BCQ Results

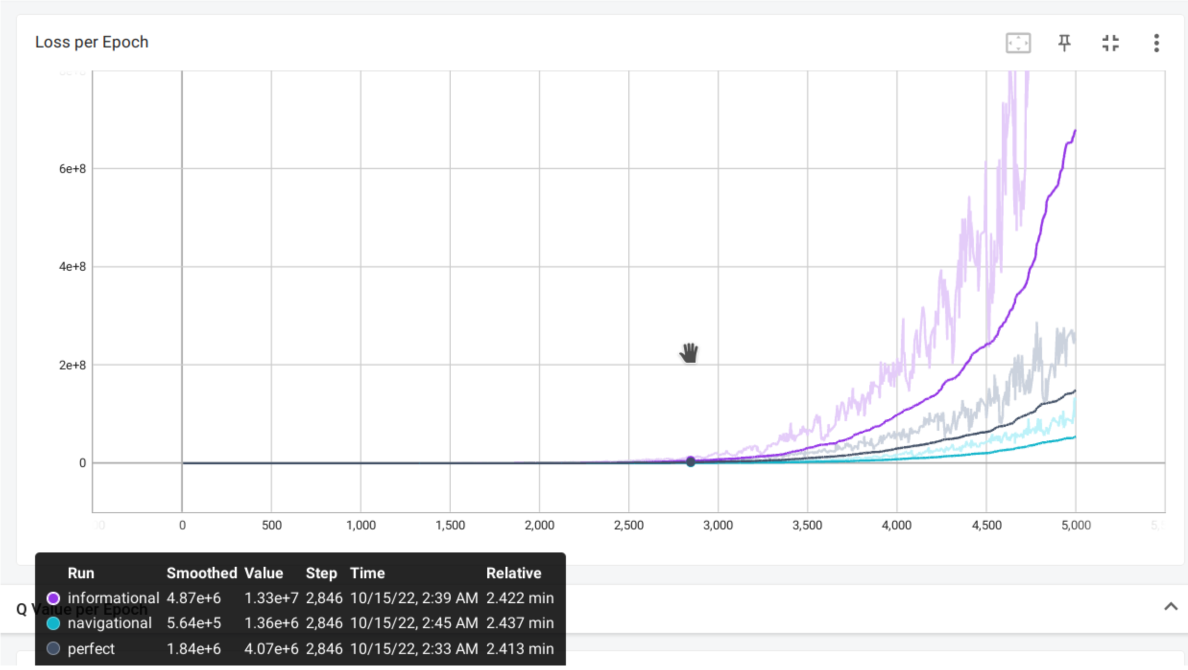
We implement ablation exps with respect to the following factors:

1. *click models*: we implement three different sets of click probability, `perfect [0, 0.5, 1]`, `informational [0.4, 0.7, 0.9]`, `navigational [0.05, 0.5, 0.95]`
2. *perturbation range*: perturbation is the key in BCQ, which constraints the searching and updating area. Intuitively, larger perturbation range leads to larger variance (loss) and greater probability; and a smaller one makes an opposite effect.

