An Extreme-Adaptive Time Series Prediction Model Based on Probability-Enhanced LSTM Neural Networks

Yanhong Li, Jack Xu, David C. Anastasiu

Santa Clara University
Santa Clara Valley Water District

AAAI 2023
Problem: Univariate Time Series with Extreme Events Forecasting

\[ [X_1, X_2, \ldots, X_T] \in R^T \rightarrow [X_{T+1}, \ldots, X_{T+H}], \in R^H \]

Challenges:
- Most values are normal and contribute significantly to the overall prediction performance
- Few extreme values but they must be precisely forecasted to avoid disastrous events

Goal:
- A model that concurrently learns extreme and normal prediction functions
- Long sequence forecasting
- Good generalization

Dataset:
- We evaluate the proposed model on the difficult 3-day ahead hourly water level prediction task applied to 9 reservoirs in California
Extreme Events:

**GEV (Generalized Extreme Value) distribution** provides a better fit, showing the presence of extreme values in our data, where \( \mu \in \mathbb{R}, \sigma > 0, \) and \( \xi \) are the location, scale, and shape parameters, respectively, conditioned on \( 1+\xi(x-\mu)/\sigma > 0. \)

\[
F(x; \mu, \sigma, \xi) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x-\mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \tag{1}
\]
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 Methods:

**NEC framework:** a framework to account for the distribution shift between normal and extreme values; composed of a *Normal*, an *Extreme*, and a *Classification* model.

**GMM Indicator:** an unsupervised clustering approach to dynamically produce distribution indicators, which improves the model’s robustness to the occurrence of severe events.

**Selected Backpropagation:** helps the models learn the positions and values of appropriate normal or extreme data better.

**Parameterized Loss Function:** helps our classifier better distinguish event severity.
NEC framework:

NEC is composed of three separate models, which can be trained in parallel:

- The **Normal (N)** model is trained to best fit normal values in the time series.
- The **Extreme (E)** model is trained to best fit extreme time series values.
- The **Classifier (C)** model is trained to detect when a certain value may be categorized as normal or extreme.

**NEC** is a purely single variate model; **NEC+** uses optional exogeneous other inputs.
GMM Indicator:

A Gaussian mixture model (GMM) is a weighted sum of \( M \) component Gaussian densities,

\[
p(x|\lambda) = \sum_{i=1}^{M} w_i g(x|\mu_i, \Sigma_i),
\]

where \( x \) is a \( D \)-dimensional continuous-valued vector, \( w_i \ \forall i = 1, \ldots, M \) are the mixture weights, and \( g(x|\mu_i, \Sigma_i) \) are the component Gaussian densities.

We compute an indicator feature as the weighted sum of all component probabilities, given the weights learned when fitting the GMM model.
Two-stage sampling policy:
1) randomly sample subsections of length $h + f$ from the series to use in training our models
2) perform stratified sampling of regions with and without extreme values, allowing the E and C models to oversample up to OS% samples with at least 1 extreme value in the prediction zone.

Selected Backpropagation:
N model: only normal values add to the loss; E model: only extreme values add to the loss.
Parameterized Loss Function:

\[ BCE(t, p) = -(t \times \log(p) + (1 - t) \times \log(1 - p)) \]

\[ L = \beta \times BCE(t, p^\alpha) + (1 - \beta) \times RMSE(t, p) \]

where \( \alpha \) and \( \beta \) are parameters that can be tuned. Values \( \alpha > 1 \) cause the model to predict \( p \) values that are higher in general in order to minimize the distance between \( t \) and \( p^\alpha \).

Problem: for datasets with a high imbalance between the two classes, BCE will favor the prominent class.

- The BCE part: can be thought of as a blunt instrument that grossly exaggerates all miss-classifications in order to more accurately predict the obscure class
- The RMSE part: allows for a more gentle penalty based on the distance between \( t \) and \( p \).
Research Questions:

1. What is the effect of adding the GMM indicator to a model?
2. What is the effect of introducing exogenous features?
3. How do the loss function parameters affect performance?
4. How does NEC+ compare against state-of-the-art baselines?

Baselines:

• ARIMA
• Facebook Prophet
• LSTM
• DNN-U: univariate state-of-the-art LSTM-based encoder-decoder hydrologic model.
• Attention-LSTM: a state-of-the-art hydrologic model for stream-flow prediction.
• N-BEATS: a state-of-the-art time series prediction method that outperformed all competitors on the standard M3, M4 and TOURISM datasets.
Research Questions:

1. What is the effect of adding the GMM indicator to a model?
2. What is the effect of introducing exogenous features?
Research Questions:

3. How do the loss function parameters affect performance?
Research Questions:

4. How does NEC+ compare against state-of-the-art baselines?
Table 1: Effectiveness Comparison (RMSE) of NEC+ Against Baselines for 9 Reservoirs

