On Clustering Cellists using Acquired Data through Performance Movies Observation

Koichi Furukawa[1], Toshiki Masuda[2], and Takeshige Nishiyama[3]

[1] Keio University kfurukawa109@gmail.com
[2] Cellist BYS03735@nifty.com
[3] SFC Research Institute, Keio University tksg004@gmail.com

Abstract. By specifying a set of more than ten characteristic attributes related to cello playing such as "right elbow height", "vertical movement of the wrist on bow reverse action" and "the degree of left-right body trunk motion", we collect a set of data from observation of 46 cellists movies on a video site and conduct clustering and decision tree analyses to identify a set of clusters. As a result, we succeeded in obtaining five clusters which may be useful in finding a cello playing style suited for each player.

Keywords: individual differences, cellists' performance, attributes selection, data acquisition by observation, clustering, decision tree analysis, cluster characterization

1 Introduction

In performance skill acquisition, the problem of individual differences is quite serious which inhibits us to find objective rules to any performers in common. This fact causes difficulties in skill acquisition process. When trainers coach to beginners, they may stick to their own performance styles regardless to their students' styles.

As a first step to overcome this problem, we try to group professional cellists in all over the world by their performance styles. This attempt should give any hints for explaining skill diversities.

To accomplish grouping, we first specify a set of more than ten characteristic attributes related to cello playing. We then collect a set of data from visual observation of 46 cellists movies on a video web site and conduct clustering and decision tree analyses to identify clusters. As a result, we succeed in obtaining five clusters which may be useful in finding a cello playing style suited for each player. This paper reports the entire process of the cellists' clustering.

In Chapter 2, we describe how to conduct attributes selection. In Chapter 3, we state the way of acquiring data by observing movies on the web site. In Chapter 4 and 5, we introduce clustering and decision tree analysis to the acquired data. In Chapter 6, we conclude our paper.

adfa, p. 1, 2011.

2 Attributes Selection

2.1 Attributes selection by means of acquired knowledge

The problem of attributes selection is crucial in cellists clustering. If we cover all important attributes, the results would be successful. However too many unnecessary attributes might prohibit us to lead success because they may contain duplicated attributes which lead to heavy computation and possible noise in data. In our experiment, we first selected 17 attributes by subjective criteria for judgement based on personal experiences of the first and second authors. Their attributes consist of general data such as nationality, sex, body size, hand size, and more particularly cello performance data such as force generation ways in bow handling, whether one bends the right hand wrist vertically or not when one changes the bowing direction, whether one moves the body heavily or not, whether one leans the cello much or not, and so on.

2.2 Review of attributes selection

After selecting the 17 attributes, we conducted data acquisition and preliminary clustering. Through the pre-experiment, we discovered excess and deficiency of attributes and proceeded to modify the attributes set. There are three reasons for the modification: (1) unnecessary attributes due to overlapping characteristics, (2) those attributes which are hard to measure, and (3) newly discovered attributes during the data acquisition. We introduce examples below.

(1) Overlapping attributes

At the beginning, we selected “arm length” and “hand size” besides “body size”. However, those three attributes showed very high similarity in data acquisition phase and we decided to eliminate the two attributes.

(2) Attributes hard to measure

Such attributes as “force generation ways in bowing: whether we push the bow by the elbow or we pull the bow by the wrist”, “the difference of posture: whether stooping or straight or backward-leaning”, “the way of position shifting: whether elbow first, hand-first or at-the-same-time” were hard to judge and we decided to eliminate them. In case of the posture difference, it seems easy to collect data but most of the performance movies were taken from the front and there were almost no scenes taken from side, which were indispensable to judge the posture.

(3) Newly discovered attributes during the data acquisition phase

In observing performance movies, we discovered several attributes which seem to be useful in clustering as listed below.

i. Wrist Dent on Top of the Bow

ii. Left Arm Angle at the 1st Position
iii. Bow-Grasp Finger Width
iv. Right Palm Shape

Three attributes, i, ii and iv, played important roles in clustering. By observing different types of players carefully, we happened to recognize the importance of those features.

The finally selected 12 attributes are listed as follows:

1. Sex
2. Body Size
3. Elbow Height at E String
4. Wrist Vertical Bend
5. Wrist Dent on Top of the Bow
6. Neck Move
7. Head-Cello Distance
8. Cello Angle
9. Body Movement
10. Left Arm Angle at 1st Position
11. Bow-Grasp Finger Width
12. Right Palm Shape

3 Data Acquisition by Observing Performance Movies

For each player, we determined attributes' values by visual judgement. Each attribute has numerical value within 0~1, almost all were 0, 1 binary values, or 0, 0.5, 1 ternary values. Exceptions were elbow heights and Head-Cello Distance and they were treated as continuous values in between 0 to 1.

