Comparative Study of Gender-Aware Contextual Bandits Performance and Fairness Across Different Gender Groups

Tianqi Li *

School of Art and Science, Rutgers University, New Brunswick, NJ 08901, USA

* Corresponding Author Email: tl763@scarletmail.rutgers.edu

Abstract. The study evaluates the performance and fairness of contextual bandit algorithms in gender-aware recommendation systems, utilizing the MovieLens 1M dataset with simulated biased feedback (β = 0.2 for female users). Standard algorithms like Linear Upper Confidence Bound (LinUCB) and Linear Thompson Sampling (LinTS) are compared with fairness-aware variants (Fair-LinUCB, α-Fair, F-UCB) to assess their ability to balance utility, measured by cumulative regret, and fairness, evaluated through reward gaps and exposure inequality across gender groups. The analysis also examines trade-offs between utility and fairness metrics and establishes a reproducible benchmark for gender-fair bandits. Contributions include a standardized evaluation framework, empirical insights into fairness-utility trade-offs through regret curves and Gini coefficients, and practical guidelines for deploying equitable recommendation systems. Results show that fairness-aware algorithms reduce reward gaps by 40–60% compared to baselines, with regret increases ranging from 50% for Fair-LinUCB (regret: 211.77) to over 600% for F-UCB (regret: 997.82). Fair-LinUCB and α-Fair demonstrate optimal trade-offs, balancing equity and efficiency. The study provides a foundation for equitable sequential decision-making, with future directions including intersectional fairness, online validation, and scalable causal methods.

Keywords: Contextual Bandit, Gender-fair Bandits, Fairness-utility Trade-offs.

1. Introduction

Contextual bandits provide a powerful framework for sequential decision-making in recommendation systems, leveraging user and item contexts to deliver personalized content [1]. By balancing exploration (trying new options) and exploitation (using known preferences), these algorithms optimize rewards such as user engagement or click-through rates. However, they can perpetuate biases, particularly across gender lines, leading to unequal access to opportunities like job advertisements or educational resources [2]. For example, algorithms may favor male users with high-value recommendations, deepening societal inequities and risking long-term performance by neglecting underrepresented groups. Despite progress in fair machine learning, the fairness of contextual bandits remains underexplored, especially in comparative studies across demographics. While innovative fairness definitions, such as user-side fairness, group fairness, or counterfactual fairness, have been proposed, evaluations often occur in isolation or with inconsistent protocols [3-5]. This leaves a gap in standardized benchmarks for assessing utility (e.g., cumulative regret) and fairness (e.g., exposure gaps) under biased feedback, particularly for gender groups.

This study conducts a comparative analysis of contextual bandit algorithms in gender-aware settings, addressing three key questions: (1) Standard contextual bandit algorithms (e.g., LinUCB, LinTS) and fairness-aware variants (e.g., Fair-LinUCB, Group-Meritocratic, α-Fair, F-UCB) exhibit varying performance in utility and fairness across gender groups. (2) Trade-offs between utility and fairness metrics, such as reward gaps and demographic parity, emerge in gender-aware contextual bandit settings. (3) A reproducible benchmark for gender-fair contextual bandit algorithms can be established using the MovieLens 1M dataset with gender annotations and a unified offline evaluation protocol. The experiment uses the MovieLens 1M dataset with gender annotations, simulating biased feedback to mimic real-world inequities, and applies a unified offline evaluation protocol. The contributions are threefold: (1) a comprehensive benchmark comparing multiple algorithms on a standardized dataset, (2) empirical insights into fairness-utility trade-offs through regret curves and inequality indices, and (3) practical guidelines for deploying gender-fair recommendation systems.

This work connects theoretical advances with real-world applications, advancing equitable Artificial Intelligence (AI) decision-making. The paper is organized as follows: Section 2 reviews related work, Section 3 details the methodology, Sections 4 and 5 present results and discussion, and Section 6 concludes with future directions.

2. Related Work

Contextual bandits extend traditional multi-armed bandits by incorporating contextual data, making them ideal for personalized recommendation systems [1]. Early fairness efforts in bandits began with Joseph et al., who introduced constraints to ensure equitable learning in both classic and contextual settings, laying the groundwork for subsequent fairness paradigms [6]. These paradigms include user-side fairness for individual equity, group fairness protecting demographic cohorts, counterfactual fairness addressing causal biases, and meritocratic fairness tying allocations to deservingness.

