading the path name of the dataset containing the faculty rating in the dataset loader is the first step. The objects in the dataset composed of the students rating in two semesters were used for pre-processing of the Slim-tree to determine objects in the middle level of the tree. The objects identified are in a sense actual cluster centers so it can conveniently use to start the clustering [3][4].

The clustering process starts with the inputting of the number of *k* clusters from the user, in this case, the number of clusters is 2 representing the higher and lower group. The Enhanced CLARANS draws initial cluster centers from the output generated by the Slim-tree. The number of executions of the Enhanced CLARANS is based on the values of numlocal = 2 and *maxneighbor* (250, 1.25% of k(n-k)) [1]. After the number of attempts of the two parameters the clusters centers were identified and objects are grouped in their respective cluster centers. The visual presentations of the clustered objects were presented using scatter plot shown in figure 3.

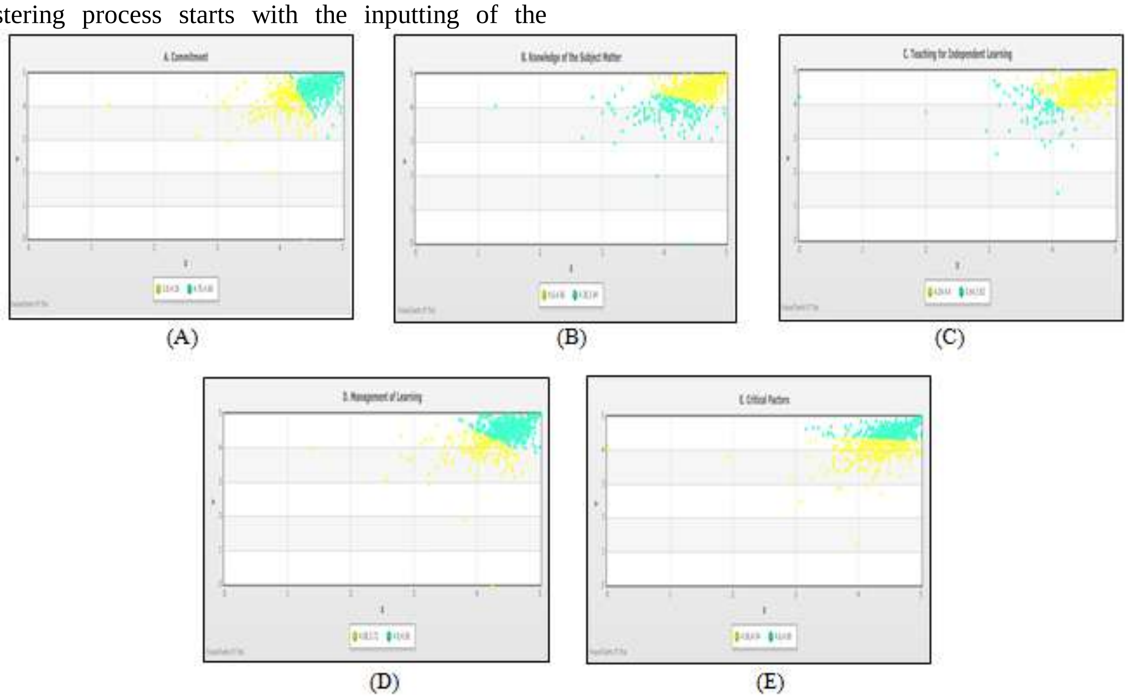


Fig. 3 Clustered objects in scatter plot of (A) Commitment, (B) Knowledge of the Subject Matter, (C) Teaching for Independent Learning and (D) Management of Leaning and (E) Critical Factors.

