Adaptive Noise Injection in Variational Autoencoders for Enhancing Fairness in Group Recommendations

Emaz Uddin Ahmad*, Maria Stratigi and Kostas Stefanidis

Data Science Research Centre, Tampere University, Finland

Abstract

This paper proposes an enhanced Variational Autoencoder (VAE) based framework that introduces adaptive noise injection into the latent space to promote fairness and satisfaction in group recommendations. In contrast to conventional VAEs that depend on static Gaussian noise, the proposed model dynamically learns data-dependent, adaptive noise from user representations, enhancing its ability to predict uncertainty and reducing bias towards dominant user preferences. The framework is further enhanced through Bayesian optimization, which is employed to fine-tune the hyperparameters of the Variational Autoencoder. Comprehensive experiments on the MovieLens 10M dataset across homogeneous, heterogeneous, and mixed groups demonstrate that the adaptive noise VAE model consistently outperforms the static noise baseline.

Keywords

Recommender Systems, Group Recommendations, Variational Autoencoder (VAE), Fairness.

1. Introduction

Fairness has emerged as a crucial requirement in recommender systems, particularly due to widespread concerns about popularity bias, unequal exposure, and the tendency of algorithms to favour highly active users. Fairness refers to the principle that users and items should receive equitable treatment across the recommendation process, without systematic favouring or marginalization of specific groups. This issue becomes even more prominent in group recommendations, where multiple users' preferences must be aggregated to produce a single ranked list. In such settings, the preferences of active or majority users can overshadow those of minority or low-activity users, leading to imbalanced group satisfaction and reduced representativeness. Additionally, the diversity of preferences within groups and the sparsity of interaction histories make it challenging to ensure fairness at both the individual and group levels.

Recent research has explored the use of deep generative models, such as Variational Autoencoders (VAEs), to improve fairness and diversity in group recommendations. VAEs have proven effective for collaborative filtering because they capture non-linear preference patterns and model user uncertainty. However, most existing VAE-based approaches rely on static noise sampling, where the Gaussian noise injected into the latent space is fixed and independent of the underlying user–item characteristics. This design limits the model's ability to adapt uncertainty based on user activity levels or preference complexity. As a result, VAEs often overfit to high-frequency users while underrepresenting users with sparse data, reducing fairness and diversity in group-level recommendations.

To address these limitations, our work proposes a modified VAE in which the noise input to the latent space is learned dynamically from user and item representations. Instead of sampling from a static prior, the model dynamically adjusts the level of injected noise according to characteristics such as user activity patterns or preference variability. This adaptive noise mechanism functions as a fairness-control component that increases uncertainty for high-activity or dominant users and provides additional flexibility in modelling sparse or underrepresented users. As a result, the model

DOLAP 2026: 28th International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data, co-located with EDBT/ICDT 2026, March 24, 2026, Tampere, Finland

*Corresponding author.
emazuddinahmad@gmail.com (E. U. Ahmad); maria.stratigi@tuni.fi (M. Stratigi); konstantinos.stefanidis@tuni.fi (K. Stefanidis)
https://homepages.tuni.fi/konstantinos.stefanidis/ (K. Stefanidis)
0009-0006-5298-105X (E. U. Ahmad); 0000-0003-2482-4605 (M. Stratigi); 0000-0003-1317-8062 (K. Stefanidis)

encourages a more balanced exposure and can manage group recommendation scenarios that require the accommodation of diverse, potentially conflicting interests. We conduct extensive experiments on the MovieLens 10M dataset. The evaluation examines not only traditional ranking metrics, such as Recall and NDCG, but also fairness- and diversity-oriented measures to assess how well the model balances accuracy with equitable treatment. Our experiments analyse performance across different group compositions, providing insights into how fairness dynamics change as preference similarity within groups varies. Additionally, we compare the adaptive-noise VAE with a static-noise baseline to quantify the benefit of learning noise from data. Finally, we employ Bayesian Optimization to identify optimal hyperparameters that effectively balance fairness, diversity, and ranking quality.

