Multi-fund2Vec: An Algorithm for Finding Similarity Among Mutual Funds

PULKIT CHOUHAN

A project report submitted

in partial fulfillment of the requirement for the award of the degree of

MASTER OF ARTS

IN

Applied Quantitative Finance

May 2025

MADRAS SCHOOL OF ECONOMICS

Chennai- 600025

Abstract

This dissertation presents Multi-Fund2Vec, a multi-stage computational approach to investigating intricate dependencies between mutual fund strategies and underlying asset behaviors. Moving beyond the constraints of conventional fund categorizations and direct holdings analysis, such as high dimensionality and noise, this effort uses a simulation-based approach. A synthetic market is created with regime-dependent assets and correlations. A heterogeneous corpus of synthetic funds (M=450) of different styles and goals is created with optimized portfolio construction . The framework initially determines fund and asset clusters with MDS and K-Means. It proceeds to quantify inter-cluster interactions with Frobenius norms on respective weight submatrices. Interestingly, a weighted bipartite graph is constructed where nodes are these clusters and edge weights are the interaction norms. Node2Vec is used on this graph at the cluster level to learn low-dimensional Multi-Fund2Vec embeddings. This hierarchical process captures higher-order structural similarities between groups, providing a computationally tractable substitute for direct fund-asset graph analysis. Embeddings are tested with V-Measure and Silhouette Scores, illustrating the framework's capability for sophisticated insights into the fund-asset ecosystem structure.