1. Click Models

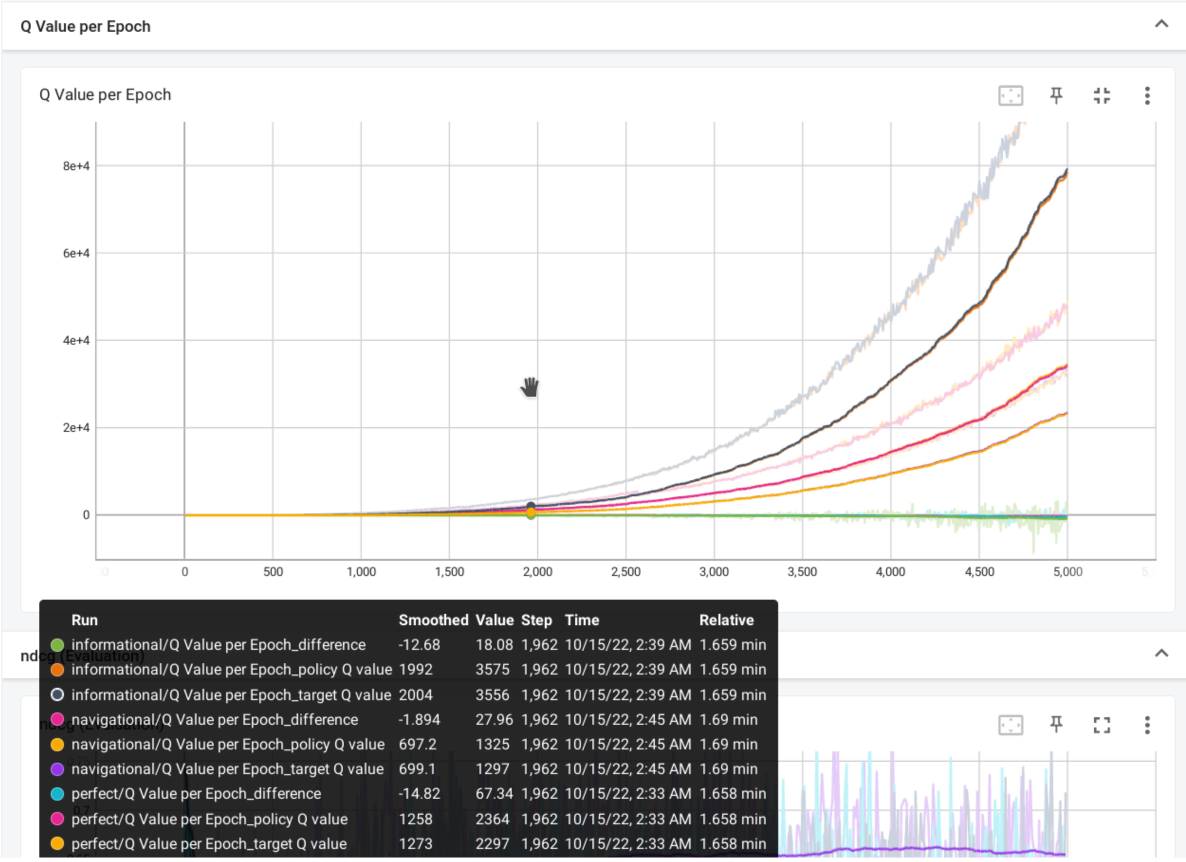
To figure out the effect of click models, we fix perturbation as a single `tanh()` function, and then compared the loss, Q value and evaluation (ndcg@10) of different click models. Here are the results:

- *Loss per epoch*



