# 论文周报 | 推荐系统领域最新研究进展，含KDD, SIGIR, ICDE等顶会论文

原创 ML\_RSer [机器学习与推荐算法](javascript:void(0);) 2023-06-05 08:10 发表于北京

收录于合集#推荐系统干货分享246个

**嘿，记得给“机器学习与推荐算法”添加星标**

本文精选了上周（0529-0604）最新发布的18篇推荐系统相关论文，主要研究方向包括大型语言模型赋能推荐系统、对话推荐系统、图推荐系统、多模态推荐系统、药物推荐等。

以下整理了论文标题以及摘要，如感兴趣可移步原文精读。

1. TransAct: Transformer-based Realtime User Action Model for  Recommendation at Pinterest, KDD2023

2. A Survey on Large Language Models for Recommendation

3. Medication Recommendation via Domain Knowledge Informed Deep Learning

4. Criteria Tell You More than Ratings: Criteria Preference-Aware Light  Graph Convolution for Effective Multi-Criteria Recommendation, KDD2023

5. Instant Representation Learning for Recommendation over Large Dynamic  Graphs, ICDE2023

6. Pure Spectral Graph Embeddings: Reinterpreting Graph Convolution for  Top-N Recommendation, PAKDD2023

7. REVAL: a semantic evaluation framework for hashtag recommendation

8. Multi-behavior Self-supervised Learning for Recommendation, SIGIR2023

9. Multimodal Recommendation Dialog with Subjective Preference: A New  Challenge and Benchmark, ACL2023

10. Sequential Condition Evolved Interaction Knowledge Graph for Traditional  Chinese Medicine Recommendation

11. AI Coach Assist: An Automated Approach for Call Recommendation in  Contact Centers for Agent Coaching, ACL2023

12. Multi-Epoch Learning for Deep Click-Through Rate Prediction Models

13. Multi-View Interactive Collaborative Filtering

14. A Survey on Fairness-aware Recommender Systems

15. Graph Exploration Matters: Improving both individual-level and  system-level diversity in WeChat Feed Recommender

16. Robust Reinforcement Learning Objectives for Sequential Recommender  Systems

17. Towards Explainable Conversational Recommender Systems

18. Choosing the Right Weights: Balancing Value, Strategy, and Noise in  Recommender Systems

### **1. TransAct: Transformer-based Realtime User Action Model for  Recommendation at Pinterest, KDD2023**

Xue Xia, Pong Eksombatchai, Nikil Pancha, Dhruvil Deven Badani, Po-Wei Wang, Neng Gu, Saurabh Vishwas Joshi, Nazanin Farahpour, Zhiyuan Zhang, Andrew Zhai

https://arxiv.org/abs/2306.00248

Sequential models that encode user activity for next action prediction have become a popular design choice for building web-scale personalized recommendation systems. Traditional methods of sequential recommendation either utilize end-to-end learning on realtime user actions, or learn user representations separately in an offline batch-generated manner. This paper (1) presents Pinterest's ranking architecture for Homefeed, our personalized recommendation product and the largest engagement surface; (2) proposes TransAct, a sequential model that extracts users' short-term preferences from their realtime activities; (3) describes our hybrid approach to ranking, which combines end-to-end sequential modeling via TransAct with batch-generated user embeddings. The hybrid approach allows us to combine the advantages of responsiveness from learning directly on realtime user activity with the cost-effectiveness of batch user representations learned over a longer time period. We describe the results of ablation studies, the challenges we faced during productionization, and the outcome of an online A/B experiment, which validates the effectiveness of our hybrid ranking model. We further demonstrate the effectiveness of TransAct on other surfaces such as contextual recommendations and search. Our model has been deployed to production in Homefeed, Related Pins, Notifications, and Search at Pinterest.

