Density-based User Representation using Gaussian Process Regression for Multi-interest Personalized Retrieval

Presenter: Haolun Wu*

Co-authors:

Ofer Meshi, Masrour Zoghi, Fernando Diaz, Steve Liu, Craig Boutilier, Maryam Karimzadehgan

** Work done while doing an internship at Google.*

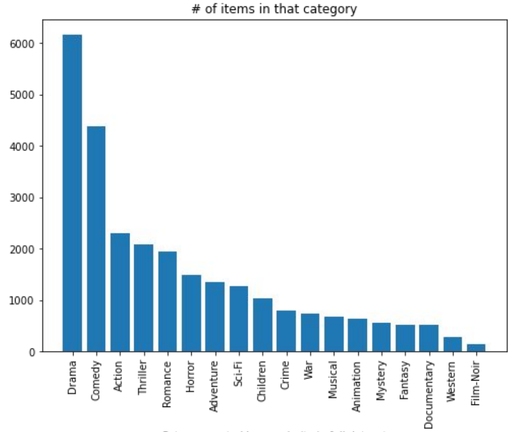
Personalization in Retrieval and Recommendation

- Personalization plays an important role in user **satisfaction**
- Requires to capture users' **multiple interests**

- **Challenges:**
  - Users have **diverse and volatile interests**
  - Hard to retrieve items from **niche interests**

- **Our goal: find good user representation that can capture multiple interests**

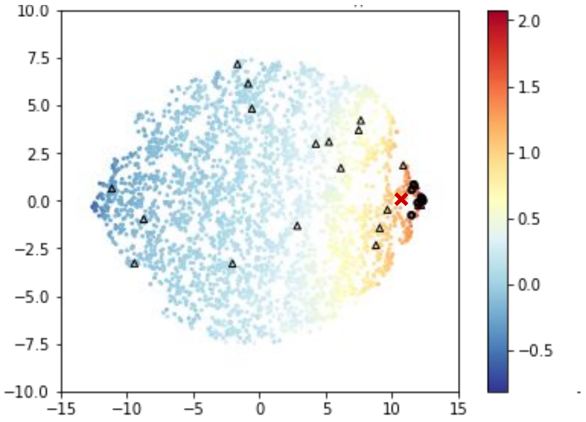
Source: Cen et al. Controllable Multi-Interest Framework for Recommendation. KDD'20.



Previous Solutions on User Representation: Point-based Representation

- Single-point User Representation (SUR)
    - Fails to cover multi-interest (unless using a very high-dimension vector)

user id: 301



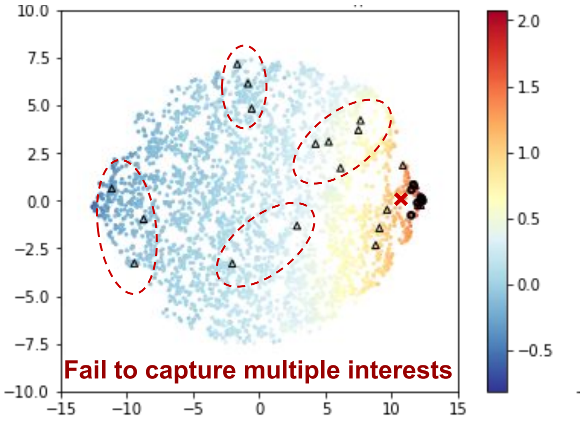
Visualization:
Learned scores on all items for a user – using SUR

Dataset: MovieLens 1M
Original dimension size: 64
Reduce to dim=2 for 2D visualization

Previous Solutions on User Representation: Point-based Representation

- Single-point User Representation (SUR)
  - Fails to cover multi-interest (unless using a very high-dimension vector)

user id: 301



Visualization:
Learned scores on all items for a user – using SUR

Dataset: MovieLens 1M
Original dimension size: 64
Reduce to dim=2 for 2D visualization

Previous Solutions on Multi-interest User Representation: Point-based Representation

- Multi-point User Representation (MUR)
  - How to choose # of points (*K*)?
    - Pre-define *K* for all users. E.g., *K=4*. *MaxMF [RecSys'13], PolyDeepWalk [KDD'19], ComiRec [KDD'20], SINE [WSDM'21], PIMI [IJCAI'21]*
    - Other heuristic rule. $K = \log_2(|\mathcal{I}_u|)$ *MIND [CIKM'19]*
    - Use the Ward clustering algorithm per-user. *PinnerSage [KDD'20]*
  - Does not model uncertainty.
    - $f(u, v) = \max_{k=1,\ldots,K} \mathbf{u}^k \cdot \mathbf{v}$ *MaxMF [RecSys'13]*
    - Retrieve N items per interest. Then choose the overall top-N items. *ComiRec [KDD'20], PinnerSage [KDD'20]*

