ECE239AS: Paper Review

Name: Ellen Wei

Paper: **Stable Neural Stochastic Differential Equations in Analyzing Irregular Time Series Data, 2024**

Link: https://arxiv.org/abs/2402.14989

Interest in this paper

I am interested in this paper because it addresses time-series data and the issue of missing data, a common problem with significant impact. I wanted to learn more about how classic deep learning models compare to Neural SDEs, which are not commonly introduced. This study applies concepts from our class on stochastic equations, delving deeper into drift and diffusion functions. I am particularly interested in the efficacy of the proposed methods in handling real-world irregular time series data and maintaining stability as the missing rate increases. With a background in time series analysis, statistics, and experience working with statistical models in finance, this paper aligns well with my interests and expertise. Additionally, it explores areas related to my background that I have not previously delved into in depth, offering a valuable opportunity to expand my knowledge and skills in a new and challenging domain.

Mathematical Framework and Rigorous Results

The paper's methodology includes several rigorous results and mathematical concepts, relevant to our stochastic systems class:

1. **Probability Space and SDEs**: The paper builds on the concept of probability spaces and stochastic processes, introducing SDEs with drift and diffusion terms designed to handle irregular data.
2. **Existence and Uniqueness**: The authors prove the existence and uniqueness of solutions for the proposed Neural SDEs, leveraging conditions like Lipschitz continuity and linear growth conditions.
3. **Robustness Analysis**: The robustness of the models under distribution shift is analyzed using stochastic stability theorems. The non-asymptotic upper bounds provided in Theorem 3.5 and 3.6 demonstrate that the proposed Neural SDEs maintain performance even with varying degrees of distribution shifts (denoted by ρ).

Connection to Decision Making in Stochastic Systems Class

The rigorous results in this paper, particularly the stability and robustness under distribution shift, directly connect to several topics from our class:

- **Lecture 1-5**: The foundational concepts of probability spaces, random variables, distribution functions, and convergence are crucial in understanding the underlying mathematical framework of Neural SDEs.
- **Lecture 6-9**: The principles of optimality and dynamic programming are mirrored in the model's ability to handle time-inhomogeneous data and maintain stability.
- **Lecture 10-14**: Stochastic integration and Brownian motion play a significant role in defining and solving the Neural SDEs.
- **Lecture 19**: The analysis of stochastic stability and Lyapunov theory in the paper directly relates to our discussions on stochastic stability in class.

Paper Summary

Introduction

Neural ODEs use neural networks with ODE solvers to learn continuous latent representations through parameterized vector fields. While ODEs involve ordinary derivatives, Neural SDEs extend this concept by incorporating a diffusion term, requiring careful design of drift and diffusion functions to maintain stability and enhance performance. Potential issues with Neural SDEs include the absence of strong solutions, stochastic destabilization, and unstable Euler discretization. This paper proposes three classes of Neural SDEs—Langevin-type SDE, Linear Noise SDE, and Geometric SDE—assessed on four benchmark datasets for interpolation, forecasting, and classification tasks, and analyzed for robustness with 30 public datasets under different missing rates, aiming to evaluate their efficacy in handling real-world irregular time series data.

Classic deep learning models like RNNs, LSTMs, and GRUs treat time series as consecutive discrete subsets, struggling with irregularly-sampled or partially-observed data. In contrast, Neural Differential Equation (NDE)-based models, including Neural ODEs, learn continuous-time dynamics and the underlying temporal structure through continuous latent state representations parameterized by neural networks. Neural controlled differential equations (Neural CDEs) introduced by Kidger et al. (2020) extend RNNs to represent irregular time series using controlled paths.

Neural SDEs, an extension of Neural ODEs, focus on gradient computation, variational inference for latent spaces, and uncertainty quantification. Despite advancements, unresolved issues remain in modeling and robustness under distribution shifts. This study extends Neural SDEs in two directions: optimal selection of drift and diffusion functions, and addressing the robustness demands of irregular time series data prone to distribution shifts. The study introduces three classes of Neural SDEs trained on theoretically well-defined SDEs to capture complex dynamics and improve robustness under distribution shifts. These models demonstrate the existence and uniqueness of solutions, particularly with Neural Geometric SDEs sharing properties with deep ReLU networks. The proposed Neural SDEs maintain performance under distribution shifts due to missing data and effectively prevent overfitting. Extensive experiments on four benchmark datasets and analysis with 30 public datasets under varying missing rates show powerful empirical performance and robustness to missing data.

Preliminaries

The paper discusses key foundational concepts necessary for understanding Neural SDEs. These include the notion of a probability space and the Wasserstein distance, along with explanations of Neural ODEs and Neural CDEs. The limitations of naive Neural SDEs are highlighted, such as the absence of unique strong solutions, stochastic destabilization, and unstable Euler discretization. These limitations underscore the need for a careful design of the diffusion term to enhance the stability and efficacy of Neural SDEs. The performance of Neural SDEs is significantly influenced by the choice of drift and diffusion functions, an area that has been underexplored and remains poorly understood.

