World Academy of Science, Engineering and Technology
International Journal of Humanities and Social Sciences
Vol:1, No:12, 2007

Cluster Analysis for the Statistical Modeling of Aesthetic Judgment Data Related to Comics Artists

George E. Tsekouras, and Evi Sampanikou

Abstract—We compare three categorical data clustering algorithms with respect to the problem of classifying cultural data related to the aesthetic judgment of comics artists. Such a classification is very important in Comics Art theory since the determination of any classes of similarities in such kind of data will provide to art-historians very fruitful information of Comics Art's evolution. To establish this, we use a categorical data set and we study it by employing three categorical data clustering algorithms. The performances of these algorithms are compared each other, while interpretations of the clustering results are also given.

Keywords—Aesthetic judgment, comics artists, cluster analysis, categorical data.

I. Introduction

Aesthetic judgment is one of the most valued human characteristics, being regarded as a remarkable display of intelligence. Popular knowledge tells us that it is purely cultural [1]. Roughly speaking, Art and Aesthetics are two different fields but still they are stronlgy related with each other. Generally, the qualitative judgment of a painting artwork is based on two major issues[1, 2]: (a) the content, and (b) the form. The former is related to the subject that is represented by the artwork, while the latter is related to the visual aesthetic value by which the content is represented (i.e. color combination, composition, shape, symmetry, etc.). Despite the fact that these two concepts are different, their impacts to the final judgment are equal. Thus, it isn't appropriate to create a dichotomy between the content and form but yet we have to accept that these two issues are totally independent. Consequently, the assessment of an artwork rests on these two issues and their interactions. Since these two issues are subjective and appear to be different between individuals, we can assume that the way in which the content influences the value of a given painting strongly depends on cultural issues.

Based on the above analysis, we can fairly say that the aesthetic judgment of artworks involves three basic concepts namely, (a) the subjectivity, (b) the cultural influence, and (c) the qualitative evaluation. All these concepts are very difficult to be quantitatively measured. Despite this difficulty, there are some attempts in the literature, which try to incorporate statistical mathematical theory to the above concepts. Such kind of works can be found in [3, 4]. On the other hand, the use of Artificial Intelligence seems to be very efficient. Baluja et al [2] were the first who tried to incorporate Artificial Neural Networks in the evaluation of digitized paintings. In [5], a hierarchical fuzzy clustering scheme was used to clasiffy artist-painters, while in [1] Machado and Cardoso developed an artificial intelligence system, which is able to evaluate (itself) art paintings based on the Graves' drawing appreciation test [6].

In this work we use two categorical data sets that are related to aesthetic judgment data of the Greek comics artists. To elaborate the data we employed three well-known categorical data clustering procedures namely, the ROCK algorithm [7], the agglomerative hierarchical clustering (AHC) scheme developed in [8], and finally the fuzzy logic based approach proposed in [9]. The scope of this study is to use the above algorithms in order to determine any possible meanigful structures underlying the data.

The rest of the paper is sythesized as follows: In Section II we present a bibliographic report of various categorical data clustering procedures and moreover, we describe the three algorithms used in this paper. Section III presents the comparative simulation study of the above algorithms, where a qualitative interpretation of the results is provided, also. Finally, the paper ends with the concluding remarks in Section IV.

G. E. Tsekouras is with the Department of Cultural Technology and Communication, University of the Aegean, Faonos and Harilaou Trikoupi str., 81100, Mytilene, Greece (tel: +310-2251-0-36631, Fax: +301-2251-0-36609, e-mail: gtsek@ ct.aegean.gr).

E. Sampanikou is with the Department of Cultural Technology and Communication, University of the Aegean, Faonos and Harilaou Trikoupi str., 81100, Mytilene, Greece (e-mail: e.sampanikou@ct.aegean.gr).

II. Categorical Data Clustering

Categorical data clustering (CDC) has been investigated by many researchers. A common approach among the various CDC procedures is to use hierarchical clustering schemes, which are based on agglomerative clustering [10] or on the use of similarity [7] and disimilarity measures [8].

To reduce the computational complexity of a CDC algorithm, Huang [11] used a simple matching dissimilarity measure and introduced the *c*-modes algorithm, which is an extension of the classical *c*-means algorithm. In their later work [12], Huang and Ng generalized the *c*-modes approach by introducing the fuzzy *c*-modes.

