

ABSTRACTS

16th Seminar D² Seminar Series

Florence Center for Data Science 'Double' Seminar Series

Andrea Barucci - IFAC-CNR Institute of Applied Physics

Title: Exploring Egyptian Hieroglyphs with Convolutional Neural Networks

Abstract: Deep Learning is expanding in every domain of knowledge, allowing specialists to build tools to support their work in fields apparently unrelated to information technology. In this study, we exploit this opportunity by focusing on Egyptian hieroglyphic texts and inscriptions. We investigate the ability of several convolutional neural networks (CNNs) to segment glyphs and classify images of ancient Egyptian hieroglyphs derived from various image datasets. Three well-known CNN architectures (ResNet-50, Inception-v3, and Xception) were considered for classification and trained on the supplied pictures. Furthermore, we constructed a specifically devoted CNN, termed Glyphnet, by changing the architecture of one of the prior networks and customizing its complexity to our classification goal. The suggested Glyphnet outperformed the others in terms of performance, ease of training, and computational savings, as judged by established measures. The hieroglyphs segmentation was faced in parallel, using a deep neural network architecture known as Mask-RCNN. This work shows how the ancient Egyptian hieroglyphs identification task can be supported by the Deep Learning paradigm, laying the foundation for developing novel information tools for automatic documents recognition, classification and, most importantly, the language translation task.

Alessandra Mattei - Department of Statistics, Computer Science, Applications "G. Parenti", University of Florence

Title: Selecting Subpopulations for Causal Inference in Regression Discontinuity Designs (Joint work with Laura Forastiere e Fabrizia Mealli)

Abstract: The Brazil Bolsa Família program is a conditional cash transfer program aimed to reduce short-term poverty by direct cash transfers and to fight long-term poverty by increasing human capital among poor Brazilian people. Eligibility for Bolsa Família benefits depends on a type of cutoff formula, which classifies the Bolsa Família study as a regression discontinuity (RD) design. Extracting causal information from RD studies is challenging. Following Li, Mattei and Mealli (2015) and Branson and Mealli (2019), we formally describe the Bolsa Família RD design as a local randomized experiment within the potential outcome approach. Under this framework, causal inference concerns Brazilian families belonging to some subpopulation where a local overlap assumption, a local SUTVA and a local ignorability assumption hold. We first discuss the potential advantages of this framework, in settings where assumptions are judged plausible, over local regression methods based on continuity assumptions, namely a) it generates treatment effects for subpopulation members rather than local average treatment effects for those at the cutoff only, making the results more easily generalizable; b) it avoids modeling assumptions on the relationship between the running variable and the outcome; c) it allows the treatment assignment mechanism to be random rather than deterministic as in typical RD analyses, so that finite population inference can be used; d) it allows to easily account for discrete running variables. A critical issue of the approach is how to choose subpopulations for which we can draw valid causal inference. We propose to use a Bayesian model-based finite mixture approach to clustering to classify observations into subpopulations where the RD assumptions hold and do not hold on the basis of the observed data. This approach has important advantages: a) it allows to account for the uncertainty about the subpopulation membership, which is typically neglected; b) it does not impose any constraint on the shape of the subpopulation (bandwidth); c) it can be used as a design phase of any analysis; d) it is scalable to high-dimensional settings; e) and it allows to account for rare outcomes. We apply the framework to assess causal effects of the Bolsa Família program on leprosy incidence in 2009, which is a rare outcome, using information on a large sample of Brazilian families who registered in the Single Registry in 2007-2008 for the first time.