Towards Long-term Fairness in Recommendation

Long-Term Fairness Project

Devika Singh, Rinshi Kumari, Varsha
210101036, 210101109, 210108040

Indian Institute of Guwahati

Contents

Abstract I

- Fairness in Recommender Systems (RS) is increasingly important due to their influence on daily life.
- Most existing approaches to fairness are static and do not adapt to changing item popularity.
- Long-term fairness necessitates the ability to accommodate dynamic changes in item popularity over time.
- This work explores long-term fairness through dynamic fairness learning.
- The focus is on fair exposure of items across groups defined by their changing popularity.
- A fairness-constrained reinforcement learning algorithm is proposed, modeled as a Constrained Markov Decision Process (CMDP).
- The model dynamically adjusts its recommendation policy to ensure ongoing fairness.
- Experiments show improved recommendation performance and fairness over time on real-world datasets.

Introduction I

- **Personalized RS:** Core for online services (e-commerce, advertising, job markets).
- **Algorithmic Bias:** RS may impact underrepresented groups negatively.
- **Matthew Effect:** Popular items gain popularity, while long-tail items receive less exposure.
- **Static Fairness:** Focuses on immediate fairness, not long-term impact.
- **Example:**
  - Four items: A, B (popular group G_0) and C, D (long-tail group G_1).
  - Static fairness can lead to unexpected long-term biases (e.g., D getting more exposure).
- **Static vs. Dynamic Fairness:**
  - Static: One-time fairness solution, ignores changes (utility, attributes, user feedback).
  - Dynamic: Adapts to changes, learns strategies over time.
- **Long-Term Fairness:**

Introduction II

- Focuses on sustained fairness, considers item exposure dynamics.

- **Key Questions:**
  - How to model long-term fairness with changing group labels?
  - How to update recommendations based on real-time data?
  - How to optimize strategies on large-scale datasets?

- **Approach:**
  - Model interactions as Markov Decision Process (MDP).
  - Convert to Constrained Markov Decision Process (CMDP) for dynamic fairness.
  - Use Constrained Policy Optimization (CPO) for learning optimal policy.

- **Results:**
  - Empirical validation on real-world datasets shows improved performance and fairness.
  - First to model dynamic fairness concerning changing group labels.

Related Works, Preliminary, Problem Formulation I

Study	Data Source	Methodology/Tools	Results
(1; 2)	Various datasets	Fairness definitions: individual	Identified biases in RS (gender, race).
(3)	E-commerce platforms	Fairness optimization	Proposed utility optimization under fairness constraints.
(4)	Ranking systems	Joint utility and fairness optimization	Developed algorithms for fairness-aware recommendations.
(5; 6)	User feedback data	Dynamic fairness models	Adjusted to changing utility with fairness constraints.
(7; 8; 9)	Contextual data	RL as contextual multi-armed bandits	Incorporated collaborative filtering methods.
(10; 11; 12)	Recommendation logs	MDP modeling	Proposed policy-based and value-based methods.
(13)	Specific fairness constraint data	RL for meritocratic fairness	Focused on long-term rewards, limited to specific constraints.
(14)	User-side fairness data	Model-free and model-based RL	Achieved demographic parity and near-optimal fairness.

- **PRELIMINARY**
  - Markov Decision Processes
  - Constrained Markov Decision Processes
  - Constrained Policy Optimization
- **PROBLEM FORMULATION**
  - CMDP for Recommendation - timestamp t, Recommendation Agent G, User u, Cost c, State S, Action A, Reward R, Cost C, Discount rates γ_r and γ_c
  - Fairness Constraints - Demographic Parity Constraints, Exact-KFairness Constraints
  - FCPO: Fairness Constrained Policy Optimization

Methodology I

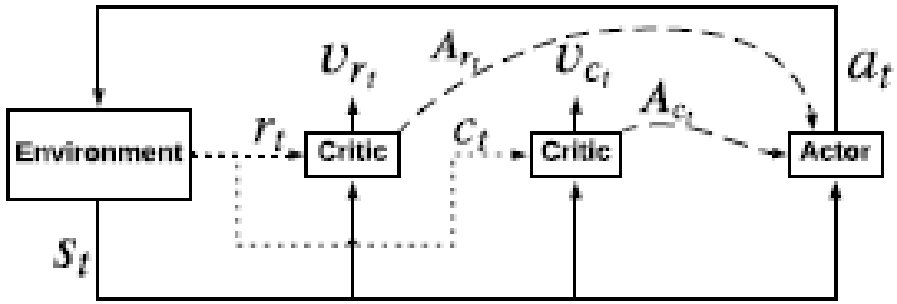

Figure 1: Illustration of the proposed method.

- **Actor-Critic Learning Framework:**
  - **The Actor:** The actor network in FCPO generates recommendations based on user states, which are represented by recent user interaction history. The state is encoded using GRU, and the actor selects items by calculating their similarity to a proposal matrix. The selected items are then recommended to the user.

Methodology II

  - **The Critics:** FCPO employs two critic networks: one for reward estimation and another for fairness evaluation. Both critics are trained using temporal-difference learning. The value critic evaluates the recommendation's accuracy, while the cost critic ensures the fairness constraint is respected.

- **Training Procedure:** The FCPO training consists of two phases: generating user-recommendation interaction trajectories and updating model parameters. The actor's policy is updated based on reward and fairness advantages, while the critics are optimized by minimizing their respective losses.
- **Testing Procedure:** In testing, FCPO evaluates recommendation performance without further parameter updates. It performs both short-term and long-term evaluations, ensuring recommendations meet fairness constraints while adapting to real-time changes in item popularity.

