Identifying Preferences in a Jury Study Using Clustering Techniques

Glenn Pietila and **Teik C. Lim**, University of Cincinnati, Cincinnati, Ohio

In the study of product sound quality using paired-comparison methods, it is generally assumed that preference within the jury population is universal. This assumption often holds true, but there are often scenarios where preference varies significantly with demographics. Identifying the presence of and partitioning the jury members with these varying criteria for preference is seldom a trivial exercise. This article presents an unsupervised clustering technique that can be applied to jury-paired preferences in an attempt to infer the number of subgroups in a jury pool. The same algorithm can then be used to classify jury members into appropriate subgroups. In addition, the approach is applied to the results of a small-engine jury study of more than 80 members. Two subgroups, corresponding to American and European jurors, were successfully identified and classified.

In analyzing product sound quality, the method of paired comparison[1,2] is often used to establish an understanding of the preference or impression of quality of a product based on the sound that the product emits during operation.[3-6] In most cases, including the commonly used Bradley-Terry model,[7] it is assumed that the entire jury pool has a common preference. Although this assumption is often true, it is not a reasonable assumption for all products with sound emission issues.

An example of varying preferences was observed during the analysis of a small informal jury study for impact harshness sound. It was found that there were two subgroups of jury members in the jury pool; those that preferred a firm short duration impact and those that preferred a low-level sound amplitude. The first group of jury members was willing to accept a higher-level impact for a short duration or "crisp" impact. In contrast, the second group preferred a lower level of impact sound even if it meant that the impact event was longer.

Therefore, it is necessary to either separate the jury population into subgroups (or demographics) prior to applying the model or to use a model that can account for the individual juror demographic.[8] In either case, it is necessary to know the number of subgroups within the jury population and classify each of the jury members into the appropriate demographic. For this reason, a reliable method that can be used to evaluate a jury population and identify subgroups based on juror voting patterns is a useful tool.

This article presents the development of an unsupervised K-means clustering approach, similar to that discussed by Lathrup and Williams,[9,10] to estimate the number of subgroups and classify each juror into the appropriate subgroup. The method uses the K-means algorithm to cluster the jury votes (data points) and evaluates the total mean square distance (error) between the data points and cluster centers. This is iterated with the number of clusters increasing from 1 to n, where n is the number of clusters that is expected to exceed the true number of clusters or jury subgroups, in the study. The total error for each iteration is then used to infer the number of independent subgroups in the jury. In addition, a Ward's agglomerative clustering approach[16] was compared to the K-means clustering method. Finally, the procedure is applied to data collected during a real jury study of small-engine sounds. Both the K-means and Ward's clustering methods are used to identify subgroups in the jury that included roughly half European and half American jurors.

Background

Clustering analysis[11,12] is an unsupervised method for organizing data into groups with similar characteristics with no prior knowledge of the classification of the data points. The goal of clustering is to group similar data and to gain an understanding of underlying characteristics or structures within the observed data.

One of the simplest and most common methods used for unsupervised clustering is the family of K-means algorithms.[13] The K-means clustering algorithm categorizes data points by attempting to minimize the total sum of the squared distances (or total squared error) between the data points and cluster centers. An introduction to clustering in general along with various examples of clustering algorithms is given by Hartigan,[14] with additional introductory discussions provided by Jain[13] and Bishop.[15]

While the goal of the K-means clustering algorithm is generally to identify cluster centers that describe and classify the data based on a given number of clusters k, the goal as applied in this approach is to infer the number of clusters (jury subgroups) based on the relationship between the number of clusters and the total squared error. This procedure, described by Lathrup and Williams[9,10] as an inverse scree test, infers the appropriate number of clusters by identifying the "elbow" in a plot of the number of clusters versus the total squared error. Once the appropriate number of clusters, or jury subgroups, is known, the K-means algorithm is then applied to classify the data points, or jury members, into the appropriate subgroups.

An alternative method found in the open literature is the use of a hierarchical approach such as that developed by Ward.[16] His approach, described as hierarchical agglomerative, essentially begins with all data points as their own respective clusters. Therefore, for n data points there will be n clusters defined, and the total sum of error will be zero. During the first iteration, two clusters (points) are merged into a cluster, and the error associated with this merger is calculated. The error, referred to as the "cost," is commonly a squared sum of the distances between all points in the cluster and the cluster center, defined as the mean of the points within the cluster. This is repeated for all possible pairs within the dataset, and the pairing with the lowest cost is selected as a cluster for that iteration. In this way, after one iteration, there will be n–1 clusters representing n data points.

