SIA's asymmetric rules approximation to hierarchical clustering in Learning Analytics: mathematical issues

Rubén Pazmiño[1], Francisco Garcia[2], Miguel Conde[3]

[1] *School of Mathematics & Data Science Research Group, Escuela Superior Politécnica de Chimborazo. Ecuador.*
E-mail: rpazmino@espoch.edu.ec

[2] *Department of Computer Science, University of Salamanca, Spain.*
E-mail: fgarcia@usal.es

[3] *Department of Mechanical, Computer Science and Aerospace Engineering, University of León, Spain.*
E-mail: miguel.conde@unileon.es

Introduction

We use the definition set out in the first international Conference on learning analytics and Knowledge and assumed by the Society for Learning Analytics Research: "Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs."[1] Bichsel, proposes an analytics maturity model used to evaluate the progress in the use of academic and learning analytics. In the progress, there are positive results, but most institutions are below 80% level. Most institutions also scored low for data analytics tools, reporting, and expertise [2]. In addition, a task with the methods of Data Mining and Learning Analytics is analyze them (precision, accuracy, sensitivity, coherence, fitness measures, cosine, confidence, lift, similarity weights) to optimize and adapt them [9]. Learning Analytics (LA) was and continues to be an emerging technology [7]. The time necessary to implement Horizon is one year or less, but how many institutions, teachers, learners and data analytics tools are ready?. The principal aim of this paper is to give mathematical issues of SIA's asymmetric rules for formal approximate to hierarchical clustering in LA.

Learning Analytics (LA) and clustering

Cluster in Learning Analytics is and remains as an emerging method, as shown in the following scientific articles: Papamitsiu [11] to examine the literature on experimental case studies conducted in the domain Learning Analytics and Educational Data Mining, from 2008 to 2013 and to find that in Learning Analytics 60% of literature using classification or cluster, and 40% regression, text mining, association rule mining, social network analysis, discovery with models, visualization or statistics. A recent study [6] show that the current methods used in Learning Analytics are decision tree, clustering, association rules, time sequence analysis and visualization techniques and [6] show that Non-hierarchical algorithm are 73% (K-means, C-means, Fuzzy K-means, K-prototypes ,Fuzzy Clustering) and hierarchical type algorithm are 27% (Agglomerative Clustering, Markov Clustering, Discrete Markov Model). The novelty of the approach is the possibility to use additional options of SIA's asymmetric rules in LA's clustering.

Statistical Implicative Analysis (SIA) and asymmetric rules.

Statistical implicative analysis is a non-symmetric method of analyzing data crossing subjects or objects with variables of any type: Boolean, numerical, modal, vectorial, sequential, interval, fuzzy

---

[1]https://tekri.athabascau.ca/analytics/

and rank[2]. Statistical Implicative Analysis [8] was created for Regis Gras [7], 48 years ago SIA is a statistical theory which provides a group of data analytics tools to extract knowledge. The approach is performed starting from the generation of asymmetric rules [5] similar to dendrograms used in the hierarchical clusters [14]. But can asymmetric rules be used like a hierarchical cluster? An intuitive approximation between asymmetric rules and hierarchical clusters was given in [13], this is a visual perception between simple white and black images, one of the conclusions is that the 69.14% the participants in the experiment agrees or strongly agrees with the kind of grouping presented by the hierarchy trees and asymmetric rules in Statistical Implicative Analysis. In Elia paper [4] is performed a comparative example between hierarchical clustering of variables, implicative statistical analysis and confirmatory factor, the concept of function is addressed by the teaching, analyzing the level of understanding that students present in this type of abstract definitions. The outcomes of the three methods were found to coincide and to complement each other. Anastasiadou in order to study the appropriate approach that a teacher should use when teaching the theory concerning probability distributions, compares two statistical tools principal components analysis and asymmetric rules, components analysis. In the conclusions she writes Hierarchical Clustering of Variables and Implicative Statistical show stable and similar results but each one has its advantages and different prospective [1]. [10] compares the implicit methods, hierarchical clustering, and confirmatory factor analysis in the study of the learning of the geometric figure by 6th graders. The paper concludes that the outcomes of the three methods were found to coincide. Some new possibilities to complement the asymmetric rules are shown in [12], we can use supplementary variables to know what are the subjects, or classes of subjects are more responsible for computed implications, contribution indicates which subjects are more representative of implication and typicality indicates the typical subjects. All previous research shows an approximation between the asymmetric rules of SIA and other hierarchical methods, but they are not formal approximations. In this paper, we want to identify the formal way to demonstrate that symmetric rules can be considered a hierarchical cluster method. We also make contributions about which formal demonstrations to perform and some alternatives.