Figure 3 presents the clustered output of the five (5) assessment areas of the NBC 461 QCE form: Student Evaluation using the scatter plot of the Enhanced CLARANS algorithm. Each object defines the performance of the faculty ratings in two semesters, represented by the *x* and *y* coordinates respectively. From the output, each area was clustered into two (2) representing the higher and lower group. The higher and lower group refers to the strengths and weaknesses of the faculty members in terms of (A) Commitment, (B) Knowledge of the Subject Matter, (C) Teaching for Independent Learning, (D) Management of Leaning and (D) Critical Factors.

The identified cluster centers for the higher group are (4.76, 4.68), (4.6, 4.56), (4.24, 4.4), (4.8, 4.56) and (4.6, 4.68) and for the lower group we have (3.8, 4.36), (4.36, 3.84), (3.64, 3.92), (4.087, 3.72) and (4.56, 4.04). Objects belong in the cluster centers of the higher group in every assessment areas are the faculty members gave a strong teaching performance, and the objects belong in the cluster centers of the lower group are faculty members who gave a weak performance. The output of the clustering presents the overall faculty performance for one academic year it provides the relationship of the faculty rating in two semesters and shows which areas that the faculty gave the strong and weak performance. Hence, faculty members in the lower group need the attention of the management to design faculty enhancement programs.

IV. Conclusion

This paper presented the application of the Enhanced CLARANS algorithm to faculty performance evaluation rating by the students. The application was made possible using the developed Enhanced CLARANS program interface. The dataset containing the faculty rating was used to simulate the clustering process of the E-CLARANS algorithm. The output of the application shows that the Enhanced CLARANS algorithm provides; new k-medoids clustering algorithm for identifying the strengths and weaknesses of the faculty members in the NBC 461 QCE form: Student Evaluation assessment areas, an alternative tool uses to group the annual performance of the faculty members; and a visual presentation of the clustered output as shown in figure 3 that can use as bases for the administrators and top management for planning and decision making.

Moreover, the output of the Enhanced CLARANS algorithm can use for designing faculty enhancement programs to sustain the quality of services rendered to the clienteles.

Acknowledgment

We thank Joe G. Lagarteja for engaging and fruitful discussions in the development of the Enhanced CLARANS algorithm's program interface.

References

[1] Han, J. and Kamber, M. (2001). Data Mining: Concepts and Techniques. Academic Press, San Diego, CA.

[2] Ng, R.T., and Han, J. (2002). Clarans: A method for clustering objects for spatial data mining. IEEE Transactions on Knowledge and Data Engineering (TKDE), 14(5):1003-1016

[3] Camungao, R.Q., Tanguilig III, B.T., Gerardo, B.D. (2016). Enhanced CLARANS Algorithm using Slim-tree Method, Proceedings of Philippine Computing Science Congress 2016, pages 57-62

[4] Barioni, M.C., Razente, H., Traina, A. J., and Traina, C. (2001). An Efficient Approach to Scale up k-medoid based Algorithms in Large Databases. XXI SimposioBrasileiro de Banco de Dados, pages 265 – 278.

[5] Vieira, M. R., Chino, C. T. F., and Traina, A. J. M. (2004). DBM tree: A metric access method sensitive to local density data. In Brazilian Symposium on Databases (SBBD), pages 163 – 177.

[6] Traina, A.J. M., Jr., C.T., Bueno, J.M., and de A. Marques, P.M. (2002). The metric histogram: A new and efficient approach for content –based image retrieval. In IFIP working Conference on Visual Database System (VDB), pages 297-311.

[7] Traina Jr., C., Traina, A., Faloustsos, C., and Seeger, B. (2002).Fast Indexing and Visualization of Metric Data Sets Using Slim-Trees.Vol 14. No.2 (244 – 260)

[8] Bacal, R. (1999). *Performance Management*. U.S.A.: McGraw –Hill.

[9] Castetter, W. B. (1981). *The Personnel Function in Educational Administration.* New York: Mc Millan Company.

[10] Galeon, G. A. (2015). Correlates of the Teaching Performance of the College Faculty Members. International Journal of Applied Psychology 2015, 5(3): 64-72.