For the next iteration, the cost is calculated for all possible pairings of the n–1 clusters. Again, the combination with the lowest cost is selected as a cluster for the second iteration, reducing the number of clusters to n–2. This is repeated until there is only a single cluster to represent all of the data points. By tracking the cost associated with each iteration, one can infer the appropriate number of clusters by identifying the point at which reducing the number of clusters by one significantly increases the error.

Funfgeld and Wang[17] discuss the use of clustering to better understand attitudes and behaviors in routine finance, and Ross[18] utilizes cluster analysis in segmenting sports fans from a marketing perspective. In both cases, the clustering approach was adopted from suggestions by Punj and Stewart[19] for using a two-stage clustering process. In this approach, a hierarchical method, such as Ward's method,[19] is used in the first stage to infer the optimal number of clusters and to establish initial cluster centers. The second stage uses a K-means algorithm, initialized with the number of clusters and cluster centers output from the first stage, to optimize the cluster assignments.

As discussed by Funfgeld,[17] the second stage uses a K-means solution to avoid a weakness of Ward's clustering approach. Since Ward's method is a deterministic and "greedy" algorithm, it has no means of resorting or optimizing the cluster assignments once a solution is reached. Punj and Stewart[19] proposed the two-stage process to address the concern that the solution to a clustering problem using a K-means approach depends on the initial cluster

Based on a paper presented at Noise-Con 2013, 26th Annual Conference on Noise Control Engineering, August 2013.

centers. For this reason, Ward's method is used to establish the initial cluster centers. Then, the K-means algorithm is used to "optimize" these clusters. While this results in a deterministic algorithm, there is still no guarantee that the solution will be a global minimum.

Xu and Wunsch[12] summarize alternative methods to addressing the initial cluster center issue with the K-means algorithm. Of the methods discussed, the strategy of solving the K-means algorithm for multiple randomly chosen initial centers and selecting the best result is used in the proposed approach employed in the present study.

The K-means algorithm was chosen instead of a model based clustering method, such as a Gaussian Mixture Model, due to the discrete nature of the training vectors. Because the jury response can be one of three options (prefer A, prefer B, or no preference) it would be inappropriate to use a continuous model. In addition, the K-means algorithm is a fairly simple algorithm to implement and has been applied successfully in many other classification problems.

Furthermore, the method of generating a K-means solution over multiple random initial centers was chosen because it is less likely to be trapped in a local minimum as compared to Ward's method described above. An additional advantage in using a K-means-only approach includes robustness to outliers[19] as compared to other clustering algorithms.

K-Means Clustering. The actual implementation of the K-means algorithm can vary slightly, but in most cases, the following structure is applied, as presented by MacKay[20]:

Initialization: Initialize the K-means ($m^{(k)}$) to a set of random values.

Assignment: Assign each data point n to the nearest mean. In this step, the initial guess for the cluster $k^{(n)}$ that the point $x^{(n)}$ belongs to is denoted as $k'^{(n)}$:

$$k'^{(n)} = \arg\min_k \left\{ d\left(m^{(k)}\right), x^{(n)} \right\} \tag{1}$$

where $\{d(m^{(k)}), x^{(n)}\}$ represents the distance between the point, $x^{(n)}$, and the cluster center $m^{(k)}$. Argument in this case represents the minimum in the set of all values of k.

Update step: The mean $m^{(k)}$ is recalculated to match the mean of the data points belonging to it.

$$m^{(k)} = \Sigma_n x^{(N)} / R^{(k)} \tag{2}$$

where $R^{(k)}$ is the number of data points belonging to $m^{(k)}$ and $\Sigma_n x^{(n)}$ represents sum of all points belonging to set n.

Iteration: Repeat the assignment and update steps until there is no more change.