<table>
<thead>
<tr>
<th>Model/Reservoir</th>
<th>4001</th>
<th>4003</th>
<th>4004</th>
<th>4005</th>
<th>4006</th>
<th>4007</th>
<th>4009</th>
<th>4010</th>
<th>4011</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>1016.32</td>
<td>1859.70</td>
<td>2501.97</td>
<td>9692.87</td>
<td>1039.38</td>
<td>5854.48</td>
<td>1060.05</td>
<td>3465.20</td>
<td>690.23</td>
</tr>
<tr>
<td>Prophet</td>
<td>8469.74</td>
<td>38827.22</td>
<td>95279.31</td>
<td>181607.50</td>
<td>20904.57</td>
<td>187603.80</td>
<td>28629.44</td>
<td>114115.4</td>
<td>2829.26</td>
</tr>
<tr>
<td>LSTM</td>
<td>1167.73</td>
<td>1514.90</td>
<td>2342.71</td>
<td>6730.93</td>
<td>959.05</td>
<td>5035.91</td>
<td>954.04</td>
<td>3734.53</td>
<td>662.48</td>
</tr>
<tr>
<td>DNN-U</td>
<td>1162.01</td>
<td>1597.72</td>
<td>3989.20</td>
<td>9878.41</td>
<td>983.27</td>
<td>4320.40</td>
<td>1411.63</td>
<td>4257.58</td>
<td>763.73</td>
</tr>
<tr>
<td>A-LSTM</td>
<td>878.71</td>
<td>1536.04</td>
<td>2548.56</td>
<td>8919.33</td>
<td>1638.65</td>
<td>13529.86</td>
<td>1064.15</td>
<td>2914.75</td>
<td>700.50</td>
</tr>
<tr>
<td>N-BEATS</td>
<td>937.24</td>
<td>1926.74</td>
<td>2280.83</td>
<td>7153.82</td>
<td>960.42</td>
<td>3153.76</td>
<td>1295.90</td>
<td>3162.17</td>
<td>514.30</td>
</tr>
<tr>
<td>NEC+</td>
<td>740.19</td>
<td>1411.44</td>
<td>1783.92</td>
<td>4352.74</td>
<td>780.46</td>
<td>2092.73</td>
<td>703.93</td>
<td>2275.48</td>
<td>632.61</td>
</tr>
</tbody>
</table>

Table 3: MAPE of NEC+ vs. Baselines for 9 Reservoirs

<table>
<thead>
<tr>
<th>Model/Reservoir</th>
<th>4001</th>
<th>4003</th>
<th>4004</th>
<th>4005</th>
<th>4006</th>
<th>4007</th>
<th>4009</th>
<th>4010</th>
<th>4011</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>1.3573</td>
<td>0.7626</td>
<td>0.8694</td>
<td>1.2560</td>
<td>1.5401</td>
<td>0.8517</td>
<td>0.9504</td>
<td>1.7871</td>
<td>3.2914</td>
</tr>
<tr>
<td>Prophet</td>
<td>16.7877</td>
<td>19.8559</td>
<td>38.9642</td>
<td>35.6662</td>
<td>56.0537</td>
<td>32.9152</td>
<td>31.8069</td>
<td>45.2579</td>
<td>15.3312</td>
</tr>
<tr>
<td>LSTM</td>
<td>1.6697</td>
<td>0.6153</td>
<td>0.7450</td>
<td>1.0092</td>
<td>1.3264</td>
<td>0.9253</td>
<td>0.9298</td>
<td>2.5520</td>
<td>3.1282</td>
</tr>
<tr>
<td>DNN-U</td>
<td>1.6509</td>
<td>0.6812</td>
<td>1.8738</td>
<td>1.9394</td>
<td>1.4551</td>
<td>0.6509</td>
<td>1.5604</td>
<td>2.1582</td>
<td>3.7131</td>
</tr>
<tr>
<td>A-LSTM</td>
<td>1.3533</td>
<td>0.6506</td>
<td>0.8424</td>
<td>1.2060</td>
<td>2.8017</td>
<td>2.1738</td>
<td>0.9705</td>
<td>1.3986</td>
<td>3.4137</td>
</tr>
<tr>
<td>N-BEATS</td>
<td>1.3346</td>
<td>0.7972</td>
<td>0.7882</td>
<td>1.1405</td>
<td>2.0061</td>
<td>0.4709</td>
<td>1.4580</td>
<td>1.7146</td>
<td>2.3108</td>
</tr>
<tr>
<td>NEC+</td>
<td>1.0319</td>
<td>0.5687</td>
<td>0.6030</td>
<td>0.6350</td>
<td>1.0662</td>
<td>0.3316</td>
<td>0.5992</td>
<td>1.2894</td>
<td>2.9237</td>
</tr>
</tbody>
</table>