MMLF: Multi-Metric Latent Feature Analysis for High-Dimensional and Incomplete Data Supplementary File

Di Wu, *Member, IEEE,* Xin Luo, *Senior Member, IEEE*, Peng Zhang, Yi He, *Member, IEEE*, and Xindong Wu, *Fellow*, *IEEE*

This is the supplementary file for the paper entitled "MMLF: Multi-Metric Latent Feature Analysis for High-Dimensional and Incomplete Data" in IEEE Transactions on Knowledge and Data Engineering. We have put some contents cited by the paper here.

S1. Cited Figures

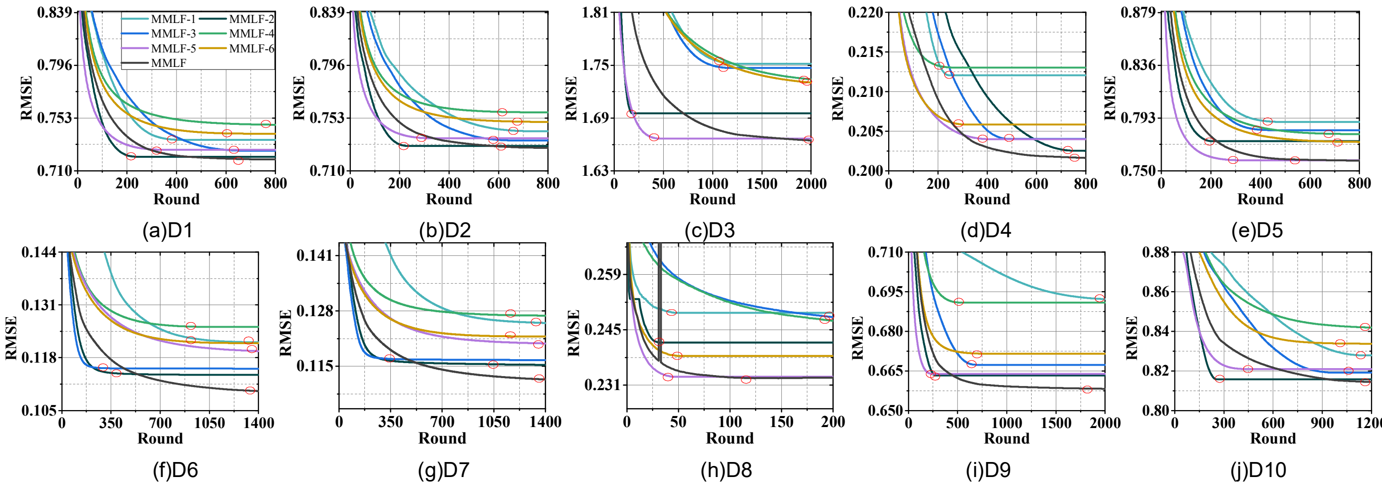


Fig. S1. The convergent siuations of RMSE of MMLF and its each base model on all the datasets, where the red circle denotes the convergent point.

