

Jiarui Jin — Research Summary

✉ jinjiarui97@sjtu.edu.cn • 🌐 jinjiarui.github.io

Short Introduction

My academic background and industrial internships mainly focus on information retrieval and conversational intelligence. My current endeavors and past achievements revolve around the following key directions:

- ✎ *How to capture **interactive** patterns between pairs of sequence data?* Instead of compressing the sequential information of users and items into a single vector before making predictions, we propose GraphHINGE [KDD 2020] (📄 **Paper**) [TOIS 2022] (📄 **Paper**) that designs a convolutional block with fast Fourier transform to get the interactive patterns between user sequence and item sequence (and sequences are defined by metapaths) in item recommendations; and TWINS [WWW 2022] (📄 **Paper**) that mines the interactions between user sequence and anchor sequence (and sequences contains user browsed items and anchor broadcast items) in anchor recommendations. TWINS has been deployed on Diantao anchor recommendation platform.
- ✎ *How to mine **relevant** patterns in (long) sequence historical data?* Long sequence issue occurs in sequential recommendations as recurrent units are always forgetting long-term message. We propose STARec [WWW 2022] (📄 **Paper**) that develops a retrieval component to extract relevant browsed items and designs a label trick treating user previous responses as input features. STARec has been deployed on China Merchants Bank item recommendation platform.
- ✎ *How to **debias** behaviors in sequence data?* Position bias always encourages users to click and purchase high-ranked items, optimizing the recommender systems in terms of biased clicks instead of unbiased relevance. We propose DRSR [SIGIR 2020] (📄 **Paper**) that incorporates survival analysis technique into recurrent units to derive true relevance; and InfoRank [Preprint 2023] (📄 **Paper**) that minimize the mutual information between observation estimation and relevance estimation to push the estimated relevance free of effects from high-ranked positions and high popularity. We further extend DRSR to HEROES [CIKM 2022] (📄 **Paper**) that jointly optimizes click-through rate and conversation rate predictions.
- ✎ *How to empower offline recommender systems with **online conversational agents**?* We propose CORE [NeurIPS 2023] (📄 **Project**), a new offline-training and online-checking paradigm that can bridge any pre-trained conversational agent (e.g., gpt-3.5-turbo) and any recommendation platforms (i.e., any recommender systems supporting either tubular data or sequence data). We also design MAGUS [Preprint 2023] (📄 **Paper**) that combines query recommendations and item recommendations into a multi-round guess-and-update system, which can be regarded as a downgraded version of free-text conversational agents.
- ✎ *How to leverage language models to **manage domain-specific database**?* We propose GUNDAM [Preprint 2023] (📄 **Project**), a data manager that uses language models to effectively select informative and personalized data from the whole data corpus and leave other common knowledge stored in pre-trained language models.

All the above publications are my first-authored papers.

We draw an illustration for my research topic in Figure 1.

