Machine Learning to predict Land Degradation in Ngong Forest -by Mwahunga
Report

Model used: Long Short- Term memory (LSTM)
What is LSTM?
LSTM is a neural network
A neural network is a model that is designed to recognise patterns and learn from data
In this case, we use LSTM due to its ability to detect trends in previous data and make predictions

Jupyter notebook code cells explained:

1.Initial cells

Importing dependencies i.e geemap,ee(earthengine-API),matplotlib,torch,numpy
Assigning roi
Obtaining sentinel images, cloudmasking
Calculate ndvi and obtain images from 2021,2022,2023
Display images in an interactive map
Create timelapse animation of ndvi images for visualization

2. Exporting NDVI arrays

Each NDVI median composite for each year is converted to a numpy array of the shape
array_2018.shape = (335, 335); This means 335 rows (North-South) and 335 columns.(East-West)

These represent the pixels which have various ndvi values i.e.

	Col 0,	col 1,	…col 334
Row 0	[0.752,	0.748, ...,	0.74]
Row 1	[0.753,	0.751, ...,	0.778]
Row 334	[0.758,	0.753, ...,	0.78]

- basically each image is converted into a list of figures based on the pixel values of each pixel

3. Creating label arrays

Define a function that creates labels
Here we first compare nparrays (numpy arrays) of two consecutive years, if pixels in these years change their value for their value drop by more than 0.05 those pixels have degraded thus assigned value 1, other pixels that are stable or have assigned 0. Thus for 3 years we will get 2 nparrays containing 0s and 1s (binary arrays) showing degradation

4. Preparing dataset

-PyTorch library used for training
-Normalization is done since Neural Networks work best with values between 0 and 1 (ndvi values from -1 to 0 are removed)
-This is done in the following lines of code; self.data = np.clip(self.data,0,1)/0.8

-total pixels in image roi =335x335=112,225 pxls
theoretical max samples 112,225
-for a pixel in 2021 and 2022, it is compared to its NDVI value in 2023, if it is reduced by >threshold 0.05 then value assigned is 1.This is done for all pixels in the time window obtaining a binary image

5. Build the model

LSTM
LSTM model for degradation prediction
What is LSTM
Long Short-Term Memory
Type of neural network
Example:
list of nparrays [0.75, 0.72, 0.68]
normal neural network thinks: "I see 0.68" and forgets 0.75 and 0.72
-> Cant detect trends
LSTM thinker: "I see declining pattern: 0.75 > 0.72 > 0.68"
-> Remembers history

6. Train the model

Epochs:
Refers to one complete pass through all training data e.g reading through a book once =1 epoch, reading through it 20 times =20 epochs
Total training pixels =112,225 x 2 = 224,450pxls (we do x2 cause its 2 years 2021 and 2022)
Epoch 1
Input pixels like [[0.75,0.72,0.68]]
Forward pass: make prediction
Calculate loss: how wrong are predictions
Validates same pixel in 2023 image
N/B: this goes on for 20 times (20 epochs) inorder to increase accuracy of prediction

A graph of training loss and validation loss is then created to demonstrate accuracy of the predictions
Loss was calculated as follows:

BCE loss = - [y * log (ŷ) + (1-y) * log (1-ŷ)]
e.g. y=1 (true label=degraded) ŷ=0.9 (model prediction)
loss = - [1 * log(0.9) + 0 * log(0.1)] = - [-0.105] ≈0.105
loss=0.105 (Low) Good prediction.

The lower the loss higher accuracy of prediction
Carried out in 20 epochs thus an improvement in prediction
Loss is obtained at each epoch. The graph shows the decline of the loss over all the 20 epochs thus an improvement in prediction
When both the losses stop declining, training is complete and we have reached the best model.

7. Predict Degradation

A risk map of high risk degradation areas is obtained

All deliverables for the project: NDVI interactive maps, Timelapse animation, graph showing trend for the mean yearly ndvi, prediction accuracy graph, Predicted Risk areas