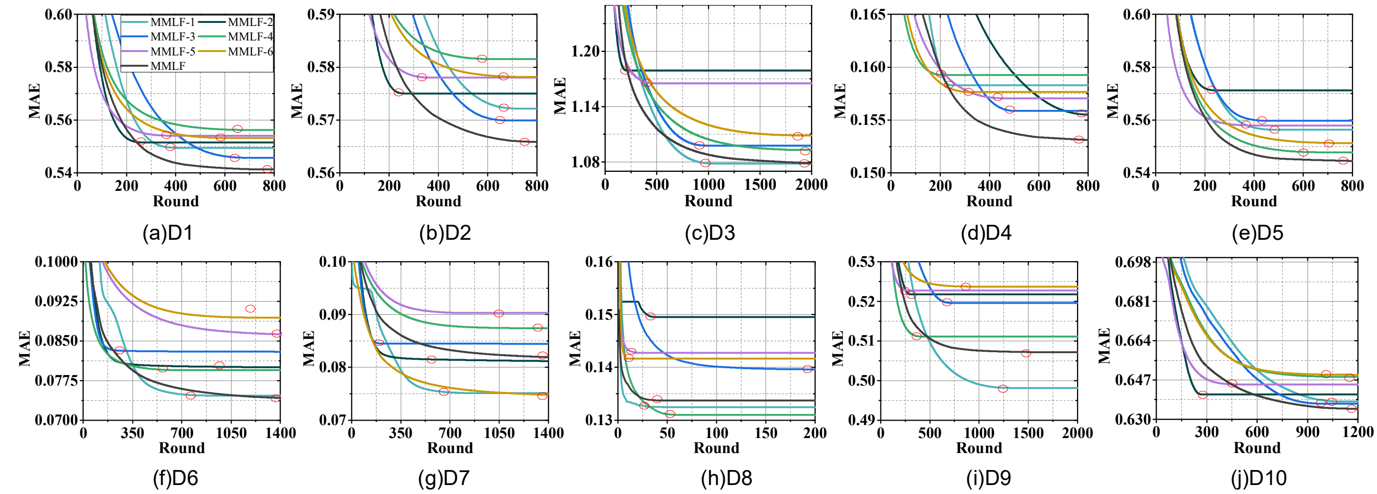


Fig. S2. The convergent siuations of MAE of MMLF and its each base model on all the datasets, where the red circle denotes the convergent point.

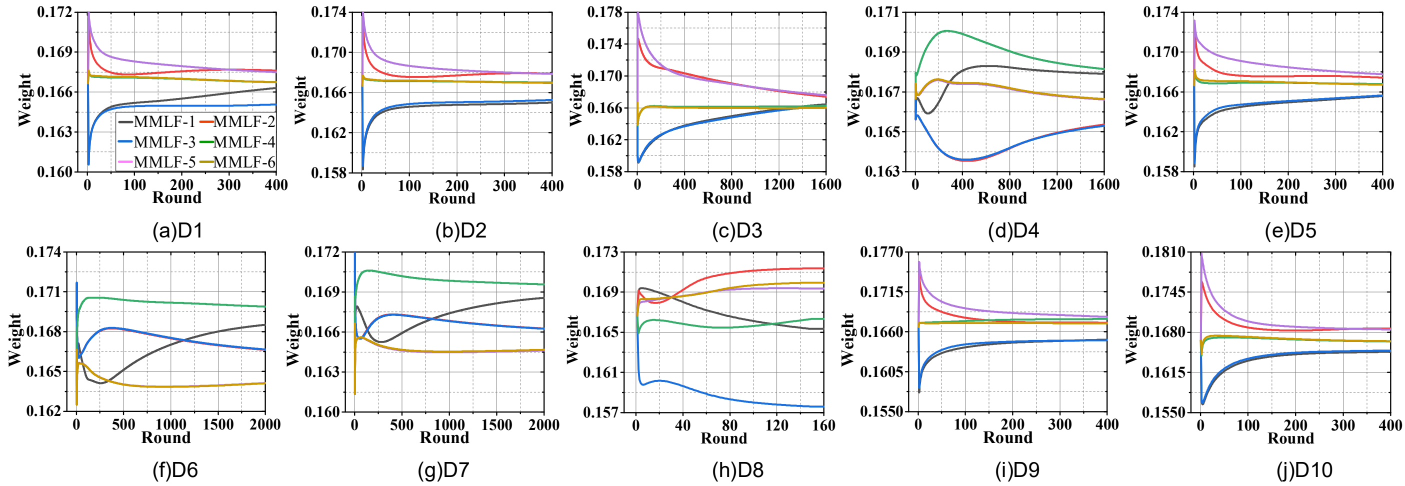


Fig. S3. The changes of ensemble weights during the training process on all the dataset.

