

How to Analyze Modeling Approach Comparison Criteria

Frédéric Mayart¹, Jean-Michel Bruel², and Brahim Hamid²

¹ Psychology M.D. and independant consultant, frederic.mayart@gmail.com

² University of Toulouse, France, bruel|hamid@irit.fr

1 Context

One possible final goal of defining a set of criteria to define modeling approaches [1] is to help people, especially from industry, picking up the good approaches or artifacts according to their own purpose. The authors of the comparison criteria have managed to get several different assessments made by defenders of particular modeling approaches. From our point of view the experiment is mature enough to support a factorial analysis of the criteria themselves. The goal of this paper is to present how such an analysis could be conducted and illustrate its usefulness. We have identified several key modeling concepts but we only focus in this document on the assessment of modeling approaches.

2 Proposed study

In this kind of exercise where a set of criteria is aimed at describing a particular topic for a certain purpose it is recommended to conduct a correlation analysis. If such purpose is to find whether *correlations exist among comparison criteria*, multivariate analysis methods like PCA (Principal Component Analysis) come to mind [2]. The idea is to explore how different subsets of methods would correlate while not doing so with other subsets due to specific comparison criteria, which is also a pre-requisite to cluster analysis. Finding underlying factors (vectors) that would best explain which variables aggregate and how subsets separate is just a first step. An even better purpose is to try to find out whether comparison criteria or subsets combine with other subsets, at different levels, in order to form “families” or types of methods. Multidimensional Statistics use two main methods: Factorial Analysis methods that consist in projecting a cloud of points on a vector space, while losing as little information as possible; and Classification Methods, that try to cluster those points. Factorial Analysis methods regroup three main techniques: Principal Component Analysis (PCA, with several quantitative variables), Correspondence Analysis (CA, two quantitative variables, represented by a contingency table) and Multiple Correspondence Analysis (MCA, more than two variables, all qualitative). In the proposed metamodel, variables describing methods are of different types, hence MCA should be preferred. Once all variables are converted in qualitative ones and put into a disjunctive table the analysis can be conducted. We propose

to study instances (methods), variables (comparison criteria) and their modalities [2]. Two methods are close if questionnaires have been answered the same way. The focus will not be on instances (methods) *per se* but more on sets: *are there groups of methods?* We are not much interested in methods themselves but groups of methods inside the whole set, that is: analyzing methods and observe how they regroup and under which factors. We want to study the relationships between variables and the associations between modalities. A modality is the “value” a variable can take. Qualitative variables and quantitative ordinal variables have discrete values, and usually a finite set of modalities. Two modalities are close if they have been taken together often. Two comparison criteria characteristics that are often cited together by a group of individuals will geometrically appear close on the plot graphs generated by the factor analysis. We are looking for such plots clouds. One or more synthetic continuous variable(s) can be looked for by a PCA to summarize the qualitative variables, and interpret the relations between them. Using a representation by modalities is easier to show how vector dimensions separate or aggregate the different criteria and gives more precision. A final goal is to characterize methods subsets by modalities of comparison criteria using a Hierarchical Cluster Analysis which is the logical and common follow-up to an MCA. We want to regroup the methods in a few number of classes corresponding to ”profiles” of Comparison Criteria. The result is a hierarchical tree easy to interpret. Methods would appear as leaves, clustering into small branches, then bigger ones, etc., up to a trunk. Classes can then be described by the criteria variables and/or their modalities, by the factorial dimensions, or the individuals/methods.

3 Conclusion

This paper presents how comparison criteria could benefit of advanced statistical methods such as MCA and Hierarchical Clustering (see [3] for more details). Such tools can help give insights about many questions relating to similarities and differences between modeling approaches, and/or comparison criteria characteristics complex relationships. Data collection can be enhanced too with a little revamping of the questionnaire so it better feeds the statistical tables and minimizes loss of information.

References

1. G. Georg, S. Ali, D. Amyot, B. H. Cheng, B. Combemale, R. France, J. Kienzle, J. Klein, P. Lahire, M. Luckey, A. Moreira, and G. Mussbacher. Modeling approach comparison criteria. Technical report, McGill U., Montreal, CAN, 2001.
2. F. Husson, J. Josse, and J. Pagès. *Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data?* Agrocampus Applied Mathematics Department, 2010.
3. F. Mayart, J.-M. Bruel, and B. Hamid. Correlation analysis of modeling approach comparison criteria: methodology proposal. Technical report, 2013. Available at <http://www.irit.fr/publis/MACAO/IRIT>