Multimodal Pre-training and Generation for Recommendation

Jieming Zhu
Huawei Noah's Ark Lab
jiemingzhu@ieee.org

Chuhan Wu
Huawei Noah's Ark Lab
wuchuhan1@huawei.com

Rui Zhang
www.ruizhang.info
rayteam@yeah.net

Zhenhua Dong
Huawei Noah's Ark Lab
dongzhenhua@huawei.com

Outline of Tutorial

- **Introduction**
- Multimodal Pre-training for Recommendation
- Multimodal generation for personalization
- Practices and Open Challenges in Multimodal Recommendation

Motivation of tutorial

- Classic recommendation models are trained by ID feature, categorical features, which are good at modeling the collaborative signals, but they always overlook the raw contents across multiple modalities such as text, image, audio and video.
- The recent advancement in pre-trained and generation multimodal models, like LLM, CLIP, ChatGPT, DALL.E, offer new opportunities in understanding item and user, and developing better recommender systems.
- We would like to share our research, survey and industrial practices in multimodal recommendations and propose challenging questions:
  - *How to enhance recommendation with multimodal pre-training technologies?*
  - *How to align and fuse the user preference modality to other content modality?*
  - *How to generate the personalized content for each individual user?*
  - *How to apply the multimodal technologies in recommender system?*

Main content and speakers

- Introduction and challenges in industrial recommender systems

By Dr. Zhenhua Dong (30 mins)

- Multimodal Pre-training and its applications in recommendation by Dr. Jieming Zhu (45mins)

- Multimodal generation for personalization by Prof. Rui Zhang (45mins)

- Practices about multimodal recommendation in products by Dr. Chuhan Wu (45mins)

Targeted audiences

- Researchers and practitioners in multimodal learning: the tutorial offers the insights into how to integrate multimodal technologies into recommender system, like practices and challenges.
- Researchers and practitioners in recommender system: the tutorial introduces the knowledge about the recent and prospective progress in multimodality and generation technologies, and how to apply them to enhance the recommendation.

About us

- Huawei’s vision: *bring digital to every person, home and organization for a fully connected, intelligent world.*
- Huawei Noah’s ark lab: *Building an intelligent world*
    - 7 labs: Computer vision, decision making & reasoning, AI theory, speech and language processing, recommendation & search, AI system, AI application
    - World wide labs: China, Singapore, U.K., France, Canada, Russia
- Recommendation & search research lab: *gets the right information to the right people*
    - Academic research and industrial practice are two wheels of horse drawn carriage[1]
    - Collaboration with product team: advanced AI for products, practical scenario for RQs
    - Collaboration with academia: learn from the best

[1] A Brief History of Recommender Systems, Zhenhua Dong et. al.

1. Industry practices

- Various recommendation scenarios, serving hundreds of millions of users in each day

Product	Scenario
App gallery	App and game recommendation & search
Instant service	Service recommendation
Ads. Platform	CTR/CVR prediction
Browser	News recommendation, search ads.
Music	Songs recommendation & search
Education	Lessons and learning method recommendation
Theme	Theme recommendation and search
GTS	Cases recommendation and search
Internal IT/HR system	Document and staff recommendation and search

2. Impactful research (10000+ citations):

- Recommendation model structure: *DeepFM, PNN, AutoML4RecSys*
- Causal Recommendation: *Counterfactual/intervention recommendations, de-bias*
- LLM4Rec: *NOVA-BERT, LLM4CTR, Survey*
- Benchmark: *FuxiCTR, BARS, SimpleX, REASONER*

Based on the great **missions**, **opportunities** for practices and our research **experiences**, we summarize 10 challenges of industrial recommender system.

Problems
1. Missing information
2. Individual treatment effect
3. Biases

Methods
4. Models reusing
5. Large language model enhancement
6. Multiple modalities
7. Simulations

Goals
8. Lifetime value modeling
9. Trustworthy
10. Win-win ecosystem

1. Missing information

- Research question (RQ): How to handle the missing information in recommender system?
  - Missing features (column data)
    - RecSys may miss some information such as item’s popularity, user’s thought about the item.
    - RecSys may don’t know the causal features: user watched a movie for her friend’s suggestion.
  - Missing samples (raw data)
    - RecSys exposes only a few items of all items in one interaction.
    - RecSys can’t collect a user’s behaviors on the items in other systems.
- Solutions:
  - Counterfactual learning [1].
  - Predict the missing features.
  - What else?

[1] Counterfactual learning for recommender system, RecSys20, Zhenhua Dong et. al.

2. Individual treatment effect

- RQ: How to find the causal attributes of a user's decision (e.g. click, rate) or preference?
- Causal attributes:
  - Example: both user A and B rate one movie C 5 star, A likes C for director, B likes C for cast. The same attribute or treatment (director/cast) may have different effects on user's decision, so it is individual treatment effect (ITE).
  - Accurate causal attribute can help in user profiling, explanation, accuracy.
- Solution:
  - Conditional counterfactual causal effect[1].
  - More novel methods for computing ITE under more generalized assumptions.