Variations to the algorithm generally involve modifications to the method for initializing the K-means, the method for calculating the distance between the means and the data points, and method for assigning the points to clusters. These approaches are summarized by Jain.[13]

Some common methods for initializing the means include randomly selecting points within the data space, randomly selecting actual data points, or selecting mean points based on data point density. Various methods are discussed by Steinley[21] and Xu and Wunsch.[12]

The distance calculations include those commonly used in measuring distances in multidimensional space, such as the Euclidean, cosine, Manhatten (city block), Mahalanobis and Hamming distance.[12] Finally, variations to the iterative approach include various strategies intended to reduce the solution time and address some of its shortcomings. Some of the approaches are presented by Xia, *et al.*[22]

The K-means algorithm behaves like a gradient-descent approach so is susceptible to getting trapped in local minimums. Also the solution depends on the starting point or the initial K-means used.[13] A common strategy applied to reduce this effect, or at least increase the likelihood of finding the global minimum, involves iterating on multiple initial means. This strategy employed by the *kmeans* function in MATLAB®[23], and discussed by Xu and Wunsch[12], is applied in this study. To implement the strategy, an error function is first defined. A common error function is the total sum of the distances between the data points and their respective cluster means. For each iteration, the initial K-means or cluster centers are initialized at random, and the error function is used to track the total error associated with the given solution after the K-means algorithm converges. After the defined number of iterations, the solution with the smallest total error is selected.

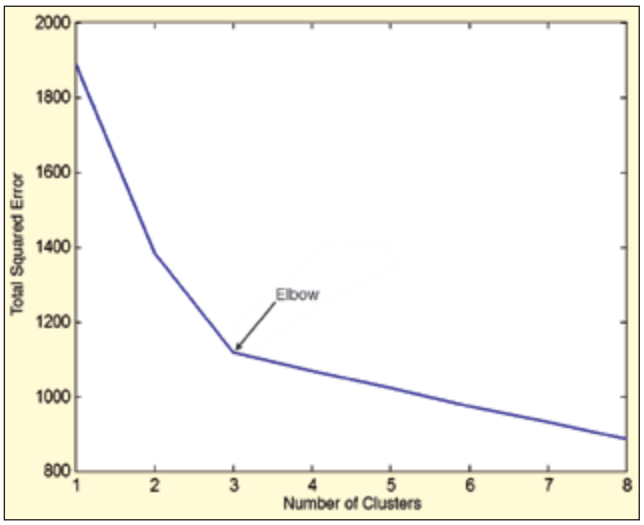

Figure 1. Number of clusters versus total squared error for a simulated jury (n=50) with three subgroups.

Although the approach of iterating the algorithm over a number of randomly selected initial K-means can be computationally intensive, especially for a large dataset, the dataset size of a typical jury study seldom exceeds 100 members.[1] This is relatively small considering that Jain, *et al.*,[13] discuss datasets that can reach millions of data points with thousands of features. So a typical jury result of less than 100 jury members (datasets) and 45 features (for paired comparison of 10 sounds) would generally be considered small. Therefore, an iterative approach would not be expected to demand an excessive amount of time.

Ward's Clustering Method. As described earlier, Ward's method[16] for clustering is also approached iteratively. In contrast to the K-means algorithm, Ward's method can be described as hierarchical agglomerative. Again there may be variations in the specific implementation, but in general, the approach is described in Reference 24:

Initialization: Initialize all points as m individual clusters (c_m).

Evaluate: For all possible pairings (c_{ij}) of the m clusters, find the pairing that results in the minimal within cluster variance,

$$c_{ij} = \arg\min_{\forall i,j \in m: i \neq j} \left\{ \sigma^2 \left(c_i \cup c_j \right) \right\} \tag{3}$$

where $\sigma^2\left(c_i \cup c_j\right)$ represents the variance of the combination of clusters c_i and c_j.

Merge: Form a set of m–1 clusters by combing clusters $\mathbf{c}_i$ and $\mathbf{c}_j$.

Iteration: Repeat the *evaluate* and *merge* steps until points belong to a single cluster.

Jury Clustering and Classification. An inverse scree test[9,10] is used to estimate the number of jury subgroups based on the total error for the K-means solution for 1 to n clusters. In this case, the total error is measured as the total squared distances between all data points and its assigned cluster center. As implemented in this study, the test is a subjective evaluation that identifies the "elbow" in comparing the total error plotted against the cluster number. To generate the scree plot, a K-means solution is found for 1 to n clusters, where n is a number that is expected to be much larger than the true number of clusters. An example of this is shown in Figure 1, where the elbow is clearly defined. The results in this plot were generated by a simulated jury with 50 jurors and three subgroups.