### **2. A Survey on Large Language Models for Recommendation**

Likang Wu, Zhi Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, Hui Xiong, Enhong Chen

https://arxiv.org/abs/2305.19860

Large Language Models (LLMs) have emerged as powerful tools in the field of Natural Language Processing (NLP) and have recently gained significant attention in the domain of Recommendation Systems (RS). These models, trained on massive amounts of data using self-supervised learning, have demonstrated remarkable success in learning universal representations and have the potential to enhance various aspects of recommendation systems by some effective transfer techniques such as fine-tuning and prompt tuning, and so on. The crucial aspect of harnessing the power of language models in enhancing recommendation quality is the utilization of their high-quality representations of textual features and their extensive coverage of external knowledge to establish correlations between items and users. To provide a comprehensive understanding of the existing LLM-based recommendation systems, this survey presents a taxonomy that categorizes these models into two major paradigms, respectively Discriminative LLM for Recommendation (DLLM4Rec) and Generative LLM for Recommendation (GLLM4Rec), with the latter being systematically sorted out for the first time. Furthermore, we systematically review and analyze existing LLM-based recommendation systems within each paradigm, providing insights into their methodologies, techniques, and performance. Additionally, we identify key challenges and several valuable findings to provide researchers and practitioners with inspiration. We have also created a GitHub repository to index relevant papers on LLMs for recommendation, https://github.com/WLiK/LLM4Rec

### **3. Medication Recommendation via Domain Knowledge Informed Deep Learning**

Sicen Liu, Xiaolong Wang, Xianbing Zhao, Hao Chen

https://arxiv.org/abs/2305.19604

Medication recommendation is a fundamental yet crucial branch of healthcare, which provides opportunities to support clinical physicians with more accurate medication prescriptions for patients with complex health conditions. Learning from electronic health records (EHR) to recommend medications is the most common way in previous studies. However, most of them neglect incorporating domain knowledge according to the clinical manifestations in the EHR of the patient. To address these issues, we propose a novel Domain Knowledge Informed Network (DKINet) to integrate domain knowledge with observable clinical manifestations of the patient, which is the first dynamic domain knowledge informed framework toward medication recommendation. In particular, we first design a knowledge-driven encoder to capture the domain information and then develop a data-driven encoder to integrate domain knowledge into the observable EHR. To endow the model with the capability of temporal decision, we design an explicit medication encoder for learning the longitudinal dependence of the patient. Extensive experiments on three publicly available datasets verify the superiority of our method. The code will be public upon acceptance.

### **4. Criteria Tell You More than Ratings: Criteria Preference-Aware Light  Graph Convolution for Effective Multi-Criteria Recommendation, KDD2023**

Jin-Duk Park, Siqing Li, Xin Cao, Won-Yong Shin

https://arxiv.org/abs/2305.18885

The multi-criteria (MC) recommender system, which leverages MC rating information in a wide range of e-commerce areas, is ubiquitous nowadays. Surprisingly, although graph neural networks (GNNs) have been widely applied to develop various recommender systems due to GNN's high expressive capability in learning graph representations, it has been still unexplored how to design MC recommender systems with GNNs. In light of this, we make the first attempt towards designing a GNN-aided MC recommender system. Specifically, rather than straightforwardly adopting existing GNN-based recommendation methods, we devise a novel criteria preference-aware light graph convolution CPA-LGC method, which is capable of precisely capturing the criteria preference of users as well as the collaborative signal in complex high-order connectivities. To this end, we first construct an MC expansion graph that transforms user--item MC ratings into an expanded bipartite graph to potentially learn from the collaborative signal in MC ratings. Next, to strengthen the capability of criteria preference awareness, CPA-LGC incorporates newly characterized embeddings, including user-specific criteria-preference embeddings and item-specific criterion embeddings, into our graph convolution model. Through comprehensive evaluations using four real-world datasets, we demonstrate (a) the superiority over benchmark MC recommendation methods and benchmark recommendation methods using GNNs with tremendous gains, (b) the effectiveness of core components in CPA-LGC, and (c) the computational efficiency.

### **5. Instant Representation Learning for Recommendation over Large Dynamic  Graphs, ICDE2023**

Cheng Wu, Chaokun Wang, Jingcao Xu, Ziwei Fang, Tiankai Gu, Changping Wang, Yang Song, Kai Zheng, Xiaowei Wang, Guorui Zhou

