Predictive and Causal Relationships: Need for Predictive Path Modeling and Probabilistic Networks for Multi-Block Data

Vincenzo Esposito Vinzi[1] & Giorgio Russolillo [2]

[1] *ESSEC Business School, Cergy, France*
E-mail : *vinzi@essec.edu*

[2] *CNAM, Paris, France*
E-mail : *giorgio.russolillo@cnam.fr*

Abstract
Latent Variable Path Models aim at studying the complexity of the real world by taking into account a network of relationships among latent concepts, where each concept is measured by a specific set of observed indicators.
Several (mostly statistical) hypotheses together with a priori theoretical knowledge are usually needed about the measurement models of the concepts, the direction of links between indicators and the latent concepts they are meant to measure, as well as the structure of the network of relationships that link the latent concepts to each other.
In order not to get lost in such a bundle of model structures, we build on knowledge stemming from theory and on conjectures based on experience, but also on optimization criteria, on statistical accuracy and prediction performance, and finally on interpretation rules and tools provided by combining features of different models and methods.

In this talk, we refer to component-based path modeling methods and probabilistic networks that adapt well to real-world dynamics where a priori knowledge happens to be weaker and various model structures are worth considering.

Keywords: latent concepts; network of relationships; linear composites; partial least squares; automatic learning; Bayesian networks.

Introduction to the talk

Covariance-based and component-based approaches to Latent Variable Path Modeling provide nowadays well-established tools for the analysis of a system of relations between observed and unobserved variables as well as of the network of predictive relationships between unobserved variables themselves. Their application is widespread and common in many fields, from Marketing and Information Systems to Genomics and Biology.
Among the component-based approaches, Partial Least Squares Path Modeling (PLS-PM) (Esposito Vinzi et al. 2010, Tenenhaus et al. 2005) is undoubtedly the most commonly and widely used. Differently from the covariance-based approach to Structural Equation Models, PLS-PM is more oriented to prediction than to hypothesis testing and validation. That is why recently researchers start to refer to PLS-PM as a Predictive Path Modeling (Esposito Vinzi & Russolillo, 2012).

However, both covariance-based and component-based approaches require the structure of the relations among the observed and unobserved variables as well as the predictive relations between the unobserved variables to be defined a priori according to theoretical hypotheses of the researcher. Further knowledge may be extracted if induction by automatic learning is merged to the evaluation of probabilistic networks.

Bayesian probabilistic networks (Jensen & Nielsen 2007, Neapolitan 2003) can be used with the objective of discovering and validating a hidden network of relationships between observed variables as well as between eventual underlying factors based on probabilistic causation. BayesiaLab (version 5.4, 2015) is undoubtedly the most representative software package for implementing Bayesian network. It allows to automatic detect associations between variables in order to define groups of highly correlated variables and to define relations between such groups, as well as to analyze a given structure of relations between variables.

During our presentation we will discuss PLS-PM, Bayesian probabilistic networks as implemented in BayesiaLab, and the relations between these two approaches either as two sequential steps of the analysis or as a unified approach where one method is integrated into the other.

References

BayesiaLab 5.4 (2015), www.bayesia.com.

Esposito Vinzi, V., Chin, W.W., Henseler, J. & Wang, H. (Eds.), 2010. Handbook of Partial Least Squares: Concepts, Methods and Applications, Springer Handbooks of Computational Statistics.

Esposito Vinzi, V., & Russolillo, G. (2013). Partial least squares algorithms and methods. Wiley Interdisciplinary Reviews: Computational Statistics, 5(1), 1-19.

Jensen, F.V. & Nielsen, T.D., 2007. Bayesian networks and decision graphs, Springer Verlag.

Neapolitan, R.E., 2003. Learning bayesian networks, Prentice Hall Upper Saddle River, NJ.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.M., Lauro, C., 2005. PLS path modeling. Computational Statistics & Data Analysis, 48(1), 159–205.