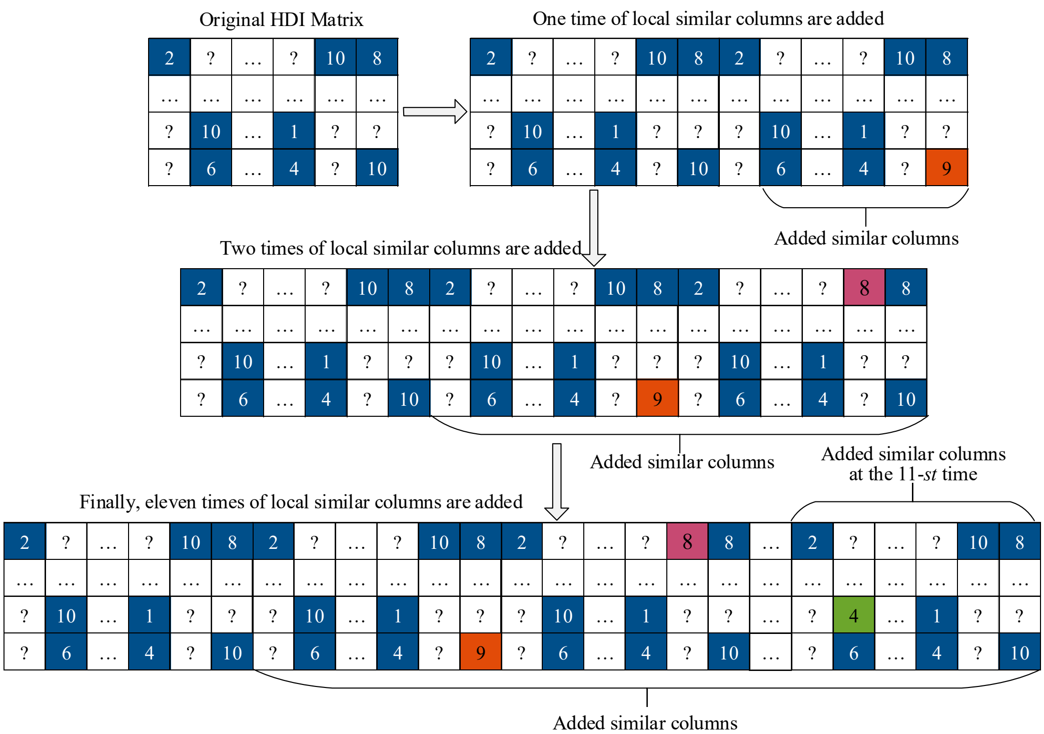


Fig. S4. An example of generating the times of local similar columns from 1 to 11.

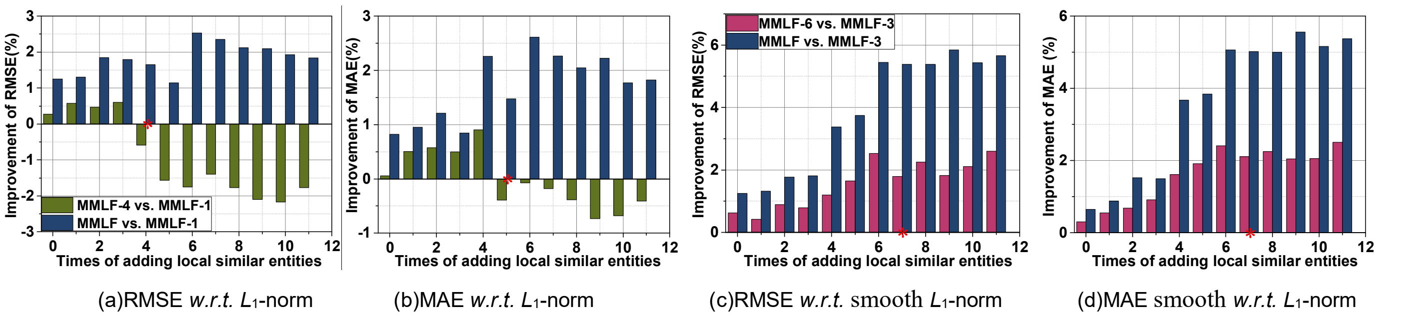


Fig. S5. The performance comparison between inner product and distance spaces *w.r.t.* L_1-norm and smooth L_1-norm, where the red $*$ denotes the inflection point of performance improvement.

