The Application Of Cluster Analysis In Ecnomics Science

Maohua Li[1, 2, a, *], Zéman Zoltán[3, b]

[1]School of Business, Xi'an Siyuan University, Xi`an, China.

[2]Faculty of Economics and Social Sciences, Szent István University, Gödöllő, Hungary.

[3]Institute of Business Studies, Faculty of Economics and Social Sciences, Szent István University, Gödöllő, Hungary.

[a]maohua.li@qq.com

*Corresponding author: Maohua Li

Key word: Cluster analysis, Comparison, Assignment.

Abstract. In order to learn how to apply cluster analysis in economics better, this paper uses comparison method to compare two articles to talk about the difference between them. First, we select two articles about the cluster analysis, they are "using cluster analysis to segment students based on self-reported emotionally intelligent leadership" and "using cluster analysis for market segmentation-typical misconceptions, established methodological weaknesses and some recommendations for improvement". For the comparison, we compare the data selection, methodology, variables, model construction, and analyses, interpretation.

Introduction

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics.

The notion of a "cluster" cannot be precisely defined, which is one of the reasons why there are so many clustering algorithms. There is a common denominator: a group of data objects. However, different researchers employ different cluster models, and for each of these cluster models again different algorithms can be given. The notion of a cluster, as found by different algorithms, varies significantly in its properties. Understanding these "cluster models" is key to understanding the differences between the various algorithms.

However, how to use cluster analysis is still a big problem. In order to help understand cluster analysis better, I use two cluster analysis papers as research sample to find the most important thing that I should pay more attention when I use the cluster analysis.

Firstly, I read the material about the cluster analysis, which is my theory preparation. Secondly, I read the two articles very carefully to understand all the parts of the two articles. Thirdly, I make a detailed comparison of them from data selection, methodology, variables, model construction, and analyses, interpretation. The total structure of my assignment can be seen as follows:
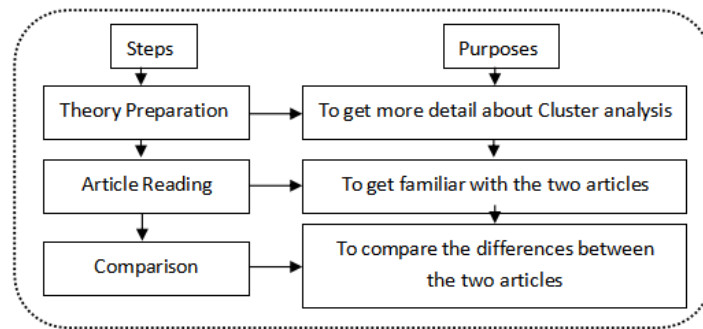
Figure 1. The total structure of my comparison

During the comparison progress, I compare the two articles "using cluster analysis to segment students based on self-reported emotionally intelligent leadership" (Article A) and "using cluster analysis for market segmentation-typical misconceptions, established methodological weaknesses and some recommendations for improvement"(Article B). The total comparison progress is shown on the following tale 1.

Table 1. The comparison progress of the Article A and Article B

Comparison steps	Comparison content
1st	Data selection comparison
2nd	Methodology comparison
3rd	Variable comparison
4th	Model construction comparison
5th	Analysis comparison
6th	Interpretation comparison
7th	Conclusion comparison

Data Selection Comparison

During the data collection progress, both article A and article B talk about the scouring of the data. Article A talks "A total of 566 students from 139 colleges and universities in the United States completed an online assessment of their leadership skills in the spring of 2009." Article B talks "The data set underlying this review consists of 243 publications in the area of business administration where data driven segments were identified or constructed".

Both article A and article B have a detailed descriptive analysis of the data that they use in the articles. And both of them use the table to make a very detailed description of the data.

The difference is that article A is more completed on the scouring of data. Article A uses the whole paragraph talking about the scouring of the data. However the article B puts the scouring of data with the selection of variables in on paragraph.

They have a huge difference in the number of the data and the year of the data they use. These can be seen in the table 2.