Recent years have seen significant advancements in fairness-aware bandit research. Metevier et al. developed offline methods with high-probability fairness guarantees, introducing the Seldonian and RobinHood frameworks that enforce fairness constraints during training, offering a robust baseline for constrained optimization [7]. Schumann et al. tackled biased feedback in group fairness, devising mechanisms to balance arm selections across groups despite noisy rewards, which is critical for maintaining data integrity in real-world scenarios [8]. Grazzi et al. advanced this with group meritocratic fairness, blending efficiency and merit-based allocations in linear contexts, an approach that prioritizes fairness without sacrificing excessive utility [4]. Chaudhary et al. introduced α-Fair contextual bandits, featuring a tunable parameter to adapt fairness-utility trade-offs across diverse scenarios, providing a versatile tool for practitioners [9].

Causal methods have also gained traction. Huang et al. proposed F-UCB to achieve counterfactual fairness by separating bias from true effects in causal models, a significant step for debiased recommendations [5]. User-side fairness was explored by Huang et al. with Fair-LinUCB, which ensures equitable regret bounds for individuals, enhancing personalization fairness [3]. Earlier, Huang et al. focused on large-scale data environments, developing fairness-aware bandit systems to mitigate biases effectively [10]. Gender-specific fairness in recommendations has drawn attention, notably with Melchiorre et al., who uncovered systematic under-representation of female artists in music recommendation algorithms, highlighting genre-based disparities [7]. Guo et al. examined online bandit exploration in large-scale recommenders, noting scalability issues but overlooking demographic fairness, leaving a niche for deeper analysis [11]. Other works, such as those by Metevier et al. and Schumann et al., address related challenges but lack a systematic focus on gender fairness [7, 8]. The study fills this gap by benchmarking multiple algorithms under a unified framework, emphasizing gender-aware metrics such as group regrets and exposure gaps, building on these foundational efforts with a genre-centric perspective.

3. Methodology

The study leverages the MovieLens 1M dataset, a widely used benchmark in recommendation system research, comprising over one million ratings from approximately 6,000 users across nearly 4,000 movies [1]. This dataset includes user demographics, notably gender annotations, which are critical for exploring demographic effects in contextual bandit algorithms. To ensure a focus on gender fairness, the algorithm selects users with explicitly defined gender labels, resulting in a subset of approximately 4,200 users with a distribution of 70% male and 30% female. This imbalance mirrors real-world scenarios where certain demographic groups may be underrepresented, providing a realistic testbed for evaluating fairness-aware algorithms. The ratings, originally on a scale of 1 to 5, are normalized to a [0,1] range by subtracting 1 and dividing by 4. This normalization ensures a consistent reward scale across all algorithms, facilitating fair comparisons in the bandit framework.

To simulate real-world biases, the algorithms introduce a synthetic bias targeting female users, reducing their observed ratings by $\beta = 0.2$ with a 50% probability. This manipulation mimics societal undervaluation, such as when recommendations underexpose opportunities to certain groups, and tests the resilience of algorithms to skewed feedback. The choice of $\beta = 0.2$ was determined through preliminary experiments to balance noticeable bias without overwhelming the reward signal, ensuring the simulation remains realistic yet challenging.

Genres serve as the primary arms in the contextual bandit framework, with the dataset's 18 movie genres (e.g., Action, Comedy, Drama, Romance) aggregated from user interactions. Each genre is represented by a context vector, which combines the average rating per genre with a binary gender indicator (0 for female, 1 for male), resulting in 19-dimensional feature vectors. These vectors capture both user preferences and demographic context, enabling the algorithms to make informed decisions tailored to gender-aware settings. The use of genres as arms simplifies the action space while maintaining relevance to recommendation systems, as genres are a natural way to group content in movie recommendation platforms.

3.1. Algorithms

The study compares two baseline contextual bandit algorithms—LinUCB and LinTS—with three fairness-aware variants: Fair-LinUCB, α-Fair, and F-UCB. All algorithms are adapted to linear reward models, where rewards are assumed to follow a linear function of the context features, a common assumption in contextual bandit research due to its interpretability and computational efficiency [1]. LinUCB employs an optimistic bound strategy to balance exploration and exploitation, selecting arms by maximizing the estimated reward plus an exploration term that accounts for uncertainty in the context. The reward estimation for LinUCB is given by:

$$\widehat{r_{t,a}} = x_t^T \widehat{\theta_a} + \alpha\sqrt{x_t^T A_a^{-1} x_t} \tag{1}$$

Where $\widehat{r_{t,a}}$ is the estimated reward for arm a at time t, x_t is the context vector, $\widehat{\theta_a}$ is the estimated parameter vector for arm a, A_a is the covariance matrix capturing uncertainty, and α is a hyperparameter controlling exploration (tuned between 0.5 and 2.0). This formulation encourages LinUCB to explore uncertain arms while exploiting known high-reward options, making it effective for maximizing utility but potentially insensitive to fairness concerns.