Methodology

The study proposes three types of Neural SDEs:

1. **Langevin-type SDE**: Features a drift function that is not explicitly dependent on time. This type is popular in MCMC and stochastic optimization but has not been explored for Neural SDEs.

   $$dz(t) = \gamma(z(t);\ \theta_\gamma)\, dt + \sigma(t; \theta_\sigma) dW(t) \text{ with } z(0) = h(x; \theta_h)$$

2. **Linear Noise SDE**: Involves a drift function that is time-dependent.

   $$dz(t) = \gamma(t,\ z(t);\ \theta_\gamma) dt + \sigma(t; \theta_\sigma) z(t) dW(t) \text{ with } z(0) = h(x; \theta_h)$$

3. **Geometric SDE**: Characterized by an exponential form, it shares properties with deep ReLU networks, such as a unique positive solution and an absorbing state of 0, which are not present in Neural LSDE and Neural LNSDE.

$$\frac{dz(t)}{z(t)} = \gamma(t,\ z(t);\ \theta_{\gamma})dt\ +\ \sigma(t;\ \theta_{\sigma})dW(t) \text{ with } z(0)\ =\ h(x; \theta_{h})$$

Where the initial condition $h(\cdot;\ \theta_{h})$ is an affine function with parameter θ_{h}, the drift function $\gamma(\cdot;\ \theta_{\gamma})$ is a neural network with parameter θ_{γ} and the diffusion function $\sigma(\cdot;\ \theta_{\sigma})$ is a neural network θ_{σ}

The proposed Neural SDEs are based on certain assumptions:

- Neural networks with activation functions like tanh, ReLU, and sigmoid generally satisfy the Lipschitz continuity condition.
- These networks also satisfy the linear growth condition, ensuring the existence and uniqueness of solutions for SDEs.
- It is assumed that the solutions to SDEs are uniformly bounded, and a diffusion approximation is applicable.
- The SDEs have unique strong solutions.

The robustness of the proposed Neural SDEs under distribution shift is heavily analyzed using stochastic stability. Theorem 3.5 (Robustness of Neural LSDE) shows that under the given assumptions, the Neural LSDE model's output distributions are robust to pertubations in the input data. The bound on the Wassertein distances W1 and W2 provide a quantitative measure. In fact, the exponential decay terms indicate that the effect of perturbations diminish over time, where it is influenced by the degree of distribution shift ρ.

Theorem 3.6 (Robustness of Neural LNSDE and Neural GSDE) shows that under the given assumptions that where the noise term is sufficiently strong relative to the neural network's sensitivity, as T becomes large enough, the bound on the Wasserertein distance W1 between the output distributions decreases exponentially over time. This demonstrates that the proposed Neural SDEs maintain robust performance even when the input distribution shifts. Experimental findings support these theoretical results.

To further improve empirical performance and effectively capture sequential observations, the authors propose incorporating a controlled path $\bar{z(t)}$ in a nonlinear manner. This path replaces $z(t)$ in the drift functions of the proposed Neural SDEs.

$$\bar{z}(t)\ -\ \zeta(t, z(t),\ X(t);\ \theta_{\zeta}$$

The effectiveness of this approach is confirmed through an ablation study, showing that when combined with , the proposed Neural SDEs have unique strong solutions.

Experiments

Superior Performance with Regular and Irregular Time Series Data: The experiments included interpolation and classification tasks using three real-world datasets: PhysioNet Mortality, PhysioNet Sepsis, and Speech Commands. For forecasting, the MuJoCo dataset was utilized.

- **PhysioNet Mortality**: Multivariate time series data from ICU records with irregular measurements taken within the first 48 hours of admission.
- **PhysioNet Sepsis**: ICU patient cases with 34 time-dependent variables aimed at classifying cases according to the sepsis-3 criteria, representing irregular time series data.
- **Speech Commands**: 1-second long audio data with 35 distinct spoken words with background noise, a classification problem.

The study compared the performance of RNN-based models (RNN, LSTM, GRU, GRU-Δt, GRU-D), attention-based models (MTAN), Neural ODEs (Neural ODE, GRU-ODE, ODE-RNN, ODE-LSTM, Latent-ODE, Augmented-ODE, ACE-NODE), Neural CDEs (Neural RDE, ANCDE, EXIT, LEAP), and Neural SDEs (Neural SDE, Latent SDE).

Results:

- **Interpolation**: Neural SDEs consistently outperformed all benchmark models at all observed time point settings, with Neural LSDE and LNSDE consistently performing the best.
- **Classification on PhysioNet Sepsis**: Neural SDEs outperformed all other methods in classification tasks, performing well even without observation interval (OI) enhancements, unlike CDE-based methods.
- **Classification on Speech Commands**: Neural SDEs performed better, especially LSDE and LNSDE, though they were less memory-efficient compared to CDE-based methods.

Robustness to Missing Data: The robustness of the proposed Neural SDEs was evaluated on 30 datasets (15 univariate, 15 multivariate) from the University of East Anglia (UEA) and the University of California Riverside (UCR) Time Series Classification Repository. The evaluation was conducted under regular (0% missing rate) and three missing rate conditions (30%, 50%, 70%).

Results:

- Neural SDEs consistently achieved top-tier performance in accuracy across different missing rates, maintaining stable accuracy even with increased missing rates.
- Traditional RNN-based models experienced noticeable performance drops as the missing rate increased.

- The proposed methods demonstrated stability and overcame the limitations of naive Neural SDEs.

Ablation Study: Ablation experiments were conducted to show the performance impact of incorporating a controlled path into the drift function and using a nonlinear neural network for the diffusion function.

Results:

- The study highlighted the effectiveness of a nonlinear neural network compared to an affine function in the diffusion function.
- There was a fundamental performance difference between the three proposed Neural SDEs and the naive Neural SDE, with Neural LSDE and Neural LNSDE performing exceptionally well.

Conclusion

The aim of the study was to capture complex dynamics and improve stability in time series data using theoretically well-defined SDEs. The proposed Neural SDEs demonstrated robustness under distribution shifts and effectiveness in handling real-world irregular time series data through extensive experiments. These models achieved state-of-the-art results across various tasks, significantly enhancing the stability of Neural SDE training and improving classification performance, although they required more computational resources compared to Neural CDE-based models.