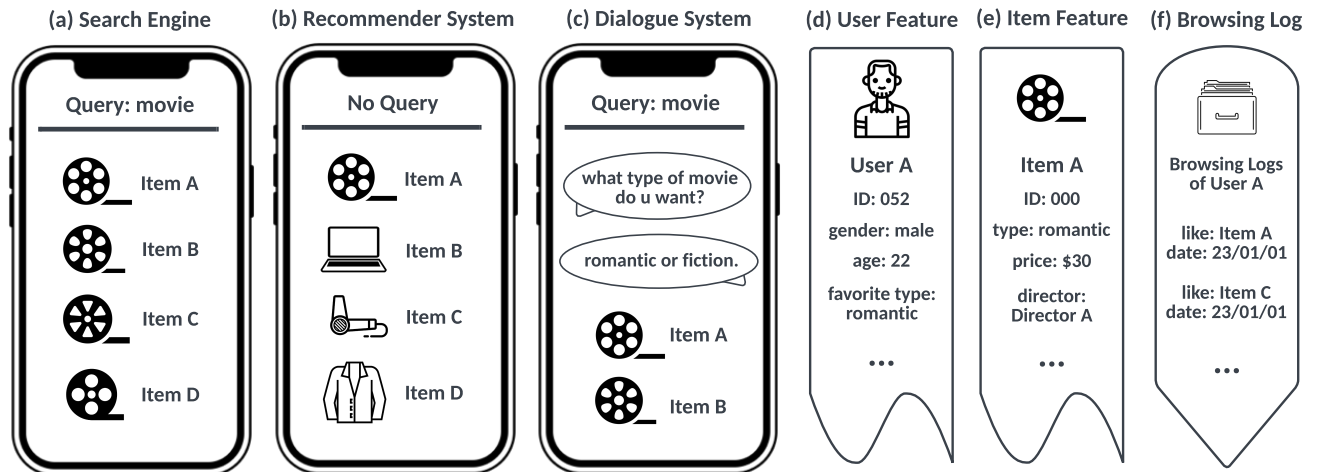










Figure 1. An illustrated example of my research topics: how to mine static features of users and items, and dynamic browsing logs of users to benefit search engines, (item) recommender systems, and dialogue systems (a.k.a., conversational recommender systems).

Mining Interactive Patterns in Sequence Data




Classical recommendation approaches based on tabular data, such as SVD++ [KDD 2008] ( **Paper**) and FM [ICDM 2010] ( **Paper**), formulate the interactions between the user’s historical records and the target item. These interactive patterns are also known as “AND” (e.g., if a user is Chinese AND the date is Chinese New Year, then she is likely to buy dumplings).

In GraphHINGE (abbr. for **H**eterogeneous **I**nteract and **aggreG**atE) [KDD 2020] ( **Paper**) [TOIS 2022] ( **Paper**), we use the predefined metapaths to sample the sequences of users and items. We design a convolutional block with fast Fourier transform to get interactions before aggregating the representations into final prediction scores for item recommendations. In TWINS (abbr. for **T**wo-side **I**nteractive **N**etworkS) [WWW 2022] ( **Paper**), we first summarize user browsed items into user sequences and anchor broadcast items into anchor sequences, and then conduct interaction operations between two sequences. TWINS is verified by online A/B tests and has been deployed on Diantao anchor recommendation platform.

Mining Relevant Patterns in Sequence Data

As pointed out in SIM [CIKM 2020] ( **Paper**), most sequential models can only perform well over sequence data with length scaling up to 1000. We propose STARec (abbr. for **S**earch-based **T**ime-Aware **R**ecommendations) [WWW 2022] ( **Paper**) that extracts relevant user-browsed items with respect to specific candidate items, mixing the retrieved items together with the user’s recently browsed ones into a time-aware sequential network considering not only the sequence orders but also the time-intervals. STARec can be simply deployed into any sequential network such as LSTM [Neural Computation 1997] ( **Paper**). STARec is verified by online A/B tests and has been deployed on China Merchants Bank item recommendation platform.

Mining Unbiased Patterns in Sequence Data

Most unbiased ranking methods, as summarized in ULTRA [CIKM 2021] ( **Paper**), are designed for tabular data. We propose DRSR (abbr. for **D**eep **R**ecurrent **S**urvival **R**anking) [SIGIR 2020] ( **Paper**) that combines survival analysis technique into the recurrent units to derive the relevance probability of each item (defined as the click probability of each item given that it has been observed). As DRSR models the relevance of click-through rate, we further extend it into HEROES (abbr. for **H**ierarchical **r**Ecurrence **R**anking **O**n the **E**ntire **S**pace) [CIKM 2022] ( **Paper**) to jointly optimize click-through rate and conversation predictions. Both DRSR and HEROES are focusing on position bias where high-ranked items are likely to receive user attention and then get clicks and purchases. Similarly,

high-popular items are also likely to have high exposure and receive user positive feedback, which is known as popularity bias. To this end, we propose InfoRank (abbr. for Unbiased **R**anking via Mutual **I**nformation Minimization) [Preprint 2023] ([📄 Paper](#)), a simple yet sufficient unbiased learning-to-rank paradigm, which first summarizes the impacts of both biases into a single observation factor, therefore providing a unified treatment of the bias problem. We then minimize the mutual information between the observation estimation and the relevance estimation to make relevance estimation free of bias.

