REPRODUCTION AND EXTENSION OF DIFFUSION RECOMMENDER MODEL

Ilayda Bilgin, Francesco Tinner, Luke Chin a Foeng

INTRODUCTION

- Learn generative process of user interactions using **Diffusion Models** (DM) instead of:
  - **GANs**: instability in training, reduced performance
  - **VAEs**: restricted representation ability, trade off between tractability and representation ability
- **Diffusion Processes**: 1) tractable forward diffusion process corrupts user interactions, 2) NN learns the reconstruction iteratively
- Objectives of Recommender Systems (RS) align well with DM
  - RS infer future interaction probabilities from noisy historical interactions.
  - The noise in interactions due to false-positive and false-negative interactions.

CONTRIBUTIONS

Diffusion Recommender Model (Original Paper)

- **DiffRec:** Employs diffusion process on interactions to introduce noise in training, which helps to reduce effect of noisy interactions.
- **Latent DiffRec:** Encodes interactions into latent space, then diffuses.
- **Temporal DiffRec:** Introduces a temporal weighting for interactions.
- **Latent Temporal DiffRec:** Includes a temporal weighting in the latent space.

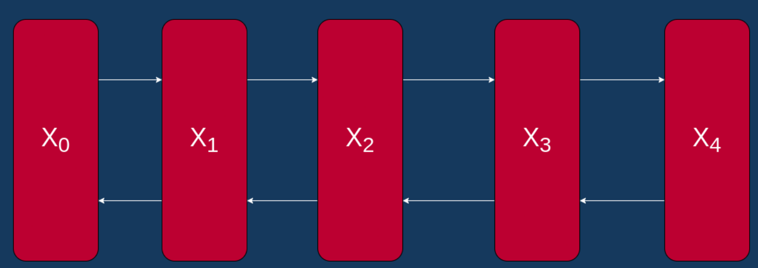
Our Extensions

- **Training Patience:** Increased training patience is introduced to improve model performance.
- **Learned Temporal Weighting (LTW):** Replaced the static temporal weighting with learnable temporal weights to better encode temporal interaction importance.
- **Clustering Ablation Study:** Exploration of better suited clustering methods cluster sizes, and their hyperparameters.
- **Overall Result Reproduction:** Result reproduction in order to validate the original paper.

METHODOLOGY

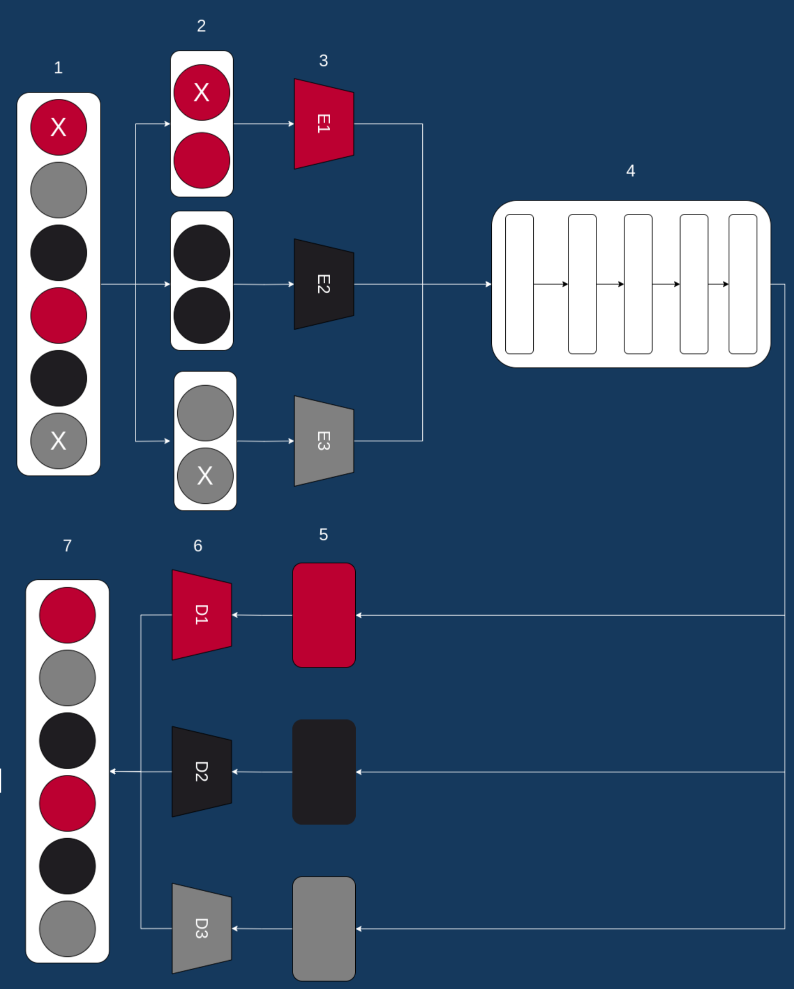
Vanilla DiffRec

Each step introduces additional noise, until the final interaction vector is pure noise, the backward pass learns the distribution of this noise through the maximization of the ELBO function.