Table 2. The difference of two articles in number of and the year the data

	Article A	**Article B**
The number of the data	566	243
The year of the data	2009	2000
Pre processing of the data	No	Yes

From the table 2 above, we can see the difference between article A and article B in the number of the data and the year of the data. The number of data in article A is 566 and the article B is 243. The year of the data in article A is 2009 and the year of data in article B is 2000. The reason for this is the time of the author who considers writing the topic.

In fact, in the cluster analysis procedures, there is no need to do data pre processing. However, from the table 2, article B does pre processing to the data before the cluster analysis. The reason is that it is better to do the cluster analysis out of the market segmentation.

Methodology Comparison

During the methodology, both of them use the cluster analysis and descriptive analysis. And article A and article B use descriptive analysis in many aspects. The detail is contained in the table 3.

Table 3. The descriptive analysis used in article A and article B

	Article A	**Article B**
Number of usage	3	4
Area of usage	Sample question, Construct Mean Scores, Demographics of variables	Sample Size and Variable Number, plot of sample size and the number of variables, Frequency
Way of usage	Table	Table, Plot

From the table3, we can see that both articles use the descriptive analysis, however there are many difference between them. Article B use it more often than article A, and they use it in different areas. Article A uses it in sample question, construct mean scores, demographics of variables, but article B uses it in sample size and variable number, plot of sample size and the number of variables, frequency. The way they use it is also different. Article A uses it only in table, but article B uses it table and plot.

Of course, both of them use cluster analysis, but they use different ways to get clusters. The comparison results are shown in the table 4.

Table 4. The type of cluster analysis used in two articles

Type of cluster analysis	Article A	Article B
Two-step cluster analysis	Yes	No
K-means cluster analysis	No	Yes

From table 4, we can see easily that article A uses the two-step cluster and article B uses K-means cluster analysis. In fact, it is difficult to tell which one is better to do the cluster analysis. It depends on the different situation and research purpose that we choose which one is better.

Variable Comparison

Because the research topics of the two articles are totally different, there is an ocean difference between the variables. In the table 5 are the variables.

Table 5. The variables used in the two articles

No.	Variables
Article A	Gender, Ethnicity, Class rank, Age, Involvement, Leadership Role
Article B	Market segments, Market dimension etc.

Table 5shows us the difference of the variables selection between two articles. In the article A, there are 6 variables such as gender, ethnicity, class rank, age, involvement, leadership role. And there are more variables in article B than article A such as market segments, market dimension etc.

Analysis Comparison

During the analysis, both of the articles use SPSS to do the cluster analysis. They use the words to tell the progress of their analysis. The comparison can been seen in the table 6.

Table 6. Analysis comparison table

Notes	Article A	Article B
Description of method and progress	Yes	Yes
Descriptive analysis of the progress	Yes	Yes
Table showing the result	Yes	Yes
Figure showing the result	Yes	No

During this stage, article A is more detailed than article B. However both of them use some words and sentences to describe the methodology and the progress of their analysis. Article A gives more detail than article B. From the table 6, we can see that both of them have a description of method they use and give a detail about the progress of the analysis. But article A gives a more detail than article B. Both of them use descriptive analysis of the progress, and both of them talk about this very carefully.

Article A uses more ways to show the results of the analysis, and it uses many tables, figures. However article B only uses some tables to show the results of its analysis. In this case, article A is better than article B.

Interpretation Comparison

During the interpretation, both of them have a good interpretation of the results that they get from the analysis. However, article B is better than article A.

The comparison of them can be seen in the table 7.

Table 7. Comparison of interpretation

Notes	Article A	Article B
Interpreting according to the analysis	Yes	Yes
Interpreting according to the table or figure	Yes	Yes
Interpreting in the numerous	Yes	Yes

From the table 7, we can see that both of them give a very detailed description of the results. They both interpret according to the analysis, table and the figure. They both interpret in the numerous. However, article B is more detailed than article A.

Conclusion Comparison

On the stage of conclusion, both of their conclusions are not of high quality. Article B is better than article A. We can judge their conclusion in the following table 8.

Table 8. The conclusion judgments of the two articles

No.	New finding	Advise	Discussion of methodology	Discussion of research limits	Future research
Article A	√	√	X	X	X
Article B	√	√	√	X	√

From the table 8, we can see that the conclusion of article A is really very short and only in one paragraph. It talks about the new finds, a little advice, and it does not talk about the methodology, research limits and future research. On the contrast, article B is better, it talk more detail and get a better conclusion. During the three paragraphs of conclusion, article B talks more detail about the new findings, advices, research methodology, and future research.

