



FYP-2021 Intermediate Presentation

4 messages

"葉志立 YIP Chi Lap [Beta]" <clyip@cs.hku.hk>

Thu, Jan 13, 2022 at 7:29 PM

To: 方均正 Fong Kwan Ching <u3556490@connect.hku.hk>

Cc: "\"葉志立 Yip Chi Lap [Beta]\"" <clyip@cs.hku.hk>, "\"蔡綺瓊 Choi Yi King [Loretta]\"" <ykchoi@cs.hku.hk>

Hi,

Here it is the notes I jotted down during your presentation.

Regards,

葉志立 YIP Chi Lap [Beta].

PhD BEng(CompEng) FHKMetSoc MACM MIEEE

Department of Computer Science,

The University of Hong Kong.



202201121120-IntermediatePresentation-FcstTCProp-Fong Kwan Ching.txt

2K

Fong, Kwan Ching <u3556490@connect.hku.hk>

Fri, Jan 14, 2022 at 3:03 PM

To: "葉志立 YIP Chi Lap [Beta]" <clyip@cs.hku.hk>

Cc: "蔡綺瓊 Choi Yi King [Loretta]" <ykchoi@cs.hku.hk>

Dear Dr Yip,

Thank you for the feedback. I would like to add a few comments to the issues you and Dr Choi raised:

- The MLP structure has been varied quite a number of times, ranging from small structures (1-3 hidden layers with 32-64 neurons each) to large ones (10 hidden layers with 1024 neurons each) and nothing seemed to work, regardless of hyperparameter choices. The same results were observed when I used [scikit-learn](#) instead of [PyTorch](#) to build the MLPs.

- I agree that some smoothing or additional preprocessing will be beneficial. At the moment, I am trying to use tools such as PCA to extract the most useful information out of the dataset, but the preliminary results seem to suggest more than 99.8% of the variance needs to be kept, otherwise the models downstream will suffer severely.

- Model comparisons are carried out in a tripartite manner:

- Statistical-dynamical model vs custom-built baseline: There are no directly/exactly comparable controls that I can use, so custom baselines need to be built in order to justify the addition of dynamical data.

- Baseline vs HKWW: While HKWW's model is only a close analogue of what I am trying to achieve, it is used to check whether a sensible baseline is built.

- I surmise this is a sufficiently reasonable method.

- Probability of impact levels = probability of some warning signals?

- In short, yes, because warning signals are taken as a proxy of the impact levels. The assumption is that the signals represent impact levels well enough. An authoritative and quantitative impact level measurement cannot be found, thus a combination of warning signals and TC closeness (<100km?) are taken as stand-ins.

- What is needed to transfer this study to other cities?

- A measure of impact level will then be needed, especially if the city in question does not have a warning signals system that works like Hong Kong and Macau's. For instance, in Taiwan, TC warnings do not necessarily encode impact levels well. The other data sources will be identical, but the calculation of feature variables may be somewhat different (the latest revision to the baseline dataset involves calculating the [radial distance and forward azimuth](#) to the city in question).

- Why are the date/time values in int64?

- The models do not accept timestamps, thus they have to be separated into MM, DD and HH, each an integer.

The default integer data type in Pandas is int64 and despite the memory wastage (only 8 bits are needed) I kept it that way to minimize potential compatibility issues when passing the data around different packages.

I hope these have answered your questions. I will take them into consideration when I write the interim report later. Should you have further questions, please simply write back.

Best regards,
FONG Kwan Ching
Student

[Quoted text hidden]

"葉志立 YIP Chi Lap [Beta]" <clyip@cs.hku.hk> Thu, Jan 20, 2022 at 2:33 PM
To: 方均正 Fong Kwan Ching <u3556490@connect.hku.hk>
Cc: "\"葉志立 Yip Chi Lap [Beta]\"" <clyip@cs.hku.hk>, "\"蔡綺瓊 Choi Yi King [Loretta]\"" <ykchoi@cs.hku.hk>

Hi,

* MLP: Interesting; can you outline in more detail what you have explored, like which NN architecture, the parameters, and the attributes used?

