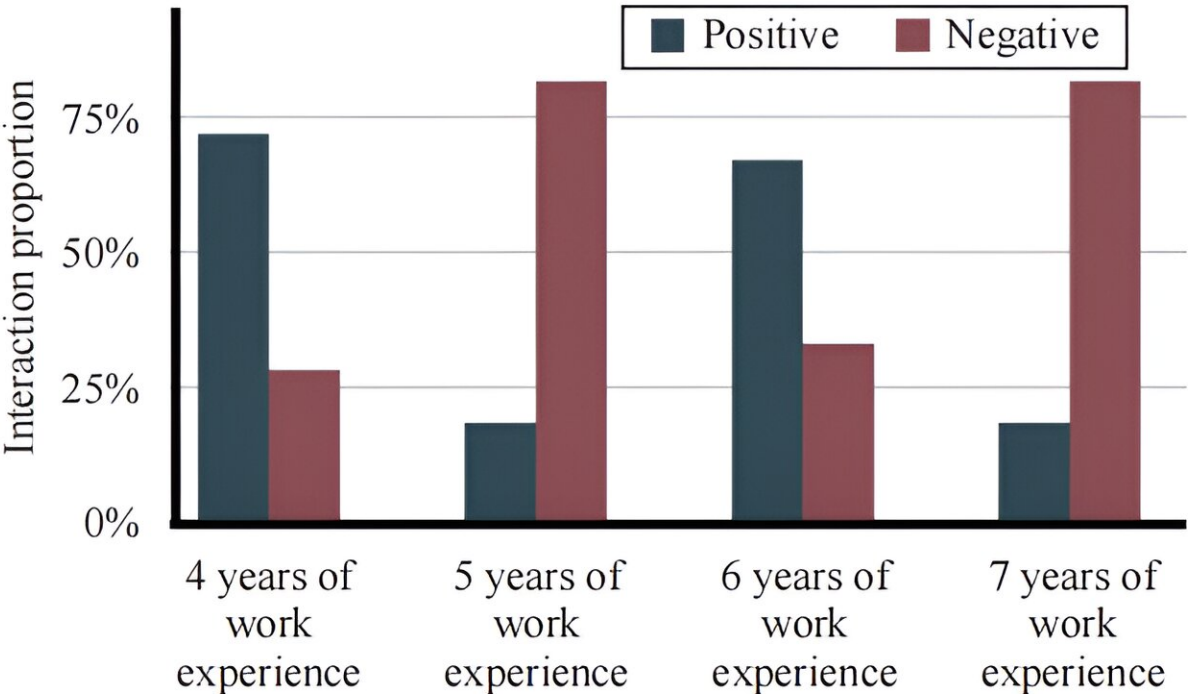


# Mitigating spurious correlations for self-supervised recommendation

August 4 2023



An example of spurious correlations in job recommendation, where the positive interactions with the full-time jobs show strong correlations with the users' 4 and 6 years of experience. Credit: Beijing Zhongke Journal Publishing Co. Ltd.

Recent years have witnessed the great success of self-supervised learning (SSL) in recommendation systems. However, SSL recommender models

are likely to suffer from spurious correlations, leading to poor generalization. To mitigate spurious correlations, existing work usually pursues ID-based SSL recommendation or utilizes feature engineering to identify spurious features.

Nevertheless, ID-based SSL approaches sacrifice the positive impact of invariant features, while feature engineering methods require high-cost human labeling. In a paper published in *Machine Intelligence Research*, a team of researchers aim to automatically mitigate the effect of spurious correlations to address the problems. This objective requires to automatically mask spurious features without supervision, and block the negative effect transmission from spurious features to other features during SSL.

Self-supervised learning (SSL) approaches have recently become state-of-the-art (SOTA) for personalized recommendation. The core idea of SSL in recommendation is to learn better user and item representations via an additional self-discrimination task, which contrasts the augmentations over user-item features or user-item interaction graphs to discover the correlation relationships among features and interactions.

