Machine learning and network techniques to investigate complex systems with an application to antifraud in the insurance sector

Michele Tumminello
Department of Economics, Business, and Statistics
University of Palermo, Italy
michele.tumminello@unipa.it

Abstract

Machine learning (ML) systems learn and make decisions based on patterns and insights extracted from data–Statistically Validated Networks [1] in this lecture. ML encompasses a variety of methods and techniques, ranging from supervised learning, where models are trained using labeled data (e.g., logistic regression and support vector machines), to unsupervised learning, which involves uncovering hidden structures in unlabeled data (e.g., clustering). As a technique that adapts well to a wide range of data types, ML is increasingly used across diverse sectors such as finance, management, economics, healthcare, and the social sciences, introducing a change of paradigm for analyzing and modeling complex systems–from understanding "why" something happens to predicting "how" and "when" it will occur. In this lecture, I will adopt a bottom-up approach to explain how statistical, machine learning, and network techniques can fruitfully be combined to tackle a complex task, such as identifying organized groups of fraudsters from the analysis of the Antifraud Integrated Archive (AIA) managed by the IVASS. Indeed, the presented methods allow one to learn about preferential patterns of connectivity [2] among subjects, vehicles, and events, by testing the local network structure surrounding each entity against a null hypothesis of random connectivity [3]. As a crucial feature, this null hypothesis suitably takes into consideration the heterogeneity of all the involved entities (individuals, professionals, companies, etc.). One of the most useful outcomes of the procedure is an integrated score of potential fraud-fraudster associated with the joint pair car accident-subject [4], which is currently evaluated and used by the IVASS.

References

[1] Tumminello, M., Miccichè, S., Lillo, F., Piilo, J., Mantegna, R.N. (2011). Statistically validated networks in bipartite complex systems. PLOS ONE, vol. 6(3), e17994.

[2] Tumminello, M., Edling, C., Liljeros, F., Mantegna, R.N., Sarnecki, J. (2013). The Phenomenology of Specialization of Criminal Suspects, PLoS one, 8(5), e64703.

[3] Tumminello, M., Petruzzella, F., Ferrara, C., Miccichè, S. (2021). Anagraphical relationships and crime specialization within Cosa Nostra, Social Networks 64, 29-41.

[4] Tumminello, M., Consiglio, A., Vassallo, P., Cesari, R., Farabullini, F. (2023). Insurance fraud detection: A statistically validated network approach. Journal of Risk and Insurance, 90(2), 381-419.