Math issues[3]

1) Let V be a finite not empty set of binary variables, prove that (V, α) is an indexed hierarchy, where $\alpha = c(a, b) = [1-(-p \log_2 p - (1-p) \log_2(1-p))^2]^{1/2}$ if $p > 0.5$, otherwise $= 0$, $c(a, b)$ is the cohesion of a R-rule $a \rightarrow b$ of degree 1.

2) For all x the binary relation R_x on V, $iR_x j$ if $i, j \epsilon C$, being $\alpha(C) \leq x$, is an equivalence relation

3) Let V be a finite not empty set of binary variables, prove that exist μ, such that (V, μ) is a ultra-metric space.

If 1), 2) and 3) are true, then we can represent (V, μ) by a dendrogram with V-ends

---

[2] https://fr.wikipedia.org/wiki/Analyse_statistique_implicative

References

[1]   ANASTASIADOU, S., 2010. Pre-service teachers' performance on the learn-ing of probability distributions and the role of projects: A multilevel statistical analysis *ASI5*, 21.

[2]   BICHSEL, J., 2012. *Analytics in higher education: Benefits, barriers, progress, and recommendations*. EDUCAUSE Center for Applied Research.

[3]   CUADRAS, C.M., 2007. *Nuevos métodos de análisis multivariante*. CMC Editions.

[4]   ELIA, I. and GAGATSIS, A., 2008. A comparison between the hierarchical clustering of variables, implicative statistical analysis and confirmatory factor analysis. In *Statistical Implicative Analysis* Springer, 131-162.

[5]   GRAS, R., COUTURIER, R., GUILLET, F., and SPAGNOLO, F., 2005. Extraction de règles en incertain par la méthode statistique implicative. *Comptes rendus des 12èmes Rencontres de la Société Francophone de Classification*, 148-151.

[6]   GWO-JEN HWANG, H.-C.C.C.Y., 2017. Objectives, methodologies and research issues of learning analytics. *INTERACTIVE LEARNING ENVIRONMENTS, 2017 25*, 2, 143–146. DOI= http://dx.doi.org/10.1080/10494820.2017.1287338.

[7]   KITCHENHAM, B., PRETORIUS, R., BUDGEN, D., BRERETON, O.P., TURNER, M., NIAZI, M., and LINKMAN, S., 2010. Systematic literature reviews in software engineering–a tertiary study. *Information and Software Technology 52*, 8, 792-805.

[8]   KOTSIANTIS, S. and KANELLOPOULOS, D., 2006. Association rules mining: A recent overview. *GESTS International Transactions on Computer Science and Engineering 32*, 1, 71-82.

[9]   LI, K.C., LAM, H.K., and LAM, S.S., 2015. A Review of Learning Analytics in Educational Research. In *International Conference on Technology in Education* Springer, 173-184.

[10]   MICHAEL, P., ELIA, I., GAGATSIS, A., and KALOGIROU, P., 2010. Examining primary school students' operative apprehension of geometrical figures through a comparison between the hierarchical clustering of variables, implicative statistical analysis and confirmatory factor analysis *ASI5*, 19.

[11]   PAPAMITSIOU, Z.K. and ECONOMIDES, A.A., 2014. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society 17*, 4, 49-64.

[12]   PAZMIÑO-MAJI, R.A., GARCÍA-PEÑALVO, F.J., and CONDE-GONZÁLEZ, M.A., 2016. Approximation of statistical implicative analysis to learning analytics: a systematic review. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* ACM, 355-376.

[13]   PAZMIÑO-MAJI, R.A., GARCÍA-PEÑALVO, F.J., and CONDE-GONZÁLEZ, M.A., 2017. Is it possible to apply Statistical Implicative Analysis in hierarchical cluster Analysis? First issues and answers. In *Congreso Internacional de Ciencia y Tecnología.*, P. GIADE Ed., ESPOCH, Riobamba, Ecuador, 63-66.

[14]   RITSCHARD, G., 2005. De l'usage de la statistique implicative dans les arbres de classification. *Troisieme Rencontre Internationale-Analyse Statistique Implicative*, 305-316.