[1] Conditional counterfactual causal effect for individual attribution, UAI23, Ruiqi Zhao et. al.

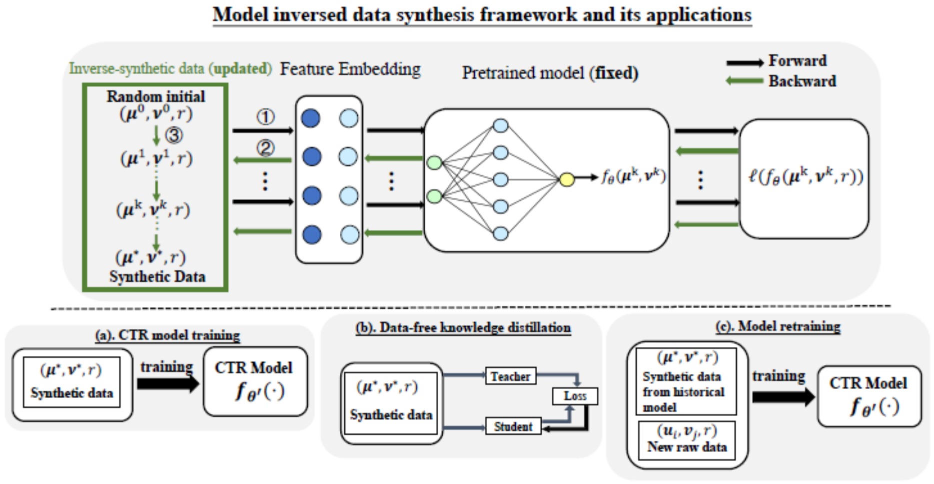
3. Biases

- RQ: How to mitigate the biases in recommender systems?
- There are so many biases [1] in recommender systems due to missing information, confounders and closed feedback loop.
- Solution:
  - Causal analysis: potential outcome model [2], structural causal model
  - Inverse propensity score, direct methods, doubly robust methods
- More research opportunities:
  - How to handle new biases such as duration bias, trust bias, confounder bias?
  - How to collect unbiased data?
  - How to train an unbiased model?
  - How to evaluate the biases?

[1] Workshop keynote -- How to De-Bias for Industrial Recommender System? A causal Perspective, SIGIR21, Zhenhua Dong
[2] On the Opportunity of Causal Learning in Recommendation Systems: Foundation, Estimation, Prediction and Challenges, IJCAI22, Peng Wu et. al.

4. Model reusing

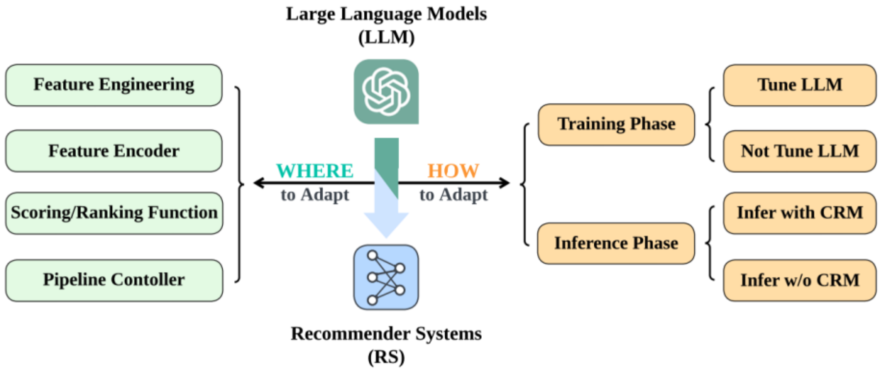
- RQ: How to reuse the historical models efficiently?
- The industrial recommendation models should be updated with recent data for better performance, but the historical models are always underutilized.
- Solutions:
  - Online learning, ensemble learning.
  - Model inversed data synthesis framework [1].
- More research opportunities:
  - Machine unlearning, learnware, etc.
  - Can we train large recommendation model like large language model?



[1] Data-free Knowledge Distillation for Reusing Recommendation Models , RecSys23, Cheng Wang et. al.

