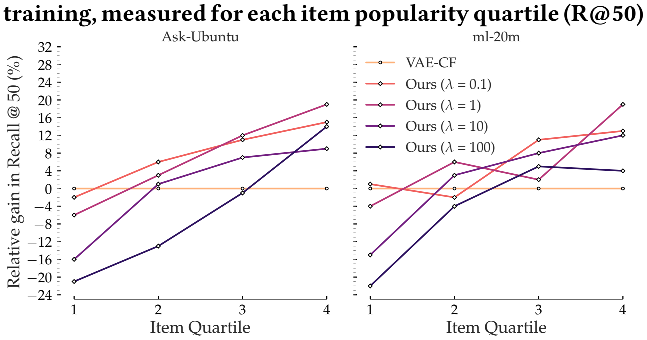
Table 1: Composition of top-100 item recommendations to users in item popularity quartiles (Q1-Most Popular Items)

Method	ml-20m				Ask-Ubuntu			
	Q-1	Q-2	Q-3	Q-4	Q-1	Q-2	Q-3	Q-4
CDAE (G_1)	74%	26%	0%	0%	97%	3%	0%	0%
D+G_1(λ = 0.1)	61%	23%	10%	6%	76%	14%	7%	3%
D+G_1(λ = 1)	62%	21%	11%	6%	73%	16%	6%	5%
D+G_1(λ = 10)	61%	19%	12%	8%	65%	19%	11%	5%
VAE-CF (G_2)	64%	24%	8%	4%	60%	25%	9%	6%
D+G_2(λ = 0.1)	58%	23%	12%	7%	53%	25%	12%	10%
D+G_2(λ = 1)	59%	21%	13%	7%	55%	21%	13%	11%
D+G_2(λ = 10)	59%	20%	13%	8%	54%	22%	14%	10%

Note that CDAE does not make *any* niche item recommendations (Q3 and Q4). Integrating our adversary to train CDAE results in a significant jump in long-tail coverage. To further dissect the above results, we will now observe our relative gains in *Recall*@50 compared to VAE-CF for each item quartile (Figure 3). We chose VAE-CF for comparison due to it's stronger long-tail performance.

Figure 3: Relative improvement over VAE-CF with adversary training, measured for each item popularity quartile (R@50)



As expected, our strongest gains are observed in Quartiles-3 and 4, which constitute long-tail items. Although there is a slight loss in popular item performance for $\lambda = 1$, this loss is not significant owing to the ease of recommending popular items with auxiliary models if required. We observe the values of tradeoff λ between 0.1 and 1 to generate balanced results.

We now analyze overall recommendation performance against VAE-CF and CDAE in Table 2 (**N** = NDCG, **R** = Recall). Even though our models recommend very different compositions of items (table 1), the results exhibit modest overall improvements for $\lambda = 0.1$ and $\lambda = 1$ over both the base recommenders. Clearly, the additional niche recommendations are coherent since there is no performance

Table 2: Overall recommender performance on ml-20m and Ask-Ubuntu datasets

Method	ml-20m			Ask-Ubuntu		
	N@100	R@20	R@50	N@100	R@20	R@50
CDAE (G_1)	0.34	0.27	0.37	0.29	0.30	0.46
VAE-CF (G_2)	0.51	0.44	0.57	0.42	0.45	0.59
D+G_2(λ = 0.1)	0.53	0.45	0.59	0.43	0.46	0.61
D+G_2(λ = 1)	0.52	0.44	0.58	0.42	0.46	0.59
D+G_2(λ = 10)	0.48	0.41	0.55	0.40	0.43	0.56
D+G_2(λ=100)	0.42	0.37	0.51	0.38	0.41	0.53

drop. However, larger λ values hurt the recommender performance. It is thus essential to balance the adversary objective and base recommender to obtain strong overall results.

5 CONCLUSION AND FUTURE WORK

In this paper, we investigated an adversarial learning framework to overcome sparsity in long-tail item recommendation. Our approach modernises conventional neighbor models, learning flexible associations guided by the feedback data. Our approach improved the long-tail performance of VAE-CF, which by itself outperforms CDAE by a significant margin. In future work, we plan several interesting directions. Integration of inter-user or cross associations with the item structure learned by the base recommender could prove valuable. Extension of our idea to retrieval problems to recover niche but relevant documents can prove impactful. Although our empirical results indicate reasonable model convergence, we plan to explore the Wasserstein metric [1] to provide a meaningful and smooth measure of the distance between the two competing distributions, with improved theoretical and empirical stability.

REFERENCES

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein gan. *arXiv preprint arXiv:1701.07875* (2017).
[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
[3] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 173–182.
[4] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. Ieee, 263–272.
[5] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009).
[6] Adit Krishnan, Ashish Sharma, and Hari Sundaram. 2017. Improving Latent User Models in Online Social Media. *arXiv preprint arXiv:1711.11124* (2017).
[7] Aaron Q Li, Amr Ahmed, Sujith Ravi, and Alexander J Smola. 2014. Reducing the sampling complexity of topic models. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 891–900.
[8] Xiaopeng Li and James She. 2017. Collaborative variational autoencoder for recommender systems. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 305–314.
[9] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. *arXiv preprint arXiv:1802.05814* (2018).
[10] Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. 2011. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 287–296.
[11] Sanjay Purushotham, Yan Liu, and C-C Jay Kuo. 2012. Collaborative topic regression with social matrix factorization for recommendation systems. *arXiv preprint arXiv:1206.4684* (2012).
[12] Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. 2000. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*. 1057–1063.
[13] Chong Wang and David M Blei. 2011. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 448–456.
[14] Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong, Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang. 2017. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 515–524.
[15] Yao Wu, Christopher DuBois, Alice X Zheng, and Martin Ester. 2016. Collaborative denoising auto-encoders for top-n recommender systems. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, 153–162.
[16] Hongzhi Yin, Bin Cui, Jing Li, Junjie Yao, and Chen Chen. 2012. Challenging the long tail recommendation. *Proceedings of the VLDB Endowment* 5, 9 (2012), 896–907.