HAIL RISK PREDICTION WITH HAILNET

A PREPRINT

Ivan Lukyanenko
Department of Control and Applied Mathematics
Moscow Institute of Physics and Technologies
Moscow
lukianenko.ia@phystech.edu

ABSTRACT

Geo-spatial time series is an open area with great potential for theoretical and practical work. Hail risk assessment is necessary to avoid the damage (agriculture, animal husbandry). The aims of the study are to understand how to recognise and predict hail events basing on days hourly time series of climatic variables and to introduce model based on this knowledge. Forecasting has been carried out in the short-term range based on the values of climate variables since 1991. The key features of the problem are: 1) rare events - over the past 30 years there have been less than 700 hail events throughout Russia, 2) the spatial structure of the series data. Hail events are rather locally in time and in space. These restrictions make us to use dense data, especially it plays role in temporal component of data, i.e. using as dense data as possible. Nowadays there are no methods in climatology which predict and asses hail risk, this study is to try filling this gap using deep learning approaches.

Keywords Hail risk prediction · spatial time series · Time Series Classification · CNN · HailNet

1 Introduction

The main goals of this research are to understand how to recognise hail events and to build model based on CNN and LSTM. Our object of research is geo-spatial time series. In common time series is a series of values of a quantity obtained at successive times with equal intervals between them. This research is about more extended concept: spatial time series. Spatial time series are almost the same as ordinary time series, but instead of values we observe some spatial objects. This work provides a solution to the hail recognising problem. The hails are extreme events. Over the past 30 years there have been less than 700 hail events throughout Russia. The assessment of hail risk is necessary because of its environment damage. According to Verisk's 2021 report [4], in 2020 year losses due to hails reached $14.2 billion in the USA.

add citation

Hail events are very locally in time and in space. According to [3], hailstones are formed when raindrops are carried upward by thunderstorm updrafts into extremely cold areas of the atmosphere and freeze. Hailstones then grow by colliding with liquid water drops that freeze onto the hailstone's surface. The hail falls when the thunderstorm's updraft can no longer support the weight of the hailstone, which can occur if the stone becomes large enough or the updraft weakens. In most studies, it is noted that most of the hail's growth occurs at a temperature of approximately -10 to -25 $\circ$C.

Hypothesis 1. *Hail event depends on certain patterns in multivariate time series of climatic variables such as wind vertical component, wind horizontal component, temperature.*

Important ingredient for creating large hailstones is time. Appreciable growth is only attainable if particles remain in an environment conducive to growth for an extended period of time. Some studies suggest that large hailstones spend as much as 10–15 min or more in growth regions of storms.

Hypothesis 2. *Hailstones formation spend approximately 20 min. Hailstones are in the clouds till the updrafts of wind prevent them from falling down.*

This is the main reason why we use such temporally dense data. Daily data or even monthly data is too smoothed and hail patterns can not be detected. How do we take in account the Hypothesis 1?

The paragraph about Dense + Conv Layers will be here

The producing architecture is inspired by works [1], [5] and [2]. [2] introduce **StrGNN** that solves the anomaly detection problem in dynamic graph. We are going to use climate data from ERA5-Land Hourly, this source gives us tensors of variety of climatic variables all over the world. Needed to be mention, that this data is modeled, because it is impossible to put climate stations in every single point in the world. The information about hail events is taken from meteo.ru.

2 Problem Statement

Let give some designations and formalize our problem:

$n \in \mathbb{R}$ – number of climatic variables;

long – longitude of considered area;

lat – latitude of considered area;

$X_{i,j,k}$ – a tensor of climatic variables at one timestamp;

$\widetilde{X}_{t,i,j,k}$ – a time-series of climatic variables corresponding to one day.

We are solving time series classification problem, our objects are time series of 24 tensors $X_{i,j,k}$ corresponding to every hour, our target are labels: 1 (hail) and 0 (no hail).

As we are solving classification problem, we have to set criterion that would be optimized during training.

$D(w, x)$ – linear/dense layer of the network;

$Conv(x)$ – convolution layer of the network;