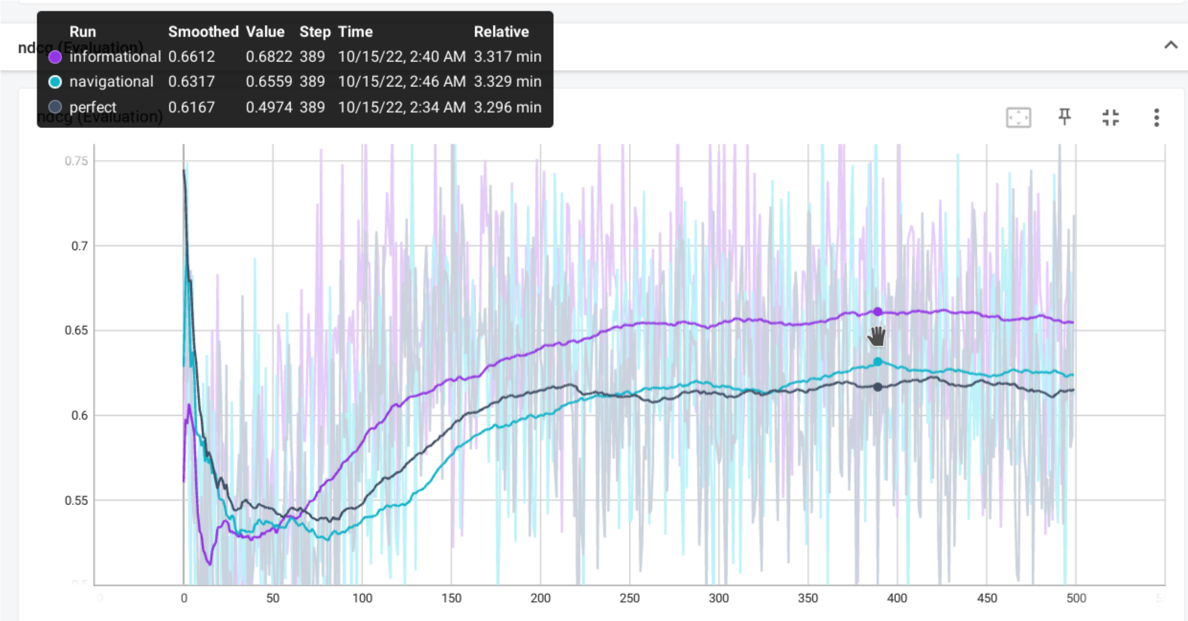
Run	Smoothed	Value	Step	Time	Relative
● informational	4.87e+6	1.33e+7	2,846	10/15/22, 2:39 AM	2.422 min
● navigational	5.64e+5	1.36e+6	2,846	10/15/22, 2:45 AM	2.437 min
● perfect	1.84e+6	4.07e+6	2,846	10/15/22, 2:33 AM	2.413 min