Mila

Google Research

Previous Solutions on Multi-interest User Representation: Point-based Representation

- Single-point User Representation (SUR)
  - Fails to cover multi-interest (unless using a very high-dimension vector)
- Multi-point User Representation (MUR)
  - How to choose # of points (*K*)?
  - Does not model uncertainty

Main research question:

Find a better way for users' multi-interest modeling

➔ Adaptive to different number of interests

➔ Be able to model uncertainty

➔ Not require very high-dimension

Mila Google Research

Previous Solutions on Multi-interest User Representation: Point-based Representation

- Single-point User Representation (SUR)
  - Fails to cover multi-interest (unless using a very high-dimension vector)
- Multi-point User Representation (MUR)
  - How to choose # of points (*K*)?
  - Does not model uncertainty

Main research question:

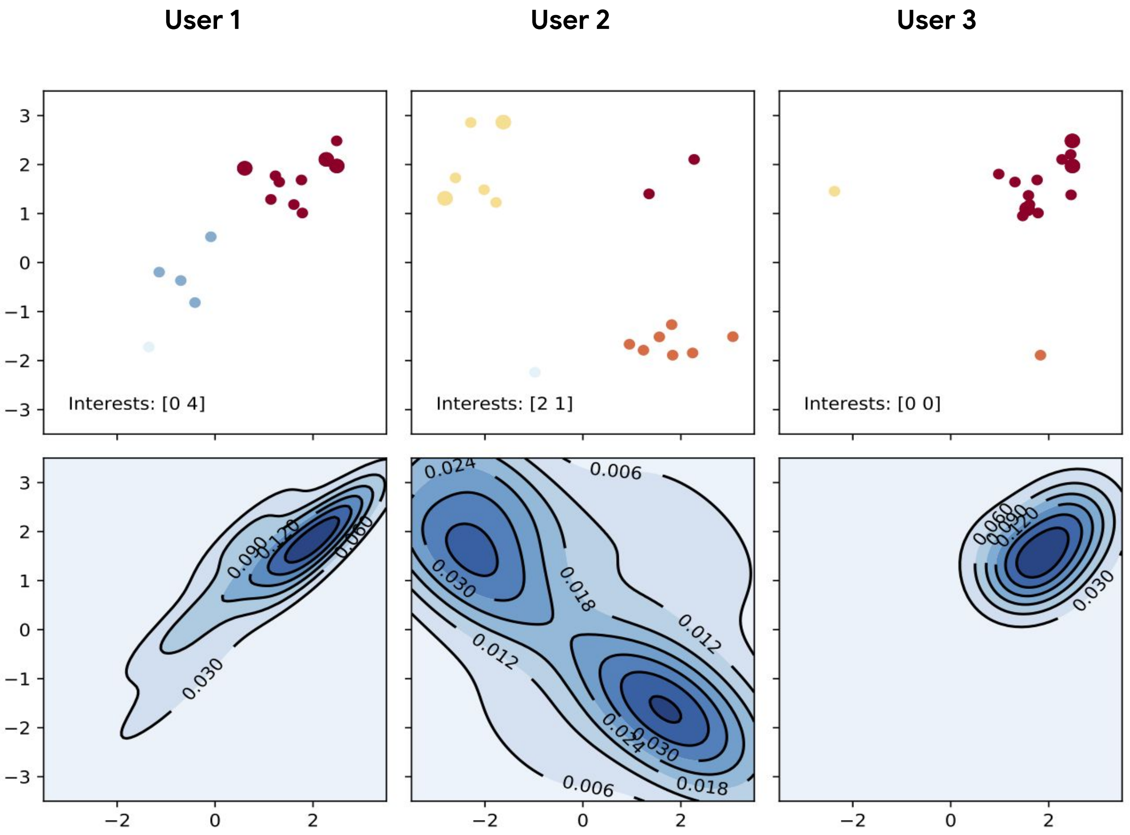
Find a better way for users' multi-interest modeling

→ Adaptive to different number of interests
→ Be able to model uncertainty
→ Not require very high-dimension

