Feature Programming for Multivariate Time Series Prediction

Alex Reneau*, Jerry Hu*, Chenwei Xu, Weijian Li, Ammar Gilani, Han Liu

Northwestern University, Computer Science

https://arxiv.org/abs/2306.06252

ICML 2023

Introduction

Problem: the challenges of feature engineering for noisy multivariate time series in the regression setting.

- Not Generic Enough: too purpose/task/domain-specific
- Handling Inherent Noise

Proposal: Feature Programming, a generic framework for programmable time series feature engineering.

- Automatically generates a large number of predictive features
- Allows users to incorporate their inductive bias with minimal effort
- Theoretically grounded with model-based fundamental principles

Introduction

Problem: the challenges of feature engineering for noisy multivariate time series in the regression setting.

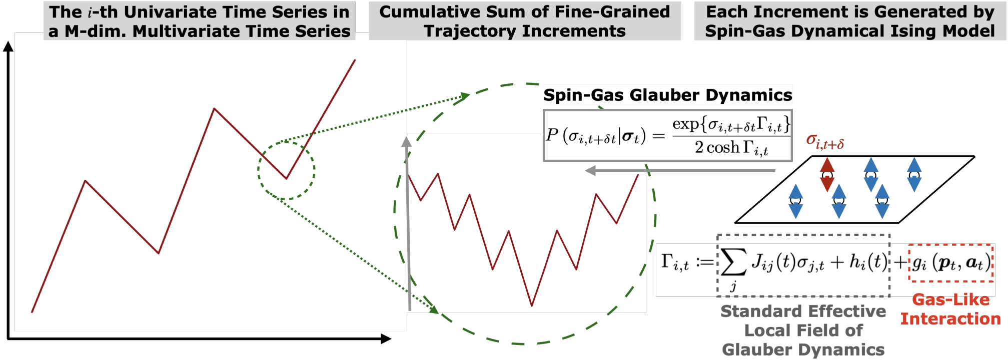
- Not Generic Enough: too purpose/task/domain-specific
- Handling Inherent Noise

Proposal: Feature Programming, a generic framework for programmable time series feature engineering.

- Automatically generates a large number of predictive features
- Allows users to incorporate their inductive bias with minimal effort
- Theoretically grounded with model-based fundamental principles

Motivation: Spin-Gas Dynamical Ising Model

We propose a novel spin-gas dynamical Ising model as the fundamental mechanism for time series generation from the fine-grained perspective.



- Any time series as a cumulative sum of fine-grained trajectory increments.
- Motivates a parsimonious set of operators that summarize multivariate time series in an abstract fashion: Difference, Window, and Shift operators.

Motivated Operators: Difference, Window, and Shift

Difference Operator: generalized derivative operation on any two series.

- performs series-wise subtraction between any two time series.
- generates curvature-like features (e.g., the momentum and acceleration.)
- characterizes features (both basic and extended) into three hierarchical classes based on their order of derivative: 0th-, 1st-, and 2nd- order features.

Window Operator: generates summary statistics for the given lookback window.

- distills information from multiple resolutions using de-noised summary statistics, such as max, min, and mean.

Shift Operator: incorporates autoregressive information.

This set of operators is versatile and complete enough to serve as building blocks for programmable feature generation:

- Motivated by the physics model from a fine-grained path-integral formulation.
- Theoretically grounded with a univariate exponential family construction.

Motivated Operators: Difference, Window, and Shift

Difference Operator: generalized derivative operation on any two series.

- performs series-wise subtraction between any two time series.
- generates curvature-like features (e.g., the momentum and acceleration.)
- characterizes features (both basic and extended) into three hierarchical classes based on their order of derivative: 0th-, 1st-, and 2nd- order features.

Window Operator: generates summary statistics for the given lookback window.

- distills information from multiple resolutions using de-noised summary statistics, such as max, min, and mean.

Shift Operator: incorporates autoregressive information.

This set of operators is versatile and complete enough to serve as building blocks for programmable feature generation:

- Motivated by the physics model from a fine-grained path-integral formulation.
- Theoretically grounded with a univariate exponential family construction.