- *Q value per epoch*



Run	Smoothed	Value	Step	Time	Relative
● informational/Q Value per Epoch_difference	-12.68	18.08	1,962	10/15/22, 2:39 AM	1.659 min
● informational/Q Value per Epoch_policy Q value	1992	3575	1,962	10/15/22, 2:39 AM	1.659 min
○ informational/Q Value per Epoch_target Q value	2004	3556	1,962	10/15/22, 2:39 AM	1.659 min
● navigational/Q Value per Epoch_difference	-1.894	27.96	1,962	10/15/22, 2:45 AM	1.69 min
● navigational/Q Value per Epoch_policy Q value	697.2	1325	1,962	10/15/22, 2:45 AM	1.69 min
● navigational/Q Value per Epoch_target Q value	699.1	1297	1,962	10/15/22, 2:45 AM	1.69 min
● perfect/Q Value per Epoch_difference	-14.82	67.34	1,962	10/15/22, 2:33 AM	1.658 min
● perfect/Q Value per Epoch_policy Q value	1258	2364	1,962	10/15/22, 2:33 AM	1.658 min
● perfect/Q Value per Epoch_target Q value	1273	2297	1,962	10/15/22, 2:33 AM	1.658 min

- *ndcg@10 (evlauted after every ten epochs)*



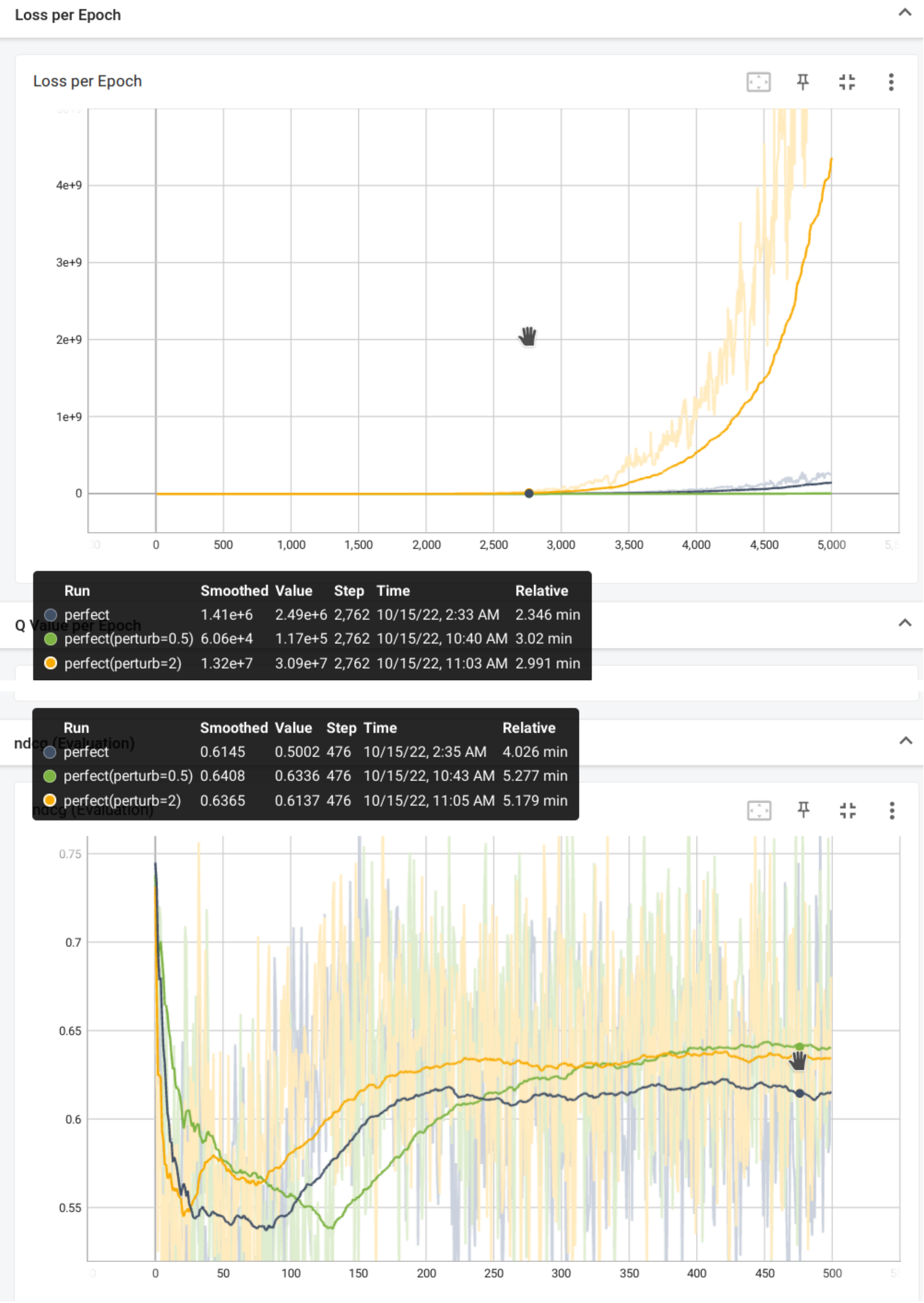
Run	Smoothed	Value	Step	Time	Relative
● informational	0.6612	0.6822	389	10/15/22, 2:40 AM	3.317 min
● navigational	0.6317	0.6559	389	10/15/22, 2:46 AM	3.329 min
● perfect	0.6167	0.4974	389	10/15/22, 2:34 AM	3.296 min

As can be seen in fig1, loss in all three models increases rapidly as the number of epochs increases, where the loss under *informational* model is the most noticeable, much higher than the other two models. Correspondingly, *informational* model has the largest Q value. However, what surprises me most is that the algorithms get the best performance under *informational* model, which is unusual as *perfect* model can unbiasedly represent the click (click probability is propotional to relevance label). Further analysis can be seen in **Analysis** section (section three).

2. Perturbation Range

Similar to exps in **Click Models** part, we fix the click model and change perturbation range by *multiplying a factor to scale the range of function* `tanh()`. We only show the loss and evaluation under each model, and here are the results:

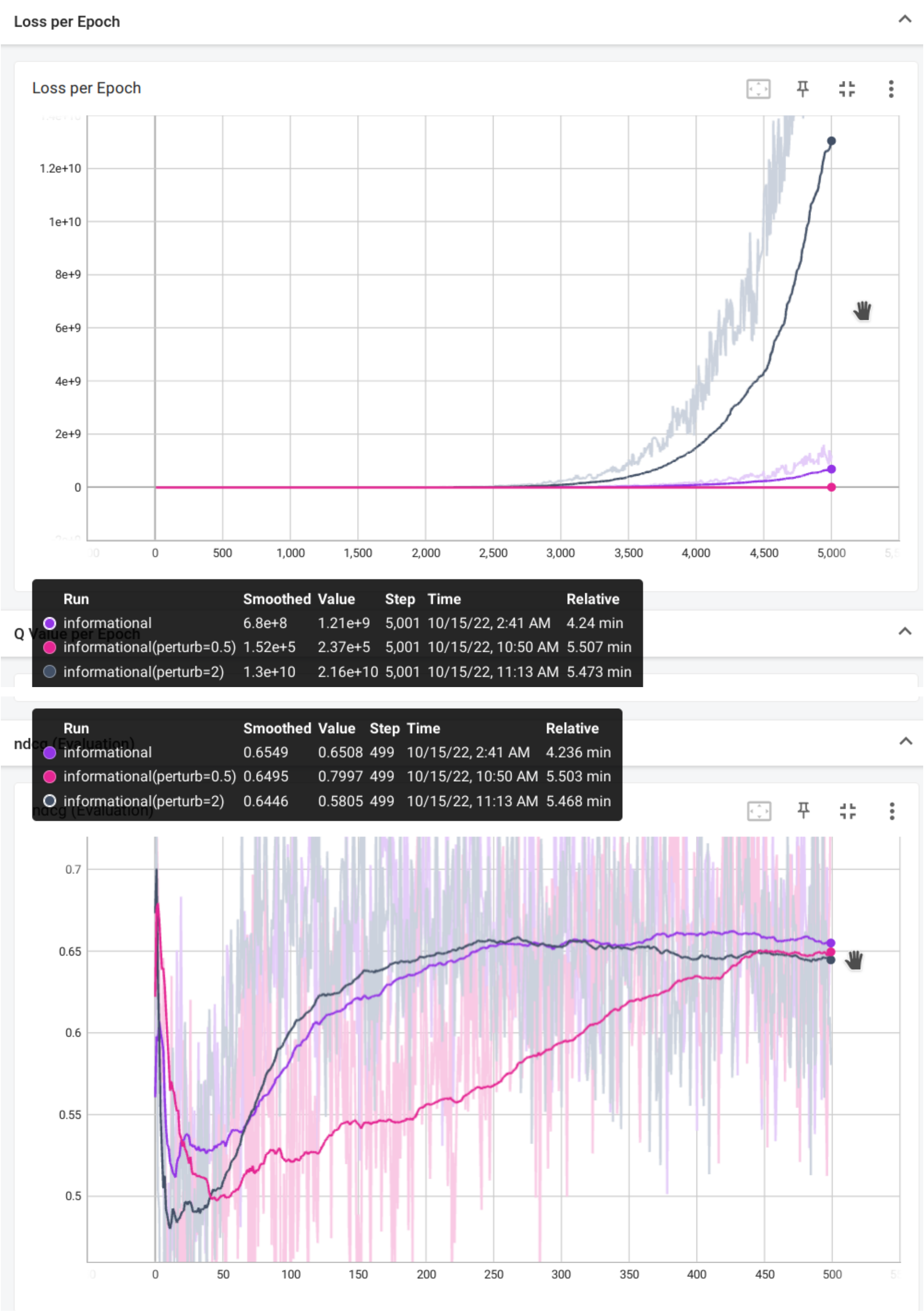
- perfect model



Run	Smoothed	Value	Step	Time	Relative
● perfect	1.41e+6	2.49e+6	2,762	10/15/22, 2:33 AM	2.346 min
● perfect(perturb=0.5)	6.06e+4	1.17e+5	2,762	10/15/22, 10:40 AM	3.02 min
● perfect(perturb=2)	1.32e+7	3.09e+7	2,762	10/15/22, 11:03 AM	2.991 min

Run	Smoothed	Value	Step	Time	Relative
● perfect	0.6145	0.5002	476	10/15/22, 2:35 AM	4.026 min
● perfect(perturb=0.5)	0.6408	0.6336	476	10/15/22, 10:43 AM	5.277 min
● perfect(perturb=2)	0.6365	0.6137	476	10/15/22, 11:05 AM	5.179 min

- informational model



Loss per Epoch

Loss per Epoch

Run	Smoothed	Value	Step	Time	Relative
○ informational	6.8e+8	1.21e+9	5,001	10/15/22, 2:41 AM	4.24 min
● informational(perturb=0.5)	1.52e+5	2.37e+5	5,001	10/15/22, 10:50 AM	5.507 min
● informational(perturb=2)	1.3e+10	2.16e+10	5,001	10/15/22, 11:13 AM	5.473 min

