Generating High-Frequency Risk Factor Collections with Transformer

Authors : Wenyan Xu, Rundong Wang, Chen Li, Yonghong Hu, Zhonghua Lu

Abstract : In the field of quantitative trading, it is common to find patterns in short-term volatile trends of the market. These patterns are known as High-Frequency (HF) risk factors, serving as effective indicators of future stock price volatility. However, in the past, these risk factors were usually generated by traditional financial models, and the validity of these risk factors is heavily based on domain-specific knowledge manually added instead of extensive market data. Inspired by symbolic regression (SR), the task of inferring mathematical laws from existing data, we take the extraction of formulaic risk factors from high-frequency trading (HFT) market data as an SR task. In this paper, we challenge the procedure of manually constructing risk factors and propose an end-to-end methodology Intraday Risk Factor Transformer (IRFT), to directly predict the full formulaic factors, constants included. Specifically, we utilize a hybrid symbolic-numeric vocabulary where symbolic tokens denote operators/stock features and numeric tokens denote constants. Then, we train a Transformer model on the HFT dataset to directly generate complete formulaic HF risk factors without relying on the skeleton, which is a parametric function using a pre-defined list of operators – typically, the math operations ($+$, $\times$, $/$) and functions ($\sqrt{x}$, $\log x$, $\cos x$). It determines the general shape of the stock volatility law up to a choice of constants, e.g., $f(x) = \tan(ax+b)$ (x is the stock price). We further refine predicted constants(a,b) using the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) as informed guesses to mitigate non-linear issues. Compared to the 10 approaches in SRBench, which is a living benchmark for SR, IRFT gains a 30% excess investment return on the HS300 and S&P500 dataset, with inference times orders of magnitude faster than theirs in HF risk factor mining tasks.

Keywords : transformer, factor-mining language model, highfrequency risk factor collections

Conference Title : ICDE 2025 : International Conference on Data Engineering

Conference Location : Athens, Greece

Conference Dates : October 21-22, 2025