Methodology III

Algorithm 1: Parameters Training for FCPO

Input: step size δ, cost limit value d, and line search ratio β
Output: parameters θ, ω and ϕ of actor network, value function, cost function
Randomly initialize θ, ω and ϕ.
Initialize replay buffer D;
for $Round = 1 \ldots M$ **do**
  Initialize user state s_0 from log data;
  for $t = 1 \ldots T$ **do**
    Observe current state s_t based on Eq. (12);
    Select an action $a_t = \{a_t^1, \ldots, a_t^K\} \in I^K$ based on Eq. (14) and Eq. (15)
    Calculate reward r_t and cost c_t according to environment feedback based on Eq. (8) and Eq. (9);
    Update s_{t+1} based on Eq. (??);
    Store transition $(s_t, a_t, r_t, c_t, s_{t+1})$ in D in its corresponding trajectory.
  end
  Sample minibatch of N trajectories $\mathcal{T}$ from D;
  Calculate advantage value A, advantage cost value A_c;
  Obtain gradient direction d_θ by solving Eq. (4) with A and A_c;
  repeat
    $\theta' \leftarrow \theta + d_\theta$
    $d_\theta \leftarrow \beta d_\theta$
  until $\pi_{\theta'}(s)$ *in trust region & loss improves & cost* $\leq$ d;
  (Policy update) $\theta \leftarrow \theta'$;
  (Value update) Optimize ω based on Eq.(16);
  (Cost update) Optimize ϕ based on Eq.(17);
end

Experiments I

- **Dataset Description:** Movielens 100K and Movielens 1M. These contain user-item interaction data, with items categorized into two groups based on popularity: popular items (top 20%) and long-tail items (remaining 80%), split into training, validation, and testing sets, and use of reinforcement learning to simulate real-time recommendations by continuously adjusting user states based on their recent interactions.
- **Experimental Setup:** Several baseline models, such as Matrix Factorization (MF), Bayesian Personalized Ranking (BPR), and Neural Collaborative Filtering (NCF), are compared to FCPO. Evaluation metrics include traditional recommendation performance measures like Recall, F1 Score, and NDCG, alongside fairness metrics like Gini Index and Popularity Rate. FCPO's performance is measured under different fairness constraints (FCPO-1, FCPO-2, FCPO-3), which control the level of fairness imposed on recommendations.

Experiments II

- **Experimental Results:** FCPO outperforms traditional models in both recommendation accuracy and fairness. FCPO achieves significantly higher recall and NDCG scores compared to baselines. Moreover, it excels in maintaining fairness, with lower Gini Index (indicating less inequality in item exposure) and better long-tail item visibility. The results demonstrate that FCPO effectively balances performance and fairness, especially in dynamically changing environments. FCPO's ability to adapt continuously allows it to handle evolving user preferences and item popularity, ensuring sustained fairness throughout the recommendation process.

CONCLUSION AND FUTUREWORK

- CONCLUSION : The paper introduces a framework for long-term fairness in recommendation systems, addressing the dynamic nature of item popularity and ensuring fairness over time. The proposed fairness-constrained reinforcement learning framework balances recommendation performance with both short-term and long-term fairness.
- FUTURE WORK :
  - **Individual Fairness:**
    Aim to generalize the framework to address other fairness dimensions, such as individual fairness
  - **Application:**
    Explore the application of paper to recommendation domains like e-commerce
  - **Enhance Model Interpretability:**
    Introduce techniques (e.g. LIME) that help in understanding how different features impact the recommendations to identify potential biases in the model, facilitating improvements in fairness and transparency.

References I

[1] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher. The unfairness of popularity bias in recommendation. 2019.

[2] Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. [n.d.].

[3] Ziwei Zhu, Xia Hu, and James Caverlee. Fairness-Aware Tensor-Based Recommendation. In Proceedings of CIKM '18 (Torino, Italy). 2018.

[4] L. Elisa Celis, Sayash Kapoor, Farnood Salehi, and Nisheeth Vishnoi. Controlling Polarization in Personalization: An Algorithmic Framework. In Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019.

[5] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. Unbiased Recommender Learning from Missing-Not-At-Random Implicit Feedback. In Proceedings of WSDM '20. 2020.

[6] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. Controlling Fairness and Bias in Dynamic Learning-to-Rank. In SIGIR. New York, NY, USA. 2020.

[7] Djallel Bouneffouf, Amel Bouzeghoub, and Alda Lopes Gançarski. A contextual-bandit algorithm for mobile context-aware recommender system. In International conference on neural information processing. Springer, 324–331. 2012.

[8] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In Proceedings of the 19th international conference on World wide web. 661–670. 2010.

References II

[9] Chunqiu Zeng, QingWang, Shekoofeh Mokhtari, and Tao Li. Online context-aware recommendation with time varying multi-armed bandit. In Proceedings of the 22nd ACM SIGKDD. 2025–2034. 2016.

[10] Tariq Mahmood and Francesco Ricci. Learning and adaptivity in interactive recommender systems. In Proceedings of the 9th international conference on Electronic commerce. 75–84. 2007.

[11] Tariq Mahmood and Francesco Ricci. Improving recommender systems with adaptive conversational strategies. In Proceedings of the 20th ACM conference on Hypertext and hypermedia. 73–82. 2009.

[12] Guy Shani, David Heckerman, and Ronen I Brafman. An MDP-based recommender system. Journal of Machine Learning Research 6, Sep (2005).

[13] Jabbari et al. Fairness in reinforcement learning. In ICML. 2017.

[14] M. Wen, Osbert Bastani, and U. Topcu. Fairness with Dynamics. ArXiv abs/1901.08568 (2019).