Despite the great success, SSL-based recommender models are vulnerable to spurious correlations due to fitting the correlations from the input features to interactions. Because of the selection bias in the data collection process, spurious correlations inevitably exist in the [training data](#), where some spurious features show strong correlations with users' positive interactions. By the self-discrimination task, SSL models tend to capture these spurious correlations, resulting in poor generalization ability.

To alleviate the harmful effect of spurious correlations on SSL models, existing solutions mainly fall into three categories: First is ID-based SSL methods, which only utilize IDs of users and items for collaborative

filtering, and thus can avoid the harmful influence of some spurious features.

However, the user and item features are still useful in the recommendation, especially for users with sparse interactions. It is necessary to consider some invariant features that causally affect the interactions. Second is feature engineering methods, which are able to identify a set of spurious features manually or using human-machine hybrid approaches. Thereafter, they can train the SSL recommender models by discarding the identified features.

Nevertheless, feature engineering methods require extensive human-labeling work and thus are not applicable to large-scale recommendations with extensive user and item features. Third is informative feature selection methods, which are capable of automatically recognizing the informative cross features and removing the redundant ones in the training process. Nevertheless, spurious features might be very informative for the interaction prediction in the training data, and thus degrade the generalization ability.

To solve the problems, researchers require the SSL models to automatically mitigate the effect of spurious correlations. In order to achieve this objective, there exist two essential challenges: The first one is that it is non-trivial to mask spurious features without supervision. The other challenge is blocking the effect transmission from spurious features to other features is of vital importance.

To address the two challenges, researchers consider learning a feature mask mechanism from multiple environments to estimate the probabilities of spurious features and then adopt the mask mechanism to guide the feature augmentation in SSL models. Specifically, researchers can cluster the interactions into multiple environments, where each environment has similar feature distributions, but the distributions shift

between environments.

The distribution shifts will guide the mask mechanism to capture invariant features across environments and exclude spurious features. Besides, they can utilize the mask mechanism to drop the spurious features as the augmented sample and then maximize the mutual information between the invariant features in the augmented sample and all the input features in the factual sample, pushing SSL models to ignore the spurious features and cut off the negative effect transmission from spurious features to invariant features.

To this end, researchers propose an invariant feature learning (IFL) framework for SSL recommender models to mitigate spurious correlations. In particular, IFL clusters the training interactions into multiple environments and leverages a masking mechanism with learnable parameters in  $[0, 1]$  to shield spurious correlations. To optimize the mask parameters, IFL adopts a variance loss to identify invariant features and achieve robust predictions across environments.

As for the self-discrimination task, researchers drop the spurious features based on the mask parameters as the augmented sample, and then maximize the mutual information between the factual and augmented samples via contrastive loss, which pushes the SSL model to ignore spurious features. They instantiate IFL on a SOTA SSL model, and extensive experiments on two real-world datasets validate the effectiveness of the proposed IFL in mitigating spurious correlations.

The contributions of this paper are summarized as follows: Firstly, researchers point out the spurious correlations in SSL recommendation and consider learning invariant features from multiple environments.

Secondly, they propose a [model](#)-agnostic IFL framework, which leverages a feature mask mechanism and mask-guided contrastive

learning to reduce spurious correlations for SSL models.

Thirdly, empirical results on two public datasets verify the superiority of their proposed IFL in masking spurious features and enhancing the generalization ability of SSL models.

**More information:** Xin-Yu Lin et al, Mitigating Spurious Correlations for Self-supervised Recommendation, *Machine Intelligence Research* (2023). [DOI: 10.1007/s11633-022-1374-8](https://doi.org/10.1007/s11633-022-1374-8)

Provided by Beijing Zhongke Journal Publishing Co.

Citation: Mitigating spurious correlations for self-supervised recommendation (2023, August 4)  
retrieved 8 May 2024 from  
<https://techxplore.com/news/2023-08-mitigating-spurious-self-supervised.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.