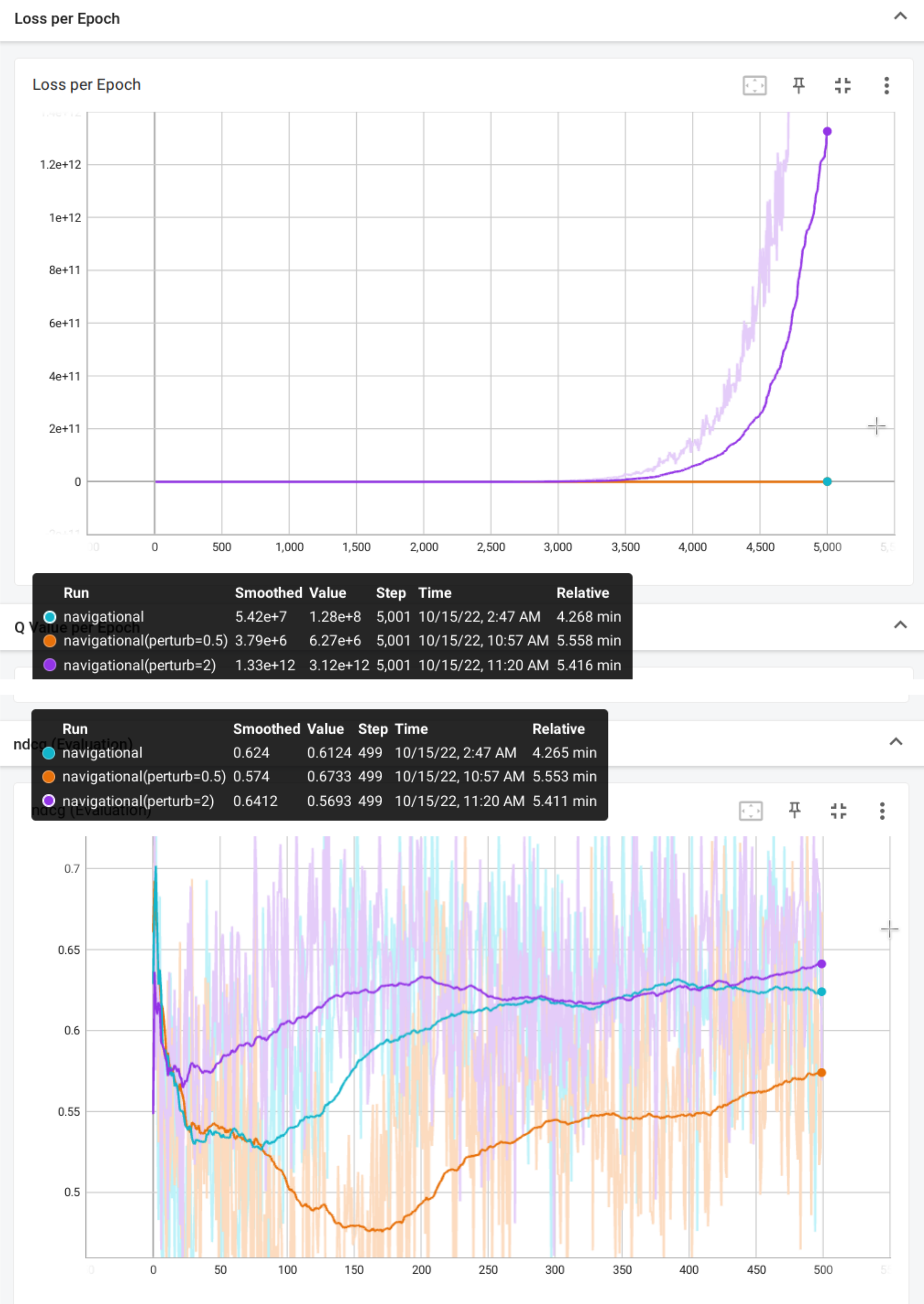
Q Value per Epoch

Run	Smoothed	Value	Step	Time	Relative
● informational	0.6549	0.6508	499	10/15/22, 2:41 AM	4.236 min
● informational(perturb=0.5)	0.6495	0.7997	499	10/15/22, 10:50 AM	5.503 min
○ informational(perturb=2)	0.6446	0.5805	499	10/15/22, 11:13 AM	5.468 min

ndcg (Evaluation)

ndcg (Evaluation)