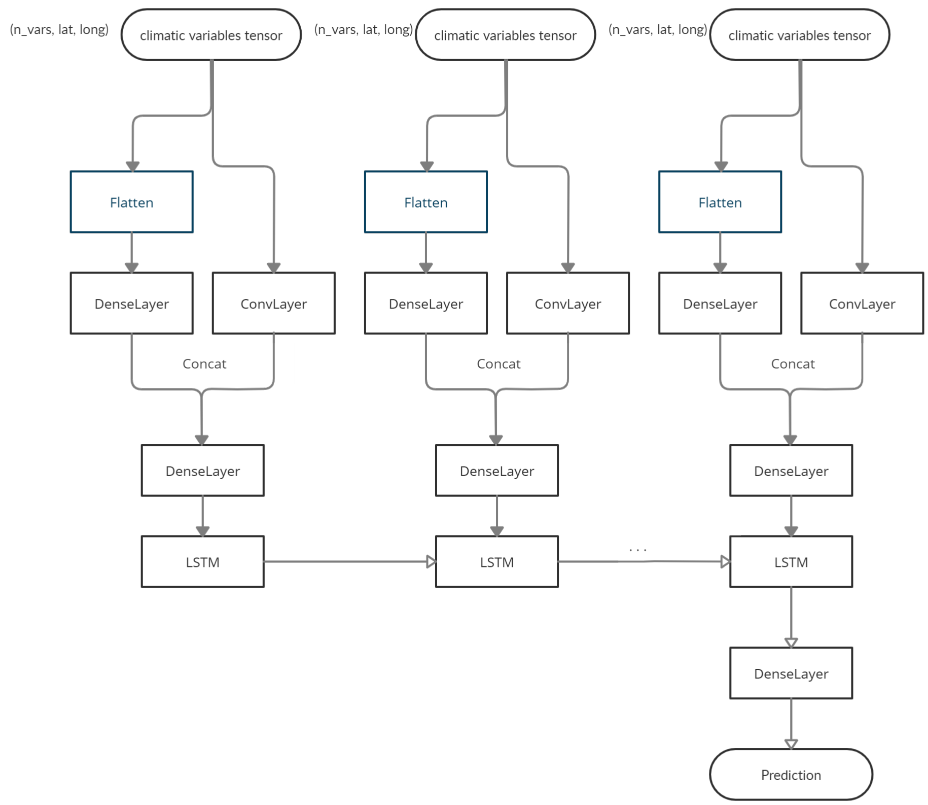
$LSTM(x, w, c, h)$ – LSTM block of the network;

$f(w, x)$ – HailNet;

Mathematically we set this optimization problem of minimizing Binary Cross Entropy for finding best model.

$$w^* = \arg\min_w \Big(-\frac{1}{N} \sum_{t=1}^{N} (y^t \log(f(w, x^t)) + (1 - y^t) \log(1 - f(w, x^t))) \Big) \tag{1}$$

Due to Hypothesis 1 provided model needs to aggregate information from different climatic variables and its spatial neighbourhood. Using Dense layers and Convolution layers we create timestamp embeddings. Due to Hypothesis 1 provided model needs to reveal temporal patterns of multivariate time series for better classification. This temporal structure is analyzed by LSTM layers in our architecture. Due to given hypotheses the result architecture is following:

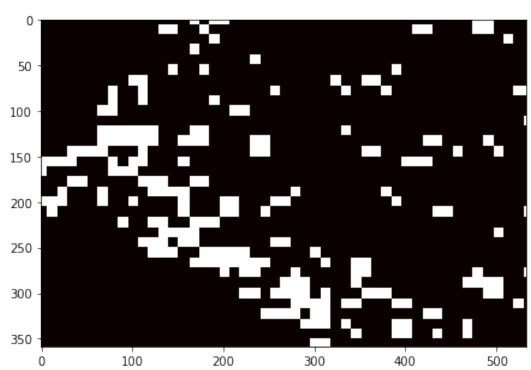


Resulting architecture

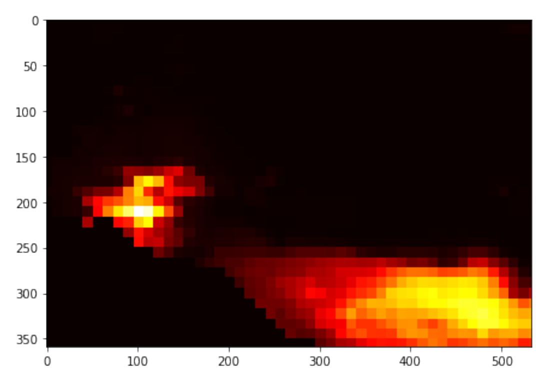
3 Computational Experiment

The experiment goal is to train and evaluate our model on TerraClimate and meteo.ru data.

At every timestamp we have tensor of climatic variables.



