Agenda

- Findings and observations
- Existing techniques by the lab
- What's next
- What other things the lab is working on?
- RA decision

Observations:

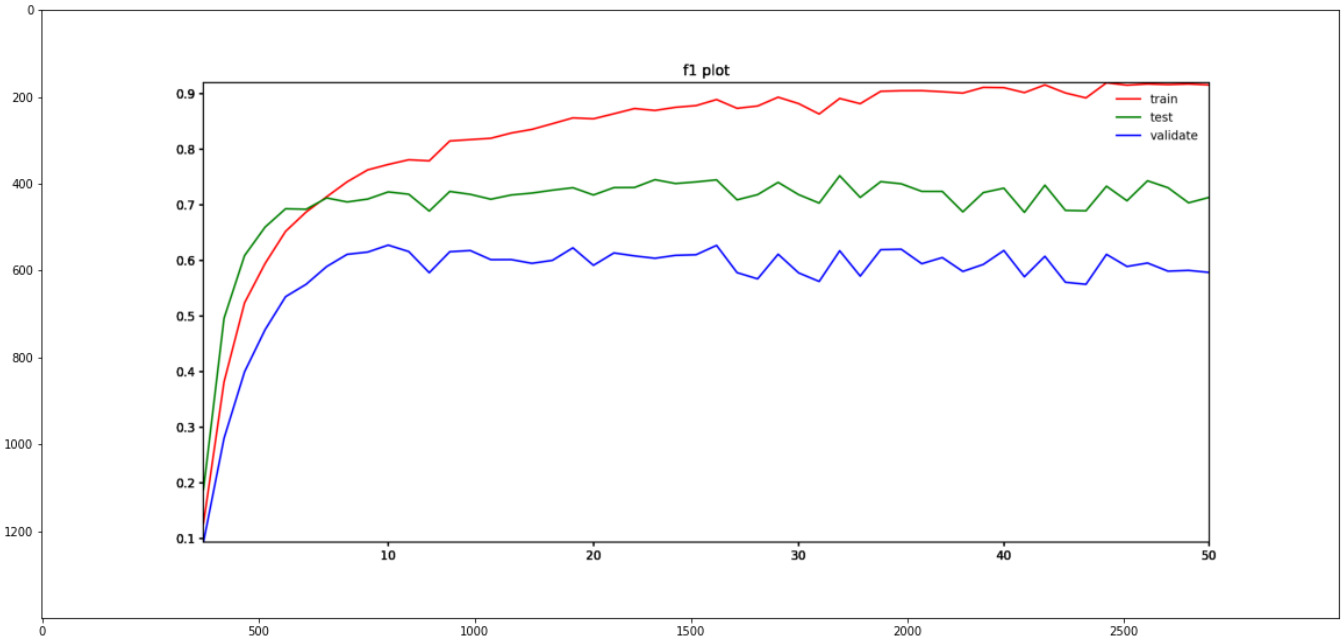
- Difficult in finding the documentation for NeuralNLP classifier
    - Especially, the taxonomy file for hierarchical classification
- After the bugfix of the taxonomy file, the precision improves significantly (recall flcutuates)
- TextRNN has (non-significantly) better performance than Transformer encoder-based classifier
- Using pre-trained Glove embeddings, performance is slightly better
- Possibilities for improvements
    - More data (currently 6k out of 22K data points are loaded perfectly)
    - HMCN-based classification
    - BERT embeddings? (not sure how)

Charts

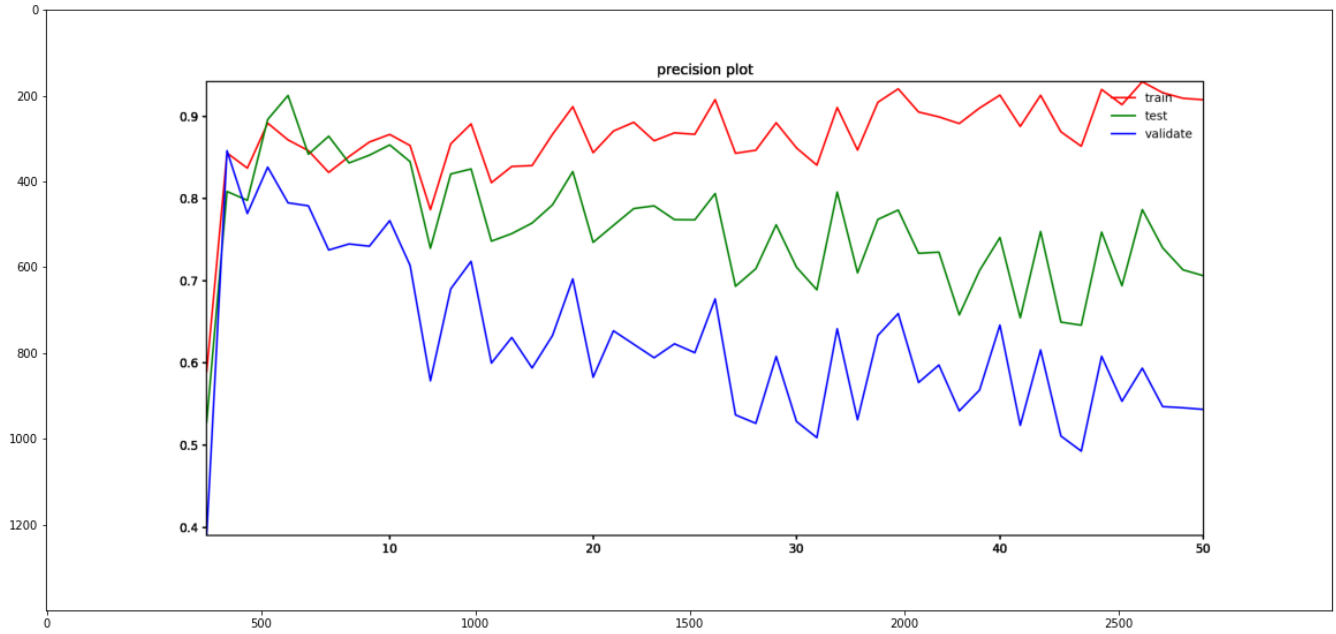
Run 1 (Transformer)