- navigational model



Run	Smoothed	Value	Step	Time	Relative
● navigational	5.42e+7	1.28e+8	5,001	10/15/22, 2:47 AM	4.268 min
● navigational(perturb=0.5)	3.79e+6	6.27e+6	5,001	10/15/22, 10:57 AM	5.558 min
● navigational(perturb=2)	1.33e+12	3.12e+12	5,001	10/15/22, 11:20 AM	5.416 min

Run	Smoothed	Value	Step	Time	Relative
● navigational	0.624	0.6124	499	10/15/22, 2:47 AM	4.265 min
● navigational(perturb=0.5)	0.574	0.6733	499	10/15/22, 10:57 AM	5.553 min
● navigational(perturb=2)	0.6412	0.5693	499	10/15/22, 11:20 AM	5.411 min

Loss under all models doesn't converge when `perturb=2`, and will be reasonably bounded to prevent overestimate if perturbation factor is small (e.g. `perturb=0.5`). Though the loss can be controlled, small perturbation factor may lead to other problems, like *slower convergence rate*, *suboptimal solution*.

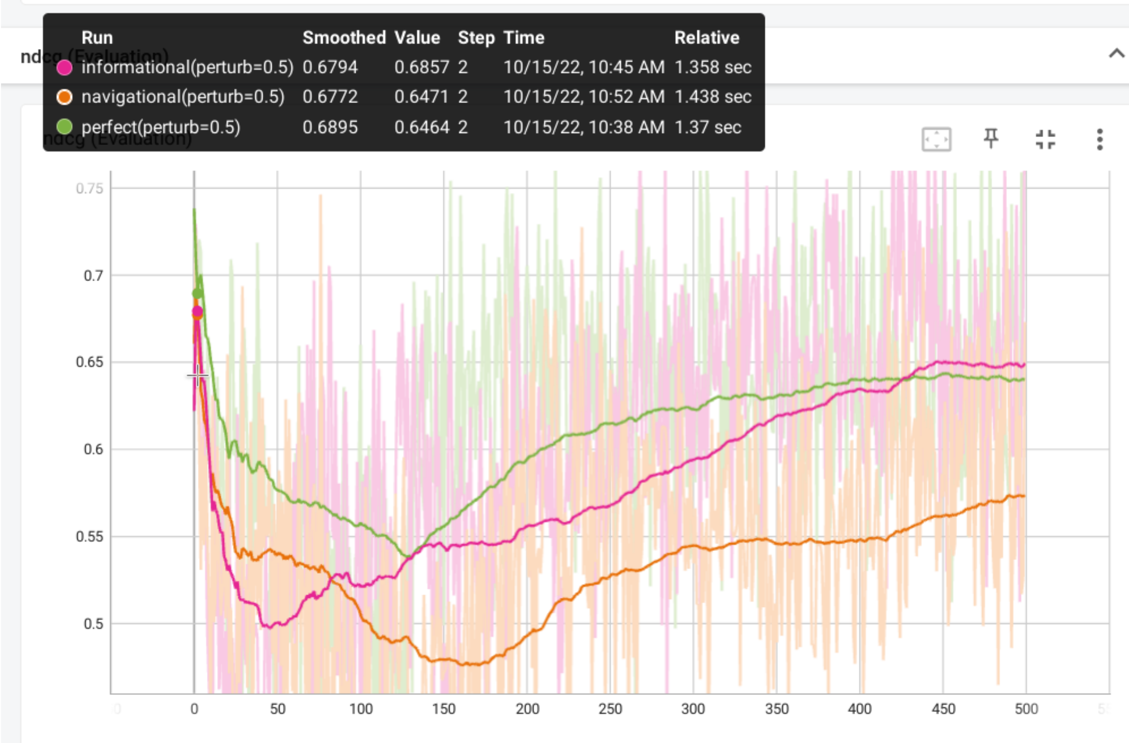
3. Analysis

Based on the two exps in the past two sections, I think that perturbation works as a ***"trade-off"*** parameter.

