Real-time Weather Forcasting using Autonomous Deep Learning

Shijia Yan

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Motivation

- Traditionally weather forcasting through physical simulations(NWP) needs huge amount of computing power

BBC

Met Office and Microsoft to build climate supercomputer

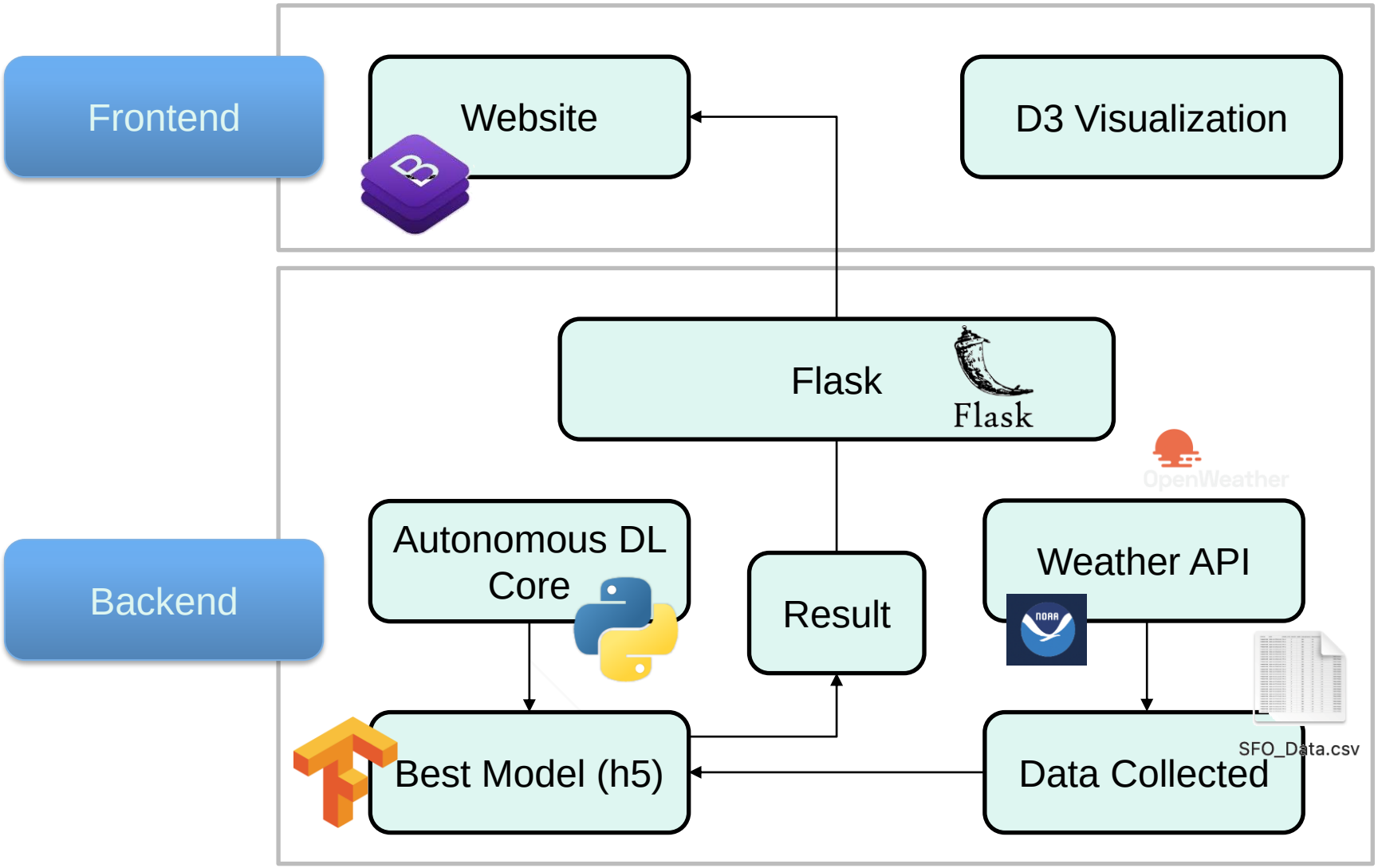
The Met Office is working with Microsoft to build a weather forecasting supercomputer in the UK. They say it will provide more accurate weather ...