The second component of this jury clustering approach involves classifying the jury into the appropriate subgroups, which can then be used as an input into a preference model. The classification step is accomplished by reapplying the K-means algorithm using the number of clusters that corresponds to the number of jury

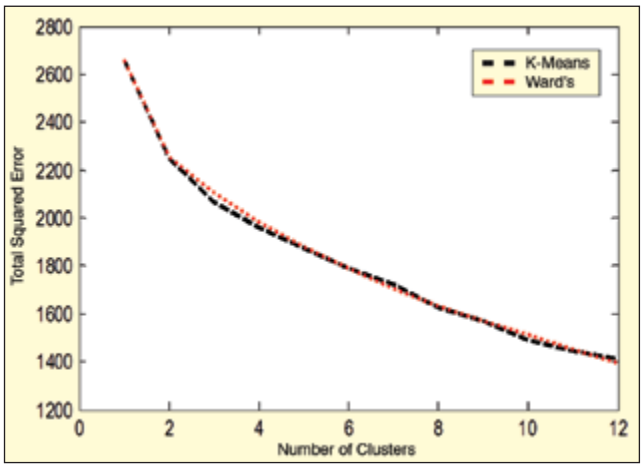

Figure 2. Scree plot for a small-engine jury study; K-means clustering solution is shown in dashed black and Ward's method is red.

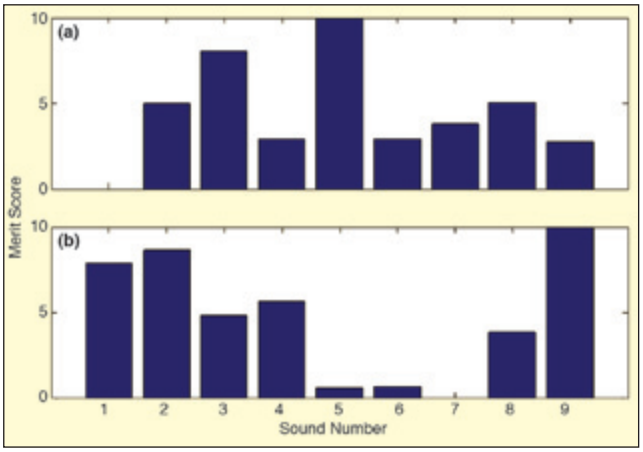

Figure 3. Merit scores for nine small-engine sounds based on (a) subgroup1 and (b) Subgroup 2 preferences.

subgroups identified by the scree test described above.

Small-Engine Jury. The two clustering algorithms were applied to real jury results from a small-engine sound quality jury study. During the study, nine small engine sounds were presented to more than 40 American and more than 40 European jurors. The jurors were informed that they would be listening to small-engine sounds and measured at a typical operator position during typical operation.

Each juror was presented with a pair of sounds, labeled Sounds A and B and asked to select the sound that they preferred. This comparison is done for all possible pairings, and so for a set of nine sounds, a total of 36 comparisons are made, and the output for that juror was a vector of length 36 consisting of values –1, 0, or 1 indicating prefering the first, no preference, or prefering the second sound. This is repeated for all jurors, and the "preference vectors" were used as the input to the jury-clustering algorithm applied in this study.

The K-means and Ward's scree tests for this jury are shown in Figure 2. This figure indicates that both methods again give nearly identical results and indicate that there are two subgroups within the jury. Based on this, the K-means solution was used to classify the jurors into two subgroups. The classification from the two subgroup K-means clustering solution is then compared to the known classification in a confusion matrix, shown in Table 1. The confusion matrix shows the percent of American and European jurors classified into Subgroups 1 and 2. In this case, the K-means clustering approach puts 98% of the American jurors and 12% of the European jurors into Subgroup 1, and 2% of the American jurors and 88% off the European jurors into Subgroup 2, based on their preferences. For reference, the subjective merit scores for the nine small engine sounds for the two subgroups are shown in Figure 3. It is not known if the European jurors classified into Subgroup 1 and the American jurors classified into Subgroup 2 are erroneously classified in that way or if they truly share preferences common with the other group.

Table 1. Confusion matrix showing percentage of American and European jurors classified into Subgroups 1 and 2.