2. Related Work

Variational Autoencoders (VAEs) have gained popularity for managing sparse and implicit feedback data in recommender systems. A probabilistic VAE framework with a multinomial likelihood goal was presented in [1], showing that VAEs perform noticeably better on a variety of real-world datasets than traditional models like matrix factorization and other neural versions. [2] presented a hybrid VAE model that uses movie content embeddings learnt from a different VAE network to supplement user-item rating data. Recently, [3] conducted a thorough analysis of the VAE-based recommender system landscape, classifying models according to their architecture, optimization approach, and data modality. According to the survey, VAEs excel at managing sparse data, simulating multi-modal inputs, and facilitating transfer learning. Even though the conventional recommenders have been very successful in several tasks, they are not appropriate to produce group recommendations [4]. To accurately anticipate the rating that a group of users would assign to an item, they apply genetic algorithms to forecast potential interactions among group members. In the context of sequential group recommendations, balancing group satisfaction with individual disagreement remains a significant challenge. To address this, [5] proposed three aggregation methods designed to optimize overall group satisfaction while minimizing intra-group disagreements. To ensure fairness and maximize the satisfaction among the group members, [6] proposed a fairness-aware group recommendation by modelling both individual satisfaction (utility) and fairness among group members. [7] proposed Group Fairness Aware Recommendations (GFAR), a technique that defines fairness in a rank-sensitive manner. The relevance of each prefix of the top-N recommendations is distributed evenly across all group members because of GFAR. SQUIRREL [8, 9] presented a framework for sequential group recommendations using reinforcement learning. It improves group satisfaction by dynamically selecting the optimal recommendation algorithm at each step, balancing satisfaction and dissatisfaction throughout the process, and was demonstrated in [10]. [11] focuses on how to produce recommendations that are fairer and more diverse among group members, by adding stochastic elements to the recommendation process.

3. Methodology

3.1. Variational Autoencoder Architecture

In this study, we use Variational Autoencoder (VAE) as the predictive model for recommendations. The VAE is a generative model that learns a probabilistic latent representation of the user–item interaction space, enabling it to capture complex, non-linear relationships between users and movies.

Encoder. The input to the VAE is a user-item interaction matrix, where rows represent users and columns represent movie items, with cell values representing ratings. This user-item interaction matrix is inserted into the VAE encoder as input. The encoder transforms this higher-dimensional data into a lower-dimensional latent space. Then the encoder predicts three parameter sets:

• Mean (μ) – representing the central tendency of the latent distribution.
• Variance (σ^2) - representing the uncertainty of the latent variables.
• Noise log-standard deviation ($\log \sigma_{\text{noise}}$) – used for adaptive fairness-oriented noise.

Given the input, the encoder generates a tractable distribution that approximates the intractable true posterior distribution of the latent variables: $q_\phi(z_u \mid x_u) = \mathcal{N}\big(\mu_\phi(x_u), \mathrm{diag}(\sigma^2_\phi(x_u))\big)$, where x_u is the observed data for user u, ϕ denotes the parameters of the encoder network in the VAE, and $q_\phi(z_u \mid x_u)$ is the variational posterior, which is the encoder's approximation of the true posterior $P(x_u \mid z_u)$.
Reparameterization Trick. The reparameterization approach is used to allow backpropagation using stochastic sampling. Latent variable z_u sampled as: $z_u = \mu_\phi(x_u) + \epsilon \odot \sigma_\phi(x_u), \quad \epsilon \sim \mathcal{N}(0, I)$. This technique preserves the stochastic nature of the model while facilitating effective training by allowing gradients to backpropagate directly through the stochastic latent variables.
Decoder. The original user–item interaction vector is reconstructed by the decoder network using the sampled latent variable z_u. The model's anticipated ratings for every item for the specified user are represented by this reconstruction. The decoder models the probability of user u interacting with item i as: $p_\theta(x_u \mid z_u) = \mathrm{Mult}\big(N_u, \pi(z_u)\big)$, where N_u is the number of items interacted with by user u and $\pi(z_u)$ represents the predicted preference distribution over all items.
Training Objective. The VAE is trained to maximize the evidence lower bound (ELBO), which balances:

1. **Reconstruction loss:** Reconstruction loss quantifies the disparity between expected and actual ratings for observed items.
2. **Kullback–Leibler (KL) divergence:** KL divergence regularizes the acquired latent distribution to align with a typical Gaussian prior to promote generalization and mitigate overfitting.

Mathematically, the objective function is: $L = \mathbb{E}_{q_\phi(z_u|x_u)}\big[\log p_\theta(x_u \mid z_u)\big] - \mathrm{KL}\,\big(q_\phi(z_u \mid x_u) \,\|\, p(z_u)\big)$.
The VAE models the user with encoder $q_\phi(z_u \mid x_u)$ and decoder $p_\theta(x_u \mid z_u)$, while the KL divergence $\mathrm{KL}\big(q_\phi(z_u \mid x_u) \,\|\, p(z_u)\big)$ regularizes the latent space.

3.2. Adaptive Noise Injection

Traditional VAEs have the ability to increase biases in recommendations by overfitting or assigning excessively confident latent encodings to particular user groups. Though in some recent studies it is shown that injecting static noise in the VAEs can improve fairness while keeping the ranking quality almost the same in both single user recommendations [12] and group recommendations [11]. We wanted to explore how VAEs work when adaptive noise is injected into it. Our proposed model introduces adaptive fairness-oriented noise injection, which alters latent variables in a learnt, data-dependent fashion. In contrast to fixed Gaussian noise, the magnitude of this modification is dynamically assessed for each user according to their encoded representation.
Noise Parameterization and Learning. The strength of the noise is learned from the hidden representation h by the encoder network. A dedicated linear layer specifically produces the logarithm of the noise standard deviation: $\log \sigma_{\text{noise}} = W_{\text{noise}}\, h + b_{\text{noise}}$. To ensure positivity of the noise, softplus activation is used: $\sigma_{\text{adaptive}} = \mathrm{softplus}\,\big(\log \sigma_{\text{noise}}\big)$. In addition, the learned noise magnitude is constrained to lie within a bounded range $[\sigma_{\min}, \sigma_{\max}]$ using a clamping operation. This prevents degenerate cases where the noise becomes excessively small, leading to overconfident and potentially biased representations, or excessively large, which could destabilize learning. The encoder directly predicts this noise parameter, which is concurrently tuned with all other model weights during training by gradient descent. The backpropagation from the reconstruction loss and KL divergence loss propagates into the latent mean, variance, and adaptive noise parameters. Rather than relying on a fixed noise value, this method enables the model to learn how much noise should be injected for each user. Thus, the noise magnitude becomes an emergent property of the optimization process:

- For users whose representations risk being overconfident, the model learns to assign a high noise variance, promoting fairness and diversity.
- For users with clear and consistent interaction patterns, the model learns to assign a lower noise variance, preserving accuracy.

Encoder-side Noise Injection. After sampling the latent vector via reparameterization, an adaptive Gaussian noise is applied: $z' = z + \epsilon_1 \odot \sigma_{\text{adaptive}}, \qquad \epsilon_1 \sim \mathcal{N}(0, I)$. This noisy z' is passed to the decoder for reconstruction. Figure 2 represents the proposed variational autoencoder architecture.