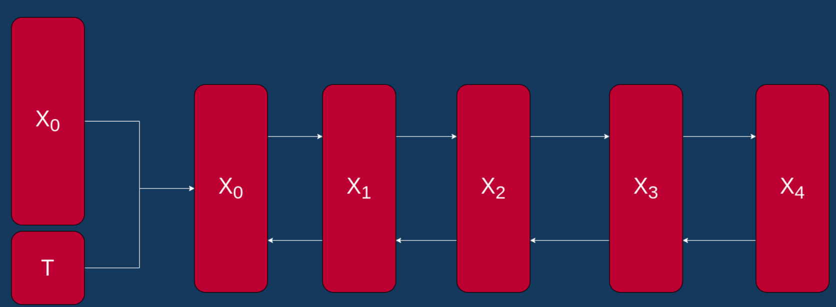
Latent DiffRec

1. Interactions
2. Encoding compression through clustering
3. Clusters are encoded via separate encoders
4. Latent space diffusion
5. Regrouping of clusters
6. Decoding of clusters
7. Final set of True-positive and True-negative interactions



Temporal DiffRec

Similar to Vanilla DiffRec, except a linear temporal encoding is introduced in order to assign each interaction a weight based on the relative time the interaction occurred.



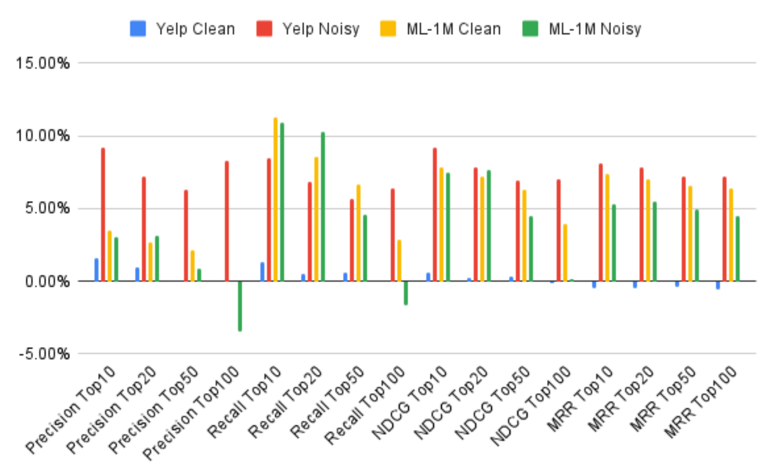
Latent Temporal DiffRec

Latent temporal DiffRec combines all the concepts of latent and temporal DiffRec into a single model.

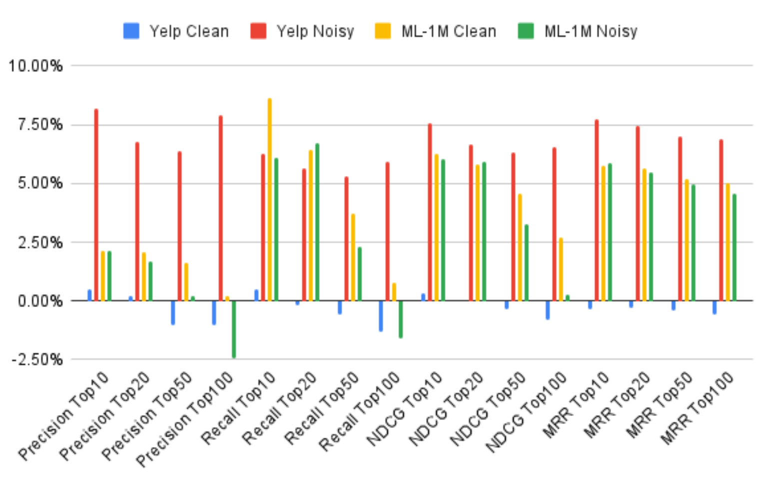
1. The input to this model is a temporally encoded set of interactions
2. Clustering of the encoded interactions
3. Latent Space Diffusion
4. Regrouping and Decoding of clusters

RESULTS

LTW + Patience improved 10%+ over baseline on some metrics, with most falling between a 5%-10% improvement. Yelp clean shows almost no improvements. **LTW** improved the original values by ~5% across all metrics, with Yelp clean showing almost no improvements. Ultimately, the improvements proved successful on most datasets, though additional validation is still required on other datasets.