2 days ago

- Machine learning applications can achieve relatively high accuracy while using much less resources – suitable for online service

Architecture	Reference	Recent Developments
convolutional neural network (CNN)	AlexNet (Alex *et al.* 2012) VGG (Simonyan & Zisserman, 2013) ResNet (He *et al.* 2015) GoogleLeNet (Szegedy *et al.* 2015)	VGG (Shi *et al.* 2018 [66]) ResNet (Pothineni *et al.* 2019 [68]) Vgg, ResNet (Wen *et al.* 2020 [70]) Inception blocks (Kleinert *et al.* 2021 [72])
long short-term memory network (LSTM)	RNN (Bengio *et al.* 1994) LSTM (Gómez *et al.* 2003)	LSTM (Qing & Niu, 2018) PhyDNet (Le Guen, 2020)
variational autoencoder (VAE)	Vanilla VAE, 2013	None
generative adversarial neural network (GAN)	Vanilla GAN (Goodfellow *et al.*, 2014)	MD-GAN (Xiong, 2018) conditional GAN (Schmidt *et al.* 2020)

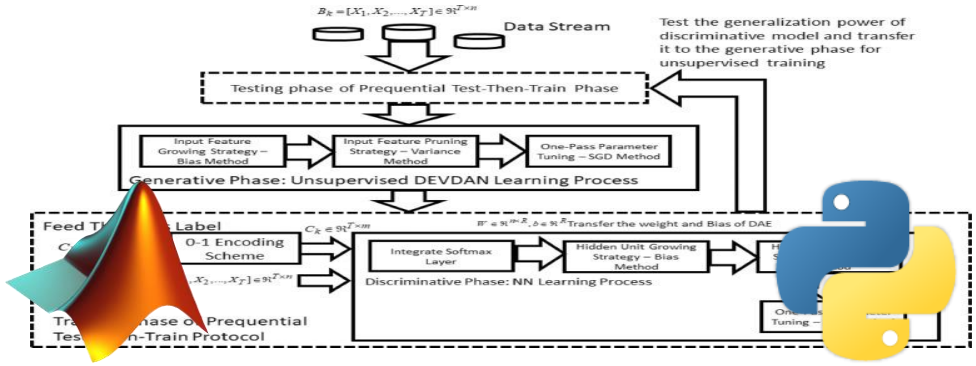
Workflow overview

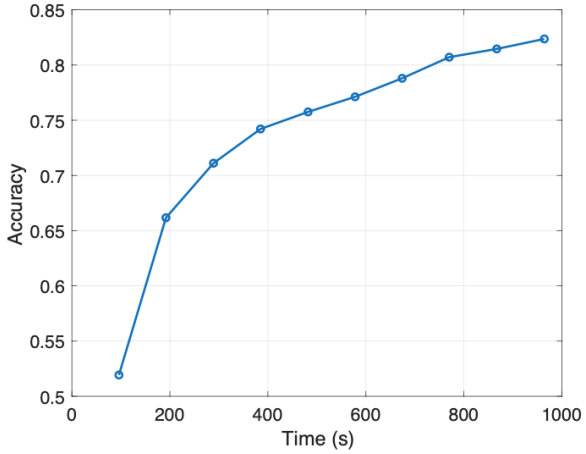


Deep Learning Core

- Implemented DEVDAN with weather data(originally in Matlab, rebuilt in Python 2.7)
- Training dataset: NOAA historical data, 3 years, 20 locations, more than 20,000 days of data

- Weather data: daily reported data, consisting of precipitation, snow, wind, sea-level pressure, cloud
- Data chunk: 14-days each chunk
- Best model saved in .h5, ready for prediction





COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

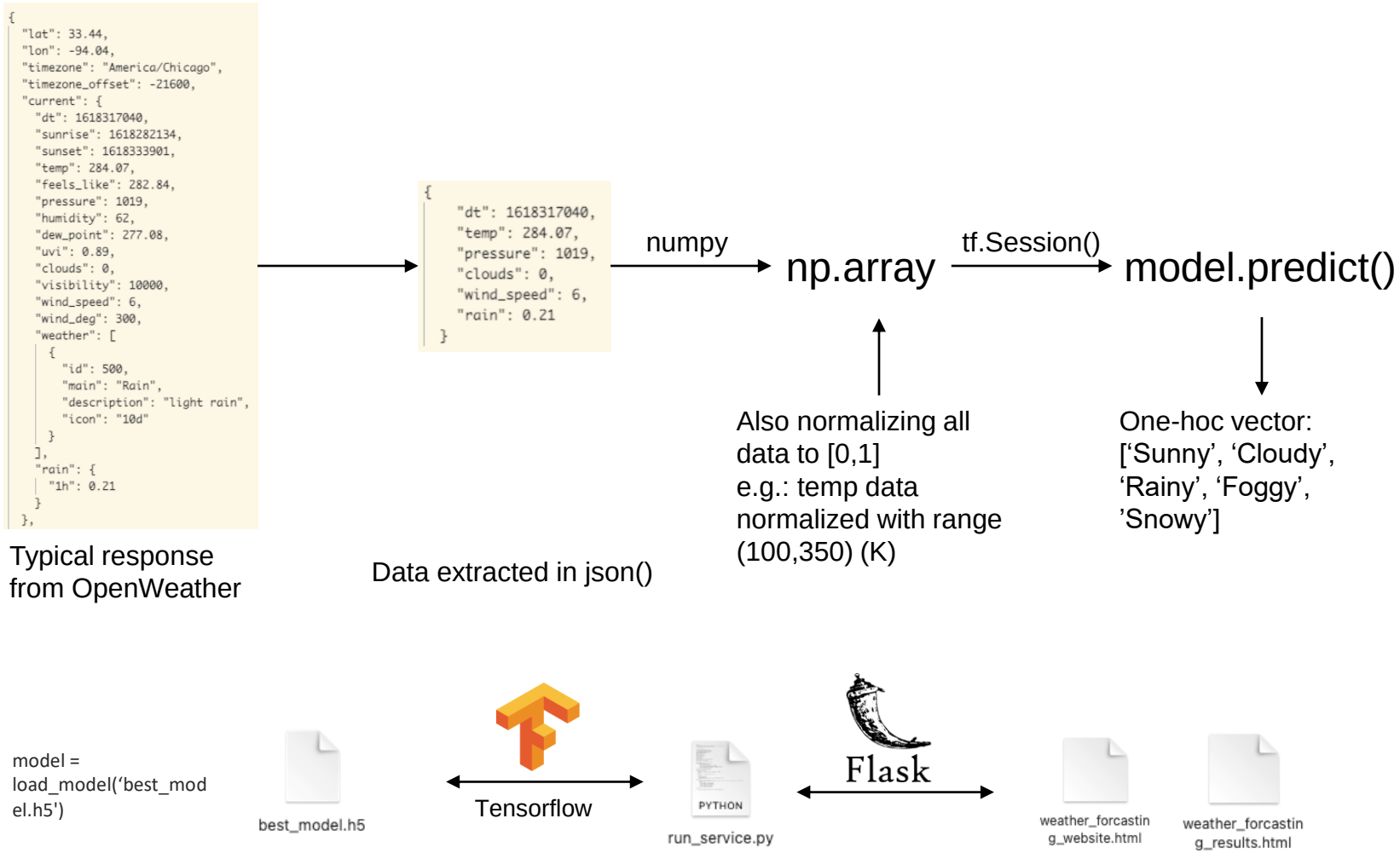
Data Fetching

- Visualcrossing provides 1000 free results per day with 50 years of history
- Open Weather only provides past 5 days data for free
- API key was obtained by creating account at these websites
- Data was then stored and filtered to only high-low temperature, wind, cloud, precipitation and pressure
- Use tf.Session() from keras to load model and generate prediction results

