HCA TECHNIQUES APPLIED TO FINANCIAL ANALYSIS

Stefan-Alexandru Ionescu, Ph.D.
The Romanian Academy, Romanian-American University, Bucharest

Abstract

Cluster analysis can be regarded as an instrument whose *purpose is the reduction of object sets*, or even of variables, to a smaller number of information entities, which are the classes or the clusters. Nevertheless, although cluster analysis, viewed as a collection of methods and techniques used for the *classification of objects*, is applied in the *space of variables*, we notice the frequent use of these techniques of analysis for the classification of objects. In this paper we have shown how these techniques can be applied in the economic-financial field and how the number of classes in which companies can be divided can be detected by observing the existing latent structure.

1. Introduction

Tyron (1939) is the first to use the term cluster analysis, and Sokal and Sneath (1963) and Lance and Williams (1967) introduce the first rigorous research in this field. In the following years, the contributions aiming to perfect this analysis multiplied and became extremely diversified; out of these, we can detach two important scientific trends, represented by *the American school* and *the French school*.

The aim of cluster analysis is to identify, inside a set of objects or forms, *classes, groups* or *clusters* with elements as *similar* as possible inside one given class (*minimum variability inside a class*) and as dissimilar as possible if these elements belong to different classes (*maximum variability among classes*). What results is that cluster analysis allows the examination of *similarities* and *dissimilarities* among the objects belonging to a certain set, to the purpose of grouping them under the form of *distinct* and *homogeneous* classes. Every object in the analyzed set is attributed to a single class, and the set of all classes is a discreet and unordered set. The classes or groups in the form of which the sets of objects are structured are called clusters.

Hierarchical Clustering is considered to be a system of ***unsupervised recognition***, as the process of classification starts without any information available regarding the number of classes and the belonging of forms to these

classes. In this case, the classes are built as the number of analyzed forms grows, the number of potential classes being determined at the end of the process of recognition. Some uncontrolled classification algorithms, such as *partitioning algorithms*, entail the prior establishment of the number of classes in which the analyzed objects will be divided. This does not mean that the number of classes is in fact known, but an assumption is made regarding this number.

The systems of uncontrolled recognition of forms resort to principles, methods, procedures and techniques known, in specialized literature, as *classification techniques*, *unsupervised classification* or *cluster analysis*.

Cluster analysis, as we will call it in what follows, entails the organization of forms or objects in clusters or groups progressively, without prior knowledge of the number of classes, while observing *two fundamental criteria*:

a) Every class should be as homogeneous as possible, namely it should contain objects or forms that are as similar as possible in relation with the characteristics taken into consideration for object classification;

b) Every class needs to contain classified objects that are as different as possible, from the point of view of the classification characteristics, from the objects included in any of the other classes.

According to the characteristics of the procedures that are being used, to the initial hypotheses and to the nature of results, the methods of hierarchical clustering can be divided into the following types of clustering:

- *Agglomerative Clustering* and
- *Divisive Clustering*.

The typical procedures in this case are: *the single-linkage clustering method*, *the complete-linkage clustering method*, *the **U**n-weighted **P**air **G**roup **M**ethod with **A**rithmetic Mean method*, *Ward's method* etc.

In the case of the analysis of a big amount of data characterized by a high degree of heterogeneity, the unsupervised recognition systems are rather used for the purpose of systematization, grouping and informational synthesis. Since these techniques, based on the use of the concept of distance, are useful and efficient in preliminary data analysis, they allow a more efficient organization of heterogeneous data, as well as easy and consistent search and interpretation of information within the framework of data structured in this way.

2. Hierarchical Clustering

The aim of ***hierarchical methods*** is to produce more cluster solutions, called *cluster hierarchies*. The main characteristic is given by the fact that the number of clusters is neither known beforehand, nor is such a parameter suggested by the user.

Cluster hierarchies are cluster structures with a variable number of clusters, of a *multilevel type*, which are differentiated through the number of clusters that they include and through their degree of agglomeration. Thus, having T objects, we will have T cluster solutions, each solution containing progressively bigger clusters, respectively, clusters with *progressively higher agglomeration levels*. A cluster hierarchy has a structure of the following type:

level 0: $\omega_1^{(0)}, \omega_2^{(0)}, \dots, \omega_{k_0}^{(0)}$

level 1: $\omega_1^{(1)}, \omega_2^{(1)}, \dots, \omega_{k_1}^{(1)}$

level 2: $\omega_1^{(2)}, \omega_2^{(2)}, \dots, \omega_{k_2}^{(2)}$

...

level T-1: $\omega_1^{(T-1)}, \omega_2^{(T-1)}, \dots, \omega_{k_{T-1}}^{(T-1)}$ (1)

where K_i is the number of clusters from the cluster solution at the "*i*" level.

Since the banal-type cluster solution, represented by the list of objects under classification, is the first partition, what results is that the possible number of solutions from a cluster structure obtained with the help of hierarchical algorithms will be smaller by 1 than the number of objects. This number is given by the following relationship:

$$N_s = T\text{-}1 \quad (2)$$

The choice of the most suitable cluster solution, from the *T-1,* is made according to the objectives of the analysis.

Two categories of hierarchical classification algorithms are known:

- *Agglomerative algorithms*. In the case of these algorithms, the number of clusters from the first partition is equal to the number of objects, namely $K_0 = T$. Also, the number of clusters from a partition at a certain level is smaller by 1 than the number of clusters from the partition positioned at the inferior level, and bigger by 1 than the

number of clusters from the partition situated at the superior level, respectively:

$$K_{T-1} = K_{T-2} - 1 = K_{T-3} - 1 = \cdots = K_2 - 1 = K_1 - 1 = K_0 - 1 \qquad (3)$$

- *Divisive algorithms*. These methods practically consist in the same operations as those used by agglomerative algorithms, but in reverse order. Thus, the first considered partition is represented by a single cluster that contains all the objects, the second partition will consist in two clusters, and so on.

The methods of hierarchical classification are considered *heuristic methods*, which comprise classification procedures that have been developed based on a certain intuitive manner of solving particular problems (*heuristics*).[1]

Among these methods we can mention: *the single-linkage clustering method*, *the complete-linkage clustering method*, *the* ***UPGMA*** *method*, *Centroid linkage clustering method, Ward's method* etc.

The Ward distance between two clusters measures the *cumulated intra cluster variability*, *induced by the linkage of two clusters* at the level of the resulting cluster configuration. The aim of linking two clusters is maximum homogeneity *at the level of all clusters* that belong to a given configuration of objects by clusters.

What results is that the Ward distance is the only one that takes into account the minimization of intra cluster variability or, in other words, the maximization of inter cluster variability, i.e. of the degree of cluster homogeneity. We need to specify that cluster homogeneity is being maximized as a result of the minimization of the total sum of the squares of intra cluster deviations.

If w_{12} is the new cluster obtained by the linkage of cluster w_1 with w_2, then the sums of intra cluster distances will be:

$$\begin{aligned} SSE_{w_1} &= \sum_{i=1}^{T_{w_1}} (y_i - \bar{y}_{w_1})^t (y_i - \bar{y}_{w_1}) \\ SSE_{w_2} &= \sum_{i=1}^{T_{w_2}} (y_i - \bar{y}_{w_2})^t (y_i - \bar{y}_{w_2}) \\ SSE_{w_{12}} &= \sum_{i=1}^{T_{w_{12}}} (y_i - \bar{y}_{w_{12}})^t (y_i - \bar{y}_{w_{12}}) \end{aligned} \qquad (4)$$

1. Heuristics are rules based on theoretical reasoning or on statistical observations.

We will link those two clusters w_1 and w_2 which minimize the increase of the sum of squares of errors defined as:

$$I_{w_{12}} = SSE_{w_{12}} - (SSE_{w_1} - SSE_{w_2}) \quad (5)$$

Ruxanda (2009) looks at the stages of cluster analysis for the classification of an object set, which he considers to be the following:

- *The choice of the characteristics* according to which the classification will be performed;
- *The choice of the type of measure* for the evaluation of proximity among objects;
- *The establishment of rules for the formation of classes or clusters*;
- *Building the classes,* i.e. distributing the objects into classes;
- *The verification of the consistency and significance* of the classification;
- *The choice of an optimum number of clusters*, according to the nature of the classification matter and to the objectives at hand;
- *The interpretation of cluster significance*.