Motivated Operators: Difference, Window, and Shift

Difference Operator: generalized derivative operation on any two series.

- performs series-wise subtraction between any two time series.
- generates curvature-like features (e.g., the momentum and acceleration.)
- characterizes features (both basic and extended) into three hierarchical classes based on their order of derivative: 0th-, 1st-, and 2nd- order features.

Window Operator: generates summary statistics for the given lookback window.

- distills information from multiple resolutions using de-noised summary statistics, such as max, min, and mean.

Shift Operator: incorporates autoregressive information.

This set of operators is versatile and complete enough to serve as building blocks for programmable feature generation:

- Motivated by the physics model from a fine-grained path-integral formulation.
- Theoretically grounded with a univariate exponential family construction.

Motivated Operators: Difference, Window, and Shift

Difference Operator: generalized derivative operation on any two series.

- performs series-wise subtraction between any two time series.
- generates curvature-like features (e.g., the momentum and acceleration.)
- characterizes features (both basic and extended) into three hierarchical classes based on their order of derivative: 0th-, 1st-, and 2nd- order features.

Window Operator: generates summary statistics for the given lookback window.

- distills information from multiple resolutions using de-noised summary statistics, such as max, min, and mean.

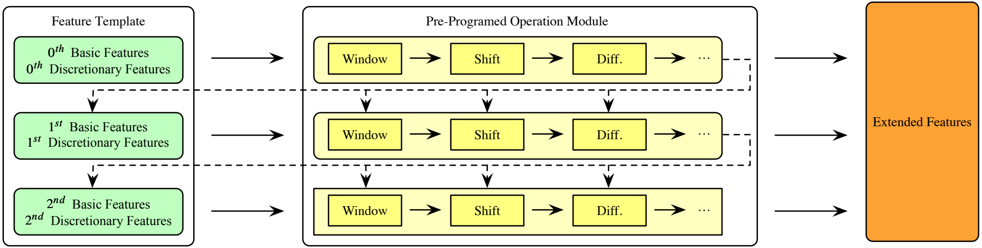
Shift Operator: incorporates autoregressive information.

This set of operators is versatile and complete enough to serve as building blocks for programmable feature generation:

- Motivated by the physics model from a fine-grained path-integral formulation.
- Theoretically grounded with a univariate exponential family construction.

Operators

Overview of a Feature Programming Framework



We propose a three-step pipeline for feature generation:

1. Design a customized three-level feature template: including both raw features from the data and user-specified (discretionary) features at each level.
2. Implementing a programmable operation module consisting of pre-specified operations by the user based on the proposed operator combinations.
3. Automatic generation of a large number of predictive features based on the template and operation module, following pre-specified feature flows (arrows).

Experimental Studies: One-Step-Back to One-Step-Ahead Prediction

- Set the feature template to default (without additional hand-crafted features).
- Pass all features from the previous order along with the basic series list of the current order to the programmed module.

Table 1. Comparison of Basic and Extended Feature Accuracy.

Metric	Dataset	MLP		CNN		LSTM	
		Basic	Extended	Basic	Extended	Basic	Extended
R^2 Score %	Synthetic	97.71 $\pm$ 0.00	**99.18** $\pm$ **0.00**	97.72 $\pm$ 0.00	**99.16** $\pm$ **0.00**	97.71 $\pm$ 0.03	**99.01** $\pm$ **0.04**
	Taxi	73.21 $\pm$ 0.00	**77.50** $\pm$ **0.01**	73.17 $\pm$ 0.00	**79.02** $\pm$ **0.06**	73.19 $\pm$ 0.01	**76.62** $\pm$ **0.04**
	Electricity	97.47 $\pm$ 0.00	**98.97** $\pm$ **0.00**	97.47 $\pm$ 0.00	**99.09** $\pm$ **0.01**	94.83 $\pm$ 0.00	**95.43** $\pm$ **0.00**
	Traffic	75.04 $\pm$ 0.00	**87.41** $\pm$ **0.00**	75.04 $\pm$ 0.00	**86.45** $\pm$ **0.01**	74.66 $\pm$ 0.00	**83.12** $\pm$ **0.00**
Pearson Correlation %	Synthetic	98.86 $\pm$ 0.00	**99.60** $\pm$ **0.00**	98.86 $\pm$ 0.00	**99.62** $\pm$ **0.00**	98.86 $\pm$ 0.01	**99.56** $\pm$ **0.01**
	Taxi	85.59 $\pm$ 0.00	**88.76** $\pm$ **0.00**	85.57 $\pm$ 0.00	**88.95** $\pm$ **0.03**	85.58 $\pm$ 0.01	**88.46** $\pm$ **0.00**
	Electricity	98.73 $\pm$ 0.00	**99.49** $\pm$ **0.00**	98.73 $\pm$ 0.00	**99.54** $\pm$ **0.00**	**98.04** $\pm$ **0.00**	97.70 $\pm$ 0.00
	Traffic	86.64 $\pm$ 0.00	**93.51** $\pm$ **0.00**	86.64 $\pm$ 0.00	**93.02** $\pm$ **0.00**	86.42 $\pm$ 0.00	**91.40** $\pm$ **0.00**