ROCK [7] is a robust hierarchical categorical data clustering algorithm, which is based on a novel similarity measure and not distances when merging clusters. Its function is based on a bottom-up procedure, where two data objects are more similar if they have more common neighbors. Since the time complexity of the ROCK algorithm is quadratic, it clusters a rendomly sampled data subset. The optimal number of clusters is obtained by maximizing a certain objective function that involves the above similarity measure. Then, it partitions the entire data set based on the clusters determined in the previous step.

In [8], Mali and Sushmita developed an agglomerative hierarchical clustering (AHC) scheme, which uses a weighted dissimilarity measure to decide which categorical objects should be merged in each step. Simultaneously, in each step, applies a cluster validity index. The iterative process continues until it exists only one cluster. Finally, the algorithm selects the partition that corresponds to the lowest value of the validity index.

In [9] it was developed a three-step fuzzy logic based procedure for clustering categorical data. The first step introduces an entropy-based categorical clustering scheme, which is used as a data pre-processor unit to group the data set into a number of initial clusters. In the second step the algorithm applies the well-known fuzzy *c*-modes [12], while the third step utilizes a certain criterion to merge the most similar clusters.

A detailed description of the last three mentioned algorithms can be found in the respsective references.

III. Modeling Aesthetic Judgmenet Data

In this section we apply the above algorithms to discover meaninful structures in a categorical data set which includes aesthetic judgmenet data. In order to create the data set, two very experienced comics art-historians permormed an aesthetic judgment of 30 Greek comics artists using three but complex conceptual judgments (attributes). These conceptual judgments are as follows:

A_1={Emphasis on Color}
A_2={Emphasis on Design}
A_3={Emphasis on Script}

On each dimension (attribute) they scored the 30 artists on a scale between 0 and 5. Thus, each attribute consists of 6 categories. The available data are depicted in Table I. Moreover, the informed comics art-historians classified the artists according to their experience into four categories. Each category represents a certain school of tendencies and culture, by which the artists have been influenced. These categories are given at the most right column of Table I.

The implementation of the ROCK algorithm gave c=4 clusters. On the other hand, to apply the validity index of the AHC algorithm, we searched for an optimal solution, within the interval [2, 4]. The implementation of the AHC algorithm showed that the lowest value of the validity index corresponds

TABLE I
Greek Comics Artists' Data Set

Name	A_1	A_2	A_3	School
1. Kailatzis	0	4	5	B
2. D. Vitalis	5	5	3	A
3. K. Vitalis	5	4	2	A
4. Navrozidou	5	5	3	A
5. Soloup	3	3	4	B
6. Zervos	4	4	4	B
7. Trashman	0	4	4	B
8. Kyriazis	3	2	2	D
9. Lolos	4	3	2	D
10. Petrou	4	4	4	B
11. Gogtzilas	1	5	3	C
12. Papamichalopoulos	0	5	4	C
13. Verykios	5	5	4	A
14. Aronis	4	2	1	D
15. Derveniotis	3	3	4	B
16. Zafeiratos	3	4	4	B
17. Kioutsioukis	0	3	3	D
18. Milioris	3	3	3	D
19. Staboulis	4	4	2	C
20. Straticopoulou	4	4	1	C
21. Tragakis	3	3	1	C
22. Chrisoulis (KON)	4	2	3	D
23. Kollias	4	2	4	B
24. Chabidis	0	5	4	B
25. Papastamos	5	4	4	A
26. Dilios	4	3	3	D
27. Konstantinou	0	4	1	C
28. Koen	5	4	0	A
29. Papaioannou	1	2	2	D
30. Iovis	0	4	2	C

to: c=4. However, we noticed that as the number of clusters was increasing beyond c=4, the validity index was still reducing its value. This result, indeed, agrees with the results obtained in [3, 4]. Nevertheless, for the AHC algorithm, we accepted the number of clusters to be equal to c=4, as a reasonable choice. Finally, the application of the fuzzy logic-based approach indicated that the appropriate number of clusters is c=5.

The results of the ROCK, the AHC, and the fuzzy logic-based algorithm are shown in Table II, Table III, and Table IV, respectively. More specifically, these tables indicate the relationship between the obtained clusters and the schools.