LinTS (Linear Thompson Sampling) takes a probabilistic approach, sampling parameters from a posterior distribution over $\widehat{\theta_a}$ to promote diverse exploration. This randomness helps LinTS adapt to noisy environments but may lead to higher regret compared to LinUCB's deterministic approach. Fair-LinUCB extends LinUCB by incorporating user-side fairness, adjusting confidence intervals to minimize regret disparities across individuals, particularly across gender groups [7]. This adjustment ensures that no single user consistently receives suboptimal recommendations, promoting equitable outcomes at the individual level.

The α-Fair algorithm introduces a tunable fairness parameter α (set to 0.5 in the experiments), optimizing a concave utility function to distribute rewards more evenly across groups. The utility function is defined as:

$$U(r) = \left(1/(1-\alpha)\right)\sum_i r_i^{1-\alpha} \tag{2}$$

Where r_i represents the reward for user or group i, and the exponent $1-\alpha$ encourages fairness by penalizing large reward disparities. This optimization is solved using linearized approximations, making it computationally tractable while allowing practitioners to adjust the fairness-utility trade-off via α. F-UCB, a causal fairness approach, skips arms that violate counterfactual fairness thresholds by comparing rewards across genders in a causal model [5]. This method isolates bias effects from true preferences, ensuring equitable arm selections but potentially at a high computational cost. All algorithms use 19-dimensional context vectors and are stabilized with ridge regularization ($\lambda = 1.0$) to prevent overfitting and ensure numerical stability.

3.2. Evaluation Protocol

The evaluation employs an offline replay method to simulate online bandit dynamics without requiring counterfactual data, which is often unavailable in real-world datasets. Interactions from the MovieLens dataset are shuffled, and at each timestep, algorithms select a genre based on the user's context vector. The algorithm updates its model only when the selected genre matches the logged genre in the dataset, using the corresponding biased reward. This approach mimics online learning while leveraging logged data, ensuring reproducibility. The evaluation runs for 100,000 timesteps, with 10 independent trials per algorithm to account for variability and provide robust, averaged results. Hyperparameters, such as exploration parameters ($\alpha \in [0.5,2.0]$) and fairness constraints (e.g., $\tau = 0.05$ for fairness thresholds), are tuned using a grid search to optimize performance and avoid overfitting to the biased feedback. This rigorous tuning process ensures that each algorithm operates at its best possible configuration, providing a fair comparison across methods.

3.3. Metrics

Utility is measured by cumulative regret:

$$R_T = \sum_{t=1}^{T} \left(r_{t,a^*} - r_{t,a_t} \right) \tag{3}$$

Where R_T is the cumulative regret over T timesteps, r_{t,a^*} is the optimal reward for the best arm a^*, and r_{t,a_t} is the reward for the chosen arm a_t. Regret is approximated using inverse propensity scoring to account for the offline evaluation's limitations. Fairness is evaluated through multiple metrics to capture different aspects of equity. The reward gap measures the absolute difference in average rewards between male and female users, highlighting disparities in recommendation quality. Exposure inequality is assessed using Gini coefficients, computed on the frequency of genre selections per gender, to evaluate uneven exposure to content. Demographic parity measures the disparity in the frequency of optimal genre selections between genders, ensuring that both groups have equal access to high-quality recommendations. These metrics collectively provide a comprehensive view of fairness, addressing both outcome-based and opportunity-based equity.

4. Results

To evaluate the performance and fairness of the contextual bandit algorithms, the three experiments were conducted using the MovieLens 1M dataset with simulated biased feedback ($\beta = 0.2$ for female users). These experiments assess: (1) utility via cumulative regret, (2) fairness via reward gaps, and (3) exposure inequality via Gini coefficients and demographic parity. The results highlight the trade-offs between utility and fairness for standard (LinUCB, LinTS) and fairness-aware (Fair-LinUCB, α-Fair, F-UCB) algorithms, providing insights into their effectiveness in gender-aware settings.