https://arxiv.org/abs/2305.18622

Recommender systems are able to learn user preferences based on user and item representations via their historical behaviors. To improve representation learning, recent recommendation models start leveraging information from various behavior types exhibited by users. In real-world scenarios, the user behavioral graph is not only multiplex but also dynamic, i.e., the graph evolves rapidly over time, with various types of nodes and edges added or deleted, which causes the Neighborhood Disturbance. Nevertheless, most existing methods neglect such streaming dynamics and thus need to be retrained once the graph has significantly evolved, making them unsuitable in the online learning environment. Furthermore, the Neighborhood Disturbance existing in dynamic graphs deteriorates the performance of neighbor-aggregation based graph models. To this end, we propose SUPA, a novel graph neural network for dynamic multiplex heterogeneous graphs. Compared to neighbor-aggregation architecture, SUPA develops a sample-update-propagate architecture to alleviate neighborhood disturbance. Specifically, for each new edge, SUPA samples an influenced subgraph, updates the representations of the two interactive nodes, and propagates the interaction information to the sampled subgraph. Furthermore, to train SUPA incrementally online, we propose InsLearn, an efficient workflow for single-pass training of large dynamic graphs. Extensive experimental results on six real-world datasets show that SUPA has a good generalization ability and is superior to sixteen state-of-the-art baseline methods. The source code is available at https://github.com/shatter15/SUPA

### **6. Pure Spectral Graph Embeddings: Reinterpreting Graph Convolution for  Top-N Recommendation, PAKDD2023**

Edoardo D'Amico, Aonghus Lawlor, Neil Hurley

https://arxiv.org/abs/2305.18374

The use of graph convolution in the development of recommender system algorithms has recently achieved state-of-the-art results in the collaborative filtering task (CF). While it has been demonstrated that the graph convolution operation is connected to a filtering operation on the graph spectral domain, the theoretical rationale for why this leads to higher performance on the collaborative filtering problem remains unknown. The presented work makes two contributions. First, we investigate the effect of using graph convolution throughout the user and item representation learning processes, demonstrating how the latent features learned are pushed from the filtering operation into the subspace spanned by the eigenvectors associated with the highest eigenvalues of the normalised adjacency matrix, and how vectors lying on this subspace are the optimal solutions for an objective function related to the sum of the prediction function over the training data. Then, we present an approach that directly leverages the eigenvectors to emulate the solution obtained through graph convolution, eliminating the requirement for a time-consuming gradient descent training procedure while also delivering higher performance on three real-world datasets.

### **7. REVAL: a semantic evaluation framework for hashtag recommendation**

Areej Alsini, Du Q. Huynh, Amitava Datta

https://arxiv.org/abs/2305.18330

Automatic evaluation of hashtag recommendation models is a fundamental task in many online social network systems. In the traditional evaluation method, the recommended hashtags from an algorithm are firstly compared with the ground truth hashtags for exact correspondences. The number of exact matches is then used to calculate the hit rate, hit ratio, precision, recall, or F1-score. This way of evaluating hashtag similarities is inadequate as it ignores the semantic correlation between the recommended and ground truth hashtags. To tackle this problem, we propose a novel semantic evaluation framework for hashtag recommendation, called #REval. This framework includes an internal module referred to as BERTag, which automatically learns the hashtag embeddings. We investigate on how the #REval framework performs under different word embedding methods and different numbers of synonyms and hashtags in the recommendation using our proposed #REval-hit-ratio measure. Our experiments of the proposed framework on three large datasets show that #REval gave more meaningful hashtag synonyms for hashtag recommendation evaluation. Our analysis also highlights the sensitivity of the framework to the word embedding technique, with #REval based on BERTag more superior over #REval based on FastText and Word2Vec.

### **8. Multi-behavior Self-supervised Learning for Recommendation, SIGIR2023**

Jingcao Xu, Chaokun Wang, Cheng Wu, Yang Song, Kai Zheng, Xiaowei Wang, Changping Wang, Guorui Zhou, Kun Gai

https://arxiv.org/abs/2305.18238

Modern recommender systems often deal with a variety of user interactions, e.g., click, forward, purchase, etc., which requires the underlying recommender engines to fully understand and leverage multi-behavior data from users. Despite recent efforts towards making use of heterogeneous data, multi-behavior recommendation still faces great challenges. Firstly, sparse target signals and noisy auxiliary interactions remain an issue. Secondly, existing methods utilizing self-supervised learning (SSL) to tackle the data sparsity neglect the serious optimization imbalance between the SSL task and the target task. Hence, we propose a Multi-Behavior Self-Supervised Learning (MBSSL) framework together with an adaptive optimization method. Specifically, we devise a behavior-aware graph neural network incorporating the self-attention mechanism to capture behavior multiplicity and dependencies. To increase the robustness to data sparsity under the target behavior and noisy interactions from auxiliary behaviors, we propose a novel self-supervised learning paradigm to conduct node self-discrimination at both inter-behavior and intra-behavior levels. In addition, we develop a customized optimization strategy through hybrid manipulation on gradients to adaptively balance the self-supervised learning task and the main supervised recommendation task. Extensive experiments on five real-world datasets demonstrate the consistent improvements obtained by MBSSL over ten state-of-the art (SOTA) baselines. We release our model implementation at: https://github.com/Scofield666/MBSSL.git

