Stock Market Modelling via Knowledge Graphs

Introduction: Understanding how stocks are mutually related plays a vital role in the risk management of portfolios. Traditional machine learning techniques for this paradigm are often based on the estimates of correlations between assets in highly uncertain market data to minimize portfolio investment risks [1]. However, these matrix based (linear algebra) estimates are often inadequate in capturing the large-dimensional, irregular, and multi-way dependencies in the stock market, as their "flat-view" matrix approaches are ill-equipped to describe their higher-order relations (e.g., business partnerships, supplier dependency, etc.). Instead, a natural way of representing these relations is through the higher-order tensor representation of knowledge graphs, where various nodes (stocks) are connected to each other via multiple edges (relations). A knowledge graph can therefore be naturally analysed through tensor decomposition techniques, which create a very effective latent space representation of each stock in a vector space that characterises the underlying relations. These latent space representations can in turn be used for various financial machine learning applications, such as portfolio optimisation, predictive modelling, and market regime detection.

Tensors: Tensors [2] are multi-linear generalisation of vectors and matrices (scalars are order-0 tensors, vectors are order-1 tensors, and matrices are order-2 tensors). Tensor Decomposition (TD) techniques, such as Canonical Polyadic Decomposition (CPD), are low-rank factorisation methods that can extract latent features from exceedingly large multi-dimensional data, which effectively bypasses the bottlenecks imposed by the Curse of Dimensionality.

Knowledge graphs: Knowledge graphs store facts about relationships between various entities through a set of triplets, (entity_i, entity_j, relation_k), which are often represented as order-3 tensors of size, (N x N x K), where each of the N entities is related to every other entity through K relations. Tensor decomposition approaches provide state-of-the-art performance in analysing knowledge graphs, both in terms of performance and complexity [3].

Project goal: The student will fetch fundamental stock data from Bloomberg Terminals available on campus to construct the knowledge graph and apply tensor decomposition to extract latent space representations of each stock. These will then be used for financial machine learning applications (e.g., portfolio optimization, predictive modelling, etc. based on e.g. S&P500, FTSE100, NASDAQ stock indices) and compared to current state-of-the-art methods.

Useful skills: Python, linear algebra, machine learning, dimensionality reduction, statistical signal processing, financial signal processing.

[1] M. L. De Prado "Advances in financial machine learning". John Wiley & Sons, 2018.
[2] A. Cichocki, D. P. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. IEEE signal processing magazine, 32(2), 145-163, 2015.
[3] M. Nickel, V. Tresp, and H. P. Kriegel. 2012. Factorizing yago: scalable machine learning for linked data. In Proceedings of the 21st international conference on World Wide Web. ACM, 271–280