{'ptype': 'best test',
 'epoch': 10.0,
 'precision': 0.865285,
 'recall': 0.621685,
 'f1': 0.723531}

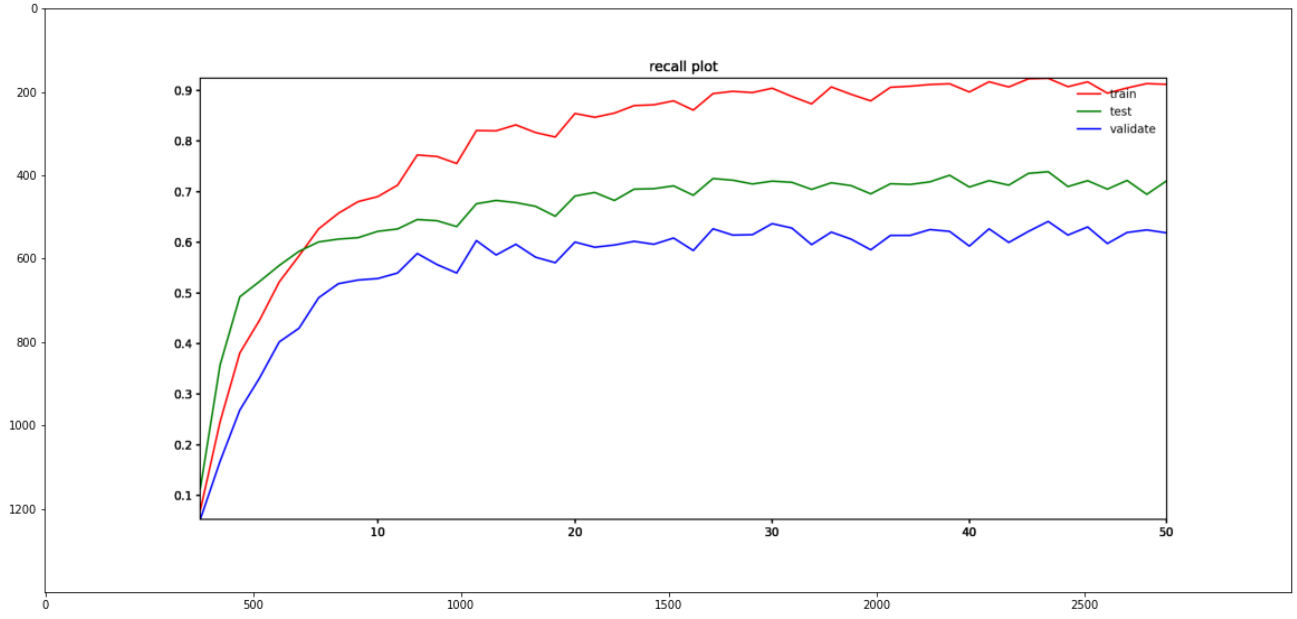
f1-plot



Precision plot



Recall Plot

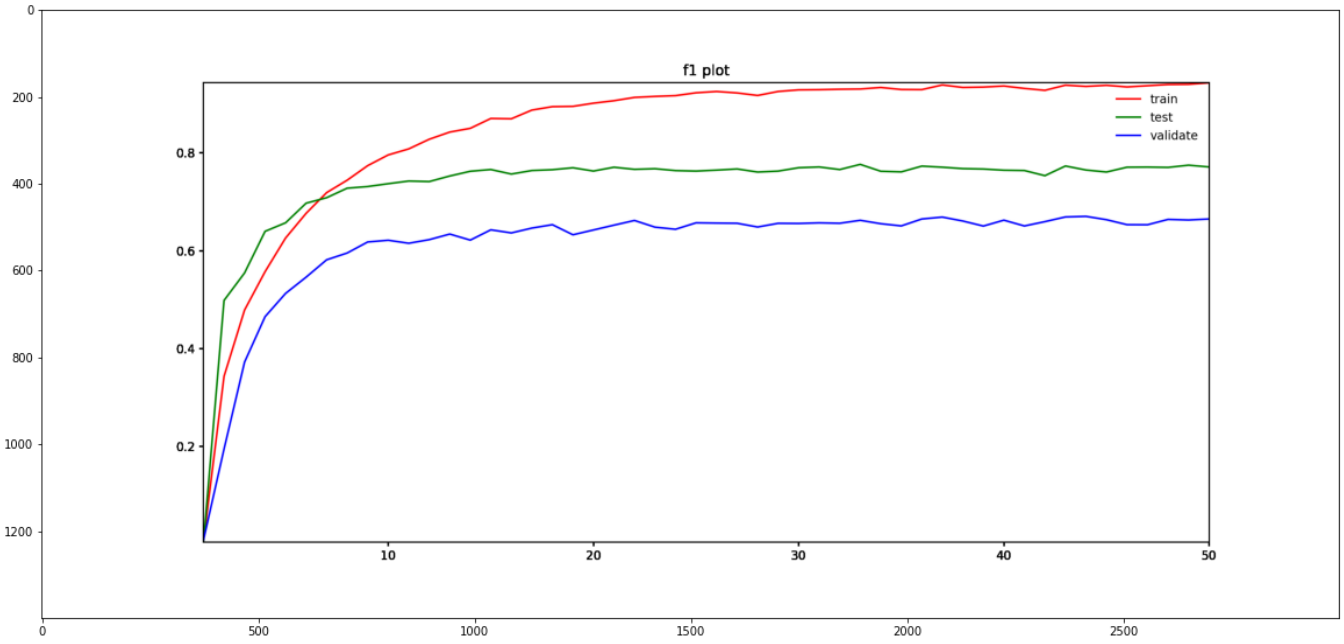


Run 2 (TextRNN)

Glove 50d

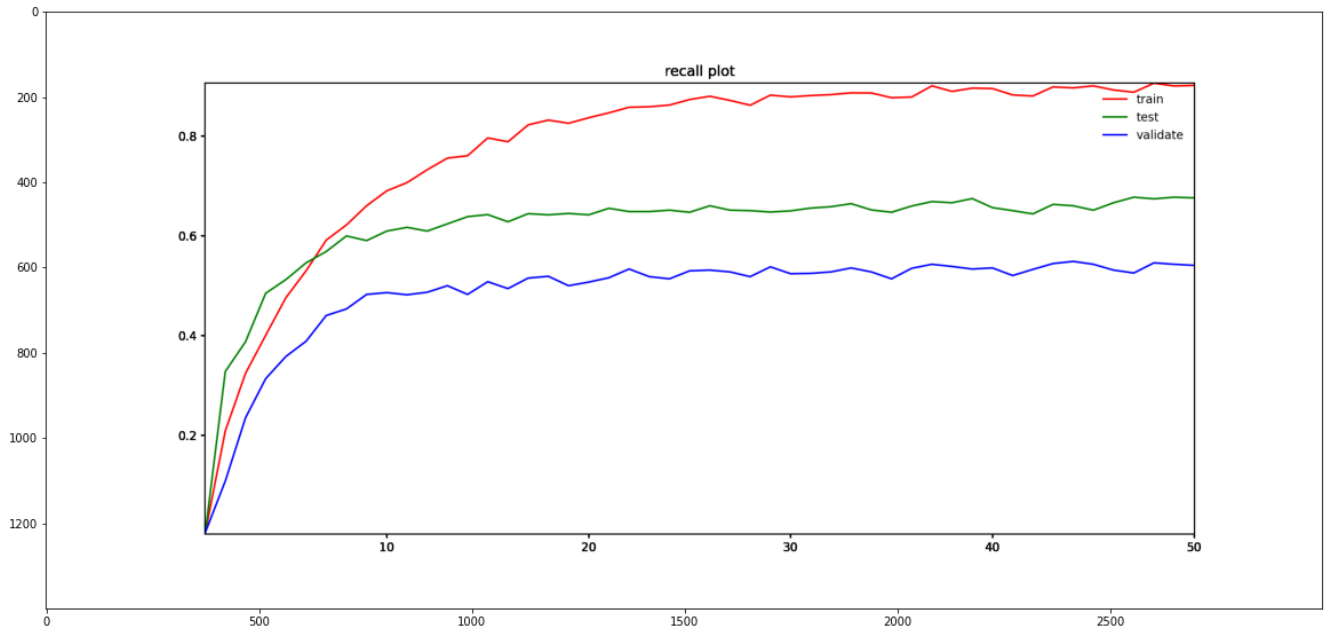
{'ptype': 'best test',
 'epoch': 44.0,
 'precision': 0.909033,
 'recall': 0.660307,
 'f1': 0.76496}

F1-plot



Precision - plot

Recall Plot


recall plot
train
test
validate
0
200
400
600
800
1000
1200
0.8
0.6
0.4
0.2
10
20
30
40
50
0
500
1000
1500
2000
2500