Address limitations of MUR

Address limitations of SUR

Motivation: Density-based User Representation



Mila

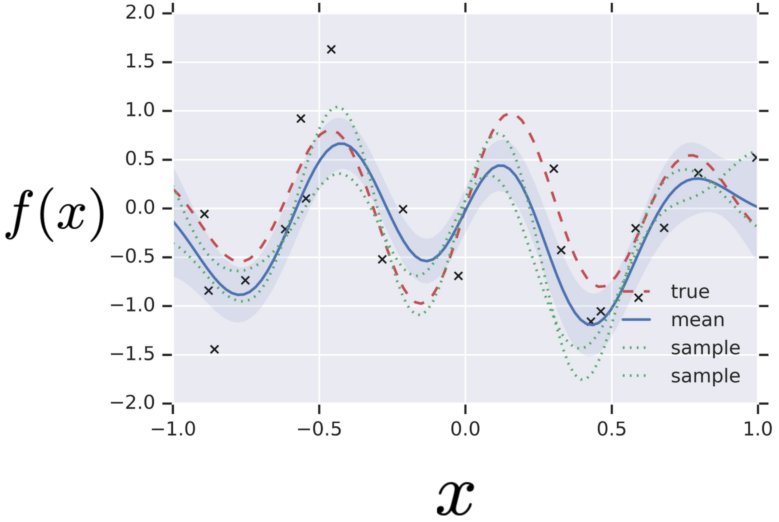
Google Research

Gaussian Process Regression (GPR)

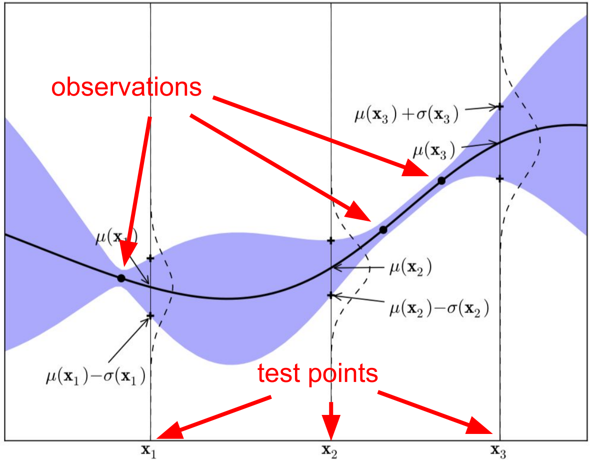
- A distribution over functions

$$f \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot))$$

- Posterior update with observations
- Can draw samples (functions)



