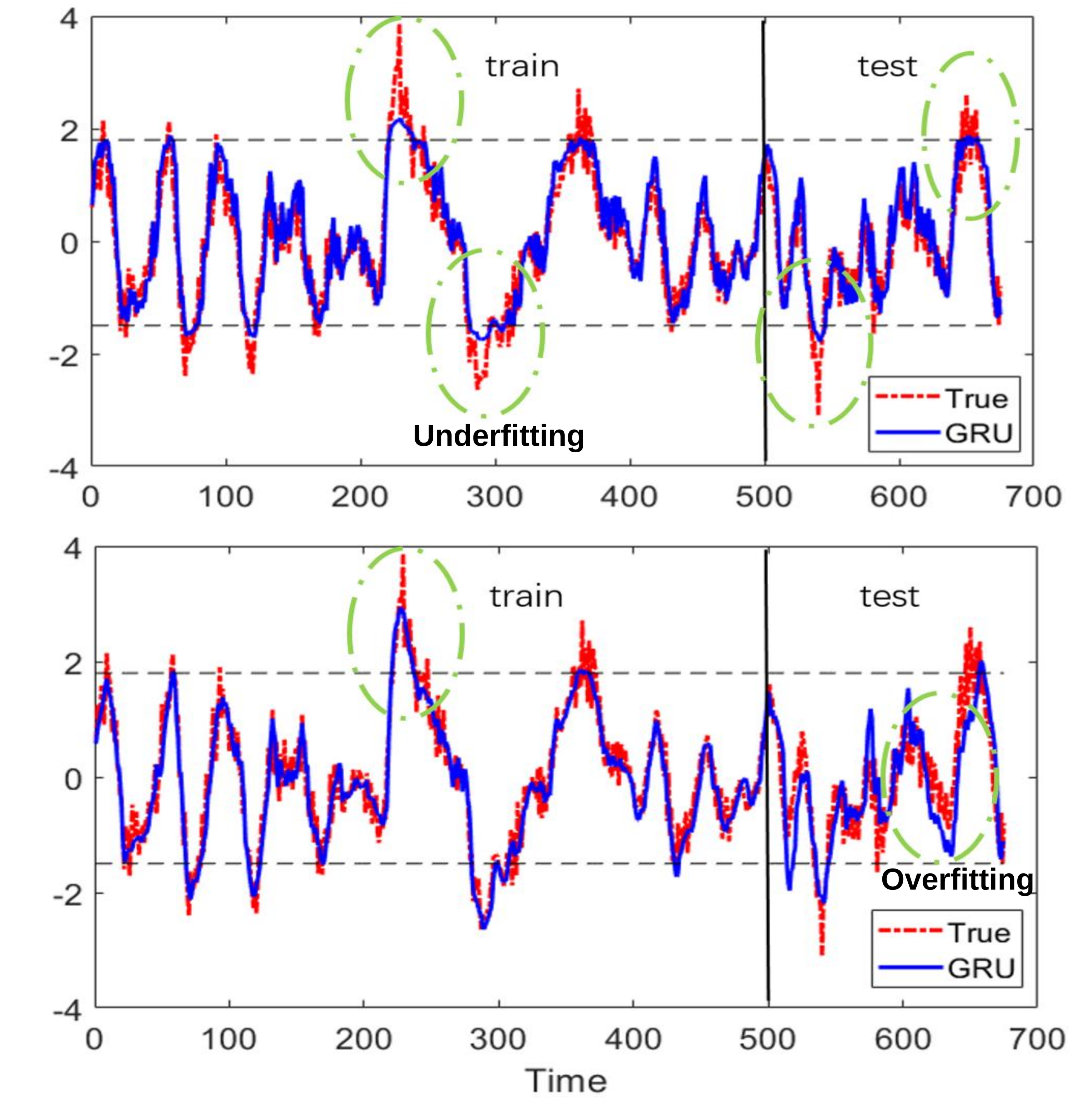
Modeling Extreme Events in Time Series Prediction