* PCA may be useful to find out the principal axes of linearly independent data, but it's always good to identify the right features by understanding the physics of the problem behind.

Regards,
葉志立 YIP Chi Lap [Beta].
PhD BEng(CompEng) FHKMetSoc MACM MIEEE
Department of Computer Science,
The University of Hong Kong.

[Quoted text hidden]

Fong, Kwan Ching <u3556490@connect.hku.hk> Fri, Jan 21, 2022 at 9:25 AM
To: "葉志立 YIP Chi Lap [Beta]" <clyip@cs.hku.hk>
Cc: "葉志立 Yip Chi Lap [Beta]" <clyip@cs.hku.hk>, "蔡綺瓊 Choi Yi King [Loretta]" <ykchoi@cs.hku.hk>

Dear Dr Yip,

Thanks for the reply. The following are several sets of parameters and NN architectures:

1.

Model structure:

```
self.fc = nn.Sequential(  
    nn.Linear(18, 128),  
    nn.ReLU(),  
    nn.Linear(128, 128),  
    nn.ReLU(),  
    nn.Linear(128, 64),  
    nn.ReLU(),  
    nn.Linear(64, 4),  
    nn.Sigmoid()  
)
```

Parameters:

```
TRAIN_RATIO = 0.85  
NUM_EPOCHS = 300  
LR = 0.01  
BATCH_SIZE = 32
```

```
criterion = nn.BCELoss()  
optimizer = SGD(model.parameters(), lr=LR)  
acc = sklearn.metrics.accuracy_score(target, preds)
```

The predictions are rounded to the nearest integer (0 or 1) before being passed to the accuracy scorer.

2.

Model:

```
self.fc = nn.Sequential(
    nn.Linear(18, 128),
    nn.ReLU(),
    nn.Linear(128, 128),
    nn.ReLU(),
    nn.Linear(128, 64),
    nn.ReLU(),
    nn.Linear(64, 32),
    nn.ReLU(),
    nn.Linear(32, 4),
    nn.Sigmoid() )
```

Parameters:

```
TRAIN_RATIO = 0.85
NUM_EPOCHS = 1000
LR = 0.01
BATCH_SIZE = 128
criterion = nn.BCELoss()
optimizer = Adam(model.parameters(), lr=LR)
acc = sklearn.metrics.accuracy_score(target, preds)
```

3.

Model:

```
self.fc = nn.Sequential(
    nn.Linear(18, 256),
    nn.ReLU(),
    nn.Linear(256, 256), # 9 such identical hidden layers
    nn.ReLU(), # 9 such activations
    nn.Linear(256, 64),
    nn.ReLU(),
    nn.Linear(64, 4),
    nn.Sigmoid()
)
```

Parameters:

```
TRAIN_RATIO = 0.85
NUM_EPOCHS = 1000
LR = 0.01
BATCH_SIZE = 128
criterion = nn.BCELoss()
optimizer = Adam(model.parameters(), lr=LR)
acc = sklearn.metrics.accuracy_score(target, preds)
```

There is also a configuration with the same parameters but the 9 layers in the middle have 1024 neurons instead. In addition, the same MLP modelling attempts were repeated using sklearn instead of pytorch, but none was ever successful in the slightest.

For feature selection using understanding of the physics behind, it is certainly an important thing. I am sorry that I forgot to mention that interaction terms were computed first before PCA was used to reduce the dimensionality, because the untransformed dataset is small enough to directly work with,

but not the one with 1300 interaction terms. Following the procedures used by Neumann when he devised the original CLIPER,

F-test was carried out to select the more useful predictors from among the 1300, but it typically chose position-related terms only (because these explain most of the variance in the data), which was far from sufficient. Therefore I switched to PCA, which can find more sensible candidates.

Regards,

Fong Kwan Ching

Student

P.S. The typesetting of this email seems to be somewhat broken, please accept my apology for that.

[Quoted text hidden]