3.1 Selection of cellists

We selected a set of cellists for the clustering from an internet video site. We excluded very young cellists since their performance were quite different from others.

3.2 Data acquisition by the examiner's visual judgement

We observed each cello performance video carefully by being aware of the target attribute and determined the value. What we need is to make the measurement precise enough. To accomplish this, we need to define the measurement attributes rigorously and to prevent the measurement fluctuation.

An example of rigorous definition is "Head-Cello Distance." At the beginning, we defined the value of this attribute as binary but later we changed it as continuous. Furthermore, we defined the distance as the distance in 3D space between the head and the cello's neck and also we took the time duration of the distance degree into account.

To avoid the measurement fluctuation, the first measurement was done through every attribute for each cellist, but the second one through every cellist for each attribute. The second attempt prevented the measurement fluctuation since the same attribute was measured across players in a short period of time.

Furthermore, we measured some attributes twice to verify the measurement accuracy and for those attributes which had different results, we measured again and conducted modification. We show the result of the measurement in Fig. 1.

Cellist	Sex	BodySize	ElbowHight	Wrist Vert Bend	WristDent	Neck Move	Head Cello Dist	Cello Angle	BodyMove	Left Arm Angle	GraspFingerWidth	Right Palm Twist
Ofla Harnoy	Female	S	0.5	No	Yes	Big	1	0.5	1	0	0.5	Flat
Madoka Hojo	Female	S	0.3	Yes	Mid	Small	0.3	0	0.5	0	0.5	Flat
Urara Urakawa	Female	S	0.6	Yes	Yes	Small	0	0.5	1	0.5	0	Flat
Tanya Anisimova	Female	S	0.5	Yes	Yes	Big	0	0.5	0.5	0	1	Flat
Mari Endoh	Female	S	1	Yes	Mid	Big	0.6	0.5	1	1	1	Flat
Osamu Kawamura	Male	S	0.5	Yes	Mid	Big	0.7	0	1	0	1	Twist
Marie-Elisabeth Heck	Female	S	0	Yes	Yes	Big	0.8	0.5	1	0	1	Twist
Yoko Hasegawa	Female	S	0.5	Yes	Yes	Big	0.5	0.5	0.5	0	0.5	Twist
Sol Gabetta	Female	L	0.5	Yes	Yes	Big	0.7	0.5	1	0	1	Twist
Tatiana Vassilieva	Female	L	1	Yes	Yes	Big	1	0.5	1	0	0.5	Twist
Hitomi Niikura	Female	S	0.5	Yes	No	Small	0.2	0.5	1	0	0.5	Twist
Yoriko Miyake	Female	S	0.7	Yes	Mid	Small	0.3	0.5	1	0.5	0.5	Twist
Jacqueline DuPre	Female	S	0.5	Yes	Yes	Big	0	0.5	1	1	0.5	Twist
Kateryna Bragina	Female	S	0.7	Yes	Yes	Big	0.5	0.5	1	1	0.5	Twist
Rinao Yaguchi	Female	S	0.5	Yes	Yes	Big	0.2	0.5	1	1	1	Twist
F.Guye	Male	L	0.7	Yes	Mid	Small	0	0	0.5	1	0.5	Flat
Jian Wang	Male	S	0.8	Yes	Mid	Small	0.5	0.5	0	0	0	Flat
Lynn Harrell	Male	L	0.6	No	Mid	Small	0.2	1	0	0	0.5	Flat
Janos Starker	Male	L	0.7	No	No	Small	0.2	0.5	0	0	0	Flat
Piatigorsky	Male	S	0.2	No	Mid	Small	0.5	0	0.5	0	0	Twist
Rostropovich	Male	L	0	No	Yes	Small	0	0	0.5	0	0	Flat
Paul Tortelier	Male	L	0.7	No	Mid	Small	0.3	0	0	0	0.5	Flat
Michaela Fukačová	Female	L	0.6	No	Yes	Small	0	0	0	0	0.5	Flat
Amit Peled	Male	L	0.7	Yes	Yes	Big	0.3	0	0	1	0	Flat
Mario Brunello	Male	S	1	No	No	Big	1	1	0.5	0	0.5	Flat
Hiroki Kashiwagi	Male	S	0.2	No	No	Big	0	0.5	1	0	1	Flat
Davide Amadio	Male	S	0	Yes	No	Big	1	1	1	0	0.5	Flat
Akiko Hasegawa	Female	S	0.5	No	No	Big	0	1	0	0	0.5	Twist
Rintaro Kaneko	Male	S	0.5	No	No	Big	0.8	1	1	0.5	0	Flat
Miklós PERÉNYI	Male	S	1	No	No	Big	1	1	1	0.5	0.5	Twist
Mischa Maisky	Male	S	0.5	No	No	Big	0	0.5	1	0	0.5	Flat
Stéphane Tétreault	Male	S	0.7	No	Mid	Big	0.5	0.5	1	1	0.5	Flat
Pierre Fournier	Male	S	0.5	No	No	Big	0.5	0	0	0.5	0.5	Flat
Ikuya Okamoto	Male	S	0.5	Yes	No	Big	0.5	0.5	0	0	0.5	Flat
Yasuiro Hasegawa	Male	S	0	No	No	Small	0.2	0.5	0	0	0	Flat
Pabro Casals	Male	S	0.5	No	Yes	Small	0	0.5	0.5	0	0.5	Flat
Michiaki Ueno	Male	S	0.7	No	No	Small	0.2	0	0.5	0	0	Flat
Dai MIYATA	Male	S	0.5	No	Yes	Big	0	0	0	0	0.5	Flat
Benedict Kloeckner	Male	S	0.5	No	Mid	Small	0.3	0.5	0.5	1	0.5	Twist
Leonard Rose	Male	L	0.5	No	Mid	Small	0.2	0	1	0.5	0	Twist
Truls Mørk	Male	L	0.8	No	Yes	Small	0	0	1	0.5	0.5	Twist
Yo-Yo Ma	Male	L	0.5	No	No	Big	0.7	0	1	1	1	Flat
Michael Schonwandt	Male	L	0.8	No	No	Big	0	0	1	1	0.5	Flat
Luka Sulic	Male	L	0.8	Yes	No	Small	0.2	0	1	1	1	Twist
Tsuyoshi Tsutsumi	Male	L	0.7	Yes	No	Small	0	0	0.5	1	0.5	Flat
Xavier Phillips	Male	L	0.7	No	No	Big	0	0.5	0	0	1	Twist