Daizong Ding, Mi Zhang, Xudong Pan, Min Yang, Xiangnan He

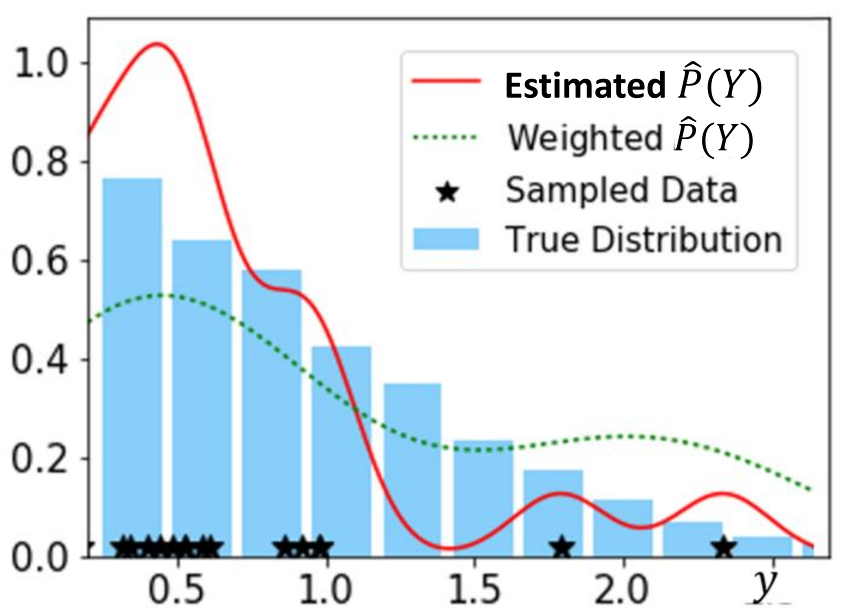
Fudan University · University of Science and Technology of China

Problem Description



Extreme Events Problem

In time series data, **extreme events** could influence the performance of deep learning model, e.g., underfitting and overfitting.



Likelihood Posterior

$$\boxed{P(Y|X)} = \frac{\boxed{P(X|Y)}\boxed{P(Y)}}{P(X)}$$

Estimated distribution of outputs $\hat{P}(Y) = \frac{1}{N}\sum_{t=1}^{T} \mathcal{N}(y_t, \hat{\tau}^2)$