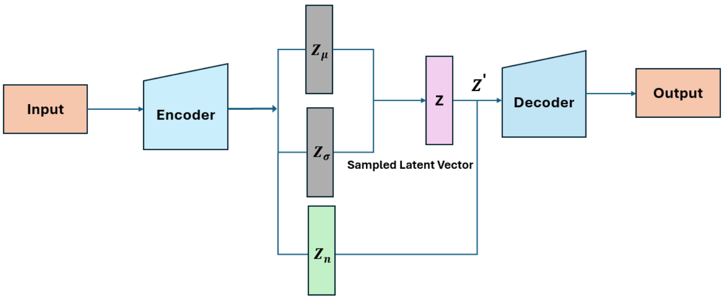

Figure 1: Proposed Variational Autoencoder Architecture.

3.3. Hyperparameter Optimization via Bayesian Optimization

We tune the VAE's key hyperparameters using Bayesian optimization (BO) to minimize the model's validation loss. For a hyperparameter vector λ, the objective is: $J(\lambda) = \frac{1}{|D_{\text{val}}|} \sum_{x \in D_{\text{val}}} \left[L_{\text{recon}}(x;\lambda) + L_{\text{KL}}(x;\lambda) \right]$, where D_{val} denotes the validation dataset, L_{recon} is the reconstruction loss and L_{KL} is the standard VAE KL term. As the Bayesian optimizer maximizes the objective, meaning it minimizes the validation loss, the implementation returns $-J(\lambda)$.

3.4. Group Recommendation Process

In this study, we employed VAEs to generate individual user recommendations. Subsequently, we applied various aggregation methods to combine the individual scores and produce group recommendations. The **Average Method** calculates the group rating for each item by using the arithmetic mean of the ratings provided by each user. This method presumes that all users' perspectives hold equal significance and aims to reflect the collective agreement of the group. **Least Misery** assigns the lowest individual rating of the group member as the group rating, ensuring that no group member encounters substantial dissatisfaction. Nevertheless, it may underestimate items that are highly favoured by the majority of group members if an individual offers a low rating. In the **Maximum Satisfaction** method, the highest rating given by any individual of the group is assigned as the group rating for that item, assuming that if at least one individual is fully satisfied with the item, it is acceptable throughout the group.

4. Evaluation Metrics

Satisfaction. Individual users' satisfaction score is calculated as: $\text{sat}(u_i, \text{Grp}) = \sum_{i \in \text{Grp}} p(u_i, i) / \sum_{i \in A(u_i)} p(u_i, i)$. Grp denotes the group recommendation list and $p(u_i, i)$ denotes the preference score for item i by user u_i. Users individual recommended list is denoted by $A(u_i)$. The group satisfaction is: $\text{grpSat}(G, \text{Grp}) = \sum_{u_i \in G} \text{sat}(u_i, \text{Grp}) / |G|$. By averaging the group members' satisfaction ratings, we ensure that the recommendations made are accepted by everyone [5].

Dissatisfaction. The disagreement, or dissatisfaction, score for a single user is calculated as: $\text{dissat}(u_i, G, \text{Grp}) = 1 - \text{sat}(u_i, \text{Grp})$ [11]. Then, the group dissatisfaction is: $\text{groupDissat}(u_i, G, \text{Grp}) = \sum_{u_i \in G} \text{dissat}(u_i, G, \text{Grp}) / |G|$.

Normalized Discounted Cumulative Gain (NDCG). NDCG@k evaluates how well the top-k recommended items are arranged. Specifically, $\text{NDCG@}k = \frac{\text{DCG@}k}{\text{IDCG@}k}$, where $\text{DCG@}k = \sum_{i=1}^{k} \frac{\text{rel}_i}{\log_2(i+1)}$ with rel_i be the relevance of the item shown at rank i in the recommended list. In turn, $\text{IDCG@}k = \sum_{i=1}^{|R|} \frac{1}{\log_2(i+1)}$, where R denotes the list of relevant items in the recommended list. By using NDCG, we can evaluate the system's ranking with the ideal configuration, giving us a strong indicator of how well it prioritizes the most relevant items at the top.