### **9. Multimodal Recommendation Dialog with Subjective Preference: A New  Challenge and Benchmark, ACL2023**

Yuxing Long, Binyuan Hui, Caixia Yuan1, Fei Huang, Yongbin Li, Xiaojie Wang

https://arxiv.org/abs/2305.18212

Existing multimodal task-oriented dialog data fails to demonstrate the diverse expressions of user subjective preferences and recommendation acts in the real-life shopping scenario. This paper introduces a new dataset SURE (Multimodal Recommendation Dialog with SUbjective PREference), which contains 12K shopping dialogs in complex store scenes. The data is built in two phases with human annotations to ensure quality and diversity. SURE is well-annotated with subjective preferences and recommendation acts proposed by sales experts. A comprehensive analysis is given to reveal the distinguishing features of SURE. Three benchmark tasks are then proposed on the data to evaluate the capability of multimodal recommendation agents. Based on the SURE, we propose a baseline model, powered by a state-of-the-art multimodal model, for these tasks.

### **10. Sequential Condition Evolved Interaction Knowledge Graph for Traditional  Chinese Medicine Recommendation**

Jingjin Liu, Hankz Hankui Zhuo, Kebing Jin, Jiamin Yuan, Zhimin Yang, Zhengan Yao

https://arxiv.org/abs/2305.17866

Traditional Chinese Medicine (TCM) has a rich history of utilizing natural herbs to treat a diversity of illnesses. In practice, TCM diagnosis and treatment are highly personalized and organically holistic, requiring comprehensive consideration of the patient's state and symptoms over time. However, existing TCM recommendation approaches overlook the changes in patient status and only explore potential patterns between symptoms and prescriptions. In this paper, we propose a novel Sequential Condition Evolved Interaction Knowledge Graph (SCEIKG), a framework that treats the model as a sequential prescription-making problem by considering the dynamics of the patient's condition across multiple visits. In addition, we incorporate an interaction knowledge graph to enhance the accuracy of recommendations by considering the interactions between different herbs and the patient's condition. Experimental results on a real-world dataset demonstrate that our approach outperforms existing TCM recommendation methods, achieving state-of-the-art performance.

### **11. AI Coach Assist: An Automated Approach for Call Recommendation in  Contact Centers for Agent Coaching, ACL2023**

Md Tahmid Rahman Laskar, Cheng Chen, Xue-Yong Fu, Mahsa Azizi, Shashi Bhushan, Simon Corston-Oliver

https://arxiv.org/abs/2305.17619

In recent years, the utilization of Artificial Intelligence (AI) in the contact center industry is on the rise. One area where AI can have a significant impact is in the coaching of contact center agents. By analyzing call transcripts using Natural Language Processing (NLP) techniques, it would be possible to quickly determine which calls are most relevant for coaching purposes. In this paper, we present AI Coach Assist, which leverages the pre-trained transformer-based language models to determine whether a given call is coachable or not based on the quality assurance (QA) questions asked by the contact center managers or supervisors. The system was trained and evaluated on a large dataset collected from real-world contact centers and provides an effective way to recommend calls to the contact center managers that are more likely to contain coachable moments. Our experimental findings demonstrate the potential of AI Coach Assist to improve the coaching process, resulting in enhancing the performance of contact center agents.