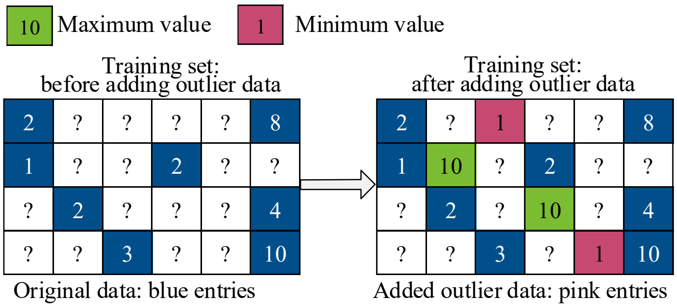
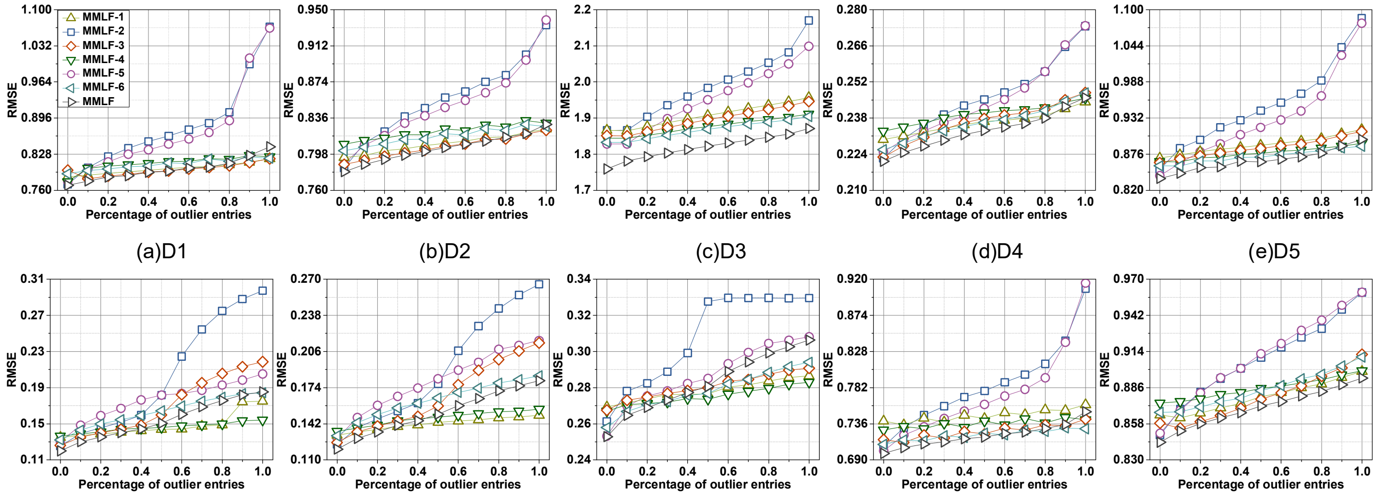


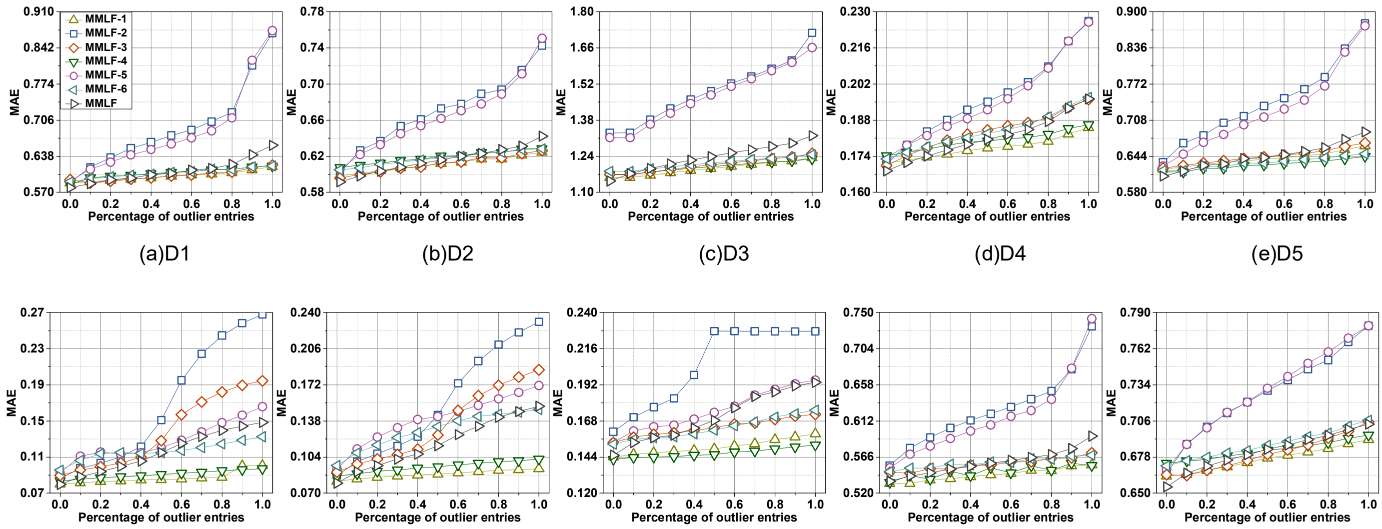
Fig. S6. An example of generating synthetic outlier data on an HDI matrix.