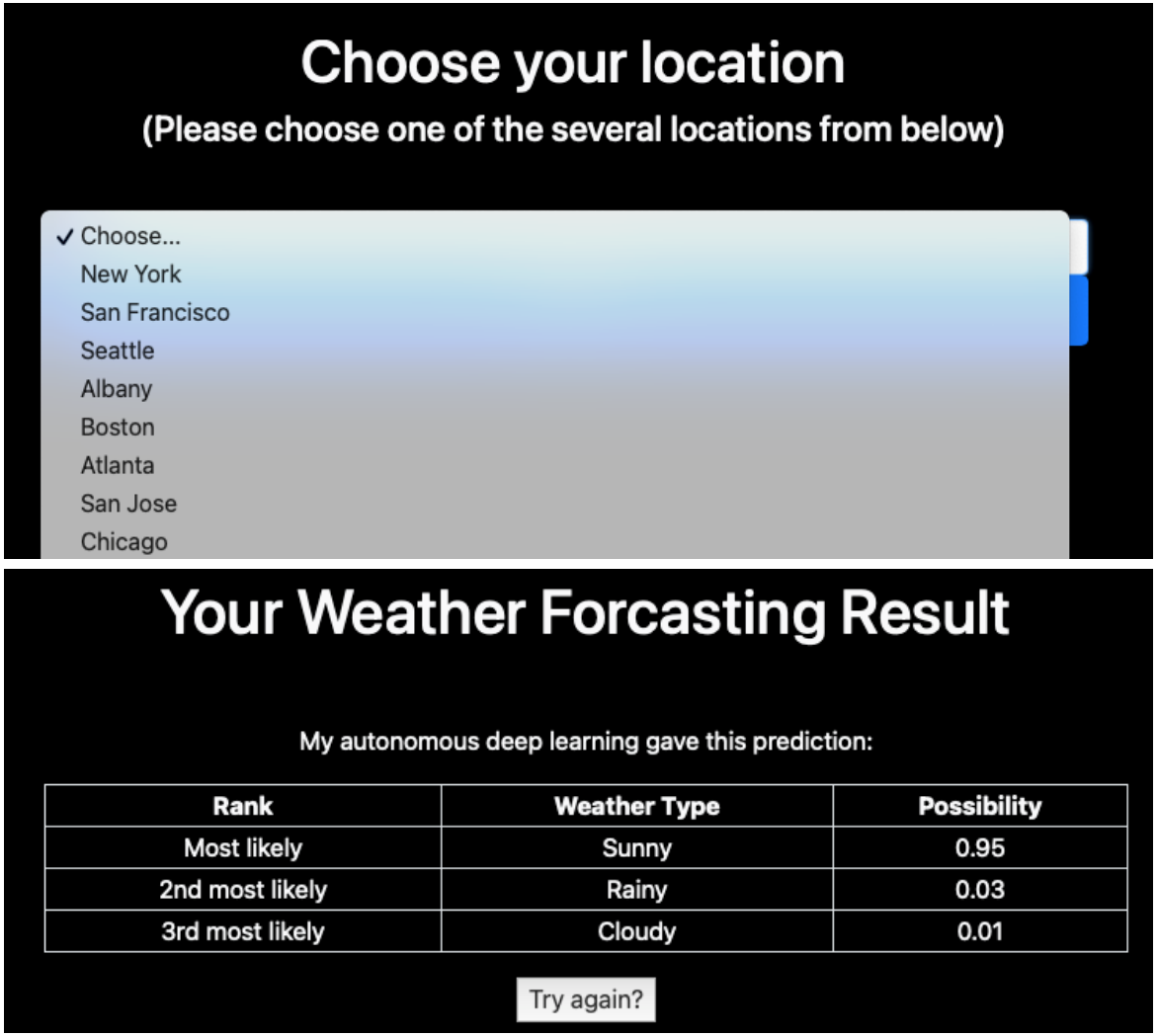
Challenges

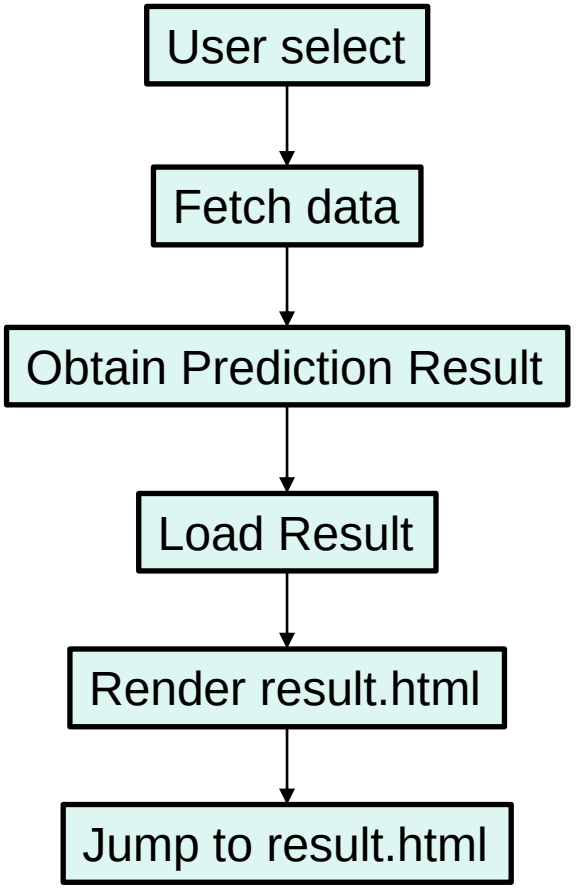
- Challenges:
  - Volume: large amount of data from weather history for training.
    - Data filtering using python, stored in csv file → ready for training
    - Store pre-trained model into .h5, avoid training on-the-fly with whole dataset
  - Velocity: long training time
    - Use pre-trained model to generate prediction results
    - Update model by training new models in the background once in a while
  - Variety: worldwide locations coverage
    - Hard to obtain worldwide weather data
    - Data in most US cities are available through multiple API websites

Online Forcasting



Weather Forcasting Page





I haven't complete the code using D3.js so we're stuck with this ☹

Places for improvements

- Result visualization with D3: creating interactive chart page
- Construct higher-dimentional data to improve prediction accuracy
- Obtain data from more sources

- Develop into online Autonomous ML Platform for various other applications

- Q & A