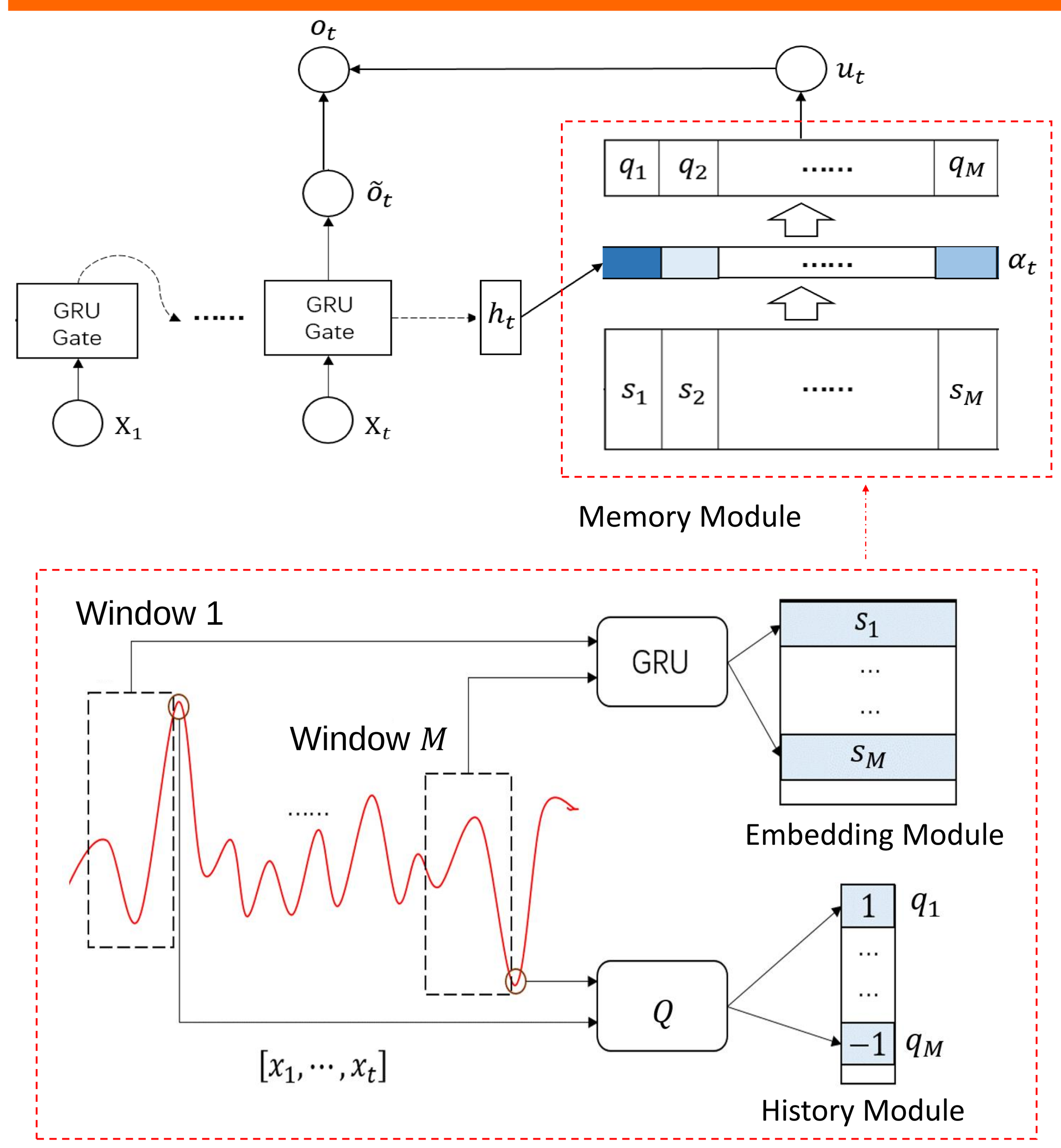
Why Deep Neural Network Could Suffer Extreme Event Problem

Through perfectly maximizing Gaussian likelihood, model will internally estimate the distribution of outputs y_t from observations. Considering that few extreme events are observed, the estimated light-tailed distribution is not accurate for those extreme events. Therefore during the prediction:

$$P(Y|X) = \frac{P(X|Y)\hat{P}(Y)}{P(X)} \le \frac{P(X|Y)P_{true}(Y)}{P(X)} = P_{true}(Y|X)$$

which could lead to underfitting phenomenon. If we increase the weight of extreme events, the overfitting phenomenon is then observed.

Model Design



Recalling Extreme Events

Considering that the freedom of extreme events is limited, we propose to use **Memory Network** to memorize these points. At each time step t,

$$\tilde{o}_t = W_o^T h_t + b_o \ , \ h_t = GRU(x_1, \cdots, x_t)$$

Meanwhile, we sample M windows for memory module. For each window j, $s_j = GRU\left(x_{t_j}, \cdots, x_{t_j+\Delta}\right)$ and we calculate $q_j = \{-1,0,1\}$ by setting threshold as extreme event indicator. The similarity between the current inputs and the extreme events in history is,

$$\alpha_{tj} = \frac{\exp(h_t^T s_j)}{\sum_{l=1}^{M}\exp(h_t^T s_l)}$$

Then the final output of our model is,

$$o_t = \tilde{o}_t + b \cdot u_t \ , \ u_t = \sum_{j=1}^{M} \alpha_{tj} \cdot q_j$$