(a)D1 (b)D2 (c)D3 (d)D4 (e)D5

(f)D6 (g)D7 (h)D8 (i)D9 (j)D10

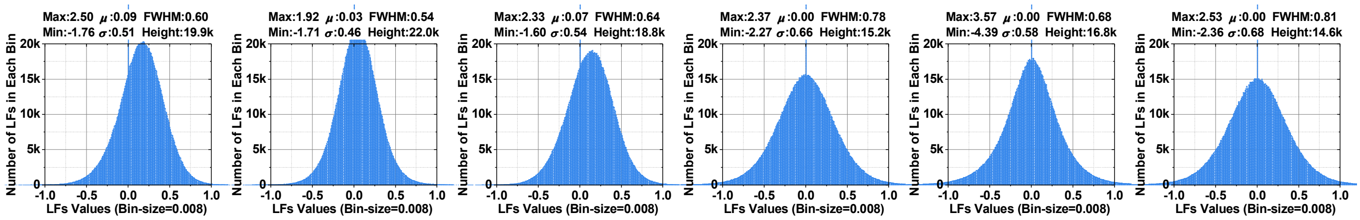
Fig. S7. The RMSE influence of outlier data on MMLF and its each base model on all the datasets.



(a)D1 (b)D2 (c)D3 (d)D4 (e)D5

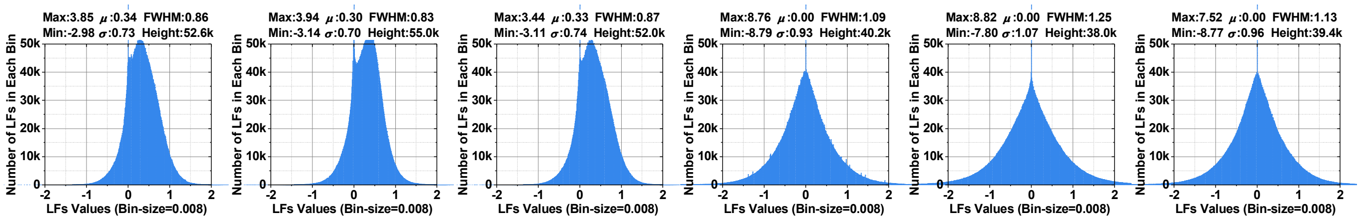
(f)D6 (g)D7 (h)D8 (i)D9 (j)D10

Fig. S8. The MAE influence of outlier data to MMLF and its each base model on all the datasets.



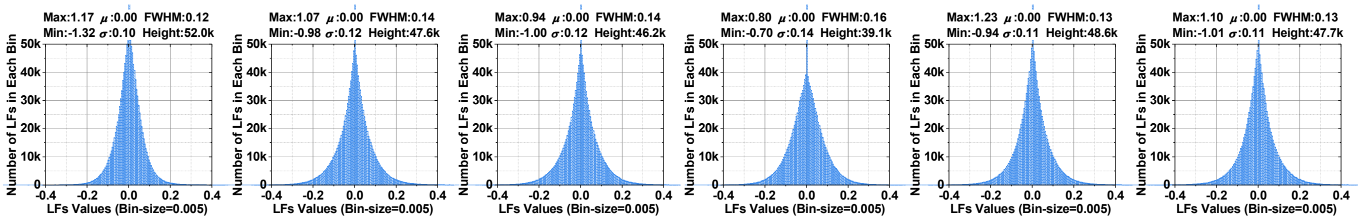
(a) MMLF-1 (b) MMLF-2 (c) MMLF-3 (d) MMLF-4 (e) MMLF-5 (f) MMLF-6

Fig. S9. The distribution histogram of latent features (LFs) of base models on D2.