- Larger perturbation means that more actions can be reached, leading to greater probability of finding optimal solution(s). However, this also means that some unrelated actions will be taken into consideration, which will cause severe *overestimation* (mismatch in the distribution of data induced by the policy and the distribution of data contained in the batch)
- Smaller perturbation add a constrain on the similarity between selected actions and actions in the batch, which will remarkably reduce the overestimation problem. However, Some other problems occurs, such as *slower convergence rate*, *suboptimal solution* , especially under click models with bias (e.g. informational and navigational).

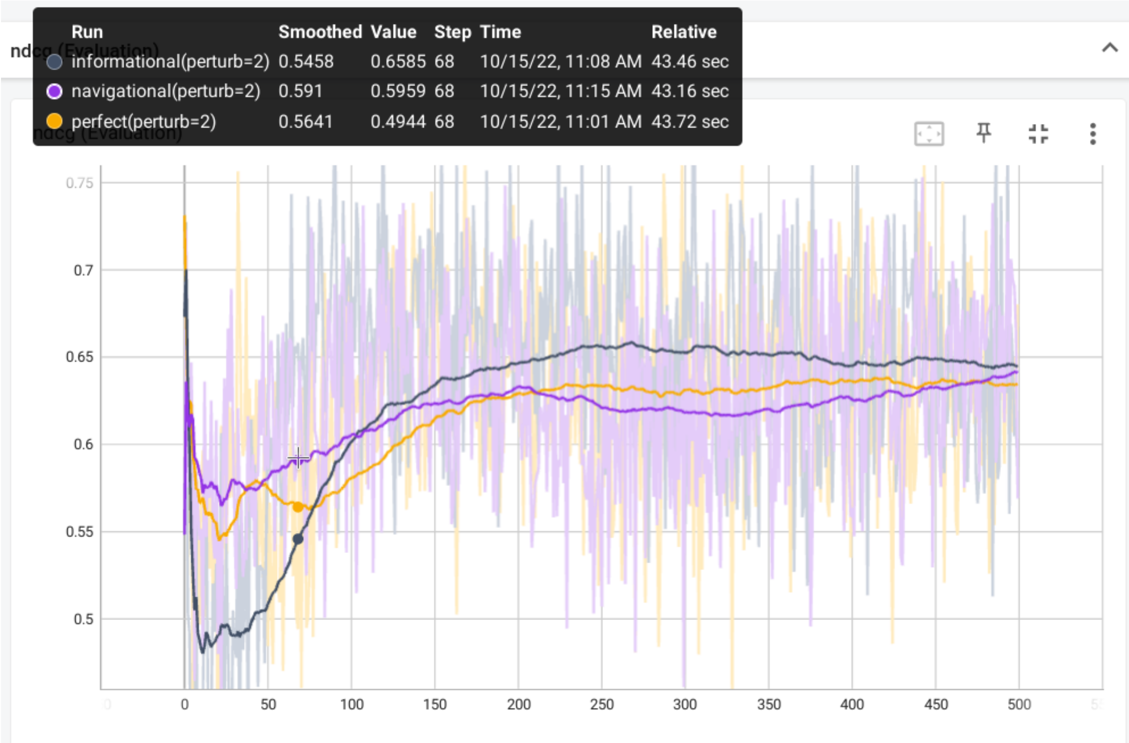
Appendix

- ndcg@10, perturb=0.5



Run	Smoothed	Value	Step	Time	Relative
● informational(perturb=0.5)	0.6794	0.6857	2	10/15/22, 10:45 AM	1.358 sec
● navigational(perturb=0.5)	0.6772	0.6471	2	10/15/22, 10:52 AM	1.438 sec
● perfect(perturb=0.5)	0.6895	0.6464	2	10/15/22, 10:38 AM	1.37 sec

- ndcg@10, perturb=2



ndcg (Evaluation)

ndcg (Evaluation)

Run	Smoothed	Value	Step	Time	Relative
● informational(perturb=2)	0.5458	0.6585	68	10/15/22, 11:08 AM	43.46 sec
● navigational(perturb=2)	0.591	0.5959	68	10/15/22, 11:15 AM	43.16 sec
● perfect(perturb=2)	0.5641	0.4944	68	10/15/22, 11:01 AM	43.72 sec