Discovering Potential Musical Instruments Teachers Using Data Clustering Approach

CHENG-FA TSAI¹, YU-TAI SU², CHIU-YEN TSAI¹, CHUN-YI SUNG¹
¹ Department of Management Information Systems, National Pingtung University of Science and Technology, Pingtung, TAIWAN
² Department of Western Music, Chinese Culture University, Taipei, TAIWAN
E-mail: cftsai@mail.npust.edu.tw

Abstract: This paper aims to introduce data clustering approach to discover potential musical instruments teachers. With a total of 5125 candidates registered respectively in 9 grades during 2000-2008. Moreover, this work presents a new data clustering algorithm named MIDBSCAN and an existing well-known neural network called self-organizing map (SOM) to perform data clustering applications for finding potential musical instruments teachers. 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.

Key-word: data mining, data clustering, neural networks, musical instrument selection, music examine, test, effective teaching

1. Introduction
Taiwan United Music Grade Test (named as music grade thereof [5]) is implemented since 2000 to evaluate the performance of various 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 [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 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?

Data clustering is a new-emerging and imperative technique for data mining applications [3]. 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 and grid-based algorithm. This paper presents a density-based MIDBSCAN algorithm [1] and a well-known self-organizing map (SOM) [4] to perform 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.

2. Related Works
This section introduces DBSCAN, IDBSCAN and KIDBSCAN density-based clustering methods, which are related to the proposed MIDBSCAN algorithm.

DBSCAN, presented by Ester et al. in 1996 [9], was the first clustering approach to employ density as a condition. DBSCAN recognizes arbitrary shapes almost perfectly with two thresholds (searching radius and minimum points), but has a very high time cost, making it unpopular for use in business applications.

IDBSCAN is developed by Borah et al. in 2004 [9]. This method applies a Marked Boundary Object to determine the data point of an expansion seed when searching for neighborhood to add in expansion seeds.

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. Experimental results verify that KIDBSAN performs data clustering quickly.
3. The proposed MIDBSCAN Scheme

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

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.

4. 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.

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.

Table 1: Taiwan United Music Grade Test Table from 2000 to 2008.

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

Table 2: The weighted score of student obtained.

<table>
<thead>
<tr>
<th>Student Score</th>
<th>Score by grade</th>
<th>Weighted Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Excellent</td>
<td>0.4</td>
</tr>
<tr>
<td>B, C</td>
<td>Pass, qualified</td>
<td>0.3</td>
</tr>
<tr>
<td>D</td>
<td>Barely pass</td>
<td>0.2</td>
</tr>
<tr>
<td>E, H</td>
<td>failure, unqualified</td>
<td>0.1</td>
</tr>
<tr>
<td>F, G</td>
<td>Exam rescheduling, absence for exam</td>
<td>No inclusion</td>
</tr>
</tbody>
</table>

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:

\[ \text{AvgScore} = \frac{\sum_{i=0}^{n} \text{StudentScore}(i)}{n} \]

Where \( \text{AvgScore} \) denotes average score of each teacher on lecturing over the years, the denominator \( (\text{StudentScore}(i)) \) 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 \( \text{AvgScore} \), 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.

4.1 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 (Epsilon) and Minimum Points. In Fig. 1, it
takes 0.263 seconds to conduct a clustering task by keying in the potential teachers data sets using MIDBSCAN algorithm (radius=0.02, minimum points=5), with 6 clusters shown in total, in which one of the clusters is noise (shown in Fig. 2), 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”.

4.2 Clustering analysis Using SOM Scheme

SOM originated from 1980, where Kohonen [10] 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. It takes 0.442 seconds for SOM to carry out data clustering in this experiment. Fig. 3 reveals the clustering result using SOM approach. In Fig. 3, 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. 3. Using SOM neural network to discover the potential teachers.

5. Conclusion
The investigation employs a new data clustering method named M-IDBSCAN 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 M-IDBSCAN 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:


