

Seminario CMM

Expositor: Cédric HEUCHENNE, Université de Liège, Belgium.

Título: Statistical Matching using KCCA, Super-OM and Autoencoders-CCA.

Abstract:

The potential to study and improve different aspects of our lives is ever growing thanks to the abundance of data available in today's modern society. Scientists and researchers often need to analyze data from different sources; the observations, which only share a subset of the variables, cannot always be paired to detect common individuals. This is the case, for example, when the information required to study a certain phenomenon is coming from different sample surveys. Statistical matching is a common practice to combine these data sets. In this talk, we investigate and extend to statistical matching three methods based on Kernel Canonical Correlation Analysis (KCCA; [2]), Super-Organizing Map (Super-OM; [1]) and Autoencoders-Canonical Correlation Analysis (A-CCA; [3]). These methods are designed to deal with various variable types, sample weights and incompatibilities among categorical variables. We use the 2017 Belgian Statistics on Income and Living Conditions (SILC) and we compare the performance of the proposed statistical matching methods by means of a cross-validation technique, as if the data were available from two separate sources.

[1] Kohonen, T. (1982), Self-organized formation of topologically correct feature map. *Biological Cybernetics*, 43 (1), 59-69.

[2] Lai, P. L. and Fyfe, C. (2000), Kernel and nonlinear canonical correlation analysis. *International Journal of Neural Systems*, 10 (05), 365-377.

[3] Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986), Learning Internal Representations by Error Propagation in Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge: MIT Press, 318-362.

Miércoles 20 abril, 15:00 a 15:45 hrs. Sala Multimedia 6to Piso, Torre Norte, Beauchef 851.

