MODEL AND ALGORITHM FOR HIERARCHICAL ANALYSIS

Dobromir Bonacin[1], Danijela Bonacin[1] and Žarko Bilić[2]

[1] *Faculty of Education, Travnik, Bosnia & Herzegovina*
[2] *Faculty of Sciences, Mostar, Bosnia & Herzegovina*

Original scientific paper

Abstract
The aim of this research was to define initial mode, i.e. paradigm for hierarchic identification in arbitrary set of data. Methodological design was organized based on identification of representative object in a total set. Results indicated that relations are in order naturally according to the hierarchic structure and can be recognized. The model limitations are not familiar because hierarchic data structuring, based on internal relations, occurred during many testing situations. Model originality is certain, since its appliance is literally unlimited, and the value of it lies in the building of new foundations towards relation determination based on internal consistency and relations.

***Key words**: hierarchy, internal consistency, extremes*

Introduction and problem

The issue of determining objectively present hierarchic structure within any defined set of data has been known in science for a long time (Bonacin et al., 2003; Espinosa, 2007; Ané, 2008). According to standard consideration, hierarchy relation determination is directly connected with distance of the object in multivariate space, which is being solved with different methods that belong to different models. Some of them are taxonomization models – multidimensional scaling, clustering, polar (Bonacin, 2004b). However, as it is well known, conduction of taxonomization can not be objective for two, not so small, limitations. The first limitation relates to the need to define (exterior) criteria depending on the purposes for which this procedure is being conducted for, which certainly leads to acceptable solutions that are not universal or have such intentions. Second limitation is defined with inconvincible determination of existing taxons correct number in a larger initial data entrance, which definitely leads to different solutions regarding individual researchers' affinity, so in the same conditions we obtain a different configuration which often leads to confusion in concluding. For this two reasons, assurance and discovery strength of such procedures is largely contaminated with subjective settings, which means there is no convincing model that can be considered as stable in different conditions. Bonacin et al., (2009) suggested Uditax taxonomization based on type extremization the taxon is formed around and this procedure is currently the most convincing model for distinct taxon optimization, based on object affiliation to one and only one taxon (type).

Model of polar taxons is party acceptable (e.g. Momirović et al., 1987), but since that is the procedure that allows non null object allocations on more then one taxon and basically it approaches to protocol of latent dimension determination in the space of variables, so it can not be completely considered as taxonomization. In any case, these two models can be accepted as the best procedures for group determination within some more or less compact sample. Existing and much applied procedures of multidimensional scaling, clustering based on criteria and other procedures (Veldman, 1967; Johnson et al., 1992; Tingquan et al., 1995; Bonacin, 2002; Kang et al., 2009) very often show very bad attributes. The most frequent is familiar "drawing" of objects on very small number of taxons, or even on only one derived taxon, which, inspection of initial data proves to be inconclusive and scientifically unsustainable (Gallizo et al., 2006; Jin et al., 2008). On the other side, there are almost no solid procedures intended for hierarchic modelling, especially if we exclude, for that purpose unacceptable factor procedures, which proves there is almost no recording on serious determination of objectively existing levels of hierarchic relations and imagined object/group allocations with arranged relations in scientific literature. For all mentioned reasons there is a need to establish logical paradigm that will define hierarchic analysis procedures, so we would equally support this important domain with solid system tools. Therefore, the aim of this research is paradigm definition, establishing foundations of hierarchic analysis and suggesting operational procedures for that purposes.

Model

The largest number of scientific discoveries in logically-philosophical basis related to hierarchy determination is generally bonded with settings, without some general consensus regarding purpose and existence of natural hierarchic relations issue. Many of these doubts are based on unsubstantiated philosophical thesis that can not endure serious criticism or serious analysis (Bonacin, 2008). Bonacin further (Bonacin, 2004a; Bonacin, 2005; Bonacin, 2006) proved it is possible to establish universal scale where objects can be projected and at the same time, that it is possible to determine Universal comprehension continuum which allocates objects according to perception in some domain (figure1).

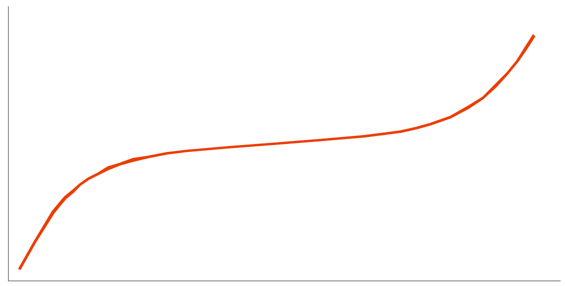

Figure 1. Continuum

Besides, Bonacin proved (2007a) that technical, biological and other processes are unique and universal and are submitted to the same rules, established Unique theory of entity existence and showed (Bonacin, 2008) that, so called, stochastic or "chaotic" models are not persistent. Finally, Bonacin et al., (2003) proved there is General expertise frame which describes discovery and possible hierarchic organizing of any entity according to constructive and destructive processes that actually exist (Bonacin, 2004a).

This is how all basic preconditions were achieved to consolidate necessary discoveries into set of information necessary for hierarchic analysis definition. The main issue that has to be solved is dualities that reflects in a discrete (dichotomy) advents in Universum (Bonacin, 2004b) but are globally generalized according to continuous type (Bonacin, 2004b) of registration. In simple terms, any generalized determination of some large sample attribute leads toward continuous model such is e.g. Comprehensive continuum (Bonacin, 2005) or simple data distribution (Bonacin, 2004b). Hierarchy however, assumes levels, in other words, discrete indicators. So there is unsolved problem of duality which disables logical establishment of serious discussion about hierarchic models.

To illustrate previous, in so called "macro" world there is almost no samples of clear continual advents except for the trace that rotation point leaves on the edge of a circle that rolls on some surface, and that advent is inquiring because of point definition as infinitesimal dimensions that the circle is composed of. We should evidently, search for the solution in some other domain different then mathematical or physical, and that is logical because it is the only one free of any undesired contamination.

Let's assume that somewhere, in any set of epiphany or advent, there must be more then one object. If they really exist, all this objects are different at least according to one parameter, which means it is necessary to determine presence of at least two objects for definition of difference. This follows that number two defines the ***difference***. In the same context number 1 is ***existence***, since it can not present anything else because in whole Universum there would be one and only one object, if it really exist and there is no other object or we don't know is there any, then we can not conclude anything else about it except that it really exists. In the same way, number 3 presents group i.e. ***similarity,*** since it is composed of three different objects and each has existence and differences. Further assorting through 4, 5, and further just presents higher degree of complex understanding of advents; we are surrounded with, and contain special attributes that are not the subject of this discussion. So as long as there are two objects, there is some kind of difference.

This rule as any other (1, 2, 3 …) remains even when the number of objects increases. Therefore, between the objects, whose existence we determined, are similarities and differences and everything else what was previously defined. It is obvious that similarities consider certain level of compatibility according to some parameters, and the differences certain level of incompatibility according to these or other parameters. When the number of individual objects in some set grows, if the level of similarity between some objects is greater regarding level of differences regarding other objects, it is clear this leads to formation of more or less homogeny sets gathered around typical – ***representative***. However, further increase of the number of objects in some set will lead to assorting of all relations, not only to a set formation. Having in mind that sets "comprise" certain comprehensions (even though, immanently according to representative) it is clear that these comprehensions will vary from group to group.

So, if there are more individual objects as well as many groups, further relation assorting, just because of present differences, will lead to positioning of these groups on a different ***levels*** and then we face formation of hierarchy. This means "vertical" structuring is a natural flow, since it's certain that extreme difference with at least two groups means one group with a few and one group with many comprehensions. Regardless relations between groups, it is certain that the group with the most comprehensions in some domain (superior) will start domination in one way or another, while the other one will be inferior, which produces, at least, two levels.

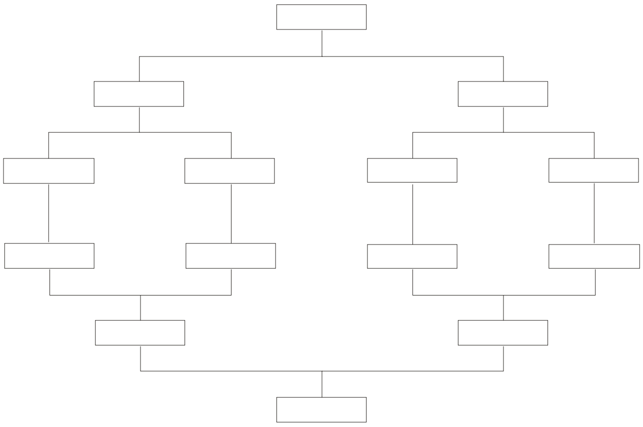

Figure 2. Logical model of hierarchy

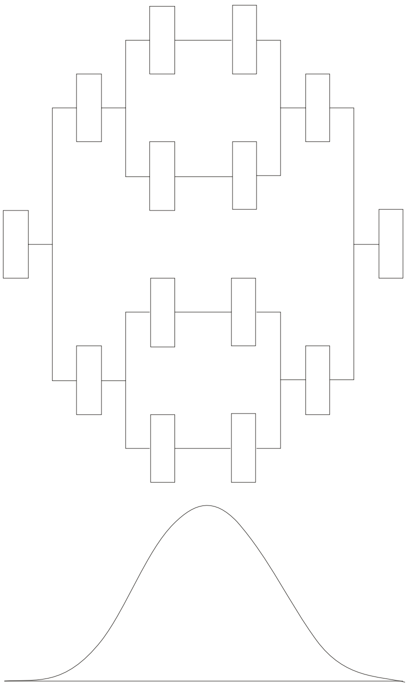

Figure 3. Normal distribution

Within each of these levels, objects are gathered on the same level and, according to the same logics, can be repositioned into inter-levels all the way to desired accuracy level.

This certainly defines hierarchy, which means that the problem of hierarchic analysis can be solved with identification of extreme representatives that gather such groups, which is mathematically relatively easily solved problem. It is also obvious there will be a small number of extremely superior object, which defines the top of the structure, but according to logical distribution it should be the same with extremely inferior (figure 2). As we can see from figure 2, maximum density of the object is in the middle and decreases towards the top and bottom. If we rotate figure 2 for 90 degrees it perfectly represents normal distribution (Bonacin, 2007b) which is almost always the rule when we talk about representative samples (figure 3).

If we assume that the objects are not static in development, but have ***dynamics in development,*** we will discover they can change their position within somehow defined "space". This enables us to, applying appropriate methods; determine their actual position, as well as previous and future states. This means that described protocol is applicable in any situation where we determine structure about some state and equally applicable if we determine changes in a structure influenced by any impact on objects in a certain time or under the influence of any comprehensions that eventually change their current position, which is undoubtedly extremely powerful tool in science and scientific methodology.

Algorithm

The procedure MOHA comprise of a few phases. In ***the first phase*** we find at least two objects mutually extremised in the set of objects described with a set of any variables that were previously assorted according to the rules of measuring theory (Bonacin, 2004b; Bonacin, 2009). In the ***second phase*** all other objects are joined with these representatives and we make (at least) two mutually extremised groups that represent basic levels (superior and inferior). In all previously described procedures and all analyzed situations, there were always identified superior and inferior objects and there is no doubt that the procedure is well founded, regardless of the way representative was determined by (which is possible to define in different ways).

In ***the third phase***, each of these groups is dividing, according to the same principle, until the final set of levels is achieved. In ***the fourth phase*** we identify each object in accordance with its level and pertinence to the unique subgroup on the level of its belonging. This is how we defined hierarchy this article is describing.

Evidently, the algorithm is extremely simple and the procedure for uncovering extreme representatives is the same one used in Uditax model of distinct taxon analysis (Bonacin et al., 2009) with the only difference being the complete relaxation of maximal representative number in the sample described with many variables. Finally, we describe each identified subgroup in the frame of each level with the help of elementary statistic indicators and it is also possible to determine differences among such formed groups for more accurate determination of details depending on the problem that is being analyzed.

Example and methods

The sample of 249 first grade pupils from elementary school was described with 26 bio-motor variables in order to determine their basic bio-motor status. This set included 14 variables for morphological status description and 12 for motor- functional status description. The sample and variables were previously described in details (Bonacin, 2004a). Out of many derived results, in regards to this work, we presented only the ones that describe final elements of hierarchic analysis. All data was brought down to universal scale (Bonacin, 2006) in the range from 1 to 5 with no information loss on the correlation between objects and variables (Bonacin, 2004b).

Results, discussion and conclusion

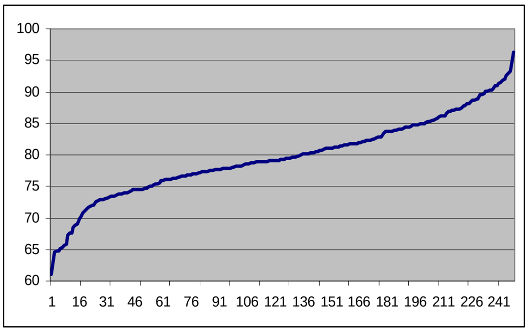

Figure 4. Sample continuum (on ordinate there is a sum according to universal scale)

After the procedure was conducted, extreme entities were identified (D095 as superior and E136 as inferior entity). Their values according to individual variables were presented in Table 1. With the procedure of joining other objects we obtained two sets of objects; the group of inferior includes 132 and the group of superior 117 objects. In the next step, 132 objects were "divided into layers" in the same way and we obtained 51 inferior and 81 superior object. The set of superior objects was also "divided into layers" and we obtained 67 inferior and 50 superior objects within this set.

The final result for 6 levels resulted in a structure where the representatives are the objects with global values as shown in table 2 in 10 groups.

Table 1. Values of extreme object per individual variable

	E136	E095
Var1	1.86	2.73
Var2	2.59	3.08
Var3	2.22	3.03
Var4	2.12	3.05
Var5	2.38	3.27
Var6	3.75	4.84
Var7	2.19	3.89
Var8	2.38	2.96
Var9	2.40	3.40
Var10	3.12	3.74
Var11	2.64	3.35
Var12	2.70	4.50
Var13	3.29	4.57
Var14	3.80	4.49
Var15	2.77	4.45
Var16	2.74	4.72
Var17	1.63	3.23
Var18	2.06	3.68
Var19	2.00	2.14
Var20	2.46	2.85
Var21	1.79	4.81
Var22	1.00	4.40
Var23	1.72	4.88
Var24	1.82	3.67
Var25	1.08	2.05
Var26	2.55	4.50

Table 2. Representatives, their values on global indicator and objects in the group:

Object	Global	Sample
E136	64.80	10
E009	69.27	9
E186	73.47	32
E194	76.49	37
E134	78.71	44
E249	80.99	35
E041	83.65	32
E059	86.61	27
E078	89.78	13
E095	92.48	10

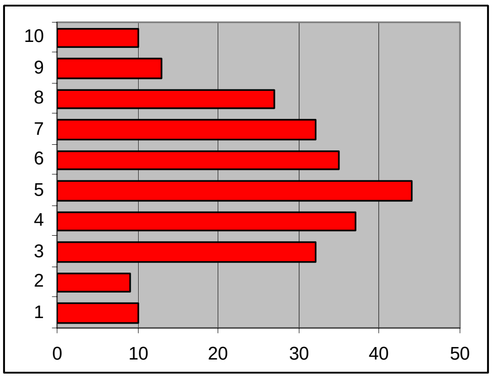

Figure 5. Sample distribution

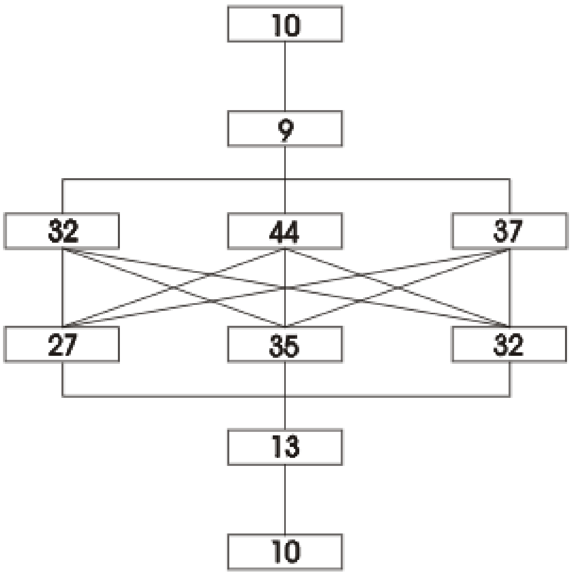

Figure 6. Final solution with number of group members

According to the results in Table 2 it is obvious there are differences in the sample. There are fewer objects on the outsides and more objects in the middle (figure 5).

It is also obvious that objects (as well as groups) are assorted according to an increasing range, resulting E136 as the most pronounced object in inferior, and E095 in superior. According to figure 6 it is evident that the preconditions and goals of this article are achieved. Also the hierarchic structure of the sample based on internal consistency of all data was identified. It is visible that the levels are clearly identified, as well as the number of object that belongs to a certain group. Certainly, the identification of each individual object or group is possible in the variable field, as well as analysis of difference and discrimination among the groups. Since this was not the intention of this article, these results are not significant, even though the inspection of individual and group indicators can be performed. Evidently the results on the complex sample with 249 objects explained with 26 variables indicated that hierarchy determination without an "outside" criteria parameter is possible. Of course, the real strength of this protocol will be confirmed through time and number of successfully solved models of hierarchic structuring in scientific research where this protocol is applied.

References

Ané, C. (2008). Analysis of comparative data with hierarchical autocorrelation. *Annals of Applied Statistics*, 2, 3, 1078-1102.

Bonacin, D. (2002). Integrativni taksonomski model kao preduvjet programiranja rada u kineziologiji.[Integrative taxonomic model as precondition for work programming in Kinesiology] /In Croatian/, *Programiranje rada u području edukacije, sporta, sportske rekreacije i kineziterapije. 11. ljetna škola kineziologa R. Hrvatske*, Rovinj, Zbornik radova: 103-104. (ed. Findak, V.).

Bonacin, D., Blažević, S., Carev, Z. (2003). Global comprehensive theory in kinesiology learning methods, *Daegu Universiade conference*, Korea, Proceedings, pp. 696-702. (ed. Park, S.).

Bonacin, D. (2004a). *Identifikacija taksona biomotoričkih dimenzija djece uzrasta 7 godina pod utjecajem transformacijskih procesa*. [Taxon identification of bio-motor dimensions with children aged 7 under influence of transformational processes]. /Dissertation. In Croatian/. Sarajevo: Fakultet sporta i tjelesnog odgoja.

Bonacin, D. (2004b). *Uvod u kvantitativne metode.* [*Introduction to quantity methods*] /In Croatian/, Kaštela: Personal edition.

Bonacin, D. (2005). Comprehensive continuum. *Homo Sporticus*, 8, 2: 16-20.

Bonacin, D. (2006). Definicija apsolutnih temelja spoznajnog kontinuuma. [Definition of absolute comprehensive continuum basis] /In Croatian/, *Kvaliteta rada u području edukacije, sporta i sportske rekreacije. 15. ljetna škola kineziologa R. Hrvatske*, Rovinj, Hrvatska, Zbornik radova, pp.405-408. (ed. Findak, V.).

Bonacin, D. (2007). Univerzalna spoznajna načela egzistencije procesa. [Universal comprehensive principle of process existence] /In Croatian/, *2nd International Symposium - New Technologies at Sport*, Sarajevo, BiH, Proceedings, pp. 48-53. (ed. Smajlović, N.).

Bonacin, D. (2007). Stvarni temelji normalne distribucije [The actual basis of normal distribution] /In Croatian/, *Acta Kinesiologica*, 1, 2: 42-47.

Bonacin, D. (2008). Determinizam i kaos. [Determinism and chaos] /In Croatian/, *Simpozij Univerziteta u Sarajevu - Čovjek i moderni svijet*, Sarajevo, Bosna i Hercegovina, Zbornik radova, pp. 13-20. (ed. Ćaklovica, F.).

Bonacin, D. (2009). Process universality. *3rd International Symposium - New Technologies at Sport*. Sarajevo, Bosnia & Herzegovina, Proceedings, pp. 31-35. (eds. Kisić, S., Kvesić, M., Rađo, I.).

Bonacin, D., Bonacin, Da., & Bilić, Ž. (2009). Uditax – a model of distinct taxons based on maximization of representative objects distances. *Acta Kinesiologica*, 3, 1, 50-53.

Espinosa, A., Hamden, R., & Walker, J. (2007). Beyond hierarchy: a complexity management perspective. *Kybernetes*, 36, 3-4, 333-347.

Gallizo, J.L. & Salvador, M. (2006). Share prices and accounting variables: a hierarchical Bayesian analysis. *Review of Accounting and Finance*, 5, 3, 268-278.

Jin, Y., Nie, B., & Xiao, Y. (2008). Theoretical model and application of multilevel modeling in the research of social system. *Kybernetes*, 37, 9-10, 1401-1408.

Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis.* Prentice-Hall.

Kang, E.L., Liu, D., & Cressie, N. (2009). Statistical analysis of small-area data based on independence, spatial, non-hierarchical, and hierarchical models. *Computational Statistics & Data Analysis*, 53, 8, 3016-3032.

Momirović, K., et al. (1987). *Metode, algoritmi i programi za analizu kvantitativnih i kvalitativnih promjena.* [*Methods, algorithms and programs for quantity and quality changes analysis*] */ In Croatian/,* Zagreb: Fakultet za fizičku kulturu.

Tingquan, Z. & Xuemou, W. (1995). A pansystems clustering approach and hierarchical analysis of complex systems. *Kybernetes*, 24, 2, 51-59.

Veldman, D, J. (1967). *Fortran programming for the behavioral sciences.* New York: Holt, Rinehart & Winston.

MODEL I ALGORITAM ZA HIJERARHIJSKU ANALIZU

Sažetak

Svrha ovog rada bila je definiranje početnog modela, tj. paradigme za identifikaciju hijarahije u proizvoljnom skupu podataka. Metodološki dizajn organiziran je na temelju identifikacije reprezentativnih objekata u totalnom skupu. Rezultati su pokazali da je hijerarhijska struktura prirodni redoslijed odnosa i da ju je moguće prepoznati. Ograničenja modela nisu poznata jer je u više testnih situacija uvijek dolazilo do jasnog i uvjerljivog hijerarhijskog strukturiranja podataka temeljem internih relacija. Originalnost modela je nedvojbena, jer je njegova primjena doslovno neograničena, a vrijednost je naročito u postavljanju potpuno novih temelja utvrđivanju relacija temeljem interne konzistencije i relacija.

Ključne riječi*: hijerarhija, interna konzistencija, ekstremi*

Received: October 15, 2008.
Accepted: May 26. 2009.
Correspondence to:
Assist. Prof. Dobromir Bonacin, Ph.D.
University of Travnik
Faculty of Education
Aleja konzula 5, 72270 Travnik, B&H
Phone: +387 (0)62 907 430
E-mail: dobromirb@sposci.com