	Subgroup 1	Subgroup 2
American	98	2
European	12	88

Conclusions

An approach to using a visual scree test in conjunction with a K-means clustering and Ward's agglomerative clustering algorithms is employed to estimate the number of jury subgroups and to classify each of the jury members into the appropriate subgroup based on voting preference.

References

1. N. Otto, S. Amman, C. Eaton, S. Lake, "Guidelines for Jury Evaluations of Automotive Sounds," *Sound & Vibration*, pp. 1-14, April 2001.
2. N. C. Otto, "Listening Test Methods for Automotive Sound Quality," *Audio Engineering Society 103rd Convention*, pp. 1-9, New York, September 1997.
3. T. C. Lim, "Correlations between Deficiencies in Power Window Systems Influencing Sound Quality and some Psychoacoustic Metrics," *Applied Acoustics*, Vol. 62, No. 9, pp.1025-1047, 2001.
4. T. Bernard, G. Cerrato-Jay, J. Dong, D. J. Pickering, L. Braner, R. Davidson, M. Jay, "The Development of a Sound-Quality-Based End-of-Line Inspection System for Powered Seat Adjusters," *SAE 2001 World Congress*, SAE Paper Number: 2001-01-0040, 2001.
5. E. Parizet, E. Guyader, V. Nosulenko, "Analysis of Car Door Closing Sound Quality," *Applied Acoustics*, Vol. 69, No. 1, 1pp. 2-22, 2008.
6. G. Cerrato, "Automotive Sound Quality – Powertrain, Road and Wind Noise," *Sound and Vibration*, pp. 16-24. April 2009.
7. H. David, *The Method of Paired Comparison*, Oxford University Press, New York, 1988.
8. J. Fry, P. Jennings, G. Dunne, R. Williams, "Jury Evaluations of Sound Quality Using Neural Networks," *4th International Conference on Recent Advances in Soft Computing (RASC2002)*, pp. 301-306, 2002.
9. R. G. Lathrop, J. E. Williams, "The Validity of the Inverse Scree Test for Cluster Analysis," *Educational and Psychological Measurement*, Vol. 50, No. 2, pp. 325-330, 1990.
10. R. G. Lathrop, J. E. Williams, "The Shape of the Inverse Scree Test for Cluster Analysis," *Educational and Psychological Measurement*, Vol. 49, No. 4, pp. 827-834, 1989.
11. J. A. Hartigan, "Clustering," *Annual Review of Biophysics and Bioengineering*, Vol. 2, pp. 81-102, 1973.
12. R. Xu, D. C. Wunsch II, *Clustering*, John Wiley & Sons, Hoboken NJ, 2009.
13. A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, Vol. 3, No. 8, pp. 651-666, 2010.
14. J. A. Hartigan, M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 28, No. 1, pp. 100-108, 1979.
15. C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, 1995.
16. J. H. Ward, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, Vol. 58, No. 301, pp. 236-244, 1963.
17. B. Funfgeld, M. Wang, "Attitudes and Behaviour in Everyday Finance: Evidence from Switzerland," *International Journal of Bank Marketing*, Vol. 27, No. 2, pp. 108-128, 2009.
18. S. D. Ross, "Segmenting Sport Fans Using Brand Associations: A Cluster Analysis," *Sport Marketing Quarterly*, Vol. 16, No. 1, pp. 15-24, 2007.
19. G. Punj, D. W. Stewart, "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research*, Vol. 20, No. 2, pp. 134-148, 1983.
20. D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*, Chapter 20, Cambridge University Press, Cambridge, 2003.
21. D. Steinley, "K-means clustering: A Half-Century Synthesis," *British Journal of Mathematical and Statistical Psychology*, Vol. 59, No. 1, pp. 1-34, 2006.
22. J. Xie, S. Jiang, W. Xie, "An efficient Global K-means Clustering Algorithm," *Journal of Computers*, Vol. 6, No. 2, pp. 271-279, 2011.
23. Mathworks, *Statistics Toolbox User's Guide, Version R2011b*, The Mathworks, Inc., Natick, MA, 2011.
24. W. H. E. Day, "Efficient Algorithms for Agglomerative Hierarchical Clustering Methods," *Journal of Classification*, Vol. 1, No. 1, pp. 7-24, 1984.
25. L. L. Thurstone, "A Law of Comparative Judgment," *Psychological Review*, Vol. 34, No. (4), 273-286, (1927).

The author can be reached at: gpietila1@hotmail.com.