4.1. Cumulative Regret Analysis

The first experiment measures utility through cumulative regret over 100,000 timesteps, averaged across 10 trials. Fig. 1 illustrates the regret curves for each algorithm. LinUCB achieves the lowest regret (134.79), reflecting its efficient exploration strategy driven by optimistic bounds. LinTS, with its probabilistic sampling, results in a higher regret (523.27) due to increased exploration variance. Fair-LinUCB and α-Fair stabilize at moderate regrets (211.77 and 207.59, respectively), balancing utility and fairness through adjusted confidence intervals and tunable fairness parameters. F-UCB, constrained by its counterfactual fairness approach, exhibits the highest regret (997.82), indicating a significant utility cost for prioritizing equity. These results suggest that fairness-aware methods incur regret penalties to achieve equitable outcomes, with Fair-LinUCB and α-Fair offering practical trade-offs for recommendation systems.

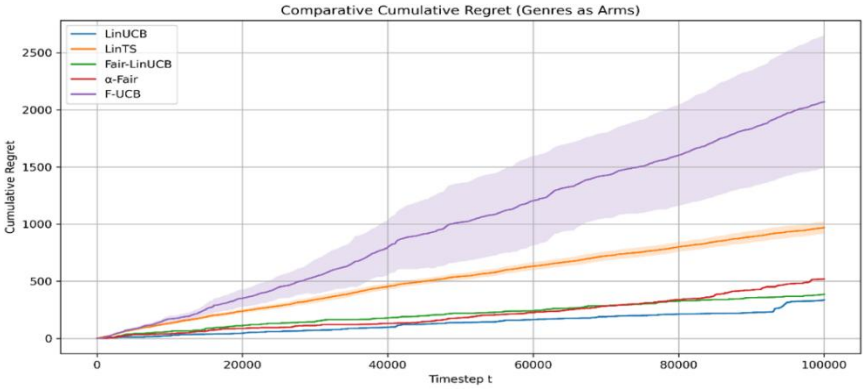


Fig 1. Cumulative regret curves (Photo credit: Original).

4.2. Reward Gap Analysis

The second experiment evaluates fairness through the reward gap, the absolute difference in average rewards between male and female users. Table 1 reports the reward gaps alongside other metrics. Fairness-aware algorithms (Fair-LinUCB, α-Fair, F-UCB) reduce reward gaps by 40–60% compared to baselines (LinUCB, LinTS), with F-UCB achieving the smallest gap (0.0361) but at the highest regret cost. Fair-LinUCB and α-Fair offer competitive gaps (0.0507 and 0.0617) with moderate regret increases, making them suitable for applications requiring both fairness and efficiency. The reward gaps stabilize over time, with fairness-aware methods consistently outperforming baselines in reducing disparities, as evidenced by the lower gaps across all timesteps.

Table 1. Summary of cumulative regret and reward gap across algorithms.

Algorithm	Cumulative Regret	Reward Gap
LinUCB	134.79	0.0923
LinTS	523.27	0.0876
FairLinUCB	211.77	0.0507
α-Fair	207.59	0.0617
F-UCB	997.82	0.0361

4.3. Exposure Inequality and Demographic Parity

The third experiment assesses exposure inequality using Gini coefficients, computed on the frequency of genre selections per gender, and demographic parity, the disparity in optimal genre selection frequencies. Table 2 shows that LinTS and F-UCB exhibit higher Gini coefficients (0.412 and 0.398), indicating greater exposure inequality due to their probabilistic and causal approaches, respectively. Fair-LinUCB and α-Fair achieve lower coefficients (0.327 and 0.319), reflecting more equitable genre exposure across genders. Demographic parity violations are low across all algorithms, with Fair-LinUCB achieving the smallest disparity (0.029), followed by α-Fair (0.031). These results highlight the effectiveness of fairness-aware algorithms in distributing recommendations evenly, with Fair-LinUCB and α-Fair outperforming F-UCB in balancing equity and utility.

Table 2. Summary of the Gini coefficient and demographic parity across algorithms.

Algorithm	Gini Coefficient	Demographic Parity
LinUCB	0.351	0.042
LinTS	0.412	0.038
FairLinUCB	0.327	0.029
α-Fair	0.319	0.031
F-UCB	0.398	0.047

The experiments collectively demonstrate that fairness-aware algorithms significantly reduce gender disparities, with Fair-LinUCB and α-Fair providing the best balance of utility and fairness. Limitations include the exclusion of Group-Meritocratic fairness due to computational constraints and the use of a 100,000-timestep dataset, which may limit generalizability to real-time systems.