5. Large language model enhanced recommendation

- RQ: ChatGPT has demonstrated great capabilities of LLMs, how to improve recommendation with large language models?
- Solutions[1]:
  - Where: feature, embedding, prediction, controller
  - How: fine tune, collaboration with classic models



- Let us embrace LLM:

"It does not matter if you love it or not
It is standing right there
With no emotion
Not going to change" By Sangs-Rgyas Rgya Mtsho

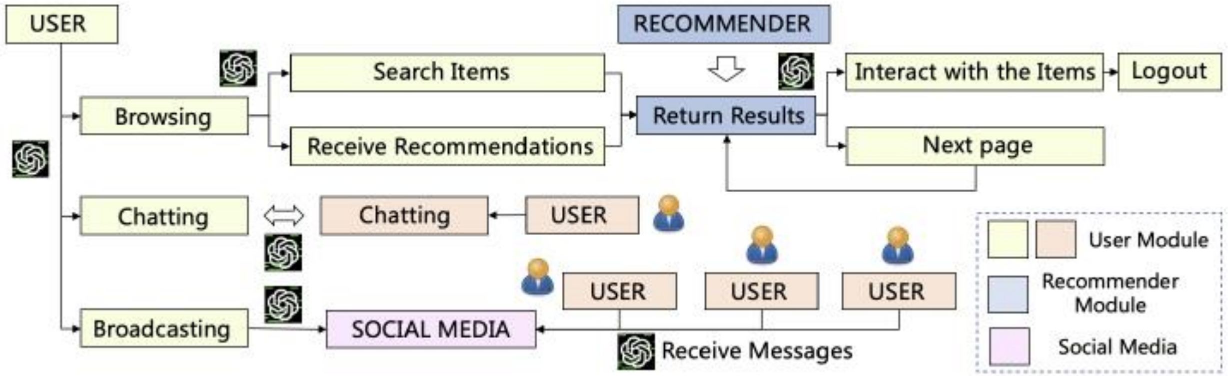
[1] How Can Recommender Systems Benefit from Large Language Models: A Survey, Jianghao Lin et. al.

6. Multiple modalities

- RQ: How to align the user preference modality to content modality?
  - Classic recommendation train models with preference modality data like user behavior, which is good at modeling the collaborative signals.
  - Content modalities like text, image and video are good at content semantic understanding.
  - There is huge gap between preference modality and content modalities leads to minor improvement in recommendation.
- Solution:
  - Deeply understanding about the different modalities and their relation to user's preferences and decisions.
  - Recommendation focused multi-modal pre-trained model that bridges the gap.

7. Simulation

- RQ: How to model user preference with simulation?
- User modeling is a classical research topic
  - ACM UMAP is 31 years old, many research topics have been studied.
  - User modeling is still hard since people and context are complex, e.g. in some ads. scenarios, the CTR is less than 1%.
- Solution: RecAgent [1]: digital twin of recommender system, simulate user behaviors, and align with real human understanding.



- More opportunities:
  - How to simulate more users and more behaviors efficient and accurately?
  - How to evaluate the simulation ?

[1] RecAgent: A Novel Simulation Paradigm for Recommender Systems, Lei Wang et. al.

8. Lifetime value modeling

- RQ: How to predict users' long term satisfaction?
- Most recommendation studies focus on optimizing short term objectives like click, rating, dwell time, which can not align the goal to improve user's long term satisfaction like deep conversion task.
- Solution: in the tutorial[1], we introduce the definitions and scenarios of LTV, some typical LTV prediction technique, and products practices.
- There are still many hard problems:
  - Delayed and sparse feedback.
  - Cold start, offline evaluation, multi-task optimizations.

[1] Tutorial -- Customer Lifetime Value Prediction: Towards the Paradigm Shift of Recommender System Objectives, RecSys23, Chuhan Wu et. al.

9. Trustworthy

- RQ: How to build trustworthy recommender systems?
- We consider 8 perspectives such as accountability, security, fairness [1], privacy, robustness, transparency, assisting or serving people, and long term enhancement of the happiness of human, society and environment.
- Most current recommender system focus on the accuracy such as AUC, Logloss, CTR, which is not enough to be a trustworthy recommender system.
- We hope more scholars can help industry to define and build trustworthy and social good recommender systems from wider perspectives such as society, economy, user-centric, ecological systems and natural environments.

[1] Workshop keynote --Two perspectives about biases in recommender system: OoD and unfairness, ICDM2023, Zhenhua Dong

10. Win-win ecosystem

- RQ: How to satisfy multi-stakeholders in dialog based IR or RecSys?
- Mainly 4 kinds stakeholders: user, content provider(CP), information system and advertisers.
- Dialog based IR or RecSys can directly providing answer or satisfying users' request. However, there are key challenges for each stakeholder:
  - Content provider: How to protect their intellectual property and benefits?
  - Advertisers: How to appropriately expose the Ads. during the dialog?
  - Users: How to ensure the generated information is objective and accountable?
  - Information system: How to design win-win interactions and mechanisms for long-term benefits?

Based on the great **missions**, **opportunities** for practices and our research **experiences**, we summarize 10 challenges of industrial recommender system.

Problems

1. Missing information
2. Individual treatment effect
3. Biases (SIGIR21)

Methods

4. Models reusing
5. Large language model enhancement
6. Multiple modalities (Web2024)
7. Simulations

Goals

8. Lifetime value modeling (RecSys23)
9. Trustworthy (ICDM23)
10. Win-win ecosystem

We desire to collaborate with the passionate, talented people and change world together!

noahlab1@huawei.com
dongzhenhua@huawei.com

WeChat