9th March 2023
ID: 1570879196

Mr. Ammar Raneez
Mr. Torin wirasingha

Informatics Institute of Technology, Sri Lanka

A Review On Breaking the Limits of Time Series Forecasting Algorithms

Agenda

Problem Background

Time Series Forecasting

- TS forecasting is a significant business issue and an area where ML could create an impact.
- Although ML and DL have outperformed classical approaches for NLP and computer vision, the domain of TS still seems to be a struggle compared to classical statistical methodologies.

Existing TS forecasting systems cannot adapt to incoming data streams with unexpected changes and characteristics that are different from the data on which they were trained.

Existing Work

Time series forecasting algorithms

- Existing forecasting solutions are all implemented using traditional deep neural net architectures (ex: LSTM, RNNs, GRU, etc.)

Existing work

- Neural ordinary differential equations [21].
- Liquid Time-Constant networks [23], solved this to some extent.

There's a limitation here: areas with greater randomness - instantaneous, miniscule changes cannot be modelled.

Existing Work - Statistical Algorithms

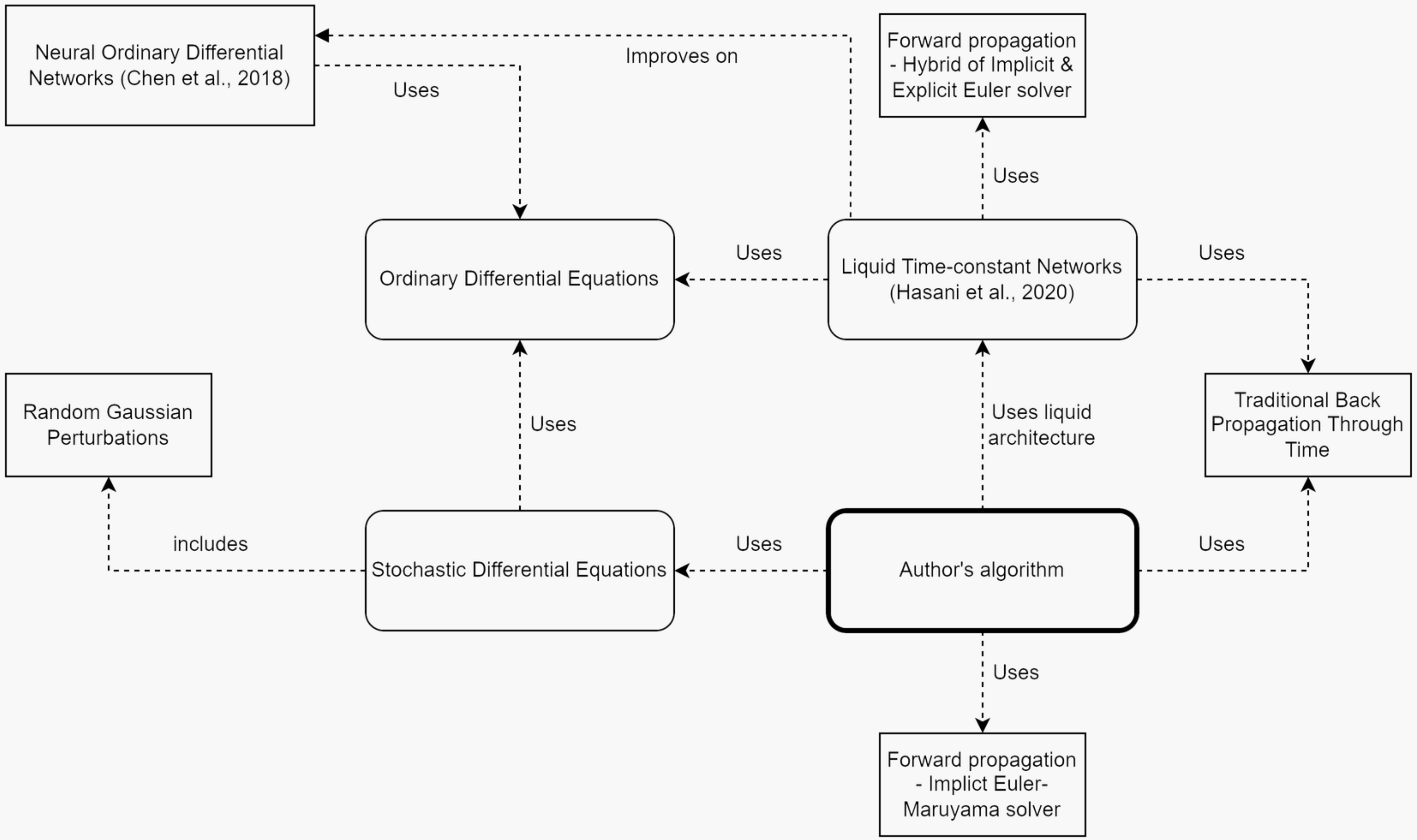
Ref.	Brief	Improvements/Contribution	Limitations/Future work
Statistical-based forecasting algorithms			
[5]	**ARIMA**. A statistical analysis model for understanding the dataset or predicting future trends. This model depends on past values to predict the future and uses lagged moving averages to smoothen the data.	Improved performance for TS forecasting data that correlate with values ahead of time.	Does not handle well with nonlinear data and long-term forecasting. Furthermore, it performs best on univariate analysis and cannot capture data volatility.
[7]	**GARCH**. A modeling technique that specializes in predicting volatility in data.	Captures volatility in datasets and boasts significant performance improvements in the family of statistical forecasting algorithms.	Needs to improve interpretability and adaptability.
[9]	**Prophet**. A modular regression model with interpretable parameters. These parameters can be adjusted according to the problem by domain experts, similar to ARIMA.	Solves forecasting at scale, where scale refers to three types. 1) A large number of people forecasting. 2) A large variety of problems. 3) A large number of forecasts being created.	It uses simple and weak assumptions and produces much poorer performance than ARIMA. And it does not model relationships between the past and future.

