Learning from Polar Representation: An Extreme-Adaptive Model for Long-Term Time Series Forecasting

Yanhong Li, Jack Xu, David C. Anastasiu

Santa Clara University
School of Engineering

Problem Description

Problem: predicting long-term streamflow values with rain off data

$$\begin{bmatrix} x_{\mathbf{1},1} & \cdots & x_{\mathbf{1},t} \\ x_{2,1} & \cdots & x_{2,t} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,t} \end{bmatrix} \in \mathbb{R}^{m\times t} \rightarrow [x_{\mathbf{1},t+1}, ..., x_{\mathbf{1},t+h}] \in \mathbb{R}^{h}$$

x_1: the ordinary series

x_2 to x_m : extraordinary indicators.

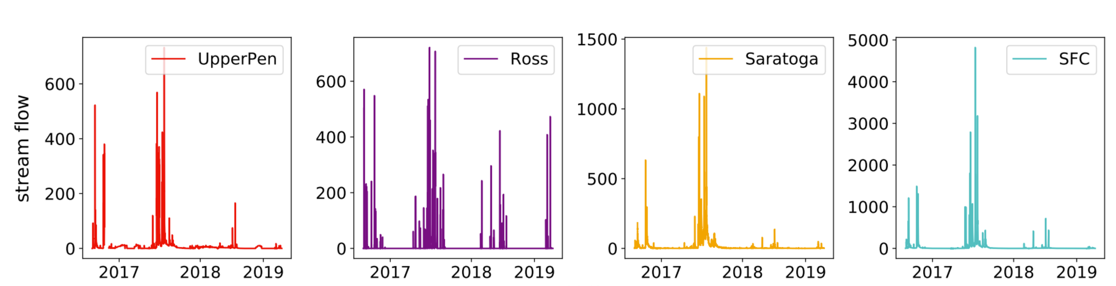
(x_2 can also be the Gaussian Mixture Model (GMM) indicator based on x_1. In such cases, the problem can be reduced to that of univariate time series forecasting.)

Challenges:
- Long-range dependencies.
- Rare but important extreme values.

Goal:
- An end-to-end model concurrently learns extreme and normal prediction functions.
- Long sequence forecasting (predicted length = 288).

Dataset:
- Four groups of hydrologic datasets from Santa Clara County, CA. Over 31 years of sensor data, 1,104,904 values.
- Namely, Ross, Saratoga, UpperPen, and SFC, named after their respective locations.
- Each group included a streamflow dataset and an associated rainfall dataset.

Extreme Events



	Ross	Saratoga	UpperPen	SFC
min	0.00	0.00	0.00	0.00
max	1440.00	2210.00	830.00	7200.00
mean	2.91	5.77	6.66	20.25
skewness	19.84	19.50	13.42	18.05
kurtosis	523.16	697.78	262.18	555.18

High skewness and kurtosis scores indicate that there is significant deviation from a normal distribution in our data!

Motivation

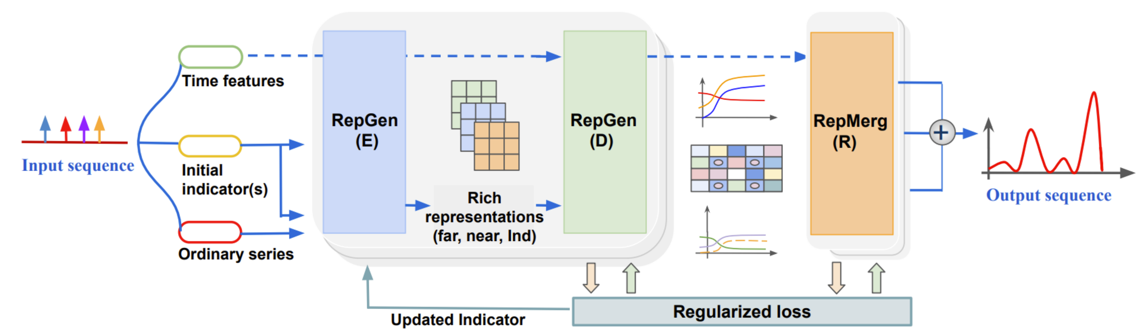
Achieving the best overall prediction performance, without sacrificing either the quality of normal or of extreme predictions.

Root Mean Square Error ***(RMSE)***

Mean Absolute Percentage Error ***(MAPE)***

Proposed Framework

- **DAN framework:** Distance-weighted Auto-regularized Neural network (DAN) uses expandable blocks to dynamically facilitate long-term prediction.



DAN's end-to-end extendable framework consists of two stages, named RepGen and RepMerg:

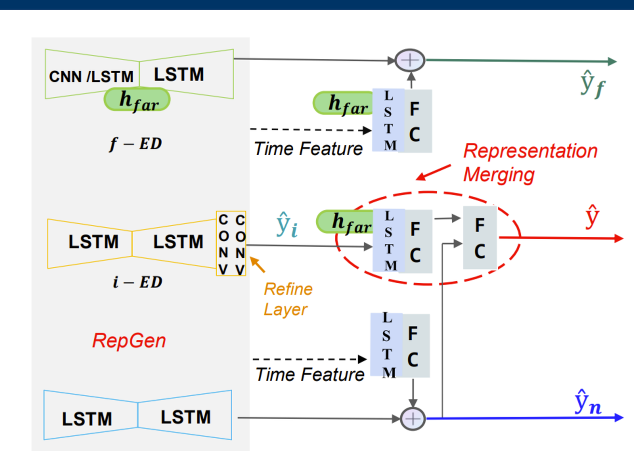
- RepGen contains three parallel encoder-decoder blocks, resulting in polar representations of ordinary series inputs and refined indicators.
- These elements are further merged in the RepMerg stack.

Kruskal-Wallis Test

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1).$$

- Examines $\boldsymbol{k}$ groups of sub series based on their medians.
- The data are first ranked, and the sum of ranks is calculated for each group. The H value is then calculated to determine if there are significant differences between the groups.
- A distribution-free test, not assume a particular distribution.
- Over-sampling regions with extreme events into training set.

Architecture Items

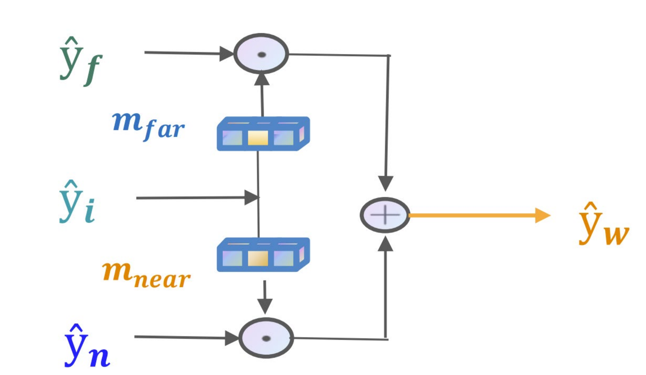


RepGen stack:
- "f-ED": representation learning of those points that are far away from the mean of the series $\hat{y}_f$. "n-ED": representation of near points $\hat{y}_n$. "i-ED": learn the indicator $\hat{y}_i$.

CONV-LSTM layers:
- Shorten the input sequence. Alleviate potential exploding or vanishing gradient.

Indicator Refine Layer:
- Made of 2×CNN. Assist in refining the expected indicator representation.

Gate control vector



Another way to hone predicted indicator:

- M_{far} is equal to sigmoid($\alpha * \hat{y}_i$), where α = 4 in our experiments, m_{near} = 1 − m_{far}.
- Doing the component-wise multiplication with predicted far values $\hat{y}_f$ and near values $\hat{y}_n$.
- Let to $\hat{y}_w$ to approach | tanh(y) | * y.

Auto-regularized Loss Function

$$\mathcal{L}_1 = RMSE((\hat{y}_f \odot w_f), (y \odot w_f)),$$
$$\mathcal{L}_2 = RMSE((\hat{y}_n \odot w_n), (y \odot w_n)),$$
$$\mathcal{L}_3 = RMSE(\hat{y}_w \odot w_p, y \odot w_p),$$
$$\mathcal{L}_4 = RMSE(\hat{y}_i \odot w_p, y_i \odot w_p),$$

where $\mathcal{L}_1$ and $\mathcal{L}_2$ are used to regulate the bipolar representation learning and $\mathcal{L}_3$ and $\mathcal{L}_4$ force the predicted indicator to reflect the change of predicted values by setting y_i equal to the first order of y. Then, the overall loss is composed as,

$$\mathcal{L} = RMSE(\hat{y}, y) + \lambda \times (\mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 + \mathcal{L}_4),$$

where λ is a multiplier ($\lambda = max(-1 \cdot e^{\frac{epoch}{45}} + 2, 0.2)$ in our experiments) applied on those regulation items, decreasing with each epoch.

Motivations:

- Multiple distance-weighted loss functions with the objective of compelling the model to learn more informative representations.
- Serve as an effective regularizer for preventing overfitting in the long-term time series prediction task.

Baselines

- DNN-U: univariate LSTM-based encoder-decoder hydrologic model.
- Attention-LSTM: a state-of-the-art hydrologic model used to predict stream-flow.
- N-BEATS: outperformed all competitors on the standard M3, M4 and TOURISM datasets.
- FEDFormer.
- InFormer.
- NLinear.
- DLinear.

Research Questions

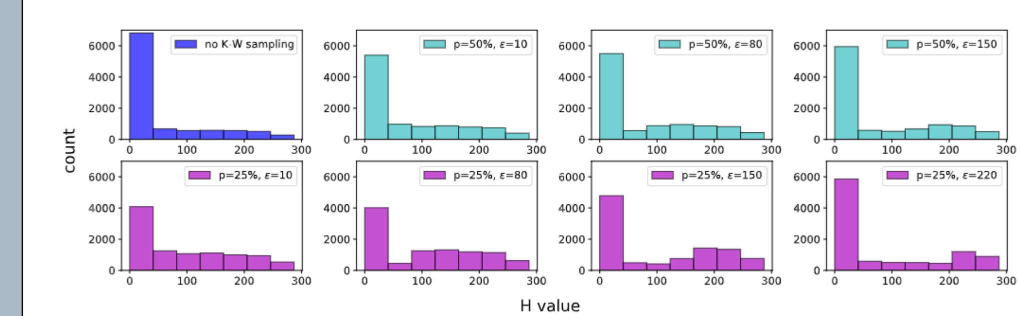
- What is the effect of DAN's extensible framework?
- What is the effect of the Kruskal-Wallis oversampling policy?
- How do the critical design elements of the framework affect performance?

Effects of Proposed Methods

- Effects of DAN's extensible framework.

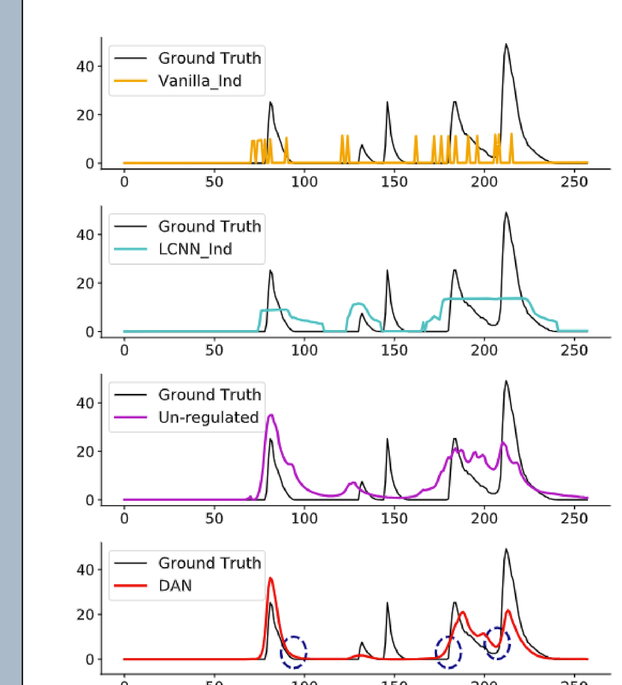
We experimented with various combinations and identified the best results as "EDEDRR", "EDR", "EDEDRR", and "EDEDR" for Ross, Saratoga, UpperPen, and SFC, respectively.

- Effects of Kruskal-Wallis oversampling policy.



Maintain the p value and increase the ε value, the training set will contain more samples with H values exceeding ε.

- Effects of the critical design elements of the framework.



- Vanilala_Ind: remove key architecture items.
- LCNN_Ind: add CNN-CNN back, refines the indicator information.
- Un-regulated: add Gate control vector back, increases the discrimination of predicted values.
- DAN: add polar representation back, enhances the accuracy of data at corners of each fluctuation, as denoted in the blue circles in the figure.

Evaluation

Table 2: Multivariate/Univariate Long-Term ($h = 288$) Series Forecasting Results on Four Datasets

Methods	Metric	Ross		Saratoga		UpperPen		SFC	
		Multi	Single	Multi	Single	Multi	Single	Multi	Single
FEDformer	RMSE	6.01	6.49	6.01	6.85	3.05	2.38	23.54	24.10
	MAPE	2.10	2.49	1.55	2.26	1.87	1.02	2.35	2.817
Informer	RMSE	7.84	9.14	5.04	4.89	5.88	5.33	39.89	19.00
	MAPE	4.05	5.45	1.43	0.73	4.10	4.21	8.64	0.54
Nlinear	RMSE	6.10	5.84	5.23	4.98	1.57	1.74	18.47	18.43
	MAPE	1.99	1.62	0.83	0.75	0.45	0.57	0.92	0.87
Dlinear	RMSE	7.16	6.90	4.33	4.06	3.53	3.25	21.62	23.64
	MAPE	3.10	2.79	1.40	1.31	2.35	2.04	2.74	4.02
NEC+	RMSE	9.44	9.33	1.88	1.95	2.22	1.94	17.00	16.39
	MAPE	4.80	4.53	0.17	0.21	0.95	0.80	1.07	0.55
LSTM-Atten / NBeats	RMSE	7.35	5.16	6.49	3.60	6.35	**1.23**	34.17	31.47
	MAPE	3.74	1.25	1.80	0.70	4.76	**0.25**	9.90	3.24
DAN	RMSE	**4.25**	**4.24**	**1.80**	**1.84**	**1.10**	1.31	**15.23**	**15.20**
	MAPE	**0.07**	**0.09**	**0.14**	**0.16**	**0.15**	0.32	**0.26**	**0.21**

Contact Information: danastasiu@scu.edu