Discounted Fairness (DFH). DFH measures how far a user's actual exposure to recommended items deviates from the exposure they ideally should receive based on relevance. Specifically:

Table 1
Combined Evaluation Results for All Group Types and Recommendation Sizes

Group / Movies	Aggregation Method	Adaptive Noise Model					Static Noise Model (Baseline)				
		Sat.	DisSat.	NDCG	DFH	Rec.	Sat.	DisSat.	NDCG	DFH	Rec.
Homogeneous Group – 20 Movies											
AVG		0.9948	0.0051	0.6466	0.7012	0.8515	0.9810	0.0189	0.6208	0.7451	0.8226
LM		0.9478	0.0521	0.4698	0.8171	0.6976	0.9012	0.0987	0.4069	0.8475	0.6560
MS		0.9948	0.0051	0.6697	0.7436	0.8088	0.9878	0.0121	0.5401	0.7806	0.8227
Homogeneous Group – 50 Movies											
AVG		0.9997	0.0003	0.6369	0.7801	0.9999	0.9370	0.0630	0.6093	0.7835	0.9991
LM		0.9995	0.0005	0.4935	0.8245	0.9998	0.9199	0.0801	0.4330	0.8287	0.9990
MS		0.9875	0.0125	0.6598	0.7827	0.9995	0.9471	0.0529	0.5454	0.8097	0.9991
Heterogeneous Group – 20 Movies											
AVG		0.6407	0.3592	0.3579	0.5834	0.0768	0.5671	0.4328	0.3422	0.5939	0.0617
LM		0.7493	0.2506	0.4461	0.7250	0.1396	0.5672	0.4329	0.3423	0.5970	0.0817
MS		0.7092	0.2907	0.4172	0.5913	0.0817	0.6189	0.3810	0.3389	0.5955	0.0680
Heterogeneous Group – 50 Movies											
AVG		0.7702	0.2297	0.4594	0.7456	0.1627	0.7299	0.2701	0.4034	0.7732	0.1030
LM		0.8549	0.1450	0.4994	0.7354	0.1604	0.7271	0.2729	0.4038	0.7357	0.1343
MS		0.8189	0.1810	0.4580	0.7431	0.1088	0.6702	0.3297	0.4225	0.7544	0.1084
Mixed Group – 20 Movies											
AVG		0.7360	0.2639	0.6089	0.6532	0.4022	0.7320	0.2679	0.5822	0.6717	0.4015
LM		0.6281	0.3718	0.4119	0.6530	0.2477	0.6222	0.3777	0.4057	0.8631	0.3002
MS		0.7340	0.2659	0.4813	0.6673	0.4026	0.7280	0.2719	0.4609	0.6821	0.3894
Mixed Group – 50 Movies											
AVG		0.8240	0.1759	0.6480	0.8344	0.5183	0.7794	0.2205	0.6295	0.8325	0.4875
LM		0.7081	0.2919	0.4853	0.8377	0.2909	0.7074	0.2925	0.4511	0.8748	0.3139
MS		0.8105	0.1894	0.5257	0.8482	0.4521	0.7582	0.2417	0.5038	0.8499	0.4338

$\text{DFH} = \frac{1}{|G|} \sum_{u \in G} |\sum_{k=1}^{n} a_{uk} - \sum_{k=1}^{n} r_{uk}|$, where the attention received by any item k by user u is expressed as a_{uk} and the relevance of item k for user u is expressed as r_{uk}. DFH measures the difference in exposure of relevant items across different users.
Recall. Recall measures the proportion of a user's truly relevant items that appear in the recommended list. For each user u, the recall is: $\text{Recall}_u\ @k = \frac{TP}{TP+FN}$. TP are the items that are both recommended and truly relevant, while FN are the items that are relevant but not recommended. We report the macro-averaged recall across the group: $\overline{\text{Recall}}@k = \frac{1}{|U|} \sum_{u \in U} \text{Recall}_u\ @k$.