It is difficult to speculate on these results without being an informed comics art-historian. The main difficulty of these data is their inherent qualitative information, which can be modeled very hardly. One thing which is apparent, however, in the three aforementioned tables is that the correspodence between the produced clusters and the schools is not fully established. This fact is more obvious in Table III, where it is easily seen that the results obtained by the AHC can establish only the correspodence between the school B and the cluster 1. The picture in Table II is totally different. By studying this table in more details we can see that the ROCK algorithm establishes a strong relation between school A and cluster 3, the school B and cluster 4, and the school D and cluster

1, while no information about the school C and the cluster 2 is provided.

TABLE II
CORRESPONDENCE BETWEEN THE SCHOOLS AND THE CLUSTERS PRODUCED BY THE ROCK ALGORITHM

	School			
Cluster	**A**	**B**	**C**	**D**
1	1		3	5
2		2	1	
3	4	1	2	2
4	1	6	1	1

TABLE III
CORRESPONDENCE BETWEEN THE SCHOOLS AND THE CLUSTERS PRODUCED BY THE AHC METHOD

	School			
Cluster	**A**	**B**	**C**	**D**
1	2	6		1
2	1	1	3	2
3	2	1	1	3
4	1	1	3	2

TABLE IV
CORRESPONDENCE BETWEEN THE SCHOOLS AND THE CLUSTERS PRODUCED BY THE FUZZY LOGIC BASED APPROACH

	School			
Cluster	**A**	**B**	**C**	**D**
1	2			2
2		7	1	
3	2			2
4	2	1	6	1
5		1		3

A similar but slightly inferior picture is given in Table IV. This table indicates that the school B can be efficiently described by the cluster 2, and the school C by the cluster 4. On the other hand, Table IV doesn't provide any information about the rest of the schools and the clusters.

Taking into account the certain difficulties involved in the specific data set, we can conclude that the ROCK algorithm significantly outperformed the other two algorithms, resulting in a more real correspodence between the resulting clusters and the schools under consideration.

IV. CONCLUSIONS

In this paper we presented a statistical modeling of cultural data related to the aesthetic judgment of comics artists. The main scope of the study was to cluster the available categorical data set and to establish relationships between the obtained clusters and the artists' schools. To maintain this we used three categorical data clustering techniques, the performances of which were compared with each other. The final results showed that only two of the three algorithms are able to partially establish such a kind of relationship.

REFERENCES

[1] P. Machado, and A. Cardoso, Computing aesthetics, *Lecture Notes in Artificial Intelligence*, 1515, 1998, 219-228.
[2] S. Baluja, D. Pomerlau, and J. Todd, Towards automated artificial evolution for computer-generated images, *Connection Science*, 6(2), 1994, 325-354.
[3] M. Davenport, and G. Studdert-Kennedy, The statistical analysis of aesthetic judgment: an exploration, *Applied Statistics*, 21, 1972, 324-332.
[4] B. S. Everitt, *Cluster analysis*, 3rd Edition, Arnold Publications, N.Y., 1993.
[5] G. E. Tsekouras, A. Kaoua, and E. Sampanikou, Potential-based fuzzy clustering and cluster validity for categorical data and its application in modeling cultural data, *3rd IEEE Conference on Computational Cybernetics*, Mauritious, May 2005, 73-78.
[6] M. Graves, *Test of Drawing Appreciation*, The Psychological Corporation, 1977.
[7] S. Guha, R. Rastogi, and K. Shim, ROCK: A robust clustering algorithm foa categorical attributes, *Information Systems*, 25(5), 2000, 345-366.
[8] K. Mali, & M. Sushmita, Clustering of symbolic data and its validation, *Lecture Notes in Artificial Intelligence*, 2275, 2002, 339-344.
[9] G. E. Tsekouras, D. Papageorgiou, S. B. Kotsiantis, C. Kalloniatis, and P. Pintelas, A fuzzy logic-based approach for detecting shifting patterns in cross-cultural data, *Lecture Notes in Artificial Intelligence*, 3533, 2005, 705-708.
[10] T. Morzy, M. Wojciechowski, and M. Zakrzewicz, Scalable hierarchical clustering method for sequences of categorical values, *Lecture Notes in Artificial Intelligence*, 2035, 2001, 282-293.
[11] Z. Huang, Extensions of the k-means algorithm for clustering large data sets with categorical values, *Data Mining and Knowledge Discovery*, 2, 1998, 283-304.
[12] Z. Huang, & M. K. Ng, A fuzzy k-modes algorithm for clustering categorical data, *IEEE Transactions on Fuzzy Systems*, 7(4), 1999, 446-452.