Results: Our generated features deliver 1.3% and 5.85% improvements in the R2 Score and the Pearson correlation, respectively.

Experimental Studies: Sequence to One-Step-Ahead Prediction

One-step-ahead time series regression problem with a lookback size of $T = 20$.

Table 4. Performance Comparison of Common Time Series Models with and without Extended Features on Synthetic and Taxi Datasets. The table demonstrates the performance enhancements consistently achieved in one-step-ahead prediction tasks using a lookback size of $T = 20$ when extended features are utilized as inputs for various machine learning (XGBoost and LightGBM) and deep neural network time series prediction models (Transformer, TCN, TFT and N-BEATS).

Metric	Dataset	XGBoost		LightGBM		Transformer		TCN		TFT		N-BEATS	
		Basic	Extended	Basic	Extended	Basic	Extended	Basic	Extended	Basic	Extended	Basic	Extended
R^2 Score %	Synthetic	99.12	**99.34**	98.92	**99.34**	96.46	**96.86**	98.87	**99.37**	96.21	**97.60**	99.26	**99.41**
	Taxi	77.34	**78.15**	77.05	**81.01**	72.64	**73.70**	75.36	**77.70**	54.58	**61.31**	76.82	**79.94**
Pearson Correlation %	Synthetic	99.56	**99.67**	99.46	**99.67**	**99.11**	99.08	99.44	**99.69**	98.41	**99.19**	97.63	**99.73**
	Taxi	87.96	**88.42**	87.80	**88.97**	85.54	**86.83**	86.87	**88.26**	76.65	**79.65**	87.84	**89.61**

Results:

- Our generated features deliver consistent performance enhancements
- Simple Models + Extended Features $\simeq$ SOTA + Raw Features

Experimental Studies: Multi-Horizon Prediction

Multi-horizon prediction with horizon sizes of 1, 2, 3, 5, 10 and 20, utilizing a lookback size of $T = 20$.

Table 5. Evaluation of Feature Quality for Multi-Horizon Prediction using Synthetic Dataset. This table presents the performance of the Transformer, TFT, TCN and N-BEATS models on multi-horizon prediction tasks with horizon sizes of 1, 2, 3, 5, 10 and 20, benchmarked on the synthetic dataset. The results showcase the effectiveness of the generated features in various multi-horizon prediction settings. Particularly, when predicting a full-length sequence (using a sequence of length 20 to predict the next 20 values), our generated features have demonstrated considerable prediction improvements across all models, suggesting that the extended features possess a greater amount of autoregressive information compared to basic features.