📌 **Bridging Online Conversational Agents and Offline Recommender Systems**

Unlike prior conversational recommendation approaches, such as EAR [WSDM 2020] ([📄 Paper](#)), which systemically combine conversational and recommender parts through a reinforcement learning framework, we propose CORE (abbr. for **C**onversational agents for **R**ecommender systems) [NeurIPS 2023] ([📄 Paper](#)), a new offline-training and online-checking paradigm that bridges any conversational agent (e.g., gpt-3.5-turbo) and any recommendation platforms (i.e., any recommender systems supporting either tubular data or sequence data) in a plug-and-play style. Here, CORE treats a recommender system as an offline relevance score estimator, while a conversational agent is regarded as an online relevance score checker to check these estimated scores. We define uncertainty as the summation of unchecked relevance scores and correspondingly develop an online decision tree algorithm to decide what to query at each turn. We release a continuously updating project ([📄 Project](#)) to support more design choices for conversational agents and recommender systems. We also design MAGUS (abbr. for **M**ulti-round **A**uto **G**uess-and-**U**pdate **S**ystem) [Preprint 2023] ([📄 Paper](#)) organizing a multi-round guess-and-update system that can be applied to any recommender system to allow the recommendation of both queries and items. In this regard, MAGUS can be regarded as a simplified version of free-text conversational agents.

📌 **Managing Offline Data via Pre-trained Language Models**

(Large) language models achieve remarkable few-shot performance by simply prompting language models with a few data points (a.k.a., demonstrations) as the input context, without the need for computationally expensive fine-tuning. With the help of any pre-trained language model, our database only needs to store a subset of high-quality data instead of the whole data corpus, as the rest data could be inferred by the subset. To this end, we propose GUNDAM (abbr. for **G**olden **p**lUg-i**N** **D**Ata **M**anager) [Preprint 2023] ([📄 Paper](#)), a novel data-centric framework that measures sufficiency and necessity of storing each data point conditioned on language models. The proposed sufficiency and necessity metrics can be operated on both demonstration instances (i.e., instance level) and demonstration sets (i.e., set level). Considering that the database would keep growing in many real-world scenarios, we develop an incremental update algorithm to avoid re-computing GUNDAM over all the changed and unchanged parts. We release a continuously updating project ([📄 Project](#)) to support more design choices for language models and demonstration selection approaches.

Short Conclusion

For ease of connecting the above papers and projects, we draw a map containing all the papers and projects mentioned above in Figure 2.

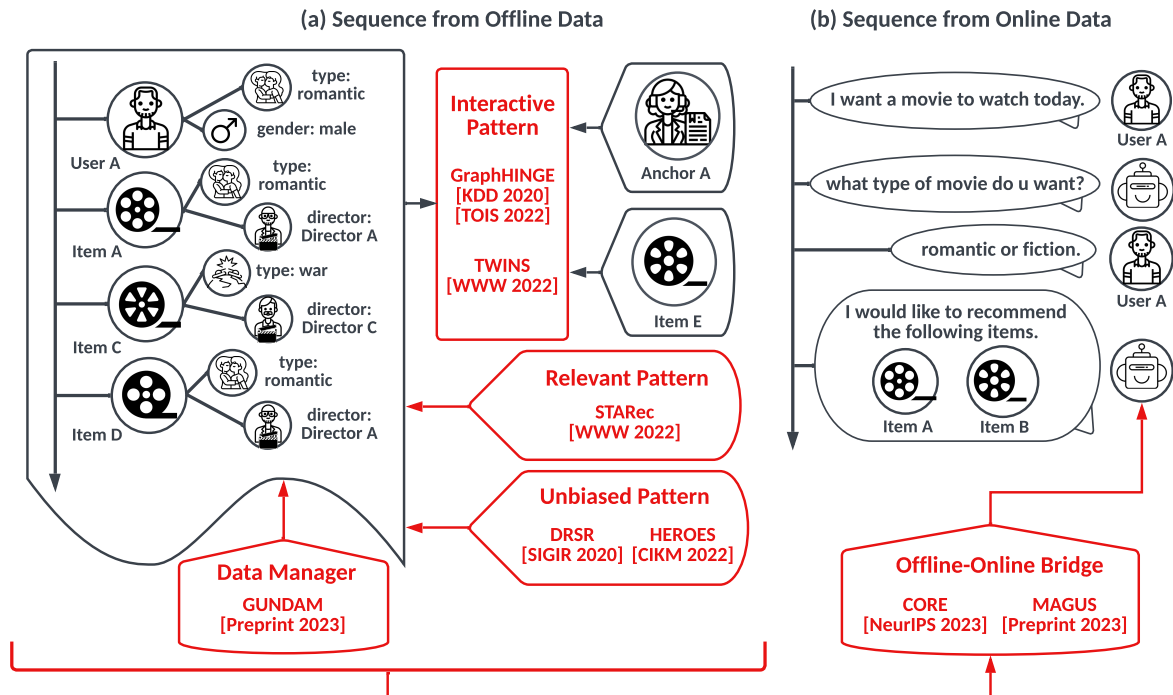


Figure 2. A map illustrating how my contributions benefit the recommender systems.

My Contributions with Short Summary

- Lending Interaction Wings to Recommender Systems with Conversational Agents**
Jiarui Jin, Xianyu Chen, Fanghua Ye, Mengyue Yang, Yue Feng, Weinan Zhang, Yong Yu, Jun Wang.
 The 37th Conference on Neural Information Processing Systems.
 NeurIPS 2023. ([Project](#)) ([Document](#)) ([Paper](#)) ([Code](#))
 TLDR: We propose CORE, a plug-and-play conversational agent allowing any recommender systems to online query user preferences.
- Multi-Scale User Behavior Network for Entire Space Multi-Task Learning**
Jiarui Jin, Xianyu Chen, Weinan Zhang, Yuanbo Chen, Zaifan Jiang, Zekun Zhu, Zhewen Su, Yong Yu.
 The 31st ACM International Conference on Information and Knowledge Management.
 CIKM 2022. ([Paper](#)) ([Code](#)) ([Slides](#))
 TLDR: We propose HEROES, a hierarchical recurrent network deriving the relevance probability of click-through rate and conversation predictions.
- Learn over Past, Evolve for Future: Search-based Time-aware Recommendation with Sequential Behavior Feedback Data**
Jiarui Jin, Xianyu Chen, Weinan Zhang, Junjie Huang, Ziming Feng, Yong Yu.
 The 2022 World Wide Web Conference.
 WWW 2022. ([Paper](#)) ([Code](#)) ([Slides](#))
Real-world Deployment at China Merchants Bank Platform
 TLDR: We propose STARec, an easy-to-implement framework, incorporating a retrieval module into any sequential network to handle long sequence behavioral data.
- Who to Watch Next: Two-side Interactive Networks for Live Broadcast Recommendation**
Jiarui Jin, Xianyu Chen, Yuanbo Chen, Weinan Zhang, Renting Rui, Zaifan Jiang, Zhewen Su, Yong Yu.
 The 2022 World Wide Web Conference.
 WWW 2022. ([Paper](#)) ([Code](#)) ([Slides](#))
Real-world Deployment at Taobao Diantao Platform

TLDR: We propose TWINS, a two-side interactive network for anchor recommendations.