Therefore, cluster analysis is an attempt to identify, in initial data, groups, classes or clusters according to the similarities and dissimilarities existing among the objects that the respective data refers to. As far as the technique that is being used is concerned, cluster analysis for object classification evaluates the distances for pairs of objects, and cluster analysis for the classification of variables evaluates distances for pairs of variables.

3. Data used in the analysis

101 firms that unfold their activity in Romania have been selected. The firms have been active and have submitted at least one credit application. This implies the fact that they also submitted all their financial statements on the 31st of December.

This sample group that we have chosen is representative for Romanian private companies that are not listed at the stock exchange. The value of the assets is between 15.000 and 30 million lei (not above the latter). Obviously, most firms are medium sized, with assets between 1 and 4 million lei. Small and big size firms appear in similar proportions in the sample group under analysis.

As specified before, primary data has been extracted from the balance sheets, profit and loss accounts submitted at the end of the year, as well as from the balances corresponding to the month of December.

Firstly, we took into account:

- assets, as well as their classification;
- debts, divided as well into categories, including those to banks and leasing companies;
- capitals and equity capitals;
- data connected with the turnover, profit, taxes and duties.

Subsequently, we took this data and we calculated a series of financial rates that offer a high degree of comparability for firms of various sizes and from various fields of activity. I was mainly interested in covering four directions, namely: liquidity, solvability (risk), activity and profitability. From the multitude of existing rates, I have chosen to take into account eight of them, namely:

- Profitability Ratios***:*** Return on Assets ***(ROA),*** Return on Equity ***(ROE),*** Return on Capital Employed ***(ROCE).***
- Efficiency ratios***:*** Total Assets Turnover ***(RAT).***
- Liquidity Ratios***:*** Current ratio ***(CR),*** Quick ratio ***(QR),*** Cash Ratio ***(CashR).***
- Solvability Ratios***:*** General Solvability ***(SP).***

Data processing has been done with the *STATISTICA 8.0* program package.

4. Results

The classification method that we have introduced is connected with hierarchical cluster analysis. As we have shown above, through this type of analysis we group the objects, in our case the 101 firms, based on the measurement of distances or similarities among them. We have taken into consideration the firms described with the help of the eight variables that we introduced previously. Such a method of amalgamation starts from the 101 clusters, represented by all the firms, which are to be linked progressively, relaxing the grouping criterion until it comes to one single cluster that contains all the objects. A desired number of clusters is not required as an input, the grouping occurs naturally and the user can observe the number of classes that appear.

In the first phase, we have calculated the distances among the 101 objects. To exemplify, we have in table 1 the distances among the first 10 firms.

City-block type distances among the first 10 objects

Table 1

	1	2	3	4	5	6	7	8	9	10
1	0.0000	7.9698	2.3494	2.9692	4.7642	5.4116	7.9730	4.0441	4.8960	7.3244
2	7.9698	0.0000	6.3338	7.0238	7.3226	3.8755	3.0325	7.4861	4.1965	7.9186
3	2.3494	6.3338	0.0000	2.4776	4.0079	3.1626	7.1092	3.2650	3.4434	5.7815
4	2.9692	7.0238	2.4776	0.0000	4.3906	5.5837	7.9746	3.6462	3.2889	7.1048
5	4.7642	7.3226	4.0079	4.3906	0.0000	5.5027	8.3231	3.9546	4.0691	2.7428
6	5.4116	3.8755	3.1626	5.5837	5.5027	0.0000	4.4133	5.1345	3.0615	5.4216
7	7.9730	3.0325	7.1092	7.9746	8.3231	4.4133	0.0000	8.9225	5.1529	8.5541
8	4.0441	7.4861	3.2650	3.6462	3.9546	5.1345	8.9225	0.0000	4.2009	4.8248
9	4.8960	4.1965	3.4434	3.2889	4.0691	3.0615	5.1529	4.2009	0.0000	4.8694
10	7.3244	7.9186	5.7815	7.1048	2.7428	5.4216	8.5541	4.8248	4.8694	0.0000

We have considered the eight-dimensional space in which we have calculated the city-block type distances. The choice was determined by the fact that this type of distance does not amplify the coordinate differences through exponentiations, thus proving more robust in relation with the presence of aberrant values in the data.

Distances appear in the form of a symmetric matrix, in which the *(i,j)* element shows the Manhattan distance between the *i* firm and the *j* firm in the eight-dimensional space defined by the eight variables. It is obvious that the elements composing the main diagonal are equal to 0, as they represent distances among objects for which $i=j$. The matrix is symmetric, i.e.: $d(i,j)=d(j,i)$. Thus, the distance between firm *1* and firm *2* is 7.9698 in the eight-dimensional space, the distance between firms *1* and *3* is 2.3494 in the same space, and so on.

I have tried to use more amalgamation methods, the one that has given the most satisfactory results being Ward's method. Through this method clusters are formed so that, at every step, the distribution of an object into a cluster minimizes variance inside the cluster.

Amalgamation program through Ward's method

Table 2

Iteration No.	Manhattan Distance	Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6
1	0.26844	27	46				
2	0.425123	56	76				
3	0.588868	57	91				
4	0.611369	31	44				
5	0.679478	27	46	51			
6	0.681315	17	96				
7	0.745417	16	77				
8	0.761216	60	81				
9	0.809053	87	94				
10	0.811932	55	57	91			
11	0.854273	12	86				
12	0.937787	87	94	92			
13	0.999241	63	82				
14	1.011493	15	53				
15	1.050489	3	24				
16	1.057668	16	77	99			
17	1.085942	38	90				
18	1.102812	29	32				
19	1.145189	39	66				
20	1.171946	34	35				
21	1.178349	42	78				
22	1.258784	61	75				
23	1.272284	47	93				
24	1.275334	21	101				
25	1.288252	49	62				
26	1.29384	22	30				
27	1.320899	1	29	32			
28	1.338401	22	30	43			
29	1.342997	5	95				
30	1.408685	8	54				
31	1.439288	56	76	87	94	92	
32	1.458543	28	36				
33	1.462615	58	60	81			

In table 2 we have exemplified the first 33 stages of agglomeration. Initially, there are 101 clusters, each containing one of the 101 firms. The smallest distance between two firms is 0.2684397. The first step of the amalgamation is represented by the formation of a cluster from these two objects. Hence, as a result of the first iteration, we will have 100 clusters: one formed by firms 27 and 46 and other 99 clusters formed by the other 99 firms. The next step is grouping

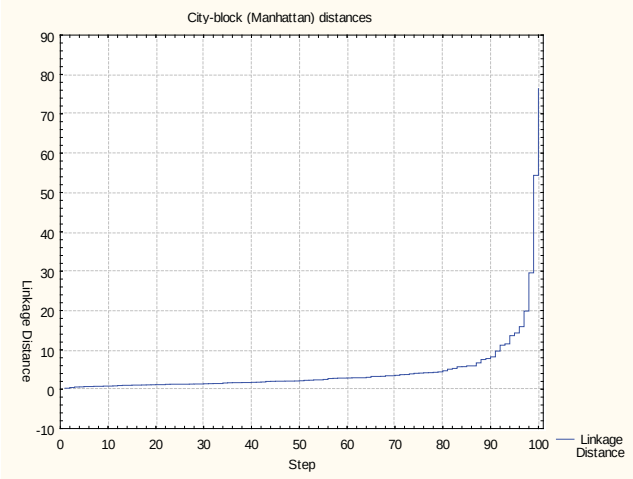
firms 56 with 76 between which there is a distance of 0.4251230. As a result of this iteration, we have 99 clusters left: the one formed at the first iteration (made up of firms 27 and 46), the one that resulted at the second iteration (formed by firms 56 and 76), as well as other 97 clusters consisting in the remaining firms. The process goes on in a similar manner.

The fifth step is assigning object 51 to the cluster that was already formed at the first step. Hence, what appears is a cluster formed by 3 firms, namely 27, 46 and 51. Every step, the sum of the squares of the deviations at the level of the newly formed cluster is the smallest in comparison with other pairs of potential clusters. At the 31st iteration, two previously-formed clusters unite in a bigger cluster. Hence, the 1.439288 distance between the cluster formed at the second iteration (made up by firms 56 and 76) and the one formed at the twelfth iteration (made up by firms 87, 94, 92) allows their linkage into a new cluster that will contain all these 5 firms. As a result of the one hundredth iteration, all the 101 firms will form a single cluster.

The distances from the first column of table 2 are represented on the *Oy* axis in chart 1. On the *Ox* axis we have the 100 iterations. Corresponding to the first iteration, we start with one point, at the level 0.2684397 on *Oy*. Corresponding to the second iteration, we draw a segment of a line, parallel with the *Oy* axis, between values 0.2684397 and 0.425123, and so on until we reach the last iteration. In each case, the superior extremity of the segment of the line corresponding to the *i* iteration gets unified with the inferior extremity of the segment of the line corresponding to the *i+1* iteration.

Graph showing agglomeration distances

Chart 1



Dendrogram of amalgamation for Ward's method

Chart 2

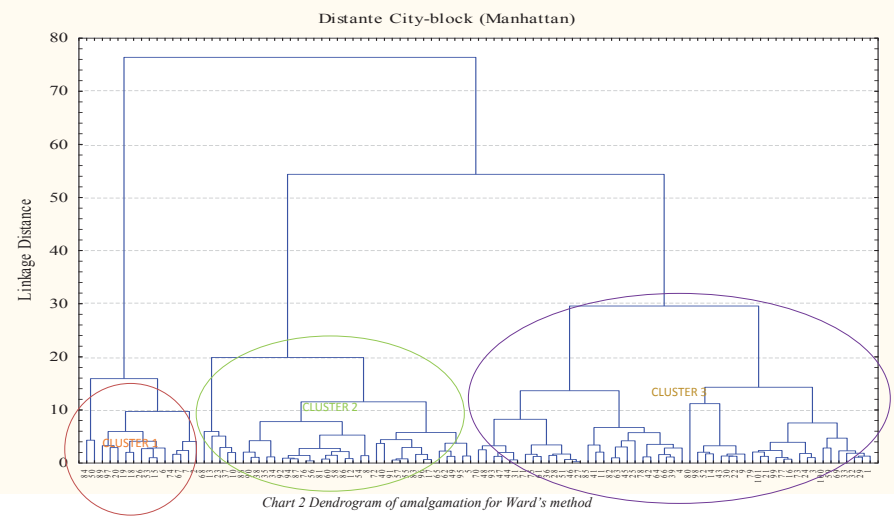


Chart 2 Dendrogram of amalgamation for Ward's method

This graph can be useful as it suggests visually where the clustering process should end naturally. As we go farther right, the distance among objects increases (the length of line segments becomes greater), bigger clusters are formed, and intra cluster variance is greater. In the first phase we notice a slow evolution, up to step 80, the increase of the distance being very small. What follows is bigger increases of distances up to step 98, the last two stages consisting in the linkage of objects with very big distances. If the distance among the objects linked on the first step is 0.2684397, the distance among the objects linked on the one hundredth step is 76.4076, namely 285 times bigger. Since the amalgamation distance from step *i* is greater than the amalgamation distance from step *i-1* (irrespective of *i*), we can say about the method that we have chosen that it fulfils the monotonicity condition and that it is ultrametric. Distance can be an optimum criterion in establishing the number of clusters that are to be kept.

The formation of 3 natural clusters is evident in chart 2 as well, where the hierarchical tree is presented. From stage 98 to stage 99 the distance almost doubles, which represents an unnatural linkage. I thus suggest keeping 3 clusters, as they are marked in chart 2.

5. Conclusions

Cluster analysis differs fundamentally from statistical procedures, in that it neither relies on nor does it entail a priori fulfillment of any specific hypothesis. Rencher (2002) considers that cluster analysis constitutes an important and efficient tool of *exploratory analysis*, the purpose of which is that of creating the so-called *taxonomies or typologies* based on the analysis of *similarities* and *dissimilarities* existing among the objects of a given set.

Cluster analysis is useful in any process of data analysis, not only in those that need a classification. For instance, in the case of a process that has in view the analysis of big amounts of data, both from the point of view of the analyzed objects and from the perspective of their characteristics, the synthesis and structuring of information can be done by resort to adequate instruments. Thus, in order to identify some categories, classes or information groups while working with a big amount of unprocessed information, cluster analysis can be used successfully.

Cluster analysis allows the inference of the evolutionary laws of some populations of phenomena, as well as of the principles of the process of knowledge, through:

- *the definition of formal classification schemes* and of *typologies,* in view of knowing and understanding better complex realities;
- *the identification of statistical-mathematical models* for the understanding, synthesis and simplification of complex and heterogeneous sets of phenomena and processes;
- *a more correct and comprehensive definition of fundamental characteristics* of populations of phenomena and processes;
- *deriving adequate numerical measures for the characterization of the dimensions of populations* of phenomena and in order to highlight the changes taking place in their structure;
- *the identification of individual entities representative* for complex classes and categories of phenomena and processes.

Acknowledgment:

This work was financially supported through the project *"Routes of academic excellence in doctoral and post-doctoral research - READ"* co-financed through the European Social Fund, by Sectoral Operational Programme Human Resources Development 2007-2013, contract no POSDRU/159/1.5/S/137926.

Bibliography

1. Aggarwal, C., & Yu, P. (2000). Finding generalized projected clusters in high dimensional spaces. *Proc. 2000 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'00)*, (pp. 70-81). Dallas, USA.
2. Arabie, P., Hubert, L., & De Soete, G. (1996). *Clustering and Classification.* New York, USA: World Scientific.
3. Back, A.D.; Weigend, A.S. Discovering Structure in Finance Using Independent Component Analysis; (1998) *Advances in Computational Management Science* Volume 2, 1998, pp 309-322
4. Beil, F., Ester, M., & Xu, X. (2002). Frequent term-based text clustering. *Proc. 2002 ACM SIGKDD Int. Conf. Knowledge Discovery in Databases (KDD'02)*, (pp. 436-442). Edmonton, Canada.
5. Bradley, P., Fayyad, U., & Reina, C. (1998). Scaling clustering algorithms to large databases. *Proc. 1998 Int. Conf. Knowledge Discovery and Data Mining (KDD'98)*, (pp. 9-15). New York, USA.
6. Chen, K.H. and Shimerda, T.A. An Empirical Analysis of Useful Financial Ratios, (1981), *Financial Management* Vol. 10, No. 1 (Spring, 1981), pp. 51-60
7. Dieckmann, S., Plank, T., Default Risk of Advanced Economies: An Empirical Analysis of Credit Default Swaps during the Financial Crisis, (2012), *Review of Finance* (2012) 16 (4):903-934.doi: 10.1093/rof/rfr015
8. Hastie, T; Tibshirani, R; Friedman, J (2009). "14.3.12 Hierarchical clustering". *The Elements of Statistical Learning* (PDF) (2nd ed.). New York: Springer. pp. 520–528. ISBN 0-387-84857-6. Retrieved2009-10-20.
9. Jain, A.K, (1999,). Data Clustering: A Review,. *ACM Computing Surveyes (CSUR), 31*, 264-323
10. Kaufmann, L., & Rousseuw, P. (2005). *Finding Groups inData: An Introduction to Cluster Analysis.* New York, USA: John Wilwy & Sons.
11. Lance, G., & Williams, W. (1967). A general theory of classificatory sorting strategies. *Computer Journal, 9*
12. Liu, B., Xia, Y., & Yu, P. (2001). Clustering through decision tree construction. *Proc. 2000 ACMCIKM Int. Conf. Information and KnowledgeManagement (CIKM'00)*, (pp. 20-29). McLean, USA.
13. Sokal, R., & Sneath, P. (1963). *Principles of numerical taxonomy.* San Francisco, USA: W.H. Freeman Co.
14. Rencher, A. (2002). *Methods of Multivariate Analysis.* New York, USA: John Wiley & Sons.
15. Ruxanda, G. (2001). *Analiza Datelor.* Bucureşti: ASE.
16. Ruxanda, G. (2010). Construirea, estimarea şi implantarea software a metodelor matematice. *Cercetarea ştiinţifică în ASE*.
17. Tyron, R. (1939). *Cluster Analysis.* Ann Arbor, USA: Edwards Brothers
18. Zhang, et al. "Graph degree linkage: Agglomerative clustering on a directed graph." 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012