Proceedings of Machine Learning Research vol 278:1–6, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

Directional Risk-Averse Integrated Loss Strategy for Time Series

Qixuan FU 202220119015@BFSU.EDU.CN
School of Information Science and Technology, Beijing Foreign Studies University, Beijing, 100089 P.R.China

Xi CHEN* CHENXI0109@BFSU.EDU.CN
Chao LIU 202220316001@BFSU.EDU.CN
Xiaosong DING DINGXIAOSONG@BFSU.EDU.CN
International Business School, Beijing Foreign Studies University, Beijing, 100089 P.R.China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

This study focuses on time series forecasting for risk-averse decision-makers, emphasizing trend direction over precise numerical predictions. Traditional methods like MSE fail to capture directional accuracy, which is crucial for risk-averse decisions. While techniques like DILATE improve time alignment, they still rely on numerical metrics. We introduce DRAILS (Directional Risk-Averse Integrated Loss Strategy for time series), a novel loss function that prioritizes directional accuracy while maintaining numerical precision. By incorporating a dynamic reward-penalty system inspired by the newsvendor model, DRAILS minimizes directional errors. Our experiments have shown that DRAILS outperforms existing methods in directional accuracy while maintaining competitive numerical results.

Keywords: Time Series Forecasting, Risk-Averse Decision-making, Directional Accuracy

1. Introduction and related work

1.1. Introduction

Time series forecasting is crucial in fields like finance, energy, and transportation, enabling data-driven decisions amid uncertainty. This paper focuses on risk-averse forecasting, where predicting trend direction (up or down) is more critical than numerical accuracy, as in financial trading. Traditional metrics like MSE and its variants focus on numerical errors but ignore directional accuracy, which is critical in risk-averse decisions. MSE treats overestimation and underestimation equally, failing to account for the importance of trend direction. Recent methods like DILATE (Le Guen and Thome, 2022) improve temporal alignment and shape preservation but still rely on numerical similarity and fail to address directional correctness, which is vital in applications like financial trading where market turning points must be predicted accurately.

To address these limitations, we propose DRAILS (Directional Risk-Averse Integrated Loss Strategy for time series), a novel loss function that prioritizes directional accuracy while maintaining numerical precision. DRAILS enhances DILATE by incorporating a newsvendor-inspired mechanism that adjusts the reward-penalty system based on directional correctness, making it particularly suitable for risk-averse forecasting. Our experiments show that DRAILS outperforms existing methods in directional accuracy while retaining competitive numerical performance.

* Corresponding author

© 2025 Q. FU, X. CHEN, C. LIU & X. DING.

Contributions The main contributions of our work are: (1) We propose DRAILS, a novel loss function for risk-averse time series forecasting that prioritizes directional accuracy without sacrificing numerical precision. (2) We introduce a dynamic reward-penalty mechanism inspired by the newsvendor model, creating an asymmetric loss that adjusts the weight of directional errors based on their real-world impact. (3) Extensive experiments show that models trained with DRAILS outperform state-of-the-art methods in directional accuracy while maintaining competitive numerical performance.

1.2. Related work

Time Series Forecasting Time series forecasting is crucial in finance, weather prediction, and energy demand. Traditional models like ARIMA capture temporal relationships but struggle with high-dimensional data and long-range dependencies. RNNs, including LSTM and GRU, handle complex patterns but are computationally expensive and limited in capturing long-term dependencies. CNNs excel in local patterns but struggle with global ones. N-BEATS (Oreshkin et al., 2021) uses a simple fully connected architecture, performing similarly to Transformer models in forecasting.

Evaluation Metrics Time series forecasting typically uses MSE, which overlooks directional accuracy, key for decision-making in financial markets. Dynamic Time Warping (DTW)(Caillault et al., 2020) measures similarity but doesn't prioritize direction. DILATE improves DTW with TDI but still lacks directional accuracy. The modified Directional Change Error (mDCE)(Nor et al., 2017) focuses on direction but doesn't balance numerical and directional accuracy. Our approach introduces a loss function that prioritizes direction while maintaining numerical performance for risk-sensitive tasks.

2. Methods

Given a time series $X_{1:T} = (X_1, ..., X_T) \in \mathbb{R}^{p \times T}$, our task is to forecast future values while capturing meaningful directional trends. We propose a forecasting framework with two key innovations: a Trend Detection Mechanism to identify significant directional changes, and a Newsvendor-inspired Asymmetric Penalty Function that applies different weights to over- and under-predictions. The proposed approach is illustrated in Figure 1. Section 2.1 reviews the DILATE framework as the foundation for our approach. Section 2.2 presents our Trend Detection Mechanism using sliding windows and significance thresholding. Section 2.3 introduces the Newsvendor-inspired Asymmetric Penalty Function with configurable parameters for different cost scenarios.

2.1. DILATE framework

The DILATE loss combines shape and temporal aspects of time series forecasting errors as shown in Equation 1:

$$\mathcal{L}_{dilate}(\hat{y}_i, y_i) := -\gamma \log \left(\sum_{\mathbf{A} \in \mathcal{A}_{k,k}} e^{-\frac{\langle \mathbf{A}, \alpha \Delta(\hat{y}_i, y_i) + (1-\alpha)\mathbf{\Omega} \rangle}{\gamma}} \right) \tag{1}$$

where $A, \Delta(\hat{y}_i, y_i), \Omega$ are the warping path, dissimilarity distance matrix, and penalization matrix, respectively.

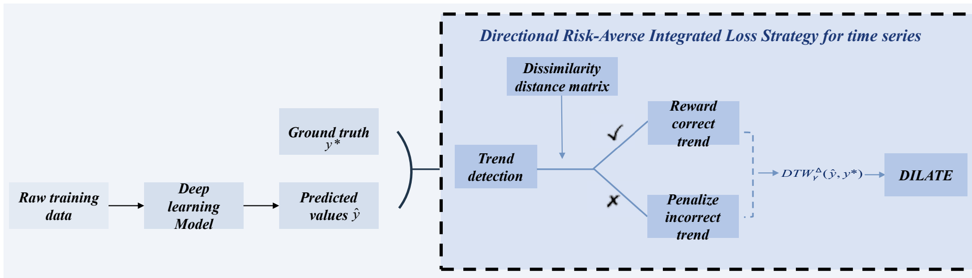

Figure 1: Overview of DRAILS. The framework combines DILATE loss with a newsvendor-inspired asymmetric penalty. It applies different weights based on trend detection using a sliding window, rewarding correct trend predictions and penalizing incorrect ones, balancing shape preservation with directional accuracy for risk-averse forecasting.

DILATE uses a differentiable approximation of Dynamic Time Warping with a temporal penalty, balancing shape similarity and temporal alignment via the hyperparameter α, addressing MSE's limitations for non-stationary time series with abrupt changes. While DILATE excels at time alignment and shape preservation, its limitations in directional accuracy motivate the development of additional components to enhance its performance in our setting.

2.2. Trend Detection Mechanism

While DILATE effectively addresses shape preservation and temporal alignment in time series forecasting, it lacks explicit mechanisms for directional correctness, which is crucial in risk-averse decision scenarios. To overcome this limitation, we propose a window-based Trend Detection Mechanism that focuses on identifying meaningful directional changes.

Given a τ-steps predictions $\hat{y} = (\hat{y}_{T+1}, \dots, \hat{y}_{T+\tau}) \in \mathbb{R}^{d \times \tau}$ for time series X, we define the actual difference at each time step i as $\Delta_{\text{target},i} = x_{i+ws} - x_i$, where ws is the window size used to measure trend changes by capturing differences over a period. Similarly, the predicted difference is given by $\Delta_{\text{pred},i} = \hat{y}_{i+ws} - x_i$. We then assess directional agreement with threshold θ to filter noise:

$$t_i = \begin{cases} 1, & \text{if } |\Delta_{\text{target},i}| > \theta \text{ and } \operatorname{sign}(\Delta_{\text{pred},i}) = \operatorname{sign}(\Delta_{\text{target},i}) \\ -1, & \text{if } |\Delta_{\text{target},i}| > \theta \text{ and } \operatorname{sign}(\Delta_{\text{pred},i}) \neq \operatorname{sign}(\Delta_{\text{target},i}) \\ 0, & \text{if } |\Delta_{\text{target},i}| \leq \theta \end{cases} \tag{2}$$

This mechanism reduces noise sensitivity through window-based comparison, filters minor fluctuations via thresholding, prioritizes directional accuracy over magnitude, and adapts to various time scales with adjustable windows. The resulting ternary signal enhances decision support where trend direction is critical.

2.3. Newsvendor-inspired Asymmetric Penalty Function

Traditional distance metrics fail to reflect the asymmetric costs of directional prediction errors. To address this limitation, we employ a newsvendor model-inspired approach with an imbalanced penalty mechanism to guide models toward risk-averse predictions.

Based on the dissimilarity distance matrix $\Delta(\hat{y}_i, y_i)$, we apply the Trend Detection Mechanism to identify meaningful correct and incorrect prediction trends. We then modify the dissimilarity distances according to the trend identification results:

$$\text{dist}'_{i,j} = \text{dist}_{i,j} \cdot (c_u \cdot \mathbf{1}_{\text{correct}} + c_o \cdot \mathbf{1}_{\text{incorrect}} + \mathbf{1}_{\text{otherwise}}) \tag{3}$$

where $c_u < c_o$ in risk-averse scenarios, effectively rewarding correct trend predictions while penalizing incorrect ones.

This module guides the model towards correct directional predictions, with the built-in trend detection mechanism intelligently identifying meaningful patterns. Penalty parameters can be dynamically adjusted based on business requirements and risk preferences, making the approach adaptable to different forecasting scenarios while maintaining computational efficiency.

3. Experiments

3.1. Experimental setup

Datasets: We conducted extensive experiments on four real-world time series datasets to evaluate our method, and compared it with MSE to assess its effectiveness and robustness.

Traffic: Hourly road occupancy recorded over 48 months from 2015 to 2016 by the California Department of Transportation. We use the last univariate series for training.

Electricity: Electricity consumption at 15-minute intervals from 2012 to 2014 for 321 clients. We convert it to hourly data and train with the last client's series.

Exchange Rate: Daily exchange rates from 1990 to 2016 for eight countries. We use data from Singapore for training.

Temperatures[1]**:** Minimum daily temperatures from 1981 to 1990 in Melbourne, Australia. The data includes 3650 observations in degrees Celsius.

In our experiments, we split each dataset into 60% training, 20% validation, and 20% testing. A sliding-window approach is used for model training, with each input sequence consisting of the previous 48 time steps to predict the next 24 steps.

Network architectures and training: For multi-step forecasting, we employ two neural network architectures: a Seq2Seq model (Zhou et al., 2022) with single-layer GRU (Cahuantzi et al., 2023) of 128 units and N-BEATS (Oreshkin et al., 2021). For N-BEATS, we implement trend and seasonality stacks for Traffic, Electricity, and Temperature datasets, while using dual generic stacks for Exchange Rate dataset. All models are trained in PyTorch with ADAM optimizer and Early Stopping, with a maximum of 1000 epochs.

3.2. Evaluation Metrics

In this experiment, we use the modified Mean Directional Accuracy (mMDA) (Nor et al., 2017) to evaluate the accuracy of directional predictions. mMDA measures the consistency between predicted and actual directional trends over k steps, with higher values indicating better accuracy. The formula is given in Equation 4, with k varying based on the dataset.

1. https://www.kaggle.com/datasets/samfaraday/daily-minimum-temperatures-in-me

$$\text{mMDA} = \frac{1}{n-k+1} \sum_{T^*=t+1}^{t+n-k+1} \left[I\left(A_{m,T^*,k} F_{m,T^*,k} > 0\right) - I\left(A_{m,T^*,k} F_{m,T^*,k} < 0\right) \right] \tag{4}$$

where $A_{m,T^*,k} = y_{T^*+k} - y_{T^*}$ and $F_{m,T^*,k} = \hat{y}_{T^*+k} - \hat{y}_{T^*}$ represent the continuous directional changes of the actual and predicted value over k steps, respectively, and $I(\cdot)$ is the indicator function, which equals 1 when the condition is satisfied and 0 otherwise. And we have converted the original [-1,1] range to [0,1] for easier interpretation.

3.3. Experimental Results and Analysis

Table 1 shows that DRAILS outperforms both GRU and N-BEATS models in long-term trend prediction across all datasets. Notably, DRAILS improved forecasting accuracy by up to **5.38%** on the Temperature dataset and **2.92%** on the Electricity dataset, both of which feature strong long-term trends. Even for the Exchange Rate dataset, which lacks clear periodicity, DRAILS achieved notable performance improvements of 1.60% and 0.82% for GRU and N-BEATS, respectively.

Table 1: Performance evaluation of two forecasting models across four real-world datasets. Numbers in parentheses indicate the size of the mMDA steps.

Models	Seq2Seq GRU		N-BEATS	
Methods	MSE	DRAILS(ours)	MSE	DRAILS(ours)
Traffic	0.9613(9)	0.9627(9)	0.9760(9)	0.9807(9)
Electricity	0.8147(7)	0.8385(7)	0.8212(7)	0.8391(7)
Temperature	0.5060(9)	0.5273(9)	0.5018(7)	0.5288(7)
Exchange	0.4990(4)	0.5070(4)	0.4890(4)	0.4930(4)

Figure 2 highlights DRAILS's superior performance in capturing long-term trends. For instance, on the Temperature dataset, DRAILS more accurately captures key inflection points compared to MSE-optimized models, providing more reliable directional forecasts, crucial for applications such as energy demand and market trend prediction.

In real-world applications like energy forecasting, temperature analysis, and financial prediction, capturing directional changes is more critical than predicting exact values. DRAILS improves directional accuracy, offering an effective solution.

4. Conclusion and future work

In this work, we introduced DRAILS, a novel loss function designed for risk-averse decision-making in time series forecasting. DRAILS prioritizes directional accuracy over numerical precision, overcoming the limitations of traditional metrics like MSE. Our experiments show that DRAILS outperforms existing methods in directional accuracy while maintaining competitive numerical performance, making it ideal for risk-sensitive applications like financial trading. For future work, we plan to integrate DRAILS with advanced deep learning models for multi-step forecasting and extend its application to other risk-sensitive domains like healthcare and supply chain management. Additionally, we will validate DRAILS on large-scale datasets to assess its scalability and robustness in real-world scenarios.

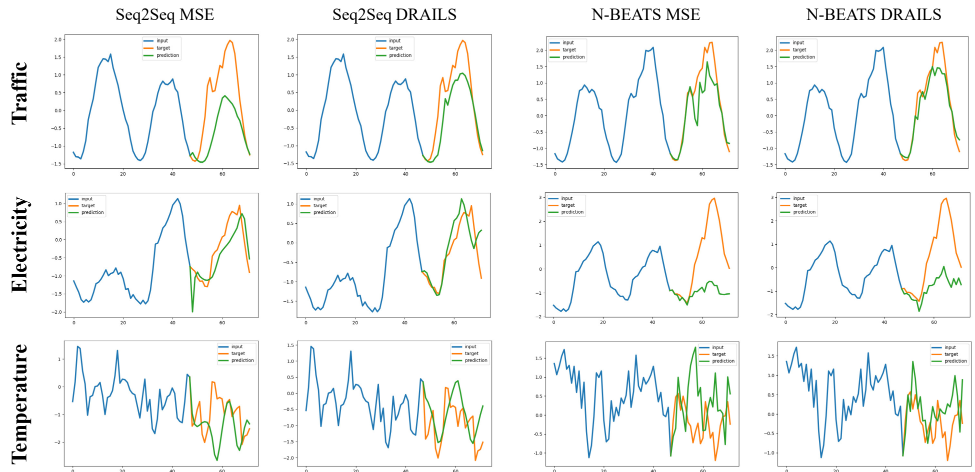

Figure 2: Qualitative forecasting results on the three real-world datasets show that DRAILS significantly outperforms MSE in capturing time series trends, regardless of the model (GRU or N-BEATS).

Acknowledgments

This work is supported by Fundamental Research Funds for Central Universities (2025ZZ009, 2023JJ002).

References

Roberto Cahuantzi, Xinye Chen, and Stefan Güttel. A comparison of lstm and gru networks for learning symbolic sequences. In *Science and Information Conference*, pages 771–785. Springer, 2023.

Émilie Poisson Caillault, Alain Lefebvre, André Bigand, et al. Dynamic time warping-based imputation for univariate time series data. *Pattern Recognition Letters*, 139:139–147, 2020.

Vincent Le Guen and Nicolas Thome. Deep time series forecasting with shape and temporal criteria. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):342–355, 2022.

Maria Elena Nor, Mohd Saifullah Rusiman, Nurul Asmaa Izzati Mohamad, and Muhammad Hisyam Lee. Directional change error evaluation in time series forecasting. In *AIP Conference Proceedings*, volume 1830. AIP Publishing, 2017.

Boris N Oreshkin, Grzegorz Dudek, Paweł Pełka, and Ekaterina Turkina. N-beats neural network for mid-term electricity load forecasting. *Applied Energy*, 293:116918, 2021.

Lijian Zhou, Lijun Wang, Zhiang Zhao, Yuwei Liu, and Xiwu Liu. A seq2seq model improved by transcendental learning and imaged sequence samples for porosity prediction. *Mathematics*, 11 (1):39, 2022.