Proposal for Final Project: Weather Forecasting for King City, Canada using combined CNN and RNN Architectures

Michael Robinson

CSC 578 Section 901

5/19/23

1. PROBLEM

I intend to predict precipitation in a 40 kilometer radius around King City, Canada. Twenty-two thousand images taken ten minutes apart for the time frame of four months will be split into training, validation and test sets. A sample image is shown (1).

A picture containing text, map, diagram, screenshot

Description automatically generated

Figure 1. Image from dataset. Weather will be predicted for the innermost circle.

Forecasting for the amount of precipitation in the 40km radius of the innermost circle will be predicted at one hour intervals (up to four hours ahead). Four images one hour apart will be used as the inputs for these predictions. A threshold precipitation amount will be used to determine if the weather for the region is considered mostly clear or mostly rain/ snow. Although I have not determined a threshold value yet, a 40km radius circle will have an area of approximately 5000 km2 and thus I anticipate using a threshold around half of this area times 0.1 (for light precipitation as shown in the legend of (1) above) to get about 250 (km2-mm of precipitation/hr).

The input for the CNN will be X = images at times [0000, 0100, 0200, 0300]. After passing through several filters/ max pooling layers, the output will be passed into a RNN. The output will be in the form of a four element binary vector corresponding to if the threshold amount of precipitation was met during the next four hours.

The labels will determined by comparing a weighted sum to the threshold value for each hour of prediction:

Y-labeled(Y(t)) = ()t = [0 or 1, 0 or 1, 0 or 1, 0 or 1]

Where

Y = four images at times [0400, 0500, 0600, 0700]

is identity function returning 1 if true and 0 if false

w is weight provided by legend for a certain color i for a pixel (mm/ hr of rain), see legend in (1) for values

p is number of pixels of certain color i in the center region (inner circle of image)

t represents the four time periods to predict over

I plan on using log-likelihood for the loss function.

1. MOTIVATION

I would like to further my understanding of CNNs and RNNs. I have explored forecasting in the financial sector (cryptocurrency trading) on my own and enjoyed, and would like to understand how this differs from weather forecasting. I also am not as familiar with CNN and although homework 7 will give me some experience, using them again in this project will improve my understanding of them. I also want to relearn some basic statistical tests (hypothesis testing/ t-test specifically to see if my model is performing better than random chance).

1. TOOLS AND DATASET

The dataset to be used will be from Kaggle: <https://www.kaggle.com/datasets/skillsmuggler/weather-radar-king-city-canada>. Images will be cropped to remove the right side (the legend). The twenty-two thousand RGB images contain 480 x 480 pixels (when the legend on the right is cropped out). I have some concerns regarding computational power, I intend to use Google Colab GPU option to speed up training. I will simplify my model as needed to speed up processing (at the cost of prediction accuracy). Cropping the images or only using part of the images will also be used to speed up training if needed.

1. REFERENCES TO BE USED

One resource, “Using CNN-LSTM Model for Weather Forecasting” (<https://www.researchgate.net/publication/367455415_Using_CNN-LSTM_Model_for_Weather_Forecasting>) by Michael Fan et al. published by IEEE in 2022 will be used as a guide to various neural net architectures. Also they used just 4 images 15 minutes apart to predict the next 8 hours, but this was a bit different (they were not predicting location of weather events, but patterns in various weather variables in a region). This helped me in choosing how many images to use as input.

Another resource includes this article ( “Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data”, <https://www.nature.com/articles/s41598-020-57897-9/> ) by Ashesh Chattopadhyat et al. from *Nature* that uses both a 4 layer and a 2 layer CNN. The article also discusses using K means clustering to classify weather patterns. It distinguished between winter and summer months when training and testing. All my data is from a four month period (April to August – I will assume these months are comparable in weather patterns).

1. EXPECTED LEARNING

I expect to gain a deeper understanding of both CNNs and RNNs. I also expect to learn more about weather forecasting, specifically how machine learning can be used in it. I will become more familiar with statistical tests and gain further experience in time series analysis.

1. WEEKLY PLAN

5/22 to 5/25:

* Identify threshold value
* Preprocess data: crop images, separate into X and Y, label Y, split into train, validation and test sets
* Determine if computational power is sufficient/ feasible 🡪 if not, further crop images
* Determine if there are sufficient training examples 🡪 if not, explore data augmentation

5/25 to 6/1:

* Make a base model?
* Perform hyperparameter tuning: number and size of filters for CNN, top layer architecture of RNN
* Time depending: explore various additional RNN types (bidirectional, LSTM, GRU)
* Work on writeup and part B

6/1 to 6/8:

* Analyze results
* Hypothesis testing / statistical tests
* Write report
* Finish Part B and writeup