### **12. Multi-Epoch Learning for Deep Click-Through Rate Prediction Models**

Zhaocheng Liu, Zhongxiang Fan, Jian Liang, Dongying Kong, Han Li

https://arxiv.org/abs/2305.19531

The one-epoch overfitting phenomenon has been widely observed in industrial Click-Through Rate (CTR) applications, where the model performance experiences a significant degradation at the beginning of the second epoch. Recent advances try to understand the underlying factors behind this phenomenon through extensive experiments. However, it is still unknown whether a multi-epoch training paradigm could achieve better results, as the best performance is usually achieved by one-epoch training. In this paper, we hypothesize that the emergence of this phenomenon may be attributed to the susceptibility of the embedding layer to overfitting, which can stem from the high-dimensional sparsity of data. To maintain feature sparsity while simultaneously avoiding overfitting of embeddings, we propose a novel Multi-Epoch learning with Data Augmentation (MEDA), which can be directly applied to most deep CTR models. MEDA achieves data augmentation by reinitializing the embedding layer in each epoch, thereby avoiding embedding overfitting and simultaneously improving convergence. To our best knowledge, MEDA is the first multi-epoch training paradigm designed for deep CTR prediction models. We conduct extensive experiments on several public datasets, and the effectiveness of our proposed MEDA is fully verified. Notably, the results show that MEDA can significantly outperform the conventional one-epoch training. Besides, MEDA has exhibited significant benefits in a real-world scene on Kuaishou.

### **13. Multi-View Interactive Collaborative Filtering**

Maria Lentini, Umashanger Thayasivam

https://arxiv.org/abs/2305.18306

In many scenarios, recommender system user interaction data such as clicks or ratings is sparse, and item turnover rates (e.g., new articles, job postings) high. Given this, the integration of contextual "side" information in addition to user-item ratings is highly desirable. Whilst there are algorithms that can handle both rating and contextual data simultaneously, these algorithms are typically limited to making only in-sample recommendations, suffer from the curse of dimensionality, and do not incorporate multi-armed bandit (MAB) policies for long-term cumulative reward optimization. We propose multi-view interactive topic regression (MV-ICTR) a novel partially online latent factor recommender algorithm that incorporates both rating and contextual information to model item-specific feature dependencies and users' personal preferences simultaneously, with multi-armed bandit policies for continued online personalization. The result is significantly increased performance on datasets with high percentages of cold-start users and items.

### **14. A Survey on Fairness-aware Recommender Systems**

Di Jin, Luzhi Wang, He Zhang, Yizhen Zheng, Weiping Ding, Feng Xia, Shirui Pan

https://arxiv.org/abs/2306.00403

As information filtering services, recommender systems have extremely enriched our daily life by providing personalized suggestions and facilitating people in decision-making, which makes them vital and indispensable to human society in the information era. However, as people become more dependent on them, recent studies show that recommender systems potentially own unintentional impacts on society and individuals because of their unfairness (e.g., gender discrimination in job recommendations). To develop trustworthy services, it is crucial to devise fairness-aware recommender systems that can mitigate these bias issues. In this survey, we summarise existing methodologies and practices of fairness in recommender systems. Firstly, we present concepts of fairness in different recommendation scenarios, comprehensively categorize current advances, and introduce typical methods to promote fairness in different stages of recommender systems. Next, after introducing datasets and evaluation metrics applied to assess the fairness of recommender systems, we will delve into the significant influence that fairness-aware recommender systems exert on real-world industrial applications. Subsequently, we highlight the connection between fairness and other principles of trustworthy recommender systems, aiming to consider trustworthiness principles holistically while advocating for fairness. Finally, we summarize this review, spotlighting promising opportunities in comprehending concepts, frameworks, the balance between accuracy and fairness, and the ties with trustworthiness, with the ultimate goal of fostering the development of fairness-aware recommender systems.

### **15. Graph Exploration Matters: Improving both individual-level and  system-level diversity in WeChat Feed Recommender**

Shuai Yang, Lixin Zhang, Feng Xia, Leyu Lin

https://arxiv.org/abs/2306.00009

There are roughly three stages in real industrial recommendation systems, candidates generation (retrieval), ranking and reranking. Individual-level diversity and system-level diversity are both important for industrial recommender systems. The former focus on each single user's experience, while the latter focus on the difference among users. Graph-based retrieval strategies are inevitably hijacked by heavy users and popular items, leading to the convergence of candidates for users and the lack of system-level diversity. Meanwhile, in the reranking phase, Determinantal Point Process (DPP) is deployed to increase individual-level diverisity. Heavily relying on the semantic information of items, DPP suffers from clickbait and inaccurate attributes. Besides, most studies only focus on one of the two levels of diversity, and ignore the mutual influence among different stages in real recommender systems. We argue that individual-level diversity and system-level diversity should be viewed as an integrated problem, and we provide an efficient and deployable solution for web-scale recommenders. Generally, we propose to employ the retrieval graph information in diversity-based reranking, by which to weaken the hidden similarity of items exposed to users, and consequently gain more graph explorations to improve the system-level diveristy. Besides, we argue that users' propensity for diversity changes over time in content feed recommendation. Therefore, with the explored graph, we also propose to capture the user's real-time personalized propensity to the diversity. We implement and deploy the combined system in WeChat App's Top Stories used by hundreds of millions of users. Offline simulations and online A/B tests show our solution can effectively improve both user engagement and system revenue.