Fig. 1. Cellists performance data acquired through performance movies observation

4 Clustering

We applied k-means clustering to find a set of clusters. The most essential parameter of the k-mean algorithm is the number of clusters. To find an appropriate number of clusters, we computed the sum of the square errors within-cluster by changing the cluster number from 3 to 7. The result attained minimum when the cluster number was five as shown in Fig. 2.

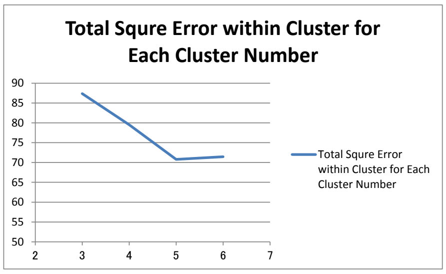

Fig. 2. The change of total square error within-cluster for each cluster number

The obtained five clusters are shown in Fig. 3, where the representative cellists of the five clusters are as follows; Cluster0: Jacqueline Du Pre, Cluster1: Mischa Maisky, Cluster2: Yo-Yo Ma, Cluster3: Leonard Rose, and Cluster4: Pabro Casals.

Cluster0		
Urara Urakawa	Sol Gabetta	Yoko Hasegawa
Tanya Anisimova	Tatiana Vassilieva	Kateryna Bragina
Mari Endoh	du Pre	Rinako Yaguchi
Marie-E. Hecker		
Cluster1		
Ofla Harnoy	Stéphane Tétreault	Rintaro Kaneko
Mario Brunello	Pierre Fournier	Miklós PERÉNYI
Hiroki Kashiwagi	Ikuya Okamoto	Mischa Maisky
Davide Amadio	Yasuhiro Hasegawa	Dai Miyata
Akiko Hasegawa	Michiaki Ueno	Xavier Phillips
Cluster2		
F.Guye	Michael Schonwandt	Yo-Yo Ma
Amit Peled	Luka Sulic	Tsuyoshi Tsutsumi
Cluster3		
Madoka Houjou	Jian Wang	Yoriko Miyake
Osamu kawamura	Piatigorsky	Leonard Rose
Hitomi Niikura	Benedict Kloeckner	
Cluster4		
Lynn Harrell	Michaela Fukačová	Paul Tortelier
Janos Starker	Pabro Casals	Truls Mørk
Rostropovich		

Fig. 3. Clustering results by K-Means Clustering

5 Characterization of Clusters by Decision Tree Analysis

In order to characterize each cluster, we conduct decision tree analysis by specifying the cluster name as the class to be decided. We use an application software called Weka, the same software for the clustering. We set the minimum numbers of the objects in each leaf of the tree as 3. The test option we adopted is "use training set". The obtained decision tree is shown in Fig. 4.