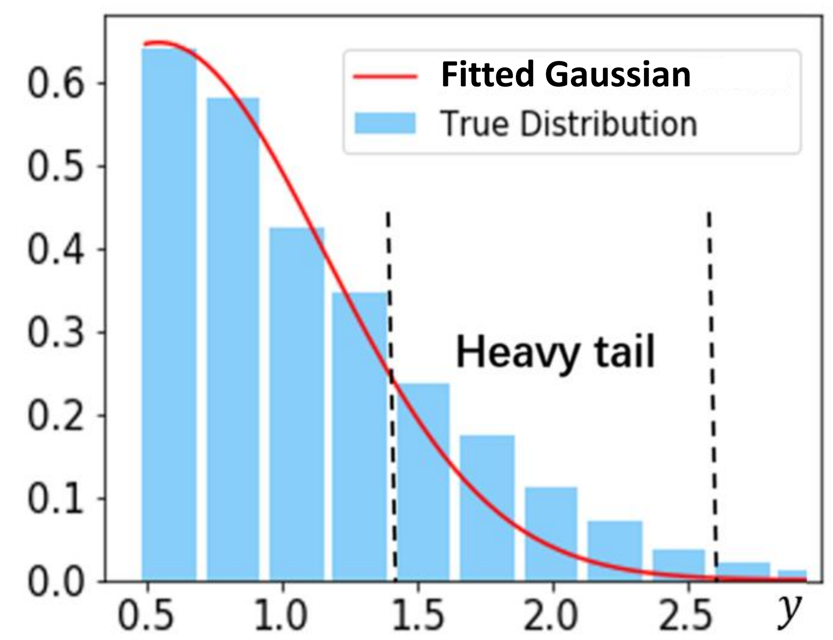
Extreme Value Loss (EVL)

The extreme events problem is caused by light-tailed loss, we propose the following loss by **extreme value theory.** For binary classification case,

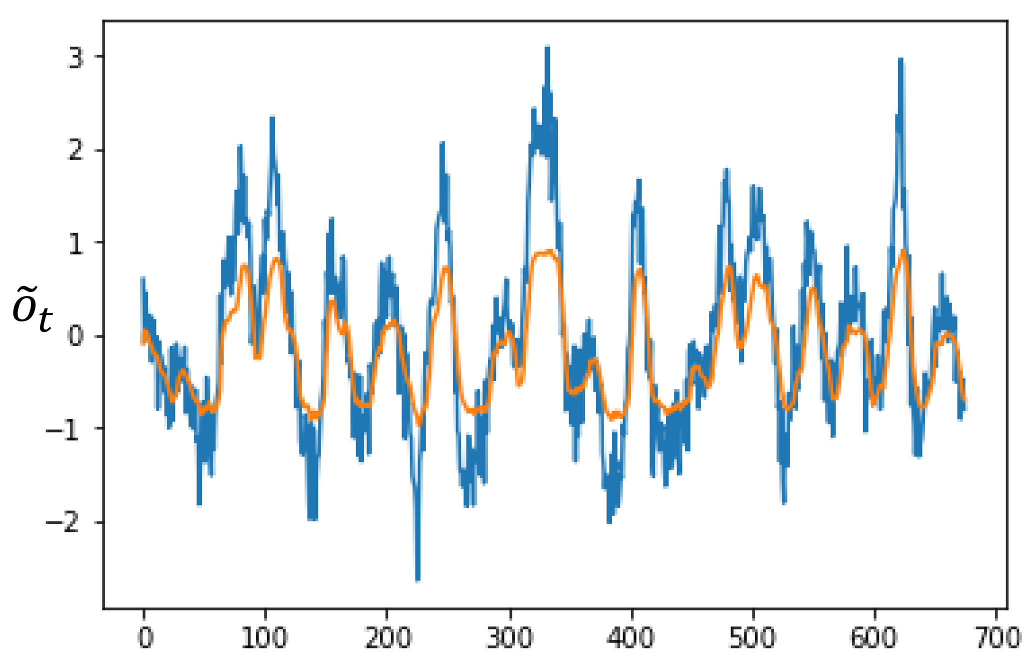
$$EVL(u_t, v_t) = -\beta_0\left[1 - \frac{u_t}{\gamma}\right]^{\gamma} v_t \log u_t - \beta_1\left[1 - \frac{1-u_t}{\gamma}\right]^{\gamma}(1 - v_t)\log(1 - u_t)$$

