INTELLIGENT COMMODITY MARKET FORECASTING SYSTEM WITH LOCAL TREND CORRECTION

Poidenko A.
Student, Educational and Scientific Institute of Applied System Analysis, Igor Sikorsky Kyiv Polytechnic Institute
Danylov V.
Professor, Department of Artificial Intelligence, Igor Sikorsky Kyiv Polytechnic Institute
https://doi.org/10.5281/zenodo.20691210

Abstract

The article considers the problem of forecasting highly dynamic commodity markets under the conditions of the "concept drift" effect. The use of recurrent neural networks (LSTM) in combination with the developed algorithm for dynamic correction of the local trend (Label Lag Correction) is investigated. It is shown that the hybrid architecture effectively compensates for the phase lag of predictions that occurs during sharp market shocks. Preliminary fractal analysis of time series allows optimizing the model's hyperparameters. The developed system was tested on real streaming data, demonstrating high accuracy and resistance to market noise. The practical value of the proposed approach is confirmed by participation in the international algorithmic trading competition Mitsui & Co. Commodity Prediction Challenge.

Keywords: forecasting, commodity markets, neural networks, LSTM, concept drift, label lag correction.

1. Introduction

Modern commodity markets are characterized by high volatility and chaotic dynamics. Traditional econometric models and classical machine learning algorithms often prove ineffective due to the "concept drift" phenomenon – a sudden change in the statistical properties of the market under the influence of macroeconomic shocks. The relevance of this study lies in the need to develop new intelligent systems that can not only detect long-term patterns in noisy data but also instantly adapt to structural shifts. Solving this problem will significantly improve the accuracy of algorithmic trading and financial risk management [1, p. 45].

2. Methodology and Research Results

2.1. Data Description and Preprocessing

The experiments were conducted on real data from the international competition Mitsui & Co. Commodity Prediction Challenge [4]. The training sample covers 1961 trading days and contains 557 input features (historical prices, trading volumes, macroeconomic indicators, and cross-market indicators). The target space consists of 424 variables (106 trading pairs × 4 lags). Missing values were filled sequentially using forward-fill and backward-fill, with the remaining gaps replaced by the historical mean of the corresponding feature. Z-standardization was applied for normalization.

2.2. Fractal Analysis

Fractal analysis was performed using the Detrended Fluctuation Analysis (DFA) method. The fractal properties of financial time series are well described in [5]. For the 557 input features, the average Hurst exponent was $H \approx 0.4628$, indicating anti-persistent noise. For the 424 target variables, $H \approx 0.6553$, indicating a trending structure. This "fractal gap" proves that direct mapping $\mathbb{R}^{557} \to \mathbb{R}^{424}$ is unstable. Therefore, an indirect approach was proposed: first, forecast the features (a more stable task due to high dimensionality), then deterministically recalculate the forecast into targets using logarithmic return and spread formulas.

2.3. LSTM Architecture

The core of the intelligent system is a Long Short-Term Memory (LSTM) recurrent neural network [2, p. 1735]. The sliding window size is 50 days. The architecture consists of one LSTM layer with 256 neurons, dropout 0.2 for regularization, and a fully connected layer projecting the output back to the 557-dimensional feature space. Training is performed with the Adam optimizer (learning rate 1e-3) and MSE loss function for 25 epochs. During inference, recursive forecasting is applied: the model predicts the next day, appends the forecast to the window, and repeats 5 times to obtain a 5-day forecast (Figure 1).

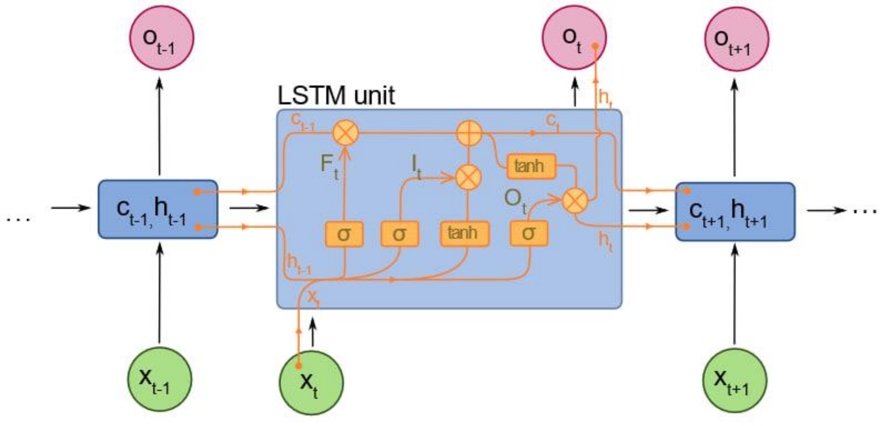

Figure 1 – Architecture of the LSTM cell (based on [2])

2.4. Phase Lag Problem and Label Lag Correction

Testing the baseline LSTM revealed a phase lag effect – the model projected past values into the future and lagged behind real trend reversals. To overcome this, the author's algorithm Label Lag Correction (LLC) was developed. The method consists of a weighted combination of the neural network forecast with the empirical mean of the latest available market values (lags). The final corrected forecast is calculated by:

$$Y_{final} = \alpha \cdot Y_{LSTM} + (1 - \alpha) \cdot Y_{lag\_mean}$$

where α is the confidence hyperparameter. On the validation sample, α was optimized by brute force (values from 0.5 to 0.9, step 0.1). The best Score (competition metric based on Spearman rank correlation) was achieved at $\alpha = 0.7$, meaning the system trusts the deep learning model by 70% and fresh market lags by 30%.

2.5. Testing Results

Comparative testing showed: the baseline LSTM achieved Score = 0.287 on validation. Adding the LLC algorithm increased Score to 0.413 – a 44% improvement. On the final leaderboard (unseen data), the system achieved Score = 0.514, which provided 6th place among over 1711 teams worldwide [3]. The hybrid approach proved effective both on validation and on real unknown data.

3. Conclusions

1. Classical linear methods are ineffective for forecasting commodity markets due to the fractal gap and concept drift.
2. A two-stage architecture is proposed: LSTM → feature forecast → deterministic transformation into targets.
3. The original Label Lag Correction algorithm improves accuracy by 44% and eliminates phase lag.
4. The system ranked 6th in the international Mitsui & Co. Commodity Prediction Challenge.

4. Further Research

Future work includes adaptive selection of the coefficient α depending on the current fractal regime, as well as ensemble learning combining LSTM with gradient boosting (XGBoost).

References:

1. Box G. E. P., Jenkins G. M., Reinsel G. C. *Time Series Analysis: Forecasting and Control*. 5th ed. Hoboken, NJ: John Wiley & Sons, 2015. 712 p.
2. Hochreiter S., Schmidhuber J. Long Short-Term Memory. *Neural Computation*. 1997. Vol. 9, No. 8. P. 1735–1780.
3. Пойденко А. В. 6th Place Solution: Recursive LSTM Feature Forecasting & Lag Correction. Kaggle. 2026. URL: https://www.kaggle.com/competitions/mitsui-commodity-prediction-challenge/writeups/artem777-in-mitsui-6th-Place (Last accessed: 07.06.2026).
4. Mitsui & Co. Commodity Prediction Challenge. Kaggle. 2025. URL: https://www.kaggle.com/competitions/mitsui-commodity-prediction-challenge (Last accessed: 07.06.2026).
5. Mandelbrot B. B. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer, 1997. 551 p.