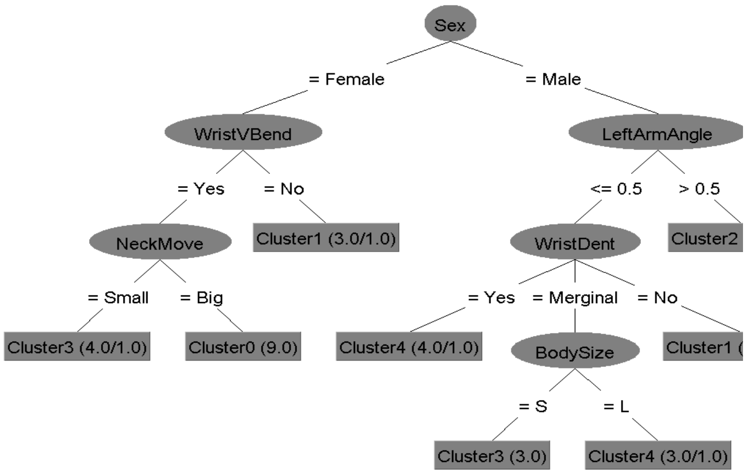

Fig. 4. Decision tree of the cellists' clusters

An oval node of the decision tree represents an attribute selected for decision at each level and the value of each arc from the node is its value. A rectangle node is a leaf node which represents the decided cluster. A rectangle node contains the decided cluster name with a number of records recognized as the cluster and also the number of miss-classified records. One measure judging the correctness of the decision tree is the accuracy rate. The accuracy rate of the decision tree sown in Fig. 4 is 84.8% and it is quite high. The decision tree shows that the top (root) node is the "sex" attribute which means it is the most important feature for the clustering. However, by careful investigation, we find that the cluster1 and the cluster3 appear both in male and female branches. To resolve the separation problem, we conduct further decision tree analysis by removing the "sex" attribute. The result is shown in Fig. 5. This decision tree shows even higher accuracy than before, attaining 87.0 %. On the contrary, the number of the leaf nodes increases from 8 to 10.

It is quite interesting that there appear three new attributes in this decision tree; they are "Head-Cello Distance", "Cello Angle", and "Right Palm Shape" among which one of the newly introduced attributes after the preliminary experiment is included. Since two other new attributes are contained in the first decision tree, three of the four new attributes are recognized as important features.

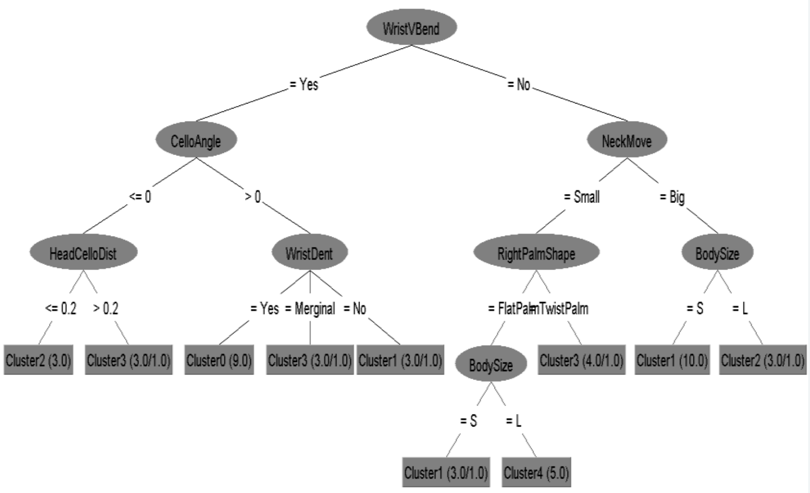

Fig. 5. Decision tree of the cellists' Clusters without the "sex" attribute

To find the similarity or dissimilarity among clusters, we computed inter-cluster distances of every pair of clusters which we defined as the average distance of each pair of records in the cluster-pair. The result is shown if Fig. 6. The two dimensional distance map is then computed by Multi-Dimensional Scaling as shown in Fig. 7. This figure shows that the cluster3 is located near the center of the gravity and may be considered as a noise cluster.

Cluster pair	Min distance	Max distance	Mean distance
2-4	2.34	6.33	4.4
3-4	1.38	7.75	4.51
1-4	1.49	8.41	4.59
0-3	2.18	7.79	4.69
1-3	2.33	7.48	4.82
1-2	2.34	8.29	5.15
2-3	2.83	7.55	5.19
0-1	2.35	8.89	5.28
0-2	4.01	7.88	5.68
0-4	2.76	9.04	6.35

Fig. 6. A list of cluster pairs sorted by average of all elements pairs square distances.

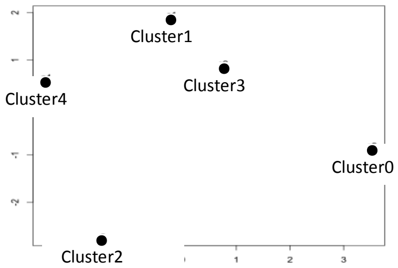

Fig. 7. Cluster locations reflecting inter-cluster distances

Now we try to characterize each cluster by carefully investigating Fig. 4 and 5.

i. Cluster0
Cluster0 is a cluster of female cellists having large amount of wrist vertical bending and also large neck motion. The representative cellist is Jacqueline duPre, who uses her small body effectively by the big body and neck motions.

ii. Cluster1
The characteristic feature of the Cluster1 cellists is that their body sizes are small and they move neck heavily and they don't bend their wrists vertically much. The last point is particularly different from Cluster0. There are both male and female players in this cluster, and the representative male player is Mischa Maisky and female player Ofla Harnoy.

iii. Cluster2
Cellists in this cluster are male, have large body and keep their left arm angle high at the 1st position. Furthermore, they lean their cellos a lot. Yo-yo Ma is the representative.

iv. Cluster3
Cellists in this cluster are generally small and they move their heads very little and keep the distance between the cello and their head far apart. They either bend their wrists vertically or twist their right palms if not. The representative player is Leonard Rose. Three Japanese female cellists are in this cluster. As indicated before, this cluster is located rather in the center of all clusters and does not have any significant characteristic features. Therefore we consider it as a noise cluster.

v. Cluster4
The characteristic feature of the Cluster4 cellists is that they don't move neck much, the left arm angle at 1st position is small, the distance between the head and the cello is small and they don't twist the right palm much. This cluster is located close to the Cluster2 as shown in Fig. 5 except the difference of the left arm height.

Those attributes which are selected as separating clusters are important to distinguish cellists types. The attribute of "Wrist Vertical Bend" is related to the way of achieving right hand flexibility, either by wrist joint or by elbow/shoulder joint. Cellists in the Cluster0 utilize mainly the vertical motion of the wrist joint and additionally neck and body vibrations. On the other hand, those in the Cluster1 utilize the horizontal move of the wrist joint and also the finger joints. The merit of the vertical motion of the wrist is its easiness to obtain large amplitude of vibration but the demerit is that the natural direction of the palm at the vertical vibration is different from the bow direction and therefore they need to twist the palm to adjust the difference of angles. It may cause deficiency of the arm torque (motion dependent torque) delivery to the bow and they tend to compensate the loss by using other parts of the body like neck and body.

The characteristic feature to keep the left elbow at high position is also notable. Since they need an extra power to keep the left elbow high, all cellists in this cluster are male and large. This posture can produce counter power to the down-bow motion and therefore it makes possible to attain strong and rapid down-bowing. On the other hand, the playing style of keeping the elbow in lower position, which is observed in Cluster1, Cluster3 and Cluster4, seems more natural and we can expect softer sound from this style.

6 Conclusion

We conducted to cluster 46 cellists. Although this number seems too small to obtain trustful outcome, the clustering result is rather surprising and very convincing. During the clustering process, the first hurdle was to determine adequate set of attributes of each player to be collected by simple visual observation. By utilizing acquired knowledge from life-long cello playing and thoughtful consideration of cello playing techniques in the skill science research, we succeeded in finding an appropriate set of attributes. Then we found the right number of clusters by simply computing the intra-clusters diversities. The obtained clusters were not known to the authors in advance and brought us a new knowledge. We knew the fact that there are several different types of cellists, but we did not know the detail. Therefore the result was very surprising.

The success of this research indicates that our clustering and decision tree analysis methods are promising in grouping performers in any fields. The utility value of the clustering result is to be investigated in future.

Acknowledgement

We deeply express of our thanks to Keita Kinjo from Okinawa International University and Tomonobu Ozaki from Nihon University for their help in data analysis.