However, there are some things in common for the two articles. For instance, they neither talk about the research limits.

Conclusion of My Comparison

The term cluster analysisencompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. A general question facing researchers in many areas of inquiry is how to organize observed data into meaningful structures, that is, to develop taxonomies. In other words cluster analysis is an exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. Given the above, cluster analysis can be used to discover structures in data without providing an explanation/interpretation. In other words, cluster analysis simply discovers structures in data without explaining why they exist.

We deal with clustering in almost every aspect of daily life. For example, a group of diners sharing the same table in a restaurant may be regarded as a cluster of people. In food stores items of similar nature, such as different types of meat or vegetables are displayed in the same or nearby locations. There is a countless number of examples in which clustering plays an important role. For instance, biologists have to organize the different species of animals before a meaningful description of the differences between animals is possible. According to the modern system employed in biology, man belongs to the primates, the mammals, the amniotes, the vertebrates, and the animals. Note how in this classification, the higher the level of aggregation the less similar are the members in the respective class. Man has more in common with all other primates (e.g., apes) than it does with the more "distant" members of the mammals (e.g., dogs), etc. For a review of the general categories of cluster analysis methods, see Joining (Tree Clustering), Two-way Joining (Block Clustering), and k-Means Clustering. In short, whatever the nature of your business is, sooner or later you will run into a clustering problem of one form or another.

There are wide areas that cluster analysis can be used. For instance, areas in the followings:

-Plant and animal ecology

Cluster analysis is used to describe and to make spatial and temporal comparisons of communities (assemblages) of organisms in heterogeneous environments; it is also used in plant systematics to generate artificial phylogenies or clusters of organisms (individuals) at the species, genus or higher level that share a number of attributes.

-Transcriptomics

Clustering is used to build groups of genes with related expression patterns (also known as coexpressed genes) as in HCS clustering algorithm . Often such groups contain functionally related proteins, such as enzymes for a specific pathway, or genes that are co-regulated. High throughput

experiments using expressed sequence tags (ESTs) or DNA microarrays can be a powerful tool for genome annotation, a general aspect of genomics.

-Sequence analysis

Clustering is used to group homologous sequences into gene families. This is a very important concept in bioinformatics, and evolutionary biology in general. See evolution by gene duplication.

-High-throughput genotyping platforms

Clustering algorithms are used to automatically assign genotypes.

-Human genetic clustering

The similarity of genetic data is used in clustering to infer population structures.

Reference

[1] Facca T M, Allen S J. Using cluster analysis to segment students based on self-reported emotionally intelligent leadership behaviors[J]. Journal of Leadership Education, 2011, 10(2): 72-182.

[2] Dolnicar S. Using cluster analysis for market segmentation-typical misconceptions, established methodological weaknesses and some recommendations for improvement[J]. 2003.

[3] Erdogmus N, Esen M. Classifying Universities in Turkey by Hierarchical Cluster Analysis[J]. Egitim ve Bilim, 2016, 41(184).

[4] Edwards M H, Robinson D E, Ward K A, et al. Cluster analysis of bone microarchitecture from high resolution peripheral quantitative computed tomography demonstrates two separate phenotypes associated with high fracture risk in men and women[J]. Bone, 2016, 88: 131-137.

[5] Tomassen P, Vandeplas G, Van Zele T, et al. Inflammatory endotypes of chronic rhinosinusitis based on cluster analysis of biomarkers[J]. Journal of Allergy and Clinical Immunology, 2016, 137(5): 1449-1456. e4.

[6] Maohua, L., & Zéman, Z. Study on the SRID Evaluation Framework of Agricultural Enterprises in China[J]. Visegrad Journal on Bioeconomy and Sustainable Development, 2016, 5(1): 36-40.

[7] Maohua, L., & Zéman, Z. The application of AHP in SRID evaluation framework of Chinese agricultural enterprise. Hungarian agricultural engineering, 2016, 30(2): 17-27.