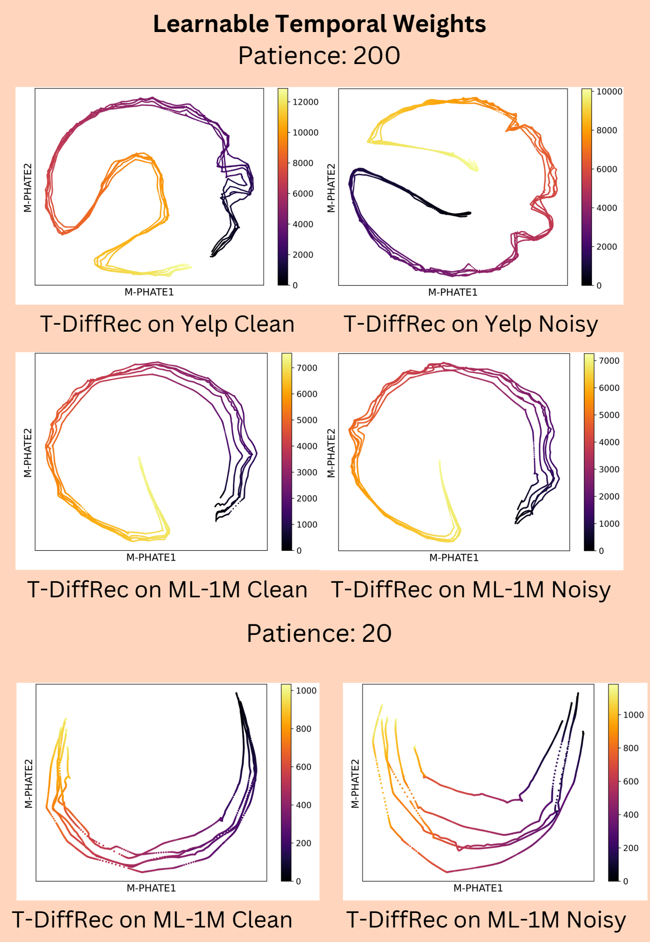
LTW+Patience Improvement over Baseline



LTW Improvement over Baseline

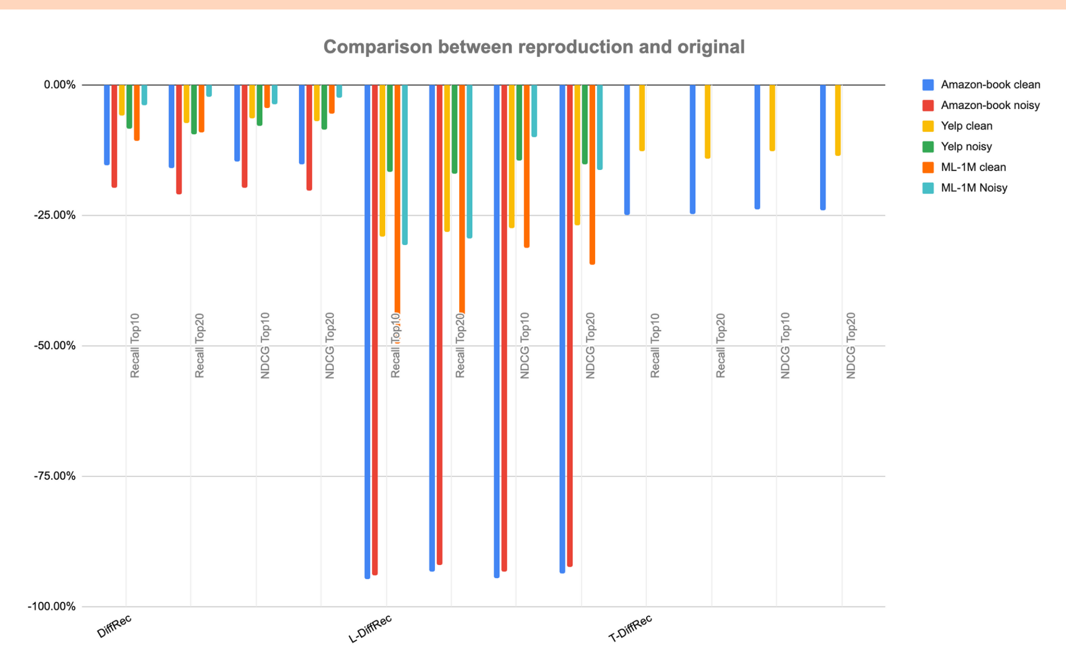
LTW VISUALIZATION

mPHATE, allows to visualize the development of the learnable parameters over the course of the training using a latent representation of the learnable parameters. The color indicates training time. Weight evolvement patterns between similar datasets (noisy/clean) are comparable, while we observe a clear distinction between Yelp and ML-1M.



DISCUSSION

- Intransparent setting of hyperparameters: not all models converge using the provided hyperparameters
- LTW improvement over baseline needs to be validated on other datasets.
- LTW improvement only visible on datasets larger than ML-1M.



CONCLUSION

- Future extensions: other datasets, investigate preprocessing
- **LTW**: leads to a performance improvement of up to 10% over the baseline
- Default parameters for L-Diffrec: underfitting
- Enhancement of patience value: ↑ Patience value ↑ Performance

REFERENCES:

Wang, W., Xu, Y., Feng, F., Lin, X., He, X., Chua, T.S.: Diffusion recommender model. arXiv preprint arXiv:2304.04971 (2023)
Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning. pp. 2256–2265. PMLR (2015)
Gao, M., Zhang, J., Yu, J., Li, J., Wen, J., Xiong, Q.: Recommender systems based on generative adversarial networks: A problem-driven perspective. Information Sciences 546, 1166–1185 (2021)