Model	Metric	1-Step Horizon		2-Step Horizon		3-Step Horizon		5-Step Horizon		10-Step Horizon		20-Step Horizon	
		Basic	Extended	Basic	Extended	Basic	Extended	Basic	Extended	Basic	Extended	Basic	Extended
Transformer	R^2 Score %	96.46	**96.86**	94.47	**96.58**	94.07	**96.92**	91.03	**92.78**	73.11	**81.17**	54.46	**73.33**
	Pearson Correlation %	**99.11**	99.08	98.02	**98.63**	97.11	**98.48**	95.42	**96.40**	85.76	**90.41**	74.31	**86.20**
TFT	R^2 Score %	96.21	**97.60**	95.81	**96.86**	93.99	**95.11**	84.76	**92.29**	60.87	**78.60**	35.35	**64.91**
	Pearson Correlation %	98.41	**99.19**	97.92	**98.57**	97.30	**98.02**	93.39	**96.11**	81.14	**88.75**	66.13	**80.85**
TCN	R^2 Score %	98.87	**99.37**	96.03	**97.98**	92.41	**95.75**	79.91	**88.32**	41.91	**66.18**	17.18*	**52.50***
	Pearson Correlation %	99.44	**99.69**	98.02	**99.01**	96.20	**97.86**	89.51	**94.11**	65.51	**81.41**	46.08*	**73.19***
N-BEATS	R^2 Score %	99.26	**99.41**	98.39	**98.87**	97.26	**98.00**	93.17	**95.69**	78.09	**86.97**	57.06	**74.65**
	Pearson Correlation %	97.63	**99.73**	99.20	**99.44**	98.62	**99.07**	96.62	**97.84**	88.90	**93.28**	75.80	**86.44**

Results: On average 88+% in R^2 and 27+% in Pearson correlation.

For larger horizons, Simple Models + Extended Features $\geq$ SOTA + Raw Features.

Conclusion

We propose a programmable automated feature engineering approach for multivariate time series modeling, inspired by a novel Ising-like dynamical model.

Theoretically, this model-based approach draws practical guidance from constructing multivariate graphical models using univariate exponential family, aligning with insights from the physics model.

Empirically, the generated features effectively improve noisy multivariate time series prediction in various settings.

Limitation: the flexibility of the method comes with the trade-off of not including any feature selection or pruning mechanism beyond user-specific programs.

Future Investigations: we plan to integrate feature selection mechanisms into feature programming and explore the joint temporal graph in the context of graphical diffusion.

Conclusion

We propose a programmable automated feature engineering approach for multivariate time series modeling, inspired by a novel Ising-like dynamical model.

Theoretically, this model-based approach draws practical guidance from constructing multivariate graphical models using univariate exponential family, aligning with insights from the physics model.

Empirically, the generated features effectively improve noisy multivariate time series prediction in various settings.

Limitation: the flexibility of the method comes with the trade-off of not including any feature selection or pruning mechanism beyond user-specific programs.

Future Investigations: we plan to integrate feature selection mechanisms into feature programming and explore the joint temporal graph in the context of graphical diffusion.

Conclusion

We propose a programmable automated feature engineering approach for multivariate time series modeling, inspired by a novel Ising-like dynamical model.

Theoretically, this model-based approach draws practical guidance from constructing multivariate graphical models using univariate exponential family, aligning with insights from the physics model.

Empirically, the generated features effectively improve noisy multivariate time series prediction in various settings.

Limitation: the flexibility of the method comes with the trade-off of not including any feature selection or pruning mechanism beyond user-specific programs.

Future Investigations: we plan to integrate feature selection mechanisms into feature programming and explore the joint temporal graph in the context of graphical diffusion.

Conclusion

We propose a programmable automated feature engineering approach for multivariate time series modeling, inspired by a novel Ising-like dynamical model.

Theoretically, this model-based approach draws practical guidance from constructing multivariate graphical models using univariate exponential family, aligning with insights from the physics model.

Empirically, the generated features effectively improve noisy multivariate time series prediction in various settings.

Limitation: the flexibility of the method comes with the trade-off of not including any feature selection or pruning mechanism beyond user-specific programs.

Future Investigations: we plan to integrate feature selection mechanisms into feature programming and explore the joint temporal graph in the context of graphical diffusion.

Thank You!

Alex Reneau*, Jerry Hu*, Chenwei Xu, Weijian Li, Ammar Gilani, Han Liu

alexreneau@u.northwestern.edu
jhu@u.northwestern.edu
cxu@u.northwestern.edu
weijianli@u.northwestern.edu
ammargilani2024@u.northwestern.edu
hanliu@northwestern.edu
http://magics.cs.northwestern.edu