COL 380 : A3

Sidharth Agarwal - 2019CS50661
Mohit Thakur - 2019CS10373

March 27, 2022

Implementation

So in *convert.cpp* which is called in *DataSetup.sh*, we simply sequentially converted all the data stored in .txt format to binary encoded .bin files. Then in *main.cpp*, which is called in *HNSWpred.sh*, we first read the binary files into memory. All the files except vect.bin, which contained binary encodings for news embedding were loaded sequentially. For vect.bin we used MPI_Read() to read certain sections of file parallely, followed by MPI_Allgatherv() to distribute missing data among each node. Then we divided users among different mpi nodes and at each node we further assigned different users to different omp threads. In the end we gathered all the recommendations from each mpi node to 0th node using MPI_Gatherv() and which was followed by printing output to the file.

Readings

The precision recall for different dimensions were on an average for different nodes and number of threads are.

Number of recommend	Precision	Recall
5	0.174	0.122
10	0.163	0.187
15	0.150	0.228

Table for time and max memory usage

N(nodes)	C(cores)	K(no. of recommend)	Time (in s)	Max memory used (in Mb)
2	5	5	74	14190
2	5	10	81	14206
2	5	15	151	14253
2	10	5	48	14243
2	10	10	64	14782
2	10	15	85	14305
2	24	5	29	14365
2	24	10	38	14432
2	24	15	52	14438
5	5	5	47	25643
5	5	10	41	25690
5	5	15	81	25770
5	10	5	38	25832
5	10	10	45	25899
5	10	15	50	25929
5	24	5	24	26011
5	24	10	34	25991
5	24	15	41	26120
10	5	5	39	29030
10	5	10	41	29110
10	5	15	50	29130
10	10	5	31	29190
10	10	10	34	29321
10	10	15	41	29312
10	24	5	23	29416
10	24	10	34	29477
10	24	15	39	29604