Winterstorm risk prediction via deep learning methods

Kornilov Nikita
Department of Control and Applied Mathematics
Moscow Institute of Physics and Technology
kornilov.nm@phystech.edu

Abstract

In article problem of harsh winterstorm prediction in short-term period is observed. This prediction is computed in large spacial grid. For practical needs model must accurately predict extreme events meanwhile on ordinary events it can make big number of mistakes.

Keywords Maximum risk prediction, GRU-CNN models, extreme events theory

1   Introduction

Extreme natural events and disasters are serious and relevant problem for every country. Having information about upcoming events, companies can better prepare, survive and recover much faster after it.

Researches in this area are attempting to predict location and time of future harsh natural event using climate data that was collected during last decades. Especially, most dangerous and rare events are called extreme. Their prediction is much harder because of lack of information and unique patterns.

In work (1) was proposed method to predict events in spacial grid using combination of LSTM and CNN neural networks architectures. Then, work (2) showed progress in extreme events prediction proposing new models and loss functions based on extreme events theory that considered unique features of .

Current task is to build model for short-term (up to 20 years) prediction of winterstorms in North America using geodata from 1991 year about snow storms in this region. The data is partially very noised and inaccurate with spacial grid.

In this article we expand extreme events prediction adding important spacial components to up-to-date methods. Also we *try* to consider geographical area features. We combine together Recurrent Neural Networks architectures with CNN and state-of-the-art models for extreme events prediction.

Goal of research is to fit model with comparable quality to models from related works, that after can be used for other regions.

2   Problem statement

The extreme events prediction problem with spacial components is both time series prediction and spacial problems. Considering this combination the formalization of problem.

2.1   Formalization

Let there are N time sequences with length T on space grid $S \subset \mathbb{R}^2$. Then i-th sequence can be described as

$$(S, X_{1:T}^{(i)}, Y_{1:T}^{(i)}, V_{1:T}^{(i)}) = \{(s, x_j^{(i)}, y_j^{(i)}, v_j^{(i)}) : s \in S, j \in \overline{1,T}\}$$

where $(s, x_j) \in \mathbb{R}^2 \times \mathbb{R}^h$, $(s, y_j) \in \mathbb{R}^2 \times \mathbb{R}^h$ are input variables and are predictable variables at j-th moment of time in s element of grid S respectively. Necessary condition of alignment in prediction is $y_t^{(i)} = x_{t+1}^{(i)}$.

$(s, v_j^{(i)})$ is special indicator of extreme event, where $v_j^{(i)} \in \{0, 1\}$ and 1 means event has happened, 0 otherwise. Further for the sake of convenience, index (i) will be omitted.

2.2 Model

Prediction model is model Memory module, attention mechanism from (2), but ordinary GRU blocks are replaced by GRU blocks with convolutions layers from CNN model in order to work with spacial grid similar with work (1), (3)

2.2.1 Memory module

For reach step t model randomly samples M windows from dataset with length Δ.

$$w_j = (x_{t_j}, \ldots, x_{t_j+\Delta}), 0 < t_j < t - \Delta$$

M, Δ are hyperparameters. Then for each window w_j GRU-CNN block is appliyed $GRU - CNN(w_j) = h_j$. Meanwhile, in training training model stores true extreme indicators $q_j = v_{t_j+\Delta+1}$.

2.2.2 Attention module

$X_{1:T}, \{h_j\}_{j=1}^M$ is module input. $X_{1:T}$ passes through the same GRU-CNN block.

$$\hat{o}_t = CNN(h_t) + b_0, h_t = GRU - CNN(X_{1:T})$$

Then model computes v_t' using linear transformation h_j and h_X - embedding from $X_{1:T}$ input.

Finally, $o_t = \hat{o}_t + bv_t'$, where $b \in \mathbb{R}^H$ - trainable parameter. Loss function and optimization problem

Let O is prediction model. It's input $(S, X_{1:T})$, output $O((S, X_{1:T})) = (S, Y_{1:T}', V_{1:T}')$ and true values are $(S, Y_{1:T}, V_{1:T})$. In particular, o_t - model output at moment t.

In order to predict extreme events in (2) was proposed special loss-function called Extreme Value Loss or EVL.

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

where $\beta_0 = Pr(v_t = 0)$ - probability of event to be normal or proportion of normal events in dataset and $\beta_1 = Pr(v_t = 1)$- proportion of extreme events.

Then loss function consists of two functions.

The first one, for extreme event classification error and time-series prediction of $Y_{1:T}$ to improve GRU-CNN blocks in attention module.

$$L_1(O, S, X_{1:T}, Y_{1:T}, V_{1:T}) = \sum_{s \in S} \left[\sum_{t=1}^T ||o_t - y_t||^2 + \lambda_1 EVL(v_t', v_t)\right]$$

This function optimizes after T batch.

The second for enchancement of GRU-CNN modules in memory module:

$$L_2(O, S, X_{1:T}, Y_{1:T}, V_{1:T}) = \sum_{s \in S} \left[\sum_{t=1}^T \sum_{j=1}^M EVL(p_j, q_j)\right]$$

where $p_j \in [0, 1]$ is computed from j-th window hidden representation output h_j, which was passed through fully-connected CNN layer. This function optimizes before v_t' computation and after Memory module phase.

Optimizer for fitting is Adam.

Список литературы

[1] E. B. Alexey Zaytsev, "Recurrent convolutional neural networks help to predict location of earthquakes,"

[2] X. P. M. Y. X. H. Daizong Ding, Mi Zhang, "Modeling extreme events in time series prediction,"

[3] Y. Lyu and X. Huang, "Road segmentation using cnn with gru,"