- **GraphHINGE: Learning Interaction Models of Structured Neighborhood on Heterogeneous Information Network**
Jiarui Jin, Kounianhua Du, Weinan Zhang, Jiarui Qin, Yuchen Fang, Yong Yu, Alexander J. Smola.
Transactions on Information Systems (Special Issue on Graph Technologies for User Modeling and Recommendation)
TOIS 2022. ([Paper](#)) ([Code](#))
TLDR: We propose GraphHINGE, a convolutional block to operate interactions among sequences.
- **An Efficient Neighborhood-based Interaction Model for Recommendation on Heterogeneous Graph**
Jiarui Jin, Jiarui Qin, Yuchen Fang, Kounianhua Du, Weinan Zhang, Yong Yu, Zheng Zhang, Alexander J. Smola.
The 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
KDD 2020. ([Paper](#)) ([Code](#)) ([Slides](#))
AWS Machine Learning Research Project Award
TLDR: We propose GraphHINGE, a convolutional block to operate interactions among sequences.
- **A Deep Recurrent Survival Model for Unbiased Ranking**
Jiarui Jin, Yuchen Fang, Weinan Zhang, Kan Ren, Guorui Zhou, Jian Xu, Yong Yu, Jun Wang, Xiaoqiang Zhu, Kun Gai.
The 43rd ACM SIGIR International Conference on Research and Development in Information Retrieval.
SIGIR 2020. ([Paper](#)) ([Code](#)) ([Slides](#))
TLDR: We propose DRSR, incorporating survival analysis techniques into recurrent units for unbiased learning-to-rank.
- **Mind Your Plug-in Data Quality for Language Models: A Data-Centric Approach**
Jiarui Jin, Yuwei Wu, Mengyue Yang, Xiaoting He, Weinan Zhang, Yiming Yang, Yong Yu, Jun Wang.
Preprint 2023. ([Project](#)) ([Document](#)) ([Paper](#)) ([Code](#))
TLDR: We propose GUNDAM, a data manager, extracting high-quality data points conditioned on a given pre-trained language model.
- **InfoRank: Unbiased Learning-to-Rank via Conditional Mutual Information Minimization**
Jiarui Jin, Zexue He, Mengyue Yang, Weinan Zhang, Yong Yu, Jun Wang, Julian McAuley
Preprint 2023. ([Paper](#))
TLDR: We propose InfoRank that summarizes the effect of multiple biases into a single observation factor, therefore providing a unified treatment of the bias problem.
- **Bridging Query Completion and Item Recommendation in a Multi-Round Guess-and-Update System**
Jiarui Jin, Xianyu Chen, Weinan Zhang, Yong Yu, Jun Wang
Preprint 2023. ([Paper](#))
TLDR: We propose MAGUS, an easy-to-implement framework that could be applied to any recommender system to enable the recommendation of both queries and items.

Contributions from Others

- **Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model**
Yehuda Koren
The 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
KDD 2008. ([Paper](#))
- **Factorization Machines**
Steffen Rendle

The 10th IEEE International Conference on Data Mining.
ICDM 2010. (Paper)

- **Search-based User Interest Modeling with Lifelong Sequential Behavior Data for Click-Through Rate Prediction**
Pi Qi, Xiaoqiang Zhu, Guorui Zhou, Yujing Zhang, Zhe Wang, Lejian Ren, Ying Fan, Kun Gai
The 29th ACM International Conference on Information and Knowledge Management.
CIKM 2020. (Paper)
- **Long Short-Term Memory**
Sepp Hochreiter, Jürgen Schmidhuber
Neural Computation 1997. (Paper)
- **ULTRA: An Unbiased Learning To Rank Algorithm Toolbox**
Anh Tran, Tao Yang, Qingyao Ai
The 30th ACM International Conference on Information and Knowledge Management.
CIKM 2021. (Paper)
- **Estimation–Action–Reflection: Towards Deep Interaction Between Conversational and Recommender Systems**
Wenqiang Lei, Xiangnan He, Yisong Miao, Qingyun Wu
The Thirteenth ACM International Conference on Web Search and Data Mining.
WSDM 2020. (Paper)