Supplementary Materials: Modality-Balanced Learning for Multimedia Recommendation

Anonymous Authors
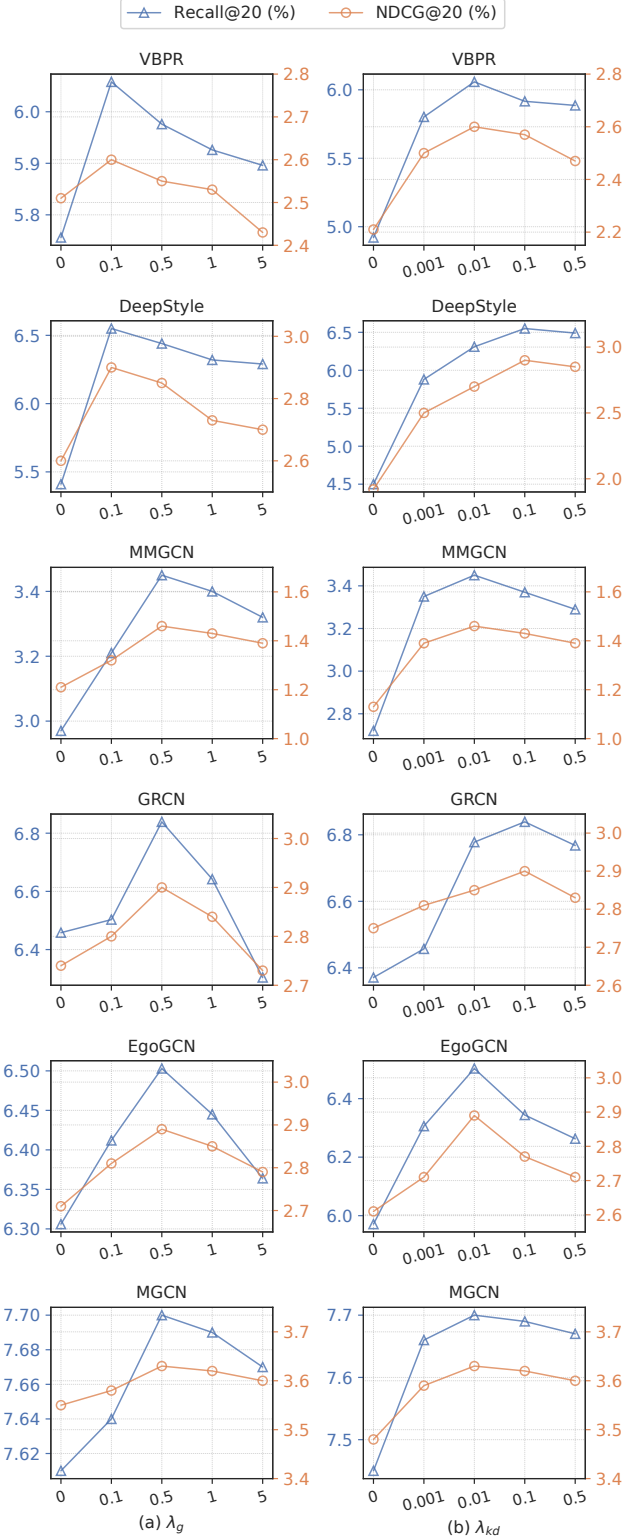

Figure 1: Performances comparison over different hyper-parameters settings on Amazon-Clothing dataset.

1 SENSITIVITY ANALYSIS

In this subsection, we examine the influence of different hyper-parameters on the performance of CKD. Specifically, we investigate the performance with respect to different λ_{kd} and λ_g values. Figure 1 reports the results of performance comparison. Our method gains significant improvement between $\lambda_{kd} = 0$ and $\lambda_{kd} = 0.001$, which validates the rationality of our proposed KD framework. When λ_{kd} is small, not enough knowledge is distilled to the student model which results in deficient performance. When λ_{kd} is too large, it hampers the optimization of the BPR loss and yield degraded performance. Similarly, a large λ_g will cover up the specific distillation loss on training triples and cause degradation.

2 COMPREHENSIVE ABLATION RESULTS

Table 1: Ablation study results on generic distillation loss and re-weight paradigm in terms of Recall@20. The best performance is highlighted in bold.

Model	Baby	Sports	Clothing	Beauty
DeepStyle + CKD	**0.0613**	**0.0684**	**0.0655**	**0.1214**
w/o. Gen.	0.0486	0.0547	0.0541	0.1009
w/o. re-weight	0.0492	0.0527	0.0412	0.0885
repl. KL-divergence Loss	0.0485	0.0522	0.0470	0.0914
repl. MSE Loss	0.0559	0.0613	0.0588	0.1121
GRCN + CKD	**0.0866**	**0.0922**	**0.0667**	0.1355
w/o. Gen.	0.0844	0.0899	0.0654	0.1333
w/o. re-weight	0.0851	0.0881	0.0649	0.1340
repl. KL-divergence Loss	0.0841	0.0915	0.0651	**0.1362**
repl. MSE Loss	0.0834	0.0857	0.0633	0.1312
MGCN + CKD	**0.0710**	**0.0937**	**0.0770**	**0.1349**
w/o. Gen	0.0689	0.0913	0.0761	0.1322
w/o. re-weight	0.0677	0.0922	0.0764	0.1331
repl. KL-divergence Loss	0.0688	0.0913	0.0759	0.1328
repl. MSE Loss	0.0691	0.0915	0.0755	0.1331

We present the ablation results of DeepStyle, GRCN, and MGCN in Table 1. We can observe that the results of these models' ablation experiments align with the results in Section 4.4. Each of these key components contributes substantially.