Call 15th Jan:

- Pay by week

- TPU v3-64 payed by year hence free

- Everything he sent me can still be used

- He obtained the data; 200 GB; 2nd Batch will be 100 GB; 3rd batch will be 2 TB ;; CommonCrawl is 1 PB

- 3rd batch actually smaller as not that clean

- He’ll upload to Google Bucket

- Taoyan is working on TabNet – They’ll upload a new paper surpassing LightGBM & TabNet

- Taoyan is also working on solving a differential equation using DL

- Luo is working on CV / RL ; not officially started

- One of the ICML Reviewer working on some theory

- Train Performer on a full-fletched

- entmaxalpha 1.5 – sparse version of softmax ask fugang - paper is called sparse-to-sparse transformers (probably worse than performer)

- contrast training with contrast loss? ; We use T5 for pre-training

- CHECK T5 PAPER - DONE

- AT HOW MUCH INPUT TOKENS IS IT WORTH USING PERFORMER?

Call Jan 11th:

- Not hastly – look for pre-training material

- TF 2.4

- **EoW:** List everything; What I think is best

- TPUs, cuz cheaper + faster > No need to worry about distributed learning

- Adam/AdamW

- T5 Alternatives?

- Layer weight-sharing – Perhaps share 0, 2, 4, 6, i.e. only every 2nd

- Embedding Layer very different for TPU

Call before Jan 11th:

- 3 years company (Ran been there for a year)

- Enough cash income to do stuff

- Ran Wang

- WeCredo is a startup focusing on FinTech

- Problem in CN: People/Big Comps do not give out their financial reports

- WeCredo goal is to predict whether small companies will default

- Lots of unstructured Data (mostly company descriptions)

- Some company located in an unknown village, wants to surpass google in some years; have media coverage as data; data on lawsuits

- 10 years time period; updated every month

- Xinhua buys the data from a legal resource; WeCredo uses it for free

- Changelog (which is very unstructured)

- how can we use the tabular data?

- CV+NLP is too limited; tabular data is more specific; face won’t tell me much

- Chinese BERT is terrible, because no common crawl (CN Wikipedia is only 100MB or so)

- Very long text, such as a statement from jurisdiction, which will be like 10K words – tough to do classification

- Performer might be interesting; Pre-training is very bad

- Train a T5

- They share compute with DeepMind

- Klue & Glue;

- They have to buy the data in CN first ;

- Model has to be sort of small; Albert perhaps layer sharing

- TensorFlow

- 2nd thing: When you do tabular data, TabNet doesn’t work so well. LightGBM

- Cost

- WeCredo;

- Build the platform for Beijing & Zhongqing

- Don’t want investments

- quite a few Kaggle gms

Perhaps:

- TaBERT (https://ai.facebook.com/blog/tabert-a-new-model-for-understanding-queries-over-tabular-data/)