$$f(x) \sim N(\mu(x), \sigma^2(x))$$



Toy Example: Maintain a GPR per user

*Item embeddings are **pre-trained** and **fixed**.*

Observed points

Observed value

1 2 -2

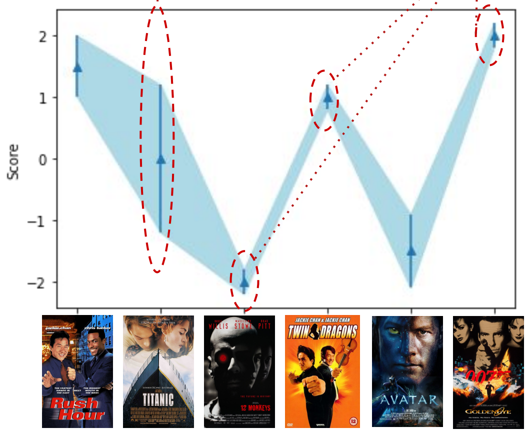
All points

GPR

- **Kernel**
- **Observation_noise**
- **Prediction_noise**

high uncertainty

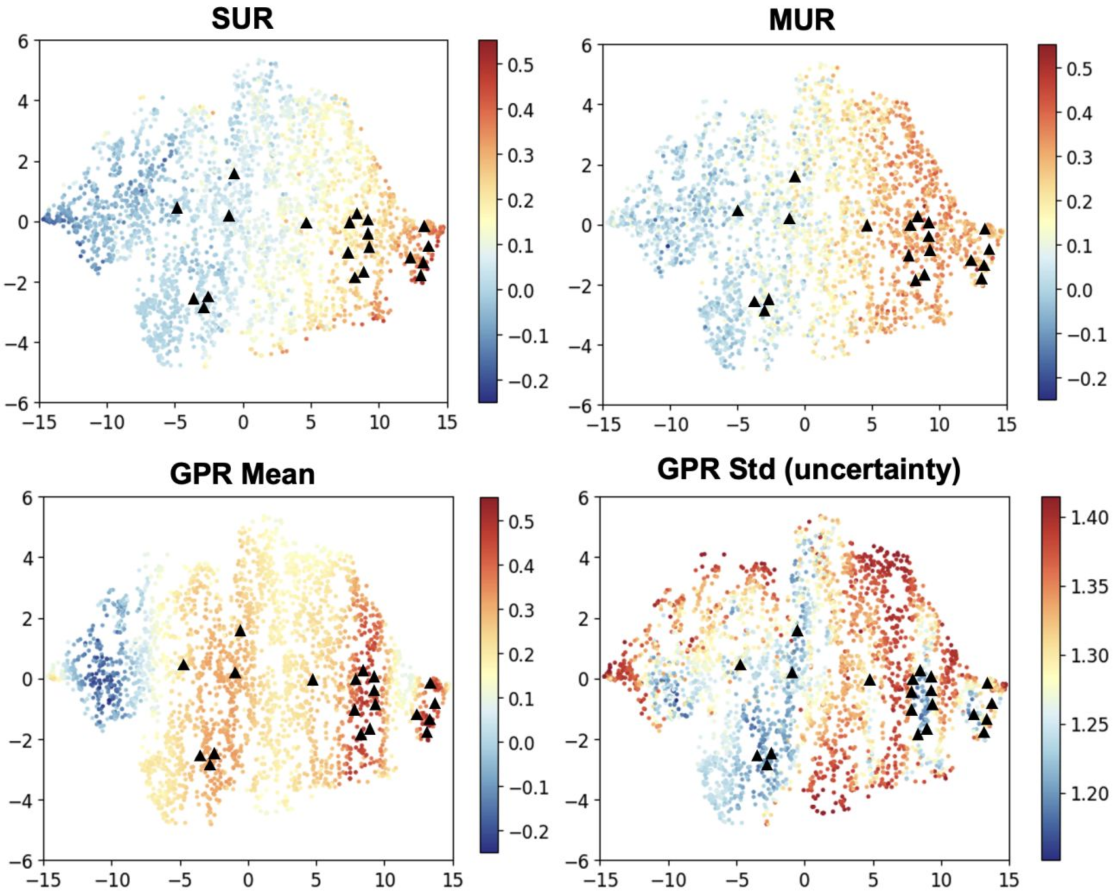
low uncertainty



UCB, sampling, ...

Retrieve top-N items

Example:
t-sne on prediction score



Experiments

- Analysis on real-world datasets

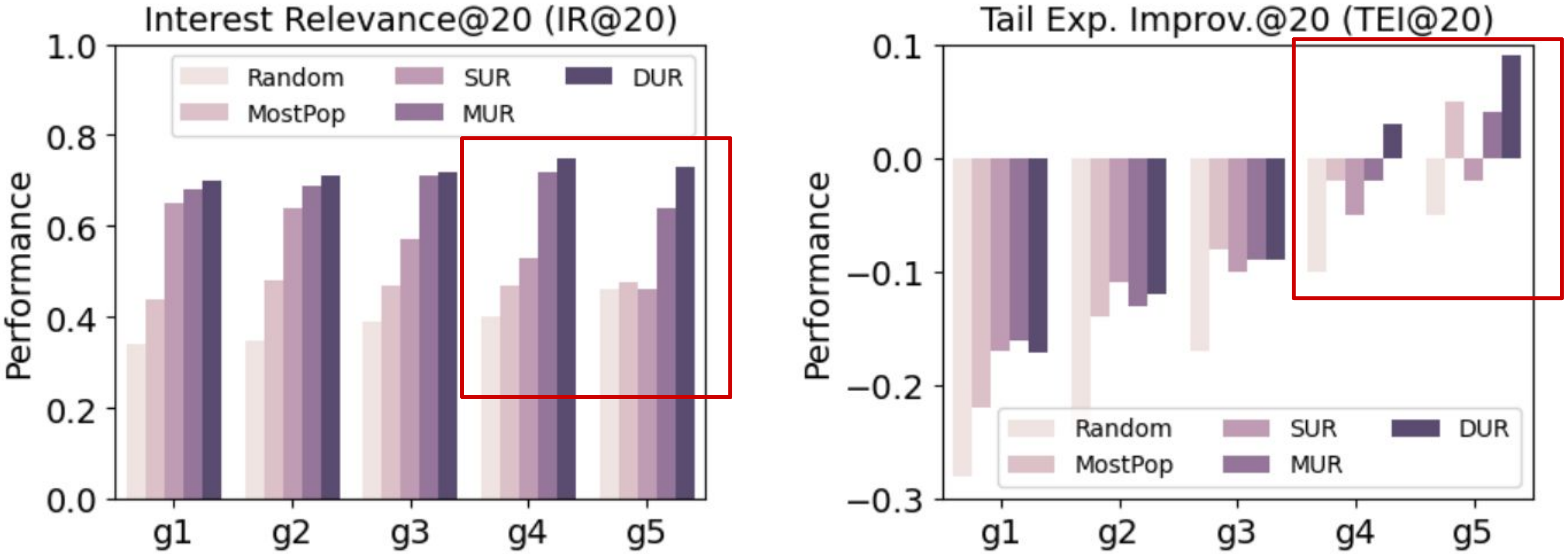
	# User	# Item	# Interac.	Density
Amazon	6,223	32,830	4M	0.18%
MovieLens	123,002	12,532	20M	1.27%
Taobao	756,892	570,350	70M	0.01%

- Baselines:
  - **Heuristics**: Random, MostPop
  - **SUR**: YoutubeDNN, