Scalene Relation in Hierarchical Clustering

Casanova del Angel, Francisco
SEPI ESIA del IPN.
fcasanova@laneta.apc.org

Abstract
Here we present the hierarchical analysis of the adherence between steel and concrete in scalene relation. We start off with a summary of the basic analysis of the mechanics of fracture applied to a set of test tubes in the laboratory. The steel-concrete hierarchical adherence is a scalene relation. We present its theorem. Finally, we present the following conclusions: the class factors show that the shorter the length is, the stronger the last adherence stress will be. Hierarchy occurs in two different ways.

Keywords: scalene relation, adherence, mechanics of fracture, cylindrical assays, allowable stress.

Introduction to theory and algorithm construction

To obtain a representation of the hierarchical relations between random variables it is necessary to define a metric structure. Let $\Re^p$ the real space of α-factors, obtained from one factorial analysis applicated to a finite set of random variables. Into the finite space of probabilities (Ω, $P(\Omega)$, P) let's define a distance that relate the factors of classes of the adherence steel-concrete, denominated $\Im^2$, as: $\Im^2(x_j, x_{j'}) = \Sigma^p_{\alpha=1} (fj/f_{j'})[F_\alpha(x_j) - F_\alpha(x_{j'})]^2$, it is the factorial distance of classes weighed between the random variables, where f_j and $f_{j'}$ are the frequency of classes j and j', and F_α the factorial values of classes x_j and $x_{j'}$.

The algorithm construction is based in a sequence of partial cluster; see [**Gordon. 1999**] denominated $C(\alpha) = C_0,..,C_h ,..., C_{\alpha-1}$, where a partial cluster is the union of two classes. The distance is calculated by $\Im^2(x_j, x_{j'})$ over the set F of dimension α after a tabular arrangement F_{JQ} hence:

$C_0 = C_0(\alpha)$ =Term[$C(\alpha)$]=$\{x_j\}$ $\forall\, j\in F_\alpha(j)$, with Term the set of last classes of the classification $C(\alpha)$,
Ver[C_0]=Term[$C(\alpha)$]=$\{x_j\}$ $\forall\, j\in F_\alpha(j)$, with Ver the vertex of the last classes of the classification $C(\alpha)$,
$n(\{x_j\}) = 0$ $\forall\, j \in F_\alpha(j)$ is the index of level of the class, $f(\{x_j\})$ is the connected frequency to x_j & $\Im^2(x_j, x_{j'}) = \delta(\{x_j\}, \{x_{j'}\})$ $\forall\, x_j, x_{j'} \in F_\alpha(j)$ is the distance between classes.

For the iteration of rank h=1, you can find the minimum of δ over the vertex of the classification C_0; Ver[C_0]. Let ($\{x_j\}$, $\{x_{j'}\}$) be an even of classes of an element to fulfill with the minimum. The first node obtained take the number Card(α)+1. Therefore for N=Card(α)+1 and h=1 it is possible to obtain the new class, denominated c made by the variables or initial classes x_j, $x_{j'}$, it is $c_1 = \{x_j, x_{j'}\}$ and the set of classes situated immediately under the class c_1 of $C(\alpha)$ is: Successor (c_1, $C(\alpha)$) = $\{x_j, x_{j'}\}$ $\forall \in$ Nodo($C(\alpha)$), the first partial classification created is: $C_1 = C_n(c_1) = C_1(\alpha) = C_0 \cup c_1$, and the new vertex is Ver[C_1] = Ver[C_0] U c_1 - $\{x_j\}$ - $\{x_{j'}\}$. The new index of classification of the class is $n(c_1) = \inf\, \{\delta^0(\{x_j\}, \{x_{j'}\})\}$ $\forall\, x_j \neq x_{j'}$ with $x_j, x_{j'} \in$ Ver[C_0]. The cardinality of the class c_1 is now 2 and the frequency of the new class is $f(c_1) = f(\{x_j\}) + f(\{x_{j'}\})$.

At the end this first iteration that construct a new partition of α, it's necessary to recalculate the entire distance among all classes, denominated Ver[C_1]. As Ver[C_1] is deduced from Ver[C_0] replaces two classes for the joint of these two classes. The over calculation of the distance among parts of clusters which permit calculate the distance between the new created class and other classes of Ver[C_1] excepting the two classes that had realized the fusion is $\delta^0(c_1, r)$ $\forall\, r \in$ Ver[C_1] with $r \neq$

x_j & $r \neq x_{j'}$. Now the iteration h. At this moment you know the partial classifications $C_0, C_1, .., C_{h-1}$. Let t_h and t'_h two classes of Ver[C_{h-1}] so to have the minimum distance δ calculated over Ver[C_{h-1}], and the cardinality is equal to Card(α)+h = N. The class h is $c_h = t_h \cup t'_h$. The successor h is Successor (c_h, C(α)) = {t_h, t'_h} where t_h and t'_h are respectively the first-born and the benjamin of the c_h. class. The h partial classification is $C_h(\alpha) = C_h = C_\nu(c_h) = C_{h-1}(\alpha) \cup c_h = C_{h-1}(\alpha) \cup \{t_h \cup t'_h\}$. The vertex of the h classification is $Ver[C_h(\alpha)] = Ver[C_{h-1}(\alpha)] \cup \{c_h\} - \{t_h\} - \{t'_h\}$. The index of the h-class is $n(c_h) = \inf\{\delta^{h-1}(t, t')\}\ \ \forall\ \ t \neq t'$ with $t, t' \in$ Vér[C_{h-1}]. The cardinality of the h-class is Card(c_h) = Card(t_h) + Card(t'_h) and the frequency of the h-class is f(c_h) = f(t_h) + f(t'_h). At last, the rank of iteration Card(α)-1. At this moment only two classes remain to add whose union is the unity of all the classes α; the classification C_{h-1}. Here the value of the cardinality of the number of elements is 2*Card(α)-1, the class is $c_h = t_{Card(\alpha)-1} \cup t'_{Card(\alpha)-1}$, The partial classification is $C_h = C(\alpha) = C_{Card(\alpha)-1}$, the vertex of this partial classification is $Ver[C_h] = Ver[C_{Card(\alpha)-1}] = \{\alpha\}$, the cardinality of the class c_h is c_h = Card $c_{Card(\alpha)-1}$ =Card(α), the frequency of the class is f(c_h) = f(α), and the index of level is $\nu(c_h) = \nu(\alpha) = \delta^{h-1}(k_{Card\alpha-1}, k'_{Card\alpha-1})$. The theoretical development of the hierarchical classification proposes the problem of election of classes when at last two even of subclasses exhibit the case of a minimum equality in the distance δ over Ver[C_h]. The mathematic solution depends of a criteria of minimum distances, and such criteria is:

Theorem: Given a sequence of a partial hierarchies of C(α), if in the following selection of classes to add exists two partial hierarchies $C_h(\alpha)$ and $C_k(\alpha)$ such as $C_h(\alpha) = C_k(\alpha)\ \forall\ h \neq k$ and present the same minimal distance: $\delta(\{x_h\}, \{x_{h'}\}) = \delta(\{x_k\}, \{x_{k'}\})$, with relation to class $C_r(\alpha)$ the hierarchical partial to add, depends on the geometrical distance δ presented by classes $C_h(\alpha)$ and $C_k(\alpha)$ in relation to the C_r class.

Dem:
Let $C(\alpha) = C_0, C_1, ..., C_h, ..., C_{\alpha-1}$ one sequence of partial hierarchies. If two hierarchies exist $C_h(\alpha)$ and $C_k(\alpha)$ such as that $C_h(\alpha) = C_k(\alpha)\ \ \forall\ \ h \neq k$ and carry out with the equality of distance:

$$\delta^{h-1}(\{x_h\},\{x_{h'}\}) = \delta^{k-1}(\{x_k\},\{x_{k'}\})$$

where Ver[C_h] = Ver[C_k] such as the index of level of classes are equal, i.e., $n(C_h) = \nu(C_k)$, then:

$$\inf\{\delta^{h-1}(C_h)\} = \inf\{\delta^{k-1}(C_k)\}$$
$$\inf\{\delta^{h-1}(\{x_h\}, \{x_{h'}\})\} = \inf\{\delta^{k-1}(\{x_k\}, \{x_{k'}\})\}$$

where x_h, $x_{h'}$ are the first-born elements and x_k, $x_{k'}$ the benjamin elements of the C_h and C_k classes respectively. Let now $C_r(\alpha) \ni C_r(\alpha) \subset C(\alpha)$ and $r < h$, $r < k$, a third class of partial hierarchy, adding one of the two partial hierarchy of order h or k, making use of ultrametric distance proprieties. The geometry form created by these three partial hierarchies are: equilateral, isosceles and scalene[1]. In agreement to geometry formed by the hierarchical classes $C_r(\alpha)$, $C_h(\alpha)$ and $C_k(\alpha)$:

$$\delta^{r-1}(C_r, C_h) \leq \sup\{\delta^{r-1}(C_r, C_k), \delta^{r-1}(C_k, C_h)\} \quad \text{and} \quad \delta^{r-1}(C_r, C_k) \leq \sup\{\delta^{r-1}(C_r, C_h), \delta^{r-1}(C_h, C_k)\}$$

to be advisable with:

$$\delta^{r-1}(C_r, C_h) \leq \delta^{r-1}(C_r, C_k) + \delta^{r-1}(C_k, C_h) \quad \text{and} \quad \delta^{r-1}(C_r, C_k) \leq \delta^{r-1}(C_r, C_h) + \delta^{r-1}(C_h, C_k)$$

it means that it doesn't matter too much if the triangular form made by the hierarchy classes $C_r(\alpha)$, $C_h(\alpha)$ and $C_k(\alpha)$: is equilateral or isosceles, choosing arbitrarily the partial hierarchy to add (case very strange have the same index of level and the same distance δ).

---

[1] : See [**Benzécri. 1976**] § 4.1, pp 138 where only talks about triangular relations equilateral and isosceles but never about the triangular scalene relation.

Now if the triangular relation is scalene, which one of two distances to class *r* is small, i.e., if

$$\delta^{r-1}(C_r, C_k) < \delta^{r-1}(C_r, C_h)$$

And this, have a demonstration very easy about the proprieties of triangular inequalities

$$\delta^{r-1}(C_r, C_k) + \delta^{r-1}(C_k, C_h) < \delta^{r-1}(C_r, C_h) + \delta^{r-1}(C_k, C_h)$$

and so the distance over *h* and *k* is the same over *k* and h, then

$$\delta^{r-1}(C_r, C_k) < \delta^{r-1}(C_r, C_h)$$

■

Introduction to application, test, variables and data in analysis

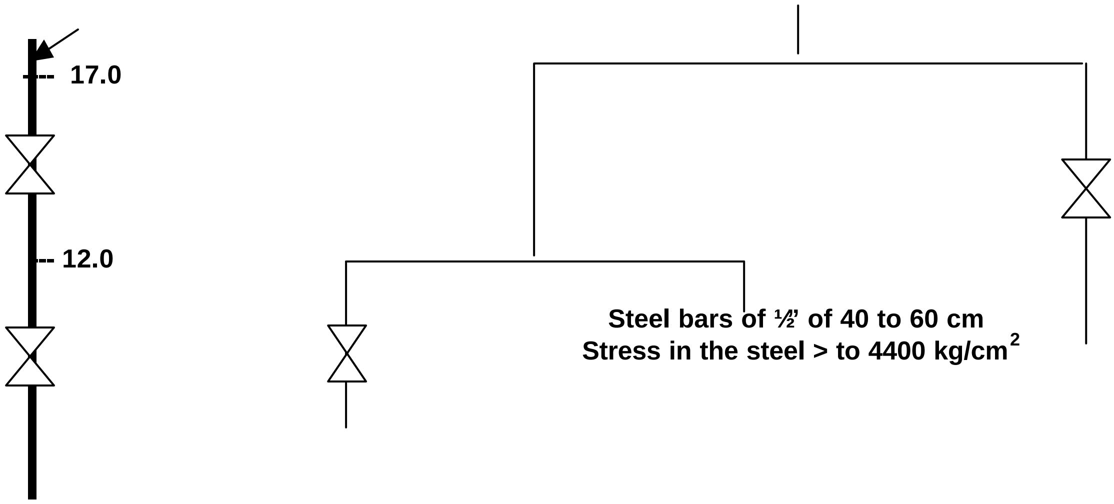
The behavior of adherence between reinforced steel and surrounded concrete is essential in the analysis of some structure of reinforced concrete. Due to low strength to the tension, the concrete elements tend to crack, that is mainly caused by the difference of deformation between steel and concrete and by stress due to tension load in the concrete when these stresses are higher rank compared to the stress supported by concrete. In order to study the fissure of some materials, the Mechanics of Fracture was developed.

Let $k_{IJ} = \{k(i, j)\}\ \forall\ i \in I,\ j \in J$ one tabular arrangement of data and k_{JJ} = card $\{k(i, j) = k(i, j') = 1\}$ $\forall\ i \in I$ with $j, j' \in J$. The logical arrangement of k_{IJ} is denominated "of Burt". *I* is the set of random test of laboratory (cylindrical specimens of hydraulic concrete) of dissimilar length and J is the set of variables between steel and concrete. Three tests were carried out with cylindrical specimens of 4" diameter with f' c = 200 kg/cm^2 and the concentrically ½' embedded stem. The length of these specimens was between 10 y 60 cm with increases of 10 cm. Based on these tests it was obtained five data for the stress of concrete f' c in kg/cm^2, the ultimate load of tension in kg, the stress of steel in kg/cm^2 and the ultimate stress of adherence in kg/cm^2. The length of each one of these 6 specimens is between 10 cm and 60 cm with an increase 10 cm.