5. Experimental Setup & Results Analysis

We use the MovieLens 10M dataset [13]. To incorporate implicit feedback, the data was converted into a binary format. Movies with ratings of 1.5 or lower and users with fewer than 5 interactions were excluded from the dataset. Any duplicate entries in the dataset were removed. After preprocessing, the dataset includes 9,402,608 ratings, 69,869 users, and 10,661 movies. We employ three different types of groups with six users to analyse the results. These groups are formed on the basis of user similarity calculated using the Pearson correlation. In the **homogeneous** setting, groups are formed based on the highest similarity scores among users. A group of six users is created by maximizing the average pairwise similarity (excluding self-similarity) and enforcing a minimum average similarity of 0.45 to ensure high similarity in preferences. In the **heterogeneous** context, groups are formed by selecting users with minimal similarity, minimizing the average pairwise similarity and enforcing a maximum average similarity of 0.20 to ensure sufficient diversity. In the **mixed** context, groups are formed by combining users with both high and low similarity scores, where half of the group consists of users with the highest similarity and the other half with the lowest, maintaining both homogeneity and variety.

Result Analysis. Table 1 presents the evaluation results of various aggregation methods applied to all group types using adaptive noise injection, and compares them with the Static Noise model, where the noise level is set to 1.5, across different recommendation-size scenarios.
Homogeneous Group. The Adaptive Noise model reliably delivers consistent satisfaction and fair exposure across 20- and 50-movie recommendation sizes, while enhancing ranking quality and mini-

mizing dissatisfaction. Overall, Adaptive Noise increases Satisfaction from the low–mid 0.90s to above 0.99 with Average aggregation, while substantially reducing Dissatisfaction (e.g., from around 0.06 to near 0.00 for larger slates). Improvements in ranking quality (NDCG) are consistent across slate sizes, with notable gains under Maximum Satisfaction, where NDCG rises from approximately 0.54 to 0.66. Recall also improves or remains near-optimal. At large slate, Satisfaction and Recall reach peak levels; however, the Adaptive Noise model continues to provide superior ranking quality and reduces user dissatisfaction, with average aggregation under the adaptive noise model demonstrating optimal fairness, and maximum satisfaction with the adaptive noise model attains the highest-ranking quality.
Heterogeneous Group. Across both 20 and 50 recommendation sizes, the Adaptive Noise model consistently outperforms the static baseline. The results highlight a clearer fairness–utility trade-off than in homogeneous settings, making aggregator choice more influential. Adaptive Noise increases Satisfaction and improves NDCG across all aggregators, with the strongest gains observed under Least Misery (e.g., NDCG rising from approximately 0.34 to 0.45 for small slates and to nearly 0.50 for larger slates). Recall also improves, while Dissatisfaction is notably reduced, particularly for larger recommendation lists. Within the Adaptive Noise model, the Least Misery aggregator attains the highest Satisfaction and NDCG across both recommendation sizes and achieves the best fairness score for larger recommendation lists, while the Average aggregator achieves the highest recall. Adaptive noise model improves ranking quality, user experience, and exposure parity for diverse groups at large slate sizes.
Mixed Group. The Adaptive Noise model consistently improves ranking quality and fairness over the static baseline, with the choice of aggregation influencing the fairness–utility trade-off. Average aggregation delivers balanced performance, improving Satisfaction, NDCG, and Recall, while maintaining stable fairness. Least Misery emphasizes higher satisfaction and fairness, though it can reduce recall (e.g., Recall drops from ≈ 0.30 to 0.25 for 20 movies). Maximum Satisfaction generally optimizes effectiveness metrics, boosting Satisfaction and NDCG, with moderate fairness and recall.
Impact of Bayesian Optimization. Employing Bayesian optimization (BO) to tune the mixed discrete–continuous hyperparameters of the VAE (learning rate, two hidden-layer widths, latent size, and dropout) had a clear, positive impact on model quality and stability. By optimizing the validation negative ELBO directly, BO efficiently navigated the search space (5 random initializations + 25 guided iterations) and identified configurations that achieved the lowest validation loss observed within our budget. Practically, BO reduced tuning time relative to grid/random search, avoided invalid architectures via index-mapped discrete choices, and yielded consistent convergence behaviour. A limitation is that the BO objective targets ELBO rather than ranking/fairness metrics directly; nonetheless, in our experiments, ELBO improvements aligned with gains in NDCG and Recall.