Existing Work - DL Algorithms

DL-based forecasting algorithms			
[13]	**LSTM**. An algorithm that learns to bridge minimal time lags by enforcing constant error flows. It learns much faster, creates more successful runs, and can solve complex tasks that have not been solved before.	Improved performance for short-sequence predictions. Overcame error back-flow problems present in conventional BPTT, where they tended to blow up or vanish.	Prediction capacity limits long sequence performance, where the MSE and RMSE rise unacceptably. Therefore, there are better solutions for predictions of the distant future. They are also prone to overfitting.
[31]	**GRU**. Similar architecture to that of LSTMs but combine the 'forget' and 'input' gates to create two gates, 'reset' and 'update,' instead of the three found in LSTMs.	Solve the vanishing gradient problem in RNNs as LSTMs, but also consume less memory and run faster.	Suitable for problems with smaller datasets and tend to be less accurate for datasets with larger sequences.
[17]	**N-BEATS**. An architecture that solves the univariate time series point forecasting problem. It carries some benefits, some of which are being understandable, easily applicable to multiple other fields, and being fast to train.	Outperformed the M4 competition winner of the previous year and improved the statistical benchmark forecast.	Tailored specifically for univariate TS analysis, therefore, would perform poorly on multivariate analysis. Additionally, Meta-learning is speculated to be a reason for the performance and must be investigated.
[19]	**TFT**. An attention-based architecture that solves multi-horizon forecasting with interpretability of the used inputs.	Demonstrate significant performance improvements over set benchmarks for a variety of datasets.	Training and inference times are expensive and require moderately extensive resources. Hardware optimizations can reduce these.
[23]	**LTC**. A novel formulation of the NODE architecture. Boasts superior expressivity that is capable of adapting to unforeseen changes.	Surpassed traditional DL and statistical models and overcame the underwhelming performance of other NODE architectures.	It cannot model uncertainty and is computationally intensive.

Proposed Algorithm Design

Stochastic differential equations can be used instead of ordinary differential equations!



Training These Networks

$$\boxed{\frac{dx(t)}{dt} = -\left[\frac{1}{\tau} + f(x(t), I(t), t, \theta) - \sigma B\right] x(t) + f(x(t), I(t), t, \theta) A}$$

To solve the state of these SDEs, we propose using implicit Euler-Maruyama solver.

And for backpropogation we propose using the traditional Backpropogation Through Time approach.

Algorithm Complexity

	BPTT	**Adjoint sensitivity**
Time	O(L)	**O(LlogL)**
Memory	O(L)	**O(1)**
Forward accuracy	High	High
Backward accuracy	**High**	Low

Future Work

Utilize different noise.

Evaluate other SDE solvers.

References

[5] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time series analysis: forecasting and control, 5th ed. John Wiley & Sons, 2015.

[7] R. F. Engle, “Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation,” Econometrica, vol. 50 iss. 4, pp. 987, Jul 1982.

[9] S. J. Taylor and B. Letham, “Forecasting at scale,” PeerJ Preprints. Rep. 5, 2017.

[13] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9 iss. 8, pp. 1735–1780, Nov 1997.

[17] B. N. Oreshkin, D. Carpov, N. Chapados and Y. Bengio, “N-BEATS: Neural basis expansion analysis for interpretable time series forecasting,” arXiv [Online], Feb 20 2020. Available: https://arxiv.org/abs/1905.10437.

References

[19] B. Lim, S. O. Arik, N. Loeff and T. Pfister, “Temporal Fusion Transformers for interpretable multi-horizon time series forecasting,” International Journal of Forecasting, vol. 37, iss. 4, pp. 1748-1764, Jun 2021.

[21] R. T. K. Chen, Y. Rubanova, J. Bettencourt and D. Duvenaud, “Neural Ordinary Differential Equations,” arXiv [Online], Dec 14 2019. Available: https://arxiv.org/abs/1806.07366.

[23] R. Hasani, M. Lechner, A. Amini, D. Rus and R. Grosu, “Liquid Time-constant networks,” arXiv [Online], Dec 14 2020. Available: https://arxiv.org/abs/2006.04439.

[31] K. Cho, et al., Learning phrase representations using RNN encoderdecoder for statistical machine translation, Sep 2014.

Thank you!