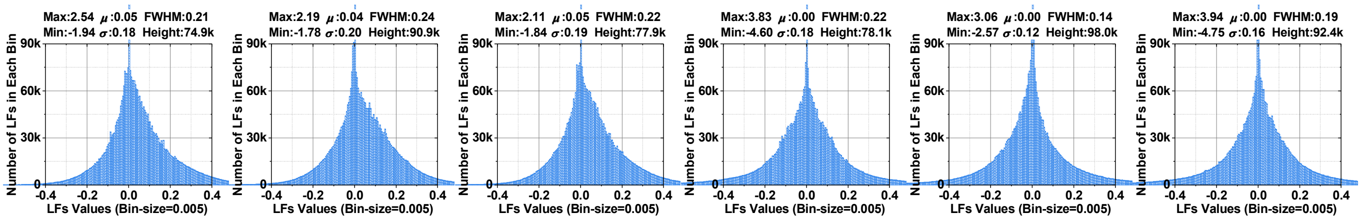
(a) MMLF-1 (b) MMLF-2 (c) MMLF-3 (d) MMLF-4 (e) MMLF-5 (f) MMLF-6

Fig. S10. The distribution histogram of latent features (LFs) of base models on D3.



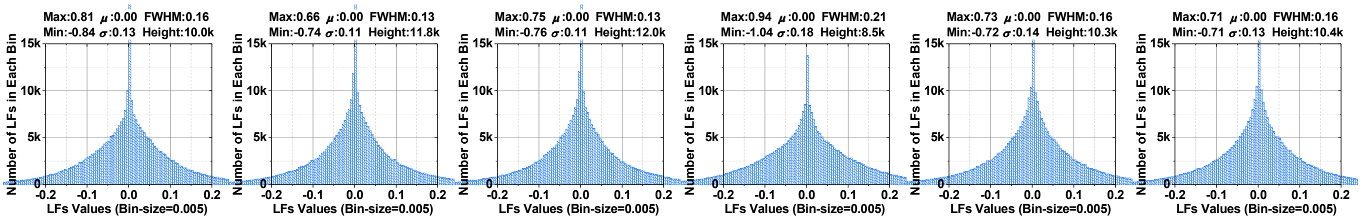
(a) MMLF-1 (b) MMLF-2 (c) MMLF-3 (d) MMLF-4 (e) MMLF-5 (f) MMLF-6

Fig. S11. The distribution histogram of latent features (LFs) of base models on D4.



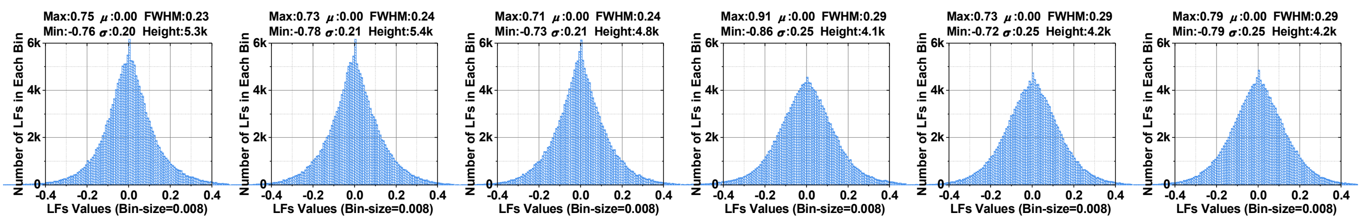
(a) MMLF-1 (b) MMLF-2 (c) MMLF-3 (d) MMLF-4 (e) MMLF-5 (f) MMLF-6

Fig. S12. The distribution histogram of latent features (LFs) of base models on D5.