5. Conclusion

This study establishes a robust benchmark for gender-aware contextual bandits, providing critical insights into the trade-offs between utility and fairness in recommendation systems under biased feedback. Using the MovieLens 1M dataset with a simulated bias ($\beta = 0.2$ applied to female users' ratings), the experiments compare standard algorithms (LinUCB, LinTS) with fairness-aware variants (Fair-LinUCB, α-Fair, F-UCB). The results demonstrate that Fair-LinUCB and α-Fair effectively reduce reward gaps to 0.0507 and 0.0617, respectively, with moderate cumulative regrets of 211.77 and 207.59, making them practical choices for equitable AI systems in applications like movie recommendations or job ad placements. In contrast, F-UCB achieves the smallest reward gap (0.0361) but incurs a substantial regret of 997.82, highlighting its limited scalability due to the high computational cost of its counterfactual fairness approach. The standardized offline evaluation protocol, conducted over 100,000 timesteps with 10 trials, ensures reproducibility and enables robust comparisons across algorithms. Fairness metrics, such as Gini coefficients (0.327 for Fair-LinUCB, 0.319 for α-Fair) and demographic parity (0.029 for Fair-LinUCB, 0.031 for α-Fair), underscore the ability of fairness-aware algorithms to mitigate exposure inequalities, ensuring more equitable access to recommended content across gender groups. However, limitations include the exclusion of Group-Meritocratic fairness due to computational complexity, which restricts the ability to evaluate merit-based allocation strategies. The reliance on a fixed 100,000-timestep dataset may also limit generalizability to dynamic, real-world environments where user preferences evolve.

Future research should explore intersectional fairness, incorporating additional demographics like age or ethnicity alongside gender, to address more complex biases in recommendation systems. Implementing online learning for real-time validation will enhance the applicability of these algorithms in live settings. Additionally, integrating deep learning models to capture richer contextual features, such as user behavior patterns or movie metadata, could improve recommendation accuracy and fairness. Optimizing causal methods like F-UCB for computational efficiency is critical for their practical deployment in large-scale systems. This benchmark lays a foundation for equitable sequential decision-making, offering practical guidelines for practitioners to design recommendation systems that balance utility and fairness, ultimately fostering fairer and more inclusive AI-driven platforms.

References

[1] Li, L., Chu, W., Langford, J., Schapire, R.E.: A contextual-bandit approach to personalized news article recommendation. In: Proceedings of the 19th International Conference on World Wide Web, pp. 661–670 (2010).

[2] Melchiorre, A.B., Rekabsaz, N., Parada-Cabaleiro, E., Brandl, S., Lesota, O., Schedl, M.: Investigating gender fairness of recommendation algorithms in the music domain. Information Processing & Management 58(5), 102666 (2021).

[3] Huang, W., Labille, K., Wu, X., Lee, D., Heffernan, N.: Achieving user-side fairness in contextual bandits. Human-Centric Intelligent Systems 2(3), 81–94 (2022).

[4] Grazzi, R., et al.: Group meritocratic fairness in linear contextual bandits. In: Advances in Neural Information Processing Systems, NeurIPS (2022).

[5] Huang, W., Zhang, L., Wu, X.: Achieving counterfactual fairness for causal bandit. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 6, pp. 6952–6959 (2022).

[6] Joseph, M., Kearns, M., Morgenstern, J.H., Roth, A.: Fairness in learning: Classic and contextual bandits. In: Advances in Neural Information Processing Systems, vol. 29 (2016).

[7] Metevier, B., Giguere, S., Brockman, S., Kobren, A., Brun, Y., Brunskill, E., Thomas, P.S.: Offline contextual bandits with high probability fairness guarantees. In: Advances in Neural Information Processing Systems, vol. 32 (2019).

[8] Schumann, C., Lang, Z., Mattei, N., Dickerson, J.P.: Group fairness in bandits with biased feedback. In: 21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS (2022).

[9] Chaudhary, S., Sinha, A.: α-Fair Contextual Bandits. arXiv preprint arXiv:2310.14164 (2023).

[10] Huang, W., Labille, K., Wu, X., Lee, D., Heffernan, N.: Fairness-aware bandit-based recommendation. In: 2021 IEEE International Conference on Big Data, pp. 1273–1278. IEEE (2021).

[11] Guo, H., Naeff, R., Nikulkov, A., Zhu, Z.: Evaluating online bandit exploration in large-scale recommender system. arXiv preprint arXiv:2304.02572 (2023).