snowmelt variable



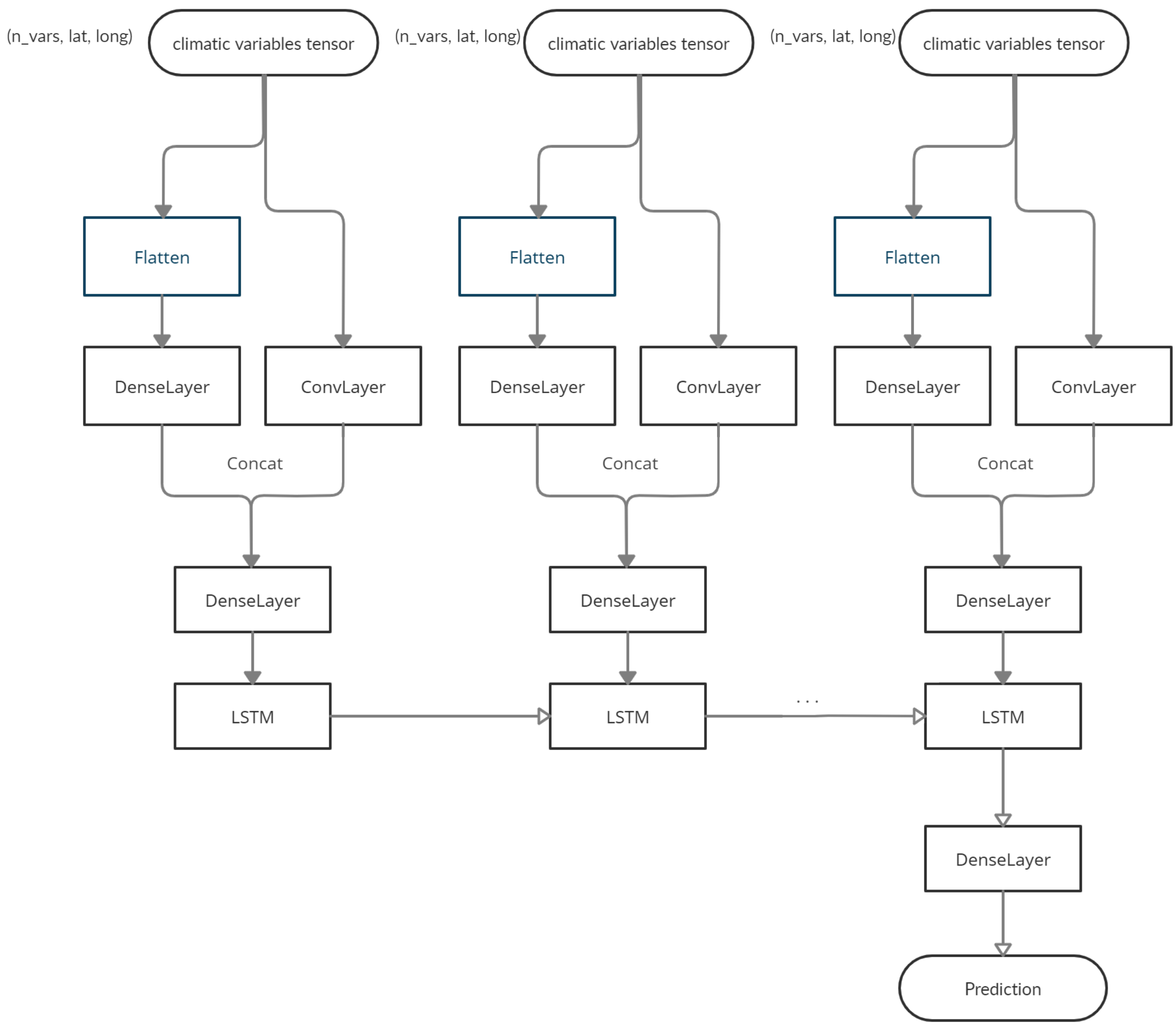
runoff variable

These variables corresponds to Krasnodar region.

Using convolution and dense layers it is possible to connect different variables and their spatial neighbourhood. Each hourly climatic tensor transforms to a vector after applying dense and convolution layers and then concatenating outputs from them. After that we have vectors that corresponds to its hour. Until this moment we have just used spatial component of the data. To take in account the temporal component of the data we use LSTM layers to analyze temporal patterns which are typical for hail days.

3.1 Architecture

HailNet architecture is visualized as following:



HailNet architecture

3.2 Experiment flow

1. Creating torch.DataLoader of downloaded climatic data from TerraClimate using Google Earth Engine and hail events data from meteo.ru.
2. Feed 4-sized batches into training algorithm for 100 epochs.
3. Updating HailNet parameters optimizing Extreme Value Loss with Adam optimizer.

4 Error analysis

Using 22 samples for training we achieve some good c

	precision	recall	f1-score	support
No Hail	1.00	1.00	1.00	12
Hail	1.00	1.00	1.00	10
accuracy			1.00	22
macro avg	1.00	1.00	1.00	22
weighted avg	1.00	1.00	1.00	22

Training classification report

	precision	recall	f1-score	support
No Hail	1.00	1.00	1.00	3
Hail	1.00	1.00	1.00	4
accuracy			1.00	7
macro avg	1.00	1.00	1.00	7
weighted avg	1.00	1.00	1.00	7

Testing classification report

Looking forward to optimize hyperparameters and solve the problem of extreme event classification.

References

[1] S. R. Cachay, E. Erickson, A. F. C. Bucker, E. Pokropek, W. Potosnak, S. Osei, and B. Lütjens. Graph neural networks for improved el niño forecasting. *CoRR*, abs/2012.01598, 2020.

[2] L. Cai, Z. Chen, C. Luo, J. Gui, J. Ni, D. Li, and H. Chen. Structural temporal graph neural networks for anomaly detection in dynamic graphs. *CoRR*, abs/2005.07427, 2020.

[3] NSSL. Nssl's hail research, 2020.

[4] Verisk. The hail hazard and its impact on property insurance, 2020.

[5] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks, 2020.