LookAlike model with TF2.x

DIN model with TF2.x code (multiGPU support)

- Train_tf2.py: trainer program
- Model_tf2.py: DIN model in tf2
- Utils_tf2.py: utility functions used, such as: tf1 -> tf2 data conversion
- __init__.py: empty file

MultiGPU with smaller dataset (5M samples)

GPU #	Batch size	Time (seconds)	Speedup
1	4,096	2,868	1.00
2	8,192	1,588	1.81
3	12,288	1,140	2.52
4	16,384	955	3.00
5	20,480	821	3.49
6	24,576	734	3.91
7	28,672	682	4.21
8	32,768	638	4.50

Nvidia GeForce GTX Titan X – 12GB memory

MultiGPU with large dataset (500M samples)

GPU #	Batch size	Time (seconds)	Speedup
1	4,096	33,005	1
4	24,576	25,255	1.3

- Computation time for each 5M batch is < 20 seconds -> total <2000 seconds
- Data-loading kills the speedup

Nvidia GeForce GTX Titan X – 12GB memory

Summary

- DIN model has converted from TF1 to TF2, same dataset generate equivalent performance.
- MultiGPU speedup
  - Close to linear speedup with smaller dataset (can be loaded to memory as a whole)
  - Poor speedup with large dataset (need to be loaded to memory batch by batch), 4GPUs generates 1.3x speedup -> bottleneck is with data-loading instead of parallel computing.