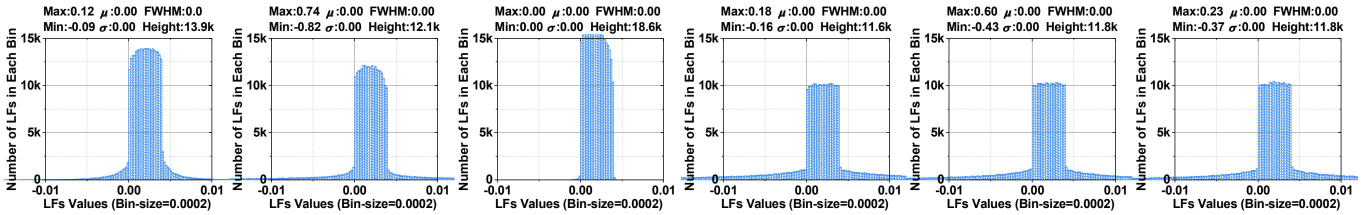
(a) MMLF-1 (b) MMLF-2 (c) MMLF-3 (d) MMLF-4 (e) MMLF-5 (f) MMLF-6

Fig. S13. The distribution histogram of latent features (LFs) of base models on D6.



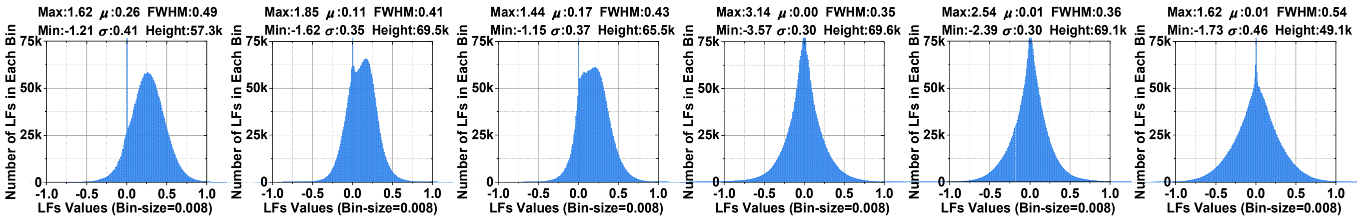
(a) MMLF-1 (b) MMLF-2 (c) MMLF-3 (d) MMLF-4 (e) MMLF-5 (f) MMLF-6

Fig. S14. The distribution histogram of latent features (LFs) of base models on D7.



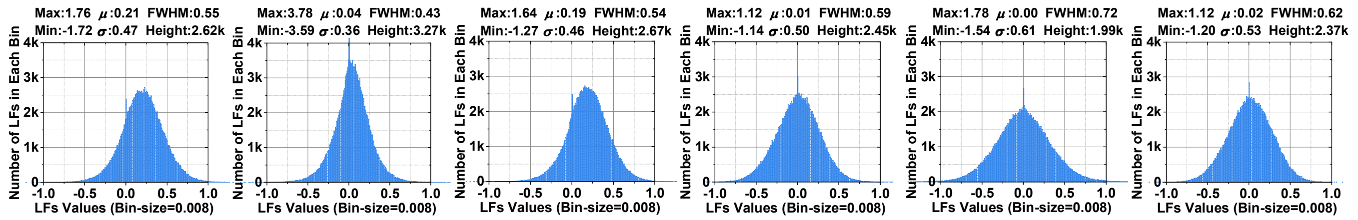
(a) MMLF-1 (b) MMLF-2 (c) MMLF-3 (d) MMLF-4 (e) MMLF-5 (f) MMLF-6

Fig. S15. The distribution histogram of latent features (LFs) of base models on D8.