Hierarchical clustering of the adherence between steel-concrete and scalene relation

Let´s start off from one classification on a finite set of aleatory variables $X_1,...,X_n$ is a partition; it means a part of certain number of no empty parts two by two and empty intersection, generally a hierarchie of embedded classes. The finite set of aleatory variables is divided in a finite number of classes where each one of those are divided at same time in another finite number of classes or sub classes. The lecture and the interpretation of the dendrogramme, figure 1, see [**Casanova. 1990**] say that its hierarchical clustering is made in three ways.

Index of level



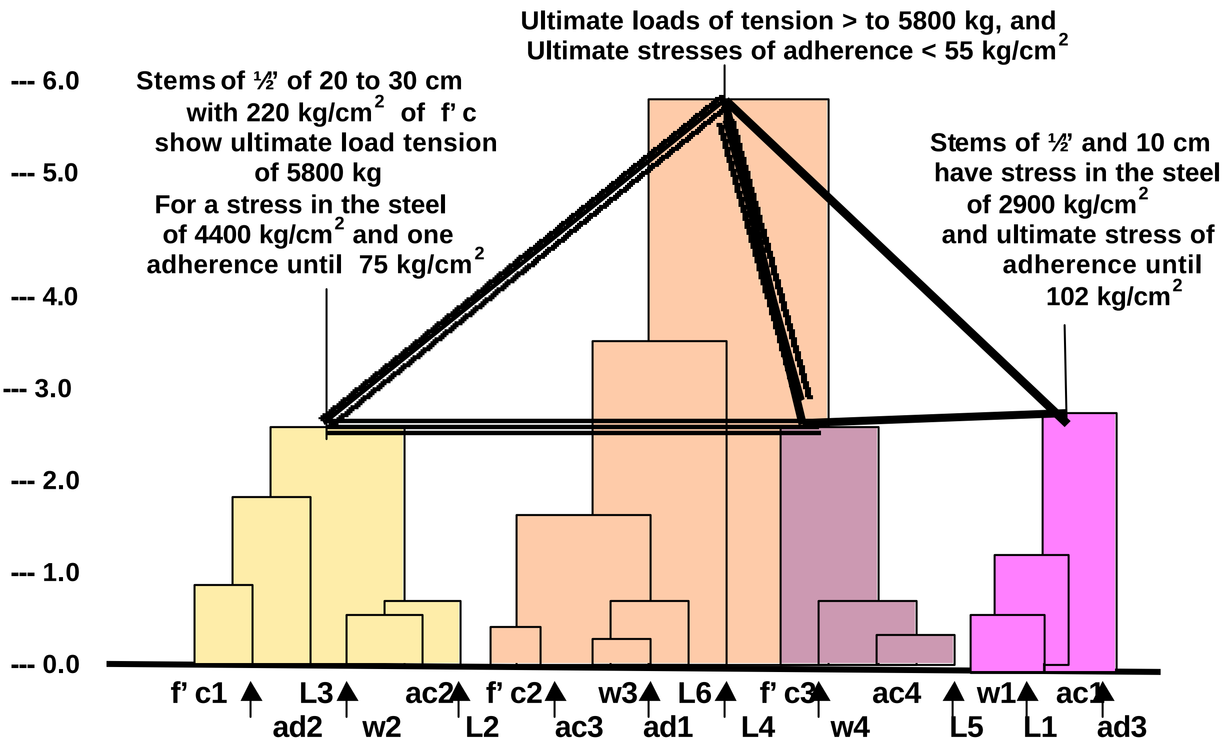


Figure 1: Hierarchical clustering of classes in the analysis of the adherence steel-concrete.

Bibliography

[**Benzécri, *et al*. 1976**]

Benzécri, J.P. & Collaborateurs. 1976. La Taxinomie. 2 l' Analyse des Données. Ed. Dunod

[**Casanova. 1990**]

Casanova del Angel, F. 1990. Análisis multidimensional de datos. Ed. LOGICIELS. México.

[**Gordon. 1999**]

Gordon A. D. 1999. Classification: Second edition. Monographs on Statistics and Applied Probability series.