### **16. Robust Reinforcement Learning Objectives for Sequential Recommender  Systems**

Melissa Mozifian, Tristan Sylvain, Dave Evans, Lili Meng

https://arxiv.org/abs/2305.18820

Attention-based sequential recommendation methods have demonstrated promising results by accurately capturing users' dynamic interests from historical interactions. In addition to generating superior user representations, recent studies have begun integrating reinforcement learning (RL) into these models. Framing sequential recommendation as an RL problem with reward signals, unlocks developing recommender systems (RS) that consider a vital aspect-incorporating direct user feedback in the form of rewards to deliver a more personalized experience. Nonetheless, employing RL algorithms presents challenges, including off-policy training, expansive combinatorial action spaces, and the scarcity of datasets with sufficient reward signals. Contemporary approaches have attempted to combine RL and sequential modeling, incorporating contrastive-based objectives and negative sampling strategies for training the RL component. In this study, we further emphasize the efficacy of contrastive-based objectives paired with augmentation to address datasets with extended horizons. Additionally, we recognize the potential instability issues that may arise during the application of negative sampling. These challenges primarily stem from the data imbalance prevalent in real-world datasets, which is a common issue in offline RL contexts. While our established baselines attempt to mitigate this through various techniques, instability remains an issue. Therefore, we introduce an enhanced methodology aimed at providing a more effective solution to these challenges.

### **17. Towards Explainable Conversational Recommender Systems**

Shuyu Guo, Shuo Zhang, Weiwei Sun, Pengjie Ren, Zhumin Chen, Zhaochun Ren

https://arxiv.org/abs/2305.18363

Explanations in conventional recommender systems have demonstrated benefits in helping the user understand the rationality of the recommendations and improving the system's efficiency, transparency, and trustworthiness. In the conversational environment, multiple contextualized explanations need to be generated, which poses further challenges for explanations. To better measure explainability in conversational recommender systems (CRS), we propose ten evaluation perspectives based on concepts from conventional recommender systems together with the characteristics of CRS. We assess five existing CRS benchmark datasets using these metrics and observe the necessity of improving the explanation quality of CRS. To achieve this, we conduct manual and automatic approaches to extend these dialogues and construct a new CRS dataset, namely Explainable Recommendation Dialogues (E-ReDial). It includes 756 dialogues with over 2,000 high-quality rewritten explanations. We compare two baseline approaches to perform explanation generation based on E-ReDial. Experimental results suggest that models trained on E-ReDial can significantly improve explainability while introducing knowledge into the models can further improve the performance. GPT-3 in the in-context learning setting can generate more realistic and diverse movie descriptions. In contrast, T5 training on E-ReDial can better generate clear reasons for recommendations based on user preferences. E-ReDial is available at https://github.com/Superbooming/E-ReDial

### **18. Choosing the Right Weights: Balancing Value, Strategy, and Noise in  Recommender Systems**

Smitha Milli, Emma Pierson, Nikhil Garg

https://arxiv.org/abs/2305.17428

Many recommender systems are based on optimizing a linear weighting of different user behaviors, such as clicks, likes, shares, etc. Though the choice of weights can have a significant impact, there is little formal study or guidance on how to choose them. We analyze the optimal choice of weights from the perspectives of both users and content producers who strategically respond to the weights. We consider three aspects of user behavior: value-faithfulness (how well a behavior indicates whether the user values the content), strategy-robustness (how hard it is for producers to manipulate the behavior), and noisiness (how much estimation error there is in predicting the behavior). Our theoretical results show that for users, upweighting more value-faithful and less noisy behaviors leads to higher utility, while for producers, upweighting more value-faithful and strategy-robust behaviors leads to higher welfare (and the impact of noise is non-monotonic). Finally, we discuss how our results can help system designers select weights in practice.