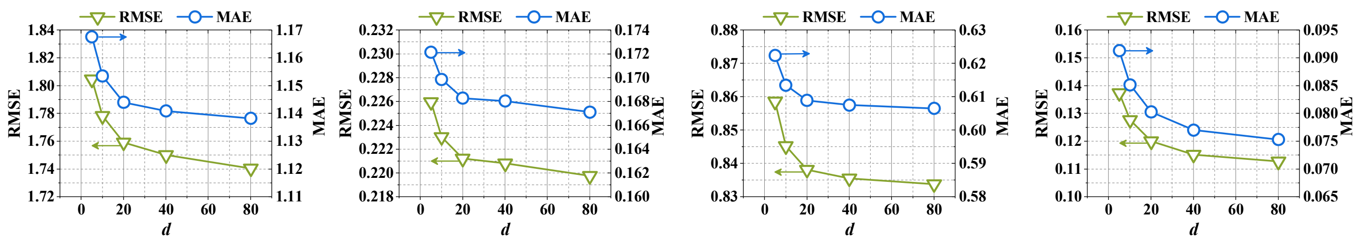
(a) MMLF-1 (b) MMLF-2 (c) MMLF-3 (d) MMLF-4 (e) MMLF-5 (f) MMLF-6

Fig. S16. The distribution histogram of latent features (LFs) of base models on D9.

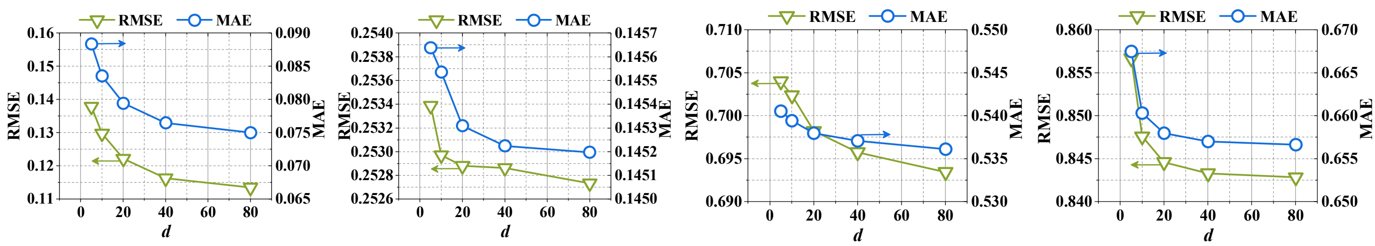


(a) MMLF-1 (b) MMLF-2 (c) MMLF-3 (d) MMLF-4 (e) MMLF-5 (f) MMLF-6

Fig. S17. The distribution histogram of latent features (LFs) of base models on D10.



(a)D3 (b)D4 (c)D5 (d)D6



(e)D7 (f)D8 (g)D9 (h)D10

Fig. S18. The RMSE/MAE of MMLF as *d* increases from 5 to 80 on D3–D10.

S2. Discussion

From Section 3, we see that MMLF is built on the known part of an HDI matrix only to extract the desired latent features for representing involved entities. Hence, the extracted latent features contain the intrinsic patterns regarding involved entities' characteristics described by the incompletely known data of the target HDI matrix. With enough known data of an HDI matrix, the overall loss function of MMLF can predict progressive unbiased variants, i.e., MMLF is an unbiased estimator. For example, we sample different ratios of known data from a real HDI matrix to train. As shown in Fig. S20, the prediction RMSE/MAE decreases as the sampling ratio increases. When the sample ratio increases to 1, the prediction RMSE/MAE tends to zero.

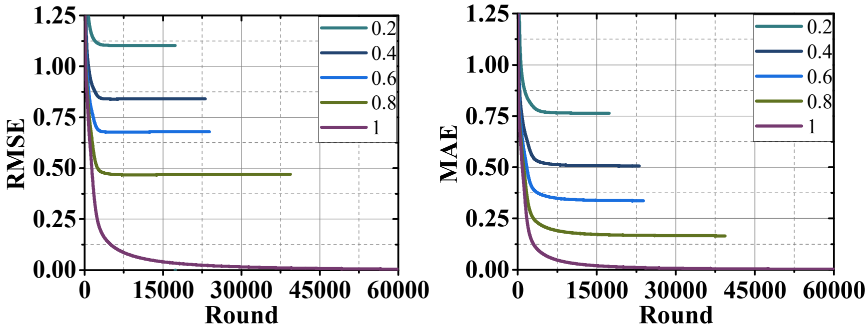


Fig. S19. The convergent siuations of MMLF when sample different ratios of known data from a real HDI matrix to train.