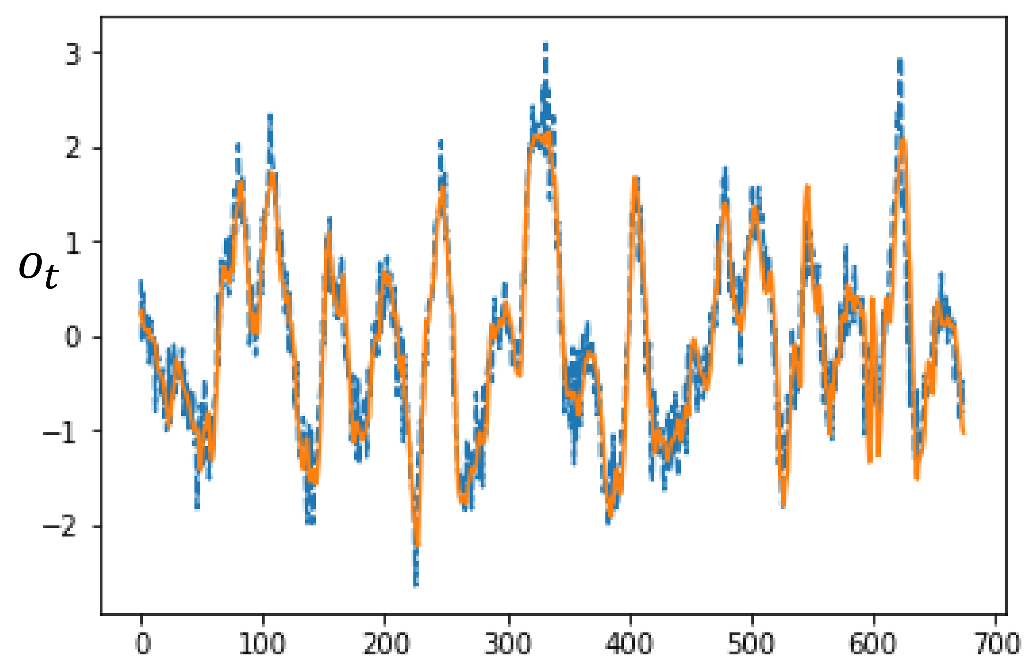
where v_t is the ground truth of extreme event indicator at time step t, β_0 is the proportion of normal outputs and γ is the hyper-parameter.

$$\ell = \sum_{t=1}^{T}(o_t - y_t)^2 + \lambda \cdot EVL(u_t, v_t)$$



Experiments





Empirical Results

- **Time Series Prediction:** The RMSE improves near 50% relatively in Pseudo dataset. Although $\tilde{o}_t$ still suffer underfitting problem, the memory network module successfully capture the characteristic of extreme events.
- **Extreme Value Loss:** We extend the EVL to multi-label classification and use classification task (predicting the occurrence of extreme events) to validate the effectiveness of the proposed loss function.

Table: F1 Value of detecting extreme events $\{-1, 0, 1\}$

Model	Climate			Stock		
	Micro	Macro	Weighted	Micro	Macro	Weighted
LSTM+CE	0.435	0.833	0.786	0.247	0.617	0.527
GRU+CE	0.471	0.717	0.733	0.250	0.617	0.547
GRU+EVL ($\gamma = 0.5$)	0.644	0.883	0.859	0.281	0.583	0.523
GRU+EVL ($\gamma = 1.0$)	**0.690**	**0.900**	**0.881**	0.267	**0.667**	0.547
GRU+EVL ($\gamma = 2.0$)	0.646	0.867	0.851	**0.324**	0.617	**0.555**
GRU+EVL ($\gamma = 3.0$)	0.508	0.867	0.825	0.295	0.617	0.548
GRU+EVL ($\gamma = 4.0$)	0.617	0.817	0.813	0.295	0.617	0.543