The Investigation of Discovering Potential Musical Instruments Teachers By Effective Data Clustering Scheme

Cheng-Fa Tsai, Yu-Tai Su, Chiu-Yen Tsai and Chun-Yi Sung

***Abstract*—**Data clustering plays an important role in various fields. Data clustering approaches have been designed in recent years. This investigation aims to present data clustering algorithm to identify potential musical instruments teachers. With a total of 5125 candidates registered respectively in 9 grades of Taiwan United Music Grade Test during 2000-2008. Moreover, this study proposes a new data clustering algorithm called MIDBSCAN and an existing well-known neural network called self-organizing map (SOM) to perform data clustering applications for discovering potential musical instruments teachers. The processing procedure of searching for neighbors (neighborhood data points) is very time consuming in the existing well-known DBSCAN and IDSCAN algorithms. Therefore, to shorten the time consumed, the proposed MIDBSCAN algorithm focuses lowering the number of expansion seeds added into the neighborhood data in this procedure, thus reducing the time cost of searching for neighbors. According to our simulation results, the proposed MIDBSCAN approach has low execution time cost, a maximum deviation in clustering correctness rate and a maximum deviation in noise data filtering rate. MIDBSCAN outperforms SOM in execution time cost. It is feasible to perform data clustering analysis in various data mining applications using the proposed MIDBSCAN algorithm.

***Keywords*—**data clustering, data mining, effective teaching, music examine

I. Introduction

DATA mining refers to the mining or discovery of new information in terms of patterns or rules from significant amounts of data. The result of mining may be to discover association rules, sequential patterns, classification trees or clusters. Data clustering in data mining for large databases is an important business application, which has recently become a highly popular topic in data mining research. Data clustering illustrates the process of grouping data into clusters such that the data in each cluster share a high degree of similarity while being very dissimilar to data from other clusters. Data clustering is a new-emerging and imperative technique for data mining applications [3]. The goal of data clustering may be to gain an insight into a structure inherent in the population or to develop a business strategy that is customized to individual customer clusters for enhanced business efficiency. Research in data clustering focuses mainly on increasing the accuracy and reducing the clustering time cost [4], [9]. Clustering approaches can be categorized into several categories, namely partitioning, hierarchical, density-based, grid-based and mixed approach. Typically, density-based methods, including DBSCAN [21], SDBSCAN [23], IDBSCAN [22] and KIDBSCAN [11] conduct expansion and clustering based on density. These approaches can filter noise, and perform clustering in disordered patterns, but take a long time to perform clustering. In general, the clustering problem can be employed for trend analysis, customer segmentation, similarity search, classification, image processing, image compression, pattern recognition, network intrusion and attack behavior analysis, data dissimilation for sensor networks and many other applications [12]-[41].

Taiwan United Music Grade Test (called as music grade thereof [5]) is performed since 2000 to evaluate the performance of numerous musical instruments as per regulations of each musical instruments and each grade to feedback teaching effectiveness immediately and to effectively enable a long-term development of music learning in Taiwan.

Many music education scholars have regarded assessment as important indicator for teaching effectiveness review [1], [2], [5], [6]. According to Goolsby's summary [5], the assessment on musical instrument learning will be divided into formative, placement, diagnostic, and summative assessments. In which formative assessment is to confirm students' understanding on the long-term learning objectives, direction, and get ready for learning action through a daily observation. The primary purpose of placement assessment is to confirm students' level to allow making proper arrangement of that student to an appropriate position, e.g. seating arrangement within a musical group, vocal of chorus, grouping by music fundamental training capabilities. Diagnostic assessment normally happens in class, to allow teachers to evaluate learners' problems and

Manuscript received March 3, 2009: Revised version received April 30, 2009. This work was supported in part by the National Science Council of Republic of China under Grant NSC 96-2221-E-020-027.

C. F. Tsai is with the Management Information Systems Department, National Pingtung University of Science and Technology, Pingtung, Taiwan 91201 R.O.C. (corresponding author to provide phone: 886-8-7703202; fax: 886-8-7740306; e-mail: cftsai@mail.npust.edu.tw).

Y. T. Su is with the Department of Western Music, the Chinese Culture University, Taipei, Taiwan 11114 R.O.C.(su.jessica@gmail.com)

C. Y. Tsai is with the Management Information Systems Department, National Pingtung University of Science and Technology, Pingtung, Taiwan 91201 R.O.C. (n9656002@mail.npust.edu.tw).

C. Y. Sung is with the Management Information Systems Department, National Pingtung University of Science and Technology, Pingtung, Taiwan 91201 R.O.C. (n9656018@mail.npust.edu.tw)

provide suggestion immediately. Summative assessment is seen frequently in concerts, tournaments, qualification exams and other occasions, which aims to demonstrate final outcomes to the public through external instrument playing. Therefore, a participation in Taiwan United Music Grade Test will help teachers to understand students' learning outcomes and is considered as a reference to teaching effectiveness.

Regarding the discrimination level of assessment, Burrack indicated that it requires a cautious institution in terms of the musical instrument ability standard by various grades and presented the objectives of students in various grades expect to achieve [2]. Taiwan United Music Grade Test regulated on number of repertoires required for each grade and each single musical instrument in detail, and through playing these set pieces, students need to demonstrate a requisite capability in playing these repertoires. A repertoire that suits the students to play will allow demonstrating a learning achievement, which expands students horizons on music and enhance their music capabilities, and foster teachers' quality of teaching. In which Persellin suggested that the repertoire selection shall consider student's age and capability when teachers conduct a teaching [7]; while in Reynolds' point of view [8], repertoire is the core of the course for musical instrument teachers, while selection of repertoire reflects teacher's overall value of music and his/her teaching direction, which helps students attain the learning objective for the next level through learning these repertoires [1].

When teachers suggest students to prepare for the repertoire of music grade test, their ultimate learning outcomes, whether their summative assessments are as what teachers expect to attain a pass level? Whether to find a teacher segment with higher pass rate on students' grade test?

To lower the clustering execution time, this work presents an efficient and effective density-based MIDBSCAN algorithm [3] and a well-known self-organizing map (SOM) [10] to conduct data clustering for expecting to understand the performance of domestic piano teaching for repertoire selection and pass rate among the various music grade tests and find out teachers with a teaching potential. According to the simulation results, the proposed MIDBSCAN approach has low execution time cost, a maximum deviation in clustering correctness rate and a maximum deviation in noise data filtering rate [3]. MIDBSCAN outperforms SOM in execution time cost. It is feasible to perform data clustering analysis in various data mining applications using the proposed MIDBSCAN algorithm.

II. Related works

This section discusses various density-based clustering schemes, involving DBSCAN, SDBSCAN, IDBSCAN and KIDBSCAN. Merits and limitations of these data clustering approaches are related to the proposed MIDBSCAN algorithm are illustrated below.

DBSCAN, presented by Ester et al. in 1996 [21], was the first clustering scheme to apply density as a condition. Although it can accurately detect any arbitrary pattern and different size clusters almost perfectly with two thresholds (searching radius and minimum points), and filters noise. However, it has a very high time cost when the database size is large, making it unpopular for use in business applications.

SDBSCAN is the first sampling-based clustering algorithm developed by Zhou et al. in 2000 [23]. The approach combines sampling technique within DBSCAN to perform clustering task in large databases. The SDBSCAN clustering algorithm must input three parameters to perform clustering, namely, sampling rate, radius and minimum points (*MinPts*). The algorithm selects sample randomly from neighborhood data to conduct expansion seed task. SDBSCAN can reduce the time cost of region queries than DBSCAN, but when it reduces sampling rate, which will increases the execution time and decreases the ratio of clustering correctness.

IDBSCAN is developed by Borah et al. in 2004 [22]. The method that selects representative points illustrates in the two-dimension database. Eight distinct points are selected as Marked Boundary Objects (MBO). This method employs a MBO to determine the data point of an expansion seed when searching for neighborhood to add in expansion seeds. IDBSCAN generates the same quality of DBSCAN but is more efficient.

KIDBSCAN is a density-based clustering method presented by Tsai and Liu in 2006 [11]. They searched for marked boundary objects with IDBSCAN, and found that inputting data sequentially from low-density database causes remnant seed searching, resulting in poor expansion results. To decrease the number of sample instances, KIDBSCAN performs expansion by inputting elite points. The major advantages of the KIDBSCAN are as follows [11]. (1) The computational time does not increase with the number of data points. (2) It is not limited by memory when dealing with large data sets. (3). It conducts excellently for arbitrary shapes. Experimental results demonstrate that KIDBSAN enhances the accuracy of the clustering result. In addition, it can perform data clustering quickly. KIDBSCAN outperforms DBSCAN and IDBSCAN.

III. The proposed MIDBSCAN scheme

The processing procedure of searching for Neighbors (neighborhood data points) is very time consuming in the DBSCAN and IDSCAN algorithms. Therefore, to shorten the time consumed, this study focuses lowering the number of expansion seeds added into the neighborhood data in this procedure, thus reducing the time cost of searching for Neighbors. This section describes the implementation procedures and algorithm for MIDBSCAN.

The parameters are set when implementing MIDBSCAN: (1) Radius (ε), (2) Minimum number of included points (*MinPts*). The MIDBSCAN clustering algorithm can be described as Fig. 1 [3].

```
Input: Datasets, Eps, MinPts
Output: Clusters

MIDBSCAN (Datasets, Eps, MinPts)
  Intialization;
  ClusterID := NextID(First);
  FOR i FROM 1 TO Datasets.Size DO
    Point := Datasets.Get(i);
    IF Point.CID = UNCLASSIFIED THEN
      IF ExpandCluster(Datasets, Point, ClusterID, Eps, MinPts) THEN
        ClusterID := NextID(ClusterID);
        UnclassfiedData.Adjust();
      END IF
    END IF
  END FOR
END;

ExpandCluster(Datasets, Point, CID, Eps, MinPts) : Boolean;
  Neighbors := UnclassfiedData.RegionQuery(Point, Eps);
  IF Neighbors.size < MinPts THEN
    Datasets.ChangeCIDs(Point, NOISE);
    RETURN False;
  ELSE
    Datasets.ChangeCIDs(Point,CID);
    Neighbors.AddMBOs();
    FOR i FROM 1 TO Neighbors.size DO
      neighborPoint := Neighbors.Get(i);
      IF neighborPoint.CID = UNCLASSIFIED || neighborPoint.CID = NOISE THEN
        Datasets.ChangeCID(neighborPoint,CID);
      END IF;
    END FOR;
    WHILE Seeds <> Empty DO
      seedPoint := Seeds.First();
      Seeds.Delete(seedPoint);
      Neighbors := UnclassfiedData.RegionQuery(seedPoint, Eps);
      IF Neighbors.size >= MinPts THEN
        Neighbors.AddMBOs();
        FOR i FROM 1 TO Neighbors.size DO
          neighborPoint := Neighbors.Get(i);
          IF neighborPoint.CID = UNCLASSIFIED || neighborPoint.CID = NOISE THEN
            Datasets.ChangeCID(neighborPoint,CID);
          END IF;
        END FOR;
      END IF;
    END WHILE;
    RETURN True;
  END IF
END;
```

Fig. 1 The proposed MIDBSCAN Algorithm

The implementation steps for the MIDBSCAN algorithm are described as follows [3]:

Step 1. Initialize all parameters, and define a new Cluster ID.

Step 2. Begin scanning all data points within the entire database. For data points belonging to the Cluster ID of those unclassified data, implement the Expand Cluster processing procedure. The database is the set of data points; the Point is the core point; the Cluster ID is the current cluster ID; ε represents the radius, and *MinPts* denotes the minimum number of included points.

Step 3. If the data point returned by the expansion procedure function is a noise data point, then go directly to Step 2, until the Datasets database has been fully scanned. If an expansion data point is returned, then update the new Cluster ID, and alter the index array of unclassified data, and then go to Step 2.

Step 4. End the algorithm when all data points have been processed.

The implementation steps for the Expand Cluster processing procedure are as follows.

Step 1. Search for Neighbors within the range of radius ε in the unclassified cluster index. If the number of Neighbors is less than *MinPts*, then leave the procedure, and return the core point as the noise data point. Otherwise, go to Step 2.

Step 2. Set the core point as the current cluster set ID.

Step 3. If the set of seeds is empty, then end the expansion processing procedure, otherwise go to Step 4.

Step 4. Search for the marking boundary point within Neighbors, and add in the expansion Seeds.

Step 5. Set all unclassified data points and Neighbors that are noise data points as the current cluster ID.

Step 6. Extract the first seed from the expansion seeds; and define it as the core point, and delete it.

Step 7. In the unclassified data index, search for Neighbors within the range of radius ε of the core point. If the number of Neighbors is greater than *MinPts*, then go to Step 3.

IV. EXPERIMENTAL RESULTS

The clustering algorithm was implemented in the C# language in Microsoft Visual Studio 2005 on a notebook computer with a 1.6 GHz Intel CPU, with 1G of RAM, running Windows XP. According to the simulation results, it is observed that the proposed MIDBSCAN outperforms the related density-based clustering approaches involving DBSCAN, IDBSCAN and KIDBSCAN in execution time cost, clustering correctness rate and noise filtering rate [3]. This study applies the MIDBSCAN and SOM clustering approaches to conduct clustering applications. The Music Grade Test Database (from 2000 to 2008) obtained data from the Music Department of Extension Education Center at the Chinese Culture University of Taiwan, as Table 1 shows, whereas the relationship diagram of database is based on the three subjects, involving the number of examinees, number of applications and number of repertoires, as Fig. 2 reveals. The center of the relationship diagram includes three Fact Tables, indicating the basic information of historical examinees applying for examinations each term from August 2000 to February 2008 (with 5,125 records), 5,231 records of Applied Item and Level Detail (sch_apply) and 12,834 records of Repertoire Selection for Application (stu_applydetail). In particular, 13 dimensions have been defined.

This research documented and introduced the musical instruments of all types by teacher code, teacher name, student's score rate in average over the years (*X*-axis) and total number of students guided by the teacher over the years (*Y*-axis) as data set type columns, with 720 data documented which uses SOM and MIDBSCAN to find out the characteristics in common and cluster of outstanding potential teachers, in which *X*-axis represents calculation of scoring and letters A~H signify the weighted scores as shown in Table 2.

Each teacher reported a score of 0.1~0.4 in average for teaching over the years. The average score of each teacher on lecturing over the years can be computed as follows:

$$AvgScore = \frac{\sum_{i=0}^{n} StudentScore(i)}{n} \quad (1)$$

Where *AvgScore* denotes average score of each teacher on lecturing over the years, the denominator (*StudentScore*(*i*)) represents total number of students lectured by that certain teacher, and *n* indicates total number of students lectured by that certain teacher. Students with more guidance and outstanding scores *X* value (*AvgScore*), the better the more positive they are on teacher's quality of teaching; *Y*-axis signifies a value of total number of students lectured by the teacher over the years/100, the bigger the *Y* value, the more students lectured by the teachers over the years, and also, reported a higher teaching experiences, awareness and higher acceptance by the students, and steady teaching quality, with central point (*x*, *y*) of "average value" lists of data sets among various clusters represented the characteristics of each segments, with type of characteristic value determined by Likert Scale, the higher (more) the type value, the higher the score is and is in sequence as below : 1. lowest－2 low－3 middle－4 high－5 highest, and finally to sum up the scores of characteristic value as "total scores for potential teacher segments rating" with various segments listed out including number of teachers, in which we expect to find out potential teachers of various clusters through MIDBSCAN, SOM clustering approaches, to understand a correlation between number of students lectured by teachers in various clusters and the score.

TABLE 1 TAIWAN UNITED MUSIC GRADE TEST TABLE FROM 2000 TO 2008

Table	Table Description	Records
stu_applystu	Examinee Information for each Term	5125
stu_apply	Applied Item and Level Details	5231
stu_applydetail	Repertoire Selection and Application	12834
stu_list	Item vs. Repertoire Level	1079
stu_cat	Category	4
stu_item	Item	15
stu_level	Level	10
stu_music	Repertoire	988
stu_zip	ZIP Code	371
stu_city	County and City	18
stu_area	Municipal County and City	4
stu_student	Examinee	2740
stu_teacher	Teacher	770
stu_yt	Term	17
stu_oldrange	Age Range	80
stu_score	Score	8

TABLE 2 THE WEIGHTED SCORE OF STUDENT OBTAINED

Student Score	Score by grade	Weighted Value
A	Excellent	0.4
B, C	Pass, qualified	0.3
D	Barely pass	0.2
E, H	failure, unqualified	0.1
F, G	Exam rescheduling, absence for exam	No inclusion

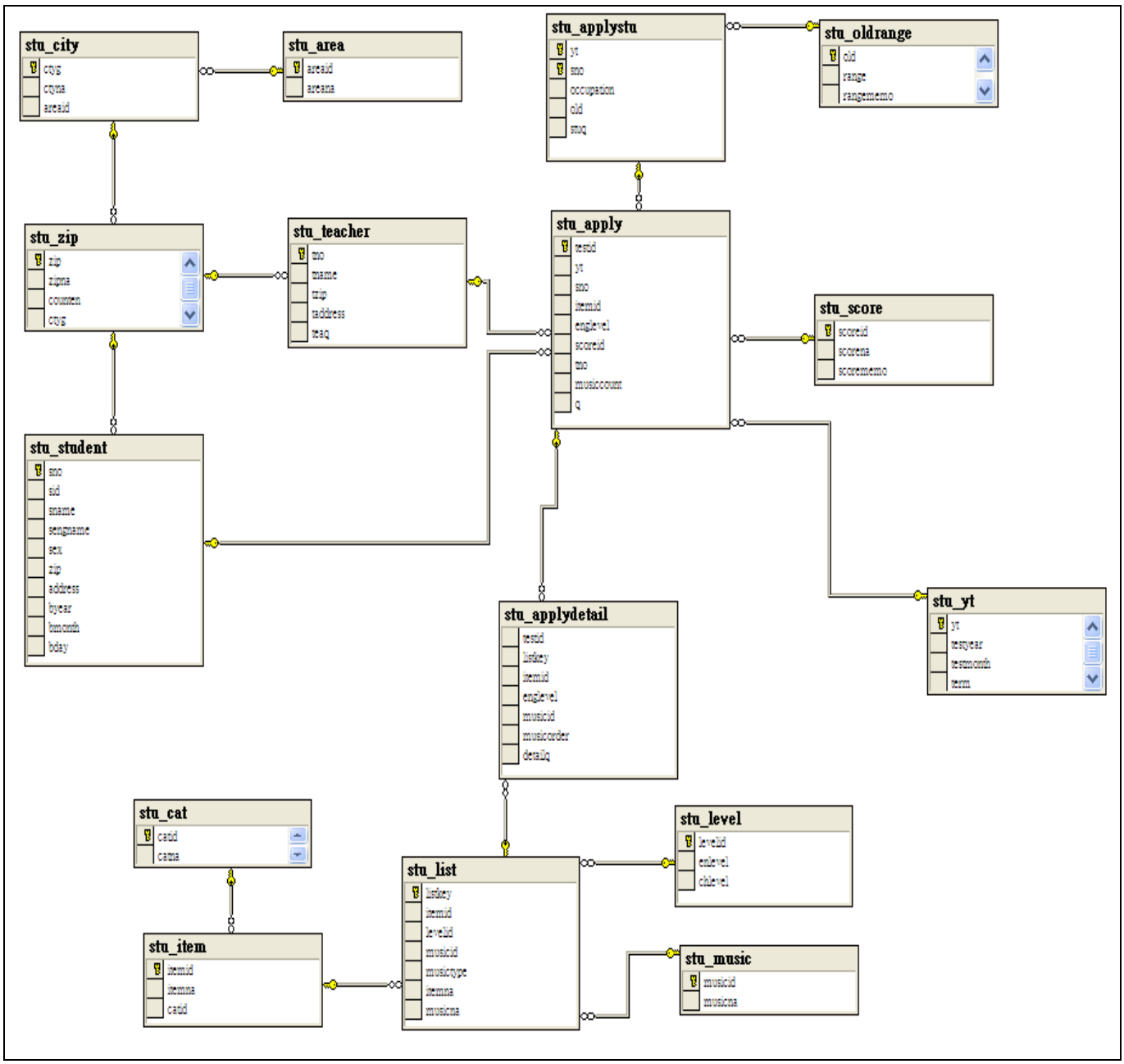

Fig. 2 Relationship diagram of Taiwan United Music Grade Test Database

V. Clustering Analysis Using MIDBSCAN Scheme

MIDBSCAN is an efficient density-based clustering algorithm [3], and it only requires two parameters when conducting a clustering task: radius (ε) and Minimum Points (*MinPts*). In Fig. 3, it takes 0.015 seconds to conduct a clustering task by keying in the potential teachers data sets using MIDBSCAN algorithm (ε=0.02, *MinPts*=5), with 6 clusters shown in total, in which one of the clusters is noise (shown in Fig. 4), an average score for the first cluster is 0.01 with 1 student lectured by teacher in average; while the second cluster reported an average score of 0.25 with 3 students lectured by teacher in average; Cluster 3 reported 0.34 scores in average with 4 students lectured by teacher in average; Cluster 4 reported 0.4 scores in average with 1 student lectured by teacher in average; Cluster 5 reported 0.26 scores in average with 14 students lectured by teacher in average; Cluster 6 reported 0.24 scores in average with 180 students lectured by teacher in average; Other clusters reported 0.26 scores in average with 430 students lectured by teacher in average, which belongs to “fair score but with the highest student numbers”.

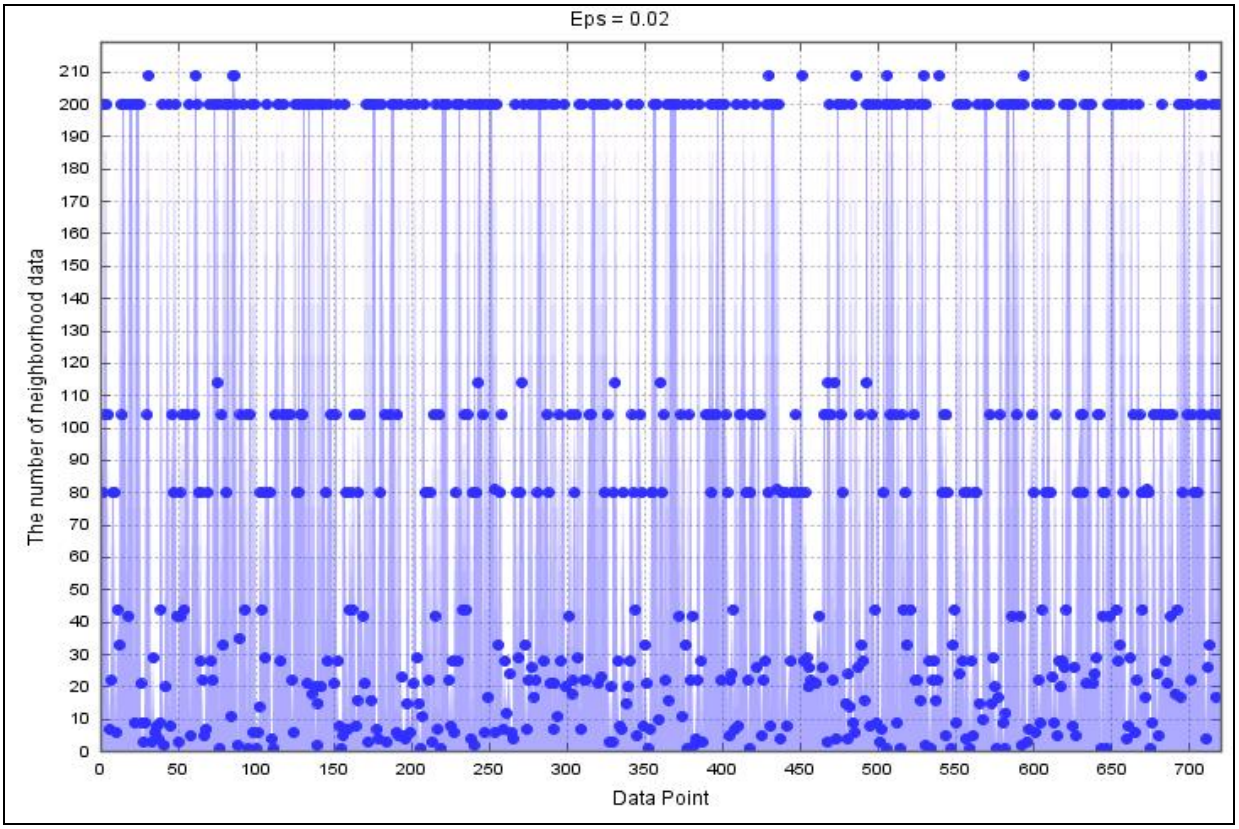

Fig. 3 Data clustering result using MIDBSCAN approach

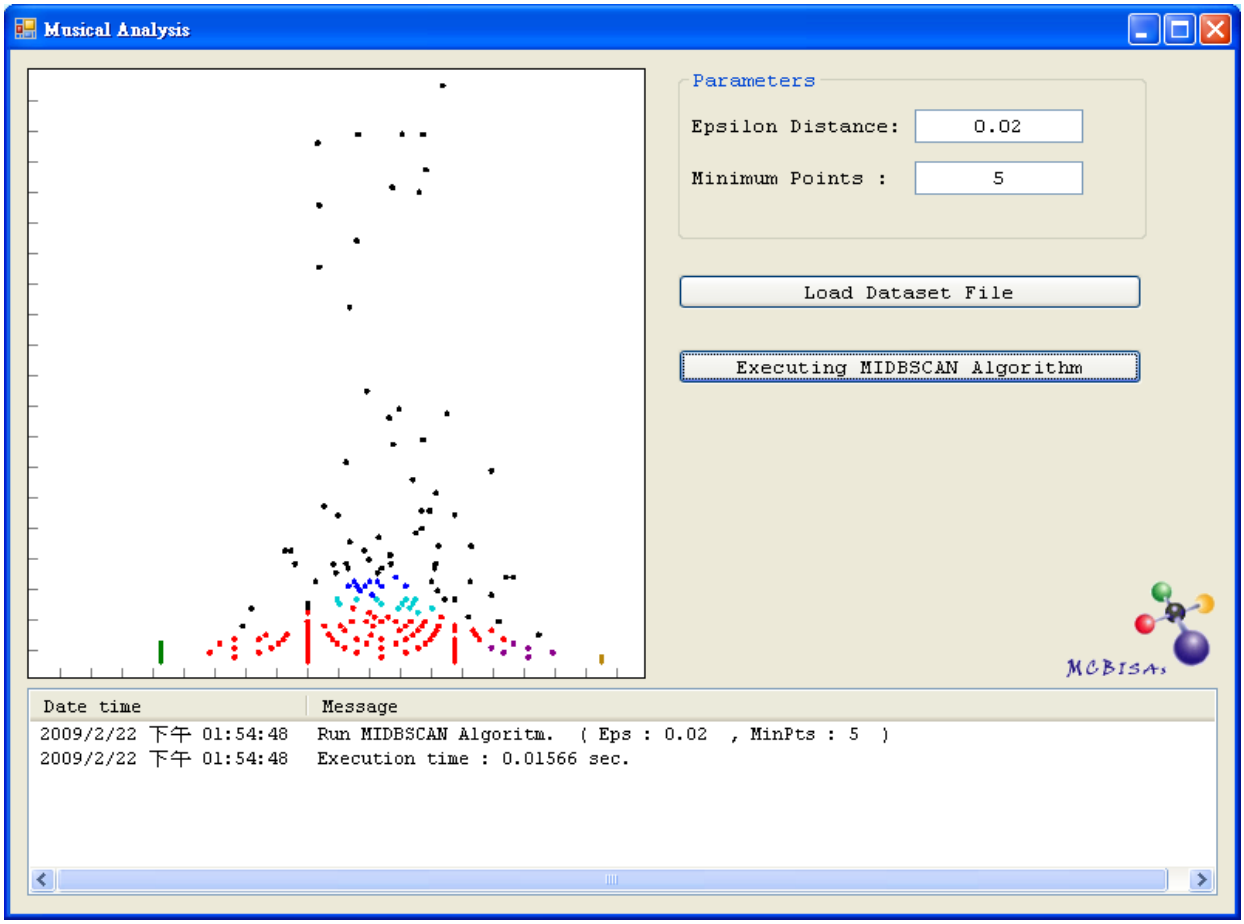

Fig. 4 Using MIDBSCAN clustering algorithm to discover the potential teachers

VI. Clustering Analysis Using SOM Scheme

Self-organizing map (SOM) originated from 1980, where Kohonen proposed a cerebrum structure that is similar to a feature of brain cell assembly, in which input units will affect one another, units that are adjacent to each other have the same function [10]. A self-organizing map comprises components named nodes or neurons. Associated with each node denotes a weight vector of the same dimension as the input data vectors and a position in the map space. Since in the training phase weights of the entire neighborhood are moved in the same direction, similar items tend to excite adjacent neurons. Therefore, SOM constructs a semantic map where similar samples are mapped close together and dissimilar apart.

The implementation steps for the SOM algorithm are described as follows [10]:

Step 1. Randomize the map's nodes' weight vectors.

Step 2. Grab an input vector.

Step 3. Traverse each node in the map.

(1) Utilize Euclidean distance formula to obtain similarity between the input vector and the map's node's weight vector.

(2) Track the node that generates the smallest distance (this node is the best matching unit, BMU).

Step 4. Update the nodes in the neighborhood of BMU by pulling them closer to the input vector.

$$Wv(t+1) = Wv(t) + \theta(t)\alpha(t)(D(t) - Wv(t)) \quad (2)$$

Step 5. Increment t and repeat from step 2 while $t < \lambda$.

Where t indicates current iteration, λ represents limit on time iteration, Wv denotes current weight vector, D is target input, $\theta(t)$ indicates restraint due to distance from BMU (usually called the neighborhood function) and $\alpha(t)$ represents learning restraint due to time.

It takes 0.442 seconds for SOM to carry out data clustering in this experiment. Fig. 5 reveals the clustering result using SOM approach. In Fig. 5, Cluster 1 reported 0.11 scores in average with 2 students lectured by teacher in average; Cluster 2 reported 0.21 scores with 3 students lectured by teacher in average; Cluster 3 reported 0.31 scores with 2 students lectured by teacher in average; Cluster 4 reported 0.25 scores with 18 students lectured by teacher in average; Cluster 5 reported 0.27 scores with 45 students lectured by teacher in average; Cluster 6 reported 0.25 scores with 109 students lectured by teacher in average.

According to experimental results, MIDBSCAN outperforms SOM in execution time cost.

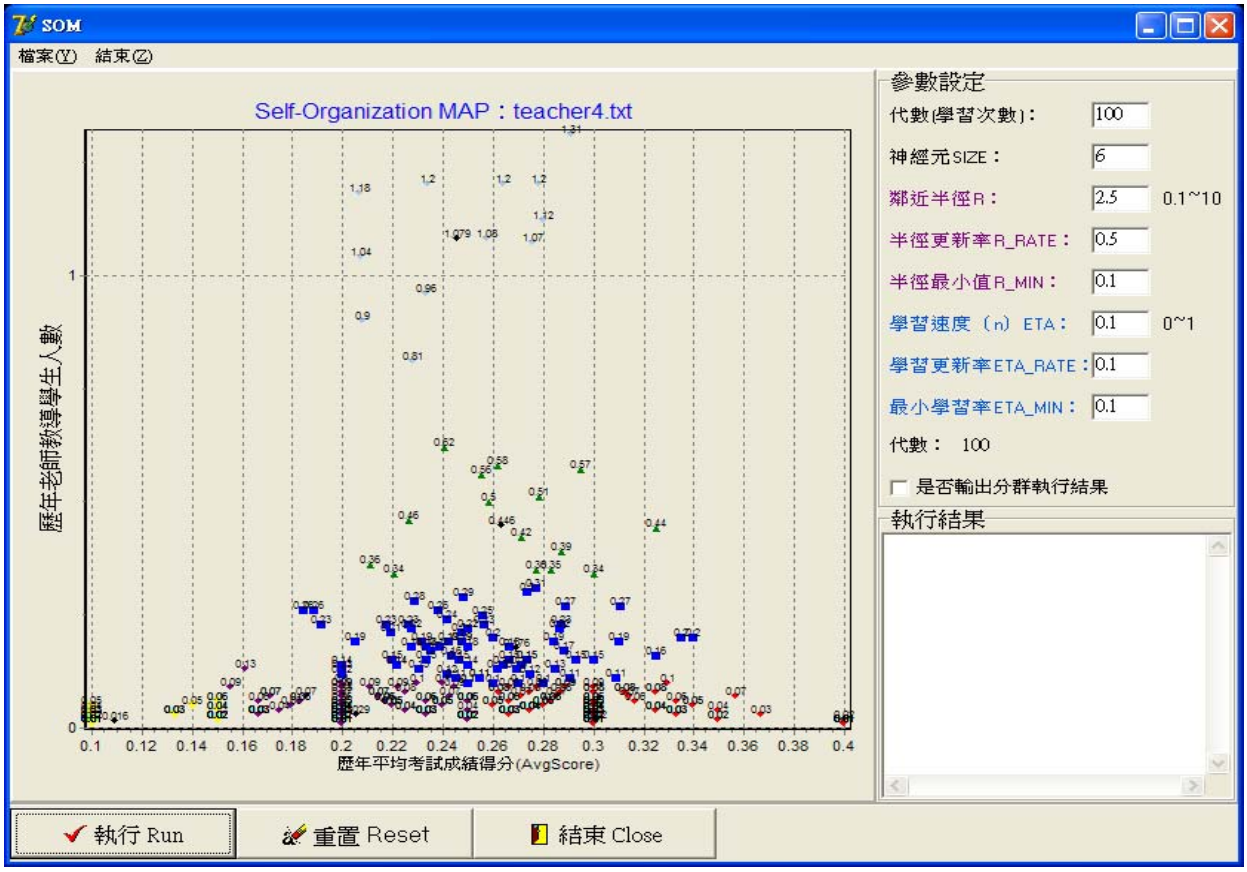

Fig. 5 Using SOM neural network to discover the potential teachers

VII. Conclusion

The investigation employs a new data clustering method named MIDBSCAN and an existing well-known scheme called SOM to conduct data clustering analysis. To enable a common practice of music learning in society, and to find out the teacher cluster with best potential through data clustering analysis, this work performs MIDBSCAN and SOM approaches to recognize those potential teachers to share their teaching experiences, and thus the teachers may discover and develop those who have a gift on music, to guide the society for music teaching and the students may enhance capabilities in music and art.

Acknowledgment

The author would like to thank the National Science Council of Republic of China, Taiwan for financially supporting this research under contract no. NSC 96-2221-E-020-027.

References

[1] Apfelstadt, Hilary, "First Things First Selecting Repertoire," *Music Educators Journal*, Vol. 87, No. 1, 2000, pp. 19.

[2] Burrack, Frederick, "Enhanced Assessment in Instrumental Programs," *Music Educators Journal*, Vol. 88, No. 6, 2002 , pp. 27.

[3] Sung, C. Y., "MIDBSCAN: An Efficient Density-Based Clustering Algorithm," Degree of Master at National Pingtung University of Science and Technology, Pingtung, Taiwan, 2009.

[4] Tsai, C. F., Tsai, C. W., Wu, H. C., Yang, T., "ACODF: A Novel Data Clustering Approach for Data Mining in Large Databases," *Journal of Systems and Software*, Vol. 73, 2004, pp. 133-145.

[5] Goolsby, Thomas W, "Assessment in Instrumental Music," *Music Educators Journal*, Vol. 86, No. 2, 1999, pp. 31.

[6] Nielsen, Siw Graabraek, "Achievement Goals, Learning Strategies and Instrumental Performance," *Music Education Research*, Vol. 10, 2008, pp. 235-47.

[7] Persellin, Diane, "The Importance of High-Quality Literature," *Music Educators Journal*, Vol. 87, No. 1, 2000, pp.17.

[8] Reynolds, H. Robert, "Repertoire Is the Curriculum," *Music Educators Journal*, Vol. 87, No. 1, 2000, pp. 31.

[9] Tsai, C.F., Yen, C.C., "ANGEL: A New Effective and Efficient Hybrid Clustering Technique for Large Databases," *Lecture Notes in Computer Science*, Vol. 4426, 2007, pp. 817–824.

[10] T. Kohonen, *Self-Organization and Associative Memory*, Springer Verlag, NY, 1984.

[11] Tsai, C.F., Liu, C.W., "KIDBSCAN: A New Efficient Data Clustering Algorithm for Data Mining in Large Databases," *Lecture Notes in Computer Science*, Vol. 4029, 2006, pp. 702–711.

[12] Kotsiantis, S., Pintelas, P., "Recent Advances in Clustering: A Brief Survey," *WSEAS Transactions on Information Science and Applications*, Vol. 1, No. 1, 2004, pp. 73-81.

[13] Kovacs, F., Ivancsy R., "A Novel Cluster Validity Index: Variance of the Nearest Neighbor Distance," *WSEAS Transactions on Computers*, Vol. 5, No. 3, 2006, pp. 477-483.

[14] Lee, H. M., "Characteristics Selecting Mode in Automatic Text Categorization of Chinese Financial Industrial News," *WSEAS Transactions on Information Science and Applications*, Vol. 3, No. 10, 2006, pp. 2016-2020.

[15] Lin, C. C., "Partitioning Capabilities of Multi-layer Perceptrons on Nested Rectangular Decision Regions Part I: Algorithm," *WSEAS Transactions on Information Science and Applications*, Vol. 3, No. 9, 2006, pp. 1674-1680.

[16] Lin, Y. H., Yu, M. N., Wu, B. L., "The investigation of developmental trends for probabilistic reasoning with fuzzy clustering on rules usage," *WSEAS Transactions on Information Science and Applications*, Vol. 3, No. 9, 2006, pp. 1674-1680.

[17] Boiangiu, C. A., Spataru, A. C., Dvornic, A. L., Cananau, D. C., "Normalized Text Font Resemblance Method Aimed at Document Image Page Clustering," *WSEAS Transactions on Computers*, Vol. 7, No. 7, 2008, pp. 1091-1100.

[18] Tsai, C. Y., Chiu, C. C., "Developing An Effective Hierarchical Clustering Method Based on Traveling Salesman Problem Model," *WSEAS Transactions on Computers*, Vol. 6, No. 3, 2007, pp. 385-393.

[19] Tsai, C. F., Yen, C. C., "G-TREACLE: A New Grid-based and Tree-alike Pattern Clustering Technique for Large Databases," *Lecture Notes in Computer Science*, Vol. 5012, 2008, pp. 739-748.

[20] Tsai, C. F., Yen, C. C., "Unsupervised Anomaly Detection Using HDG-Clustering Algorithm," *Lecture Notes in Computer Science*, Vol. 4985, 2008, pp. 356-365.

[21] Ester, M., Kriegel H., Sander, J., Xu, X., "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226-231.

[22] Borah, B., Bhattacharyya, D.K., "An Improved Sampling-Based DBSCAN for Large Spatial Databases," *Proceedings of International Conference on Intelligent Sensing and Information*, 2004, pp. 92-96.

[23] Zhou, S., Zhou, A., Cao, J., Wen, J., Fan, Y., Hu, Y., "Combining Sampling Technique with DBSCAN Algorithm for Clustering Large Spatial Databases," *Lecture Notes in Artificial Intelligence*, Vol. 1805, 2000, pp. 169-172.

[24] Nancy P. Lin, Chung-I Chang, Hao-En Chueh, Hung-Jen Chen, Wei-Hua Hao, "A Deflected Grid-based Algorithm for Clustering Analysis," *WSEAS Transactions on Computers*, Vol. 7, No. 4, 2008, pp. 125-132.

[25] Lin, N. P., Chang, C.I, Jan, N. Y., Chueh, H. E., Chen, H. J., Hao, W. H., "An Axis-shifted Crossover-Imaged Clustering Algorithm," *WSEAS Transactions on Systems*, Vol. 7, No. 3, 2008, pp. 175-184.

[26] Wang, S. C., Huang, P. H., "Power system output feedback controller design using fuzzy c-means clustering reduced model," *WSEAS Transactions on Systems*, Vol. 6, No.3, 2007, pp. 475-480.

[27] Hichem Frigui and Raghu Krishnapuram, "A Robust Competitive Clustering Algorithm With Applications in Computer Vision", *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 21, No.5, 1999, pp. 450-465.

[28] Tsai, C. Y., Chiu, C. C., "Developing an effective hierarchical clustering method based on traveling salesman problem model," *WSEAS Transactions on Computers*, Vol. 6, No. 3, 2007, pp. 385-393.

[29] Tsai, H. R., Chen T., "Wafer Lot Output Time Prediction with a Hybrid Artificial Neural Network," *WSEAS Transactions on Computers*, Vol. 5, No. 5, 2006, pp. 817-823.

[30] Hung, M. C., Weng, S. Q., Wu, J. P., Yang, D. L., "Efficient Mining of Association Rules Using Merged Transactions Approach," *WSEAS Transactions on Computers*, Vol. 5, No. 5, 2006, pp. 916-923.

[31] Liu, H. C., Wu, D. B., Yih, J. M., & Liu, S. W., "Fuzzy Possibility C-Mean Based on Mahalanobis Distance and Separable Criterion," *WSEAS Transactions on Biology and Biomedicine*, Vol. 4, No. 7, 2007, pp. 93-98.

[32] Liu, H. C., Wu, D. B., Ma, H. L, "Fuzzy Clustering with New Separable Criterion," *WSEAS Transactions on Biology and Biomedicine*, Vol. 4, No. 7, 2007, pp. 99-102.

[33] Wang, J.-H., Rau, J.-D., "VQ-Agglomeration: a novel approach to clustering," *IEE Proceedings-Vision, Image and Signal Processing*, Vol. 148, No.1, 2001, pp. 36-44.

[34] Sudipto Guha , Rajeev Rastogi and Kyuseok Shim, "CURE: An Efficient Clustering Algorithm For Large Database," *Information Systems*, Vol.26, No.1, 2001, pp. 35-58.

[35] Kohonen, T., "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, Vol. 43, 1982, pp. 59-69.

[36] Kohonen, T., "The self-organizing map," Proceedings of. IEEE, Vol. 78, No. 9, 1990, pp.1461-1480.

[37] K.Obu-Cann, K.Iwamoto,H. Tokutaka and K. Fujimura, "Clustering by SOM (self-organizing maps), MST (minimal spanning tree) and MCP (modified counter-propagation)," 6th International Conference on Neural Information Processing (IEEE ICONIP '99), Vol. 3, pp.986 -991, 1999.

[38] Philipp Tomsich ,Andreas Rauber and Dieter Merkl, "Optimizing the parSOM neural network implementation for data mining with distributed

memory systems and cluster computing," *IEEE 11th International Database and Expert Systems Applications*, pp. 661-665, 2000.

[39] Masahiro Endo, Masahiro Ueno, Takaya Tanabe and Manabu Yamamoto, M., "Clustering method using self-organizing map", *2000 IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing*, Vol.1, pp.261-270, 2000.

[40] Su, M. C., Chou, C. H., "A Modified Version of the K-Means Algorithm with a Distance Based on Cluster Symmetry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, JUNE 2001, pp. 674-680.

[41] Hichem Frigui and Raghu Krishnapuram, "A Robust Competitive Clustering Algorithm with Applications in Computer Vision," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 21, No. 5, 1999, pp. 450-465.

Cheng-Fa Tsai, an IEEE member, is full professor of the Management Information Systems Department at National Pingtung University of Science and Technology (NPUST), Pingtung, Taiwan. His research interests are in the areas of data mining and knowledge management, database systems, mobile communication and intelligent systems, with emphasis on efficient data analysis and rapid prototyping. He has published over 150 well-known journal papers and conference papers and several books in the above fields. He holds, or has applied for, three U.S. patents and twenty-seven ROC patents in his research areas. Dr. Tsai is a member of IEEE, WSEAS and TAAI (Taiwanese Association for Artificial Intelligence).

Jessica Yu-Tai Su is Assistant Professor of the Department of Western Music at the Chinese Culture University, where she is also the Director of the Music Education Institute at the School of Continuing Education, Taipei, Taiwan. She holds an Ed.D. in music education from Columbia University. Her research interests include cognitive psychological development in music, and the enhancement of teaching/learning situations in music education, especially with sociological perspectives.

Chiu-Yen Tsai is a master student of Department of Management Information Systems at National Pingtung University of Science and Technology, Pingtung, Taiwan.

Chun-Yi Sung received the BS degree from Kao-yuan University of Science and Technology, Kaohsiung, Taiwan, in 2005. He is currently working toward the MS degree in the Department of Management Information Systems, National Pingtung University of Science and Technology (NPUST), Pingtung, Taiwan. His research interests include data mining and intelligent systems.