6. Conclusions

In this work, we introduce a VAE-based group recommender with adaptive, learned noise in the latent space and demonstrate that, across multiple group compositions, it consistently outperforms a static-noise baseline on the MovieLens 10M dataset. The adaptive model delivered higher ranking quality and coverage with no increase in user dissatisfaction, and it typically improved fairness (lower DFH). The gains were strongest for NDCG and Recall, indicating that data-dependent stochastic regularization helps surface relevant items earlier and in greater number. Methodologically, we clarified that dual latent noise is redundant; a single encoder-side injection matches or exceeds its effect while simplifying optimization and improving interpretability. Bayesian optimization over learning rate, hidden/latent dimensions, and dropout reliably found optimal hyperparameters.

Declaration on Generative AI

During the preparation of this work, the author(s) used Chat-GPT-4 and Grammarly in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication’s content.

References

[1] D. Liang, R. G. Krishnan, M. D. Hoffman, T. Jebara, Variational autoencoders for collaborative filtering, in: Proceedings of the 2018 world wide web conference, 2018, pp. 689–698.

[2] K. Gupta, M. Y. Raghuprasad, P. Kumar, A hybrid variational autoencoder for collaborative filtering, arXiv preprint arXiv:1808.01006 (2018).

[3] S. Liang, Z. Pan, W. Liu, J. Yin, M. De Rijke, A survey on variational autoencoders in recommender systems, ACM Computing Surveys 56 (2024) 1–40.

[4] Y.-L. Chen, L.-C. Cheng, C.-N. Chuang, A group recommendation system with consideration of interactions among group members, Expert systems with applications 34 (2008) 2082–2090.

[5] M. Stratigi, E. Pitoura, J. Nummenmaa, K. Stefanidis, Sequential group recommendations based on satisfaction and disagreement scores, Journal of Intelligent Information Systems 58 (2022) 227–254.

[6] L. Xiao, Z. Min, Z. Yongfeng, G. Zhaoquan, L. Yiqun, M. Shaoping, Fairness-aware group recommendation with pareto-efficiency, in: Proceedings of the eleventh ACM conference on recommender systems, 2017, pp. 107–115.

[7] M. Kaya, D. Bridge, N. Tintarev, Ensuring fairness in group recommendations by rank-sensitive balancing of relevance, in: Proceedings of the 14th ACM Conference on recommender systems, 2020, pp. 101–110.

[8] M. Stratigi, E. Pitoura, K. Stefanidis, Squirrel: A framework for sequential group recommendations through reinforcement learning, Information Systems 112 (2023) 102128.

[9] M. M. Hasan, S. Pervez, M. Stratigi, K. Stefanidis, Squirrel 2.0: fairness & explanations for sequential group recommendations, in: International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data, CEUR-WS, 2024, pp. 63–67.

[10] E. Chrysostomaki, M. Stratigi, V. Efthymiou, K. Stefanidis, D. Plexousakis, Fair sequential group recommendations in squirrel movies, in: International Conference on Very Large Data Bases (VLDB), CEUR, 2023.

[11] M. S. Ali, K. Stefanidis, Fairness in group recommender systems using variational autoencoders, in: International Database Engineered Applications Symposium, Springer, 2024, pp. 297–311.

[12] W. Imtiaz, Fairness in variational autoencoders recommenders, Tampere, Finland: Tampere University (2020).

[13] F. M. Harper, J. A. Konstan, The movielens datasets: History and context, Acm transactions on interactive intelligent systems (tiis) 5 (2015) 1–19.