Hierarchical and Non-Hierarchical clustering methods to study students algebraic thinking in solving an open-ended questionnaire

Benedetto Di Paola[1] Rosario Battaglia[2] and Claudio Fazio[3]

[1]Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, Palermo, Italy; benedetto.dipaola@unipa.it

[2]Dipartimento di Fisica e Chimica Università degli Studi di Palermo, Palermo, Italy; onofriorosario.battaglia@unipa.it

[3]Dipartimento di Fisica e Chimica Università degli Studi di Palermo, Palermo, Italy; onofriorosario.battaglia@unipa.it

The problem of taking a data set and separating it into subgroups, where the members of each subgroup are more similar to each other than they are to members outside the subgroup, has been extensively studied in science and mathematics education research. Student responses to written questions and multiple-choice tests have been characterised and studied using several qualitative and/or quantitative analysis methods. However, there are inherent difficulties in the categorisation of student responses in the case of open-ended questionnaires. Very often, researcher bias means that the categories picked out tend to find the groups of students that the researcher is seeking out. In our contribution, we discuss an example of application of hierarchical and non-hierarchical analysis method, to interpret the answers given by 118 Tenth Grade students in Palermo (Italy), to six open-ended questions about algebraic thinking. We show that the parallel use of the two quantitative analyses allows us to interpret in deep way the reasoning of students solving different mathematical problems using Algebra. These clustering methods also allow us to highlight different students groups, that can be recognised and characterised by common traits in their answers, without any prior knowledge on the part of the researcher.

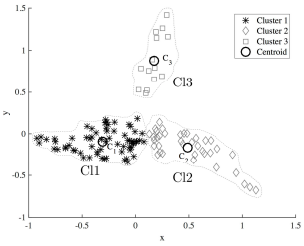
Keywords: Algebraic thinking, Clustering, k-means method, dendrograms.

Introduction

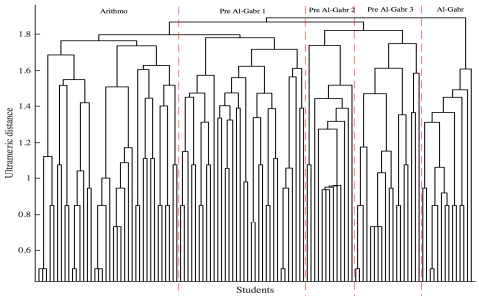
In recent years, some papers have tried to develop detailed models of the reasoning competences of the student populations tested, or to subdivide a sample of students into intellectually similar subgroups, by using quantitative or qualitative analysis methods. (Everitt, Landau, Leese & Stahl, 2011; Prediger, Bikner-Ahsbahs & Arzarello, 2008) It is worth noting that research papers using quantitative analysis methods to study student responses to open-ended questionnaire can be found in Science education; not many research work can be trace in Mathematics education, especially on the application of clustering analysis. In this paper we focus on the application of hierarchical and non-hierarchical clustering methods referred to dendrograms representation and k-means algorithm (Everitt, Landau, Leese & Stahl, 2011), trying to make sense to answers given by 118 Tenth Grade Italian students to six open-ended questions on algebraic thinking. The questionnaire was answered by students in a maximum of 45 minutes. It was administered to the students at the beginning of the school year, before any discussion about Algebra had taken place.

In particular we discuss the results on the study of typical students behaviour in tackling the algebraic resolution of word problems and, at the same time, at understanding how the student semantically and syntactically control questions containing symbolic algebraic expressions (Radford & Puig, 2007). Our decision to refer to word problems, according to the Programme for International Student Assessment (PISA), can allow us to study student literacy (PISA) in using algebra (Bohlmann, Straehler-Pohl & Gellert, 2014) and in the transition from arithmetic to the modelling of problems expressed in a not-symbolic language, called "natural language – (NL)" (Arzarello, Robutti & Bazzini, 2005). K-means approach and dendrograms representation allowed us to partition and characterize our student sample, without making any a priori assumptions and giving as output student's behaviour interesting for the researcher in Education.

Clustering results



The k-means method (showed in Figure 1) allowed us to simply group and characterize the common students traits related to their solution strategies of the open-ended questions about algebraic thinking (procedure choices, mistakes, failings etc.). This gives us the opportunity to safely partition students into three groups: these are characterized by centroids C_i (called *Arithmo, Pre Al-gabr and l-gabr*) that represent the answering strategies given with maximum frequency by the students who are part of the cluster.



The Hierarchical clustering method (showed in Figure 2), obtained using the Weighted Average linkage, identified five groups of students (called *Arithmo, Pre Al-gabr 1, Pre Al-gabr 2, Pre Al-gabr 3 and Al-gabr*) allowing us to better highlight their difficulties in the answering strategies related to the transition between the NL (typical of word problem) and the symbolic one. The results we found are largely coherent with the ones already reported in the literature obtained by means of qualitative methods. For this reason, we can consider the use of both hierarchical and non-hierarchical clustering a valid tool to complement the use of qualitative analysis to study a large number of students with respect to the way they give answer to the questionnaire.

References

Arzarello, F., Robutti, O. & Bazzini, L. (2005). Acting is learning: focus on the construction of mathematical concepts. *Cambridge Journal of Education*, 35(1):55-67. doi: 10.1080/0305764042000332498

Bohlmann, N., Straehler-Pohl, H. & Gellert, U. (2014). Deconstructing the filtration of reality in word problems. Quaderni di Ricerca in Didattica, 24 (Supplement 1). P*roceedings of CIEAEM 66):142–145.*

Everitt, B.S., Landau, S., Leese, M. & Stahl, D. (2011). *Cluster analysis* (5th ed). NY, J.Wiley & Sons, Ltd.

Prediger S, Bikner-Ahsbahs A & Arzarello F 2008. Networking strategies and methods for connecting theoretical approaches: first steps towards a conceptual framework. ZDM: The Int. J. on Math. Ed., 40(2):165-178.

Radford, L. & Puig, L. (2007). Syntax and meaning as sensuous, visual, historical forms of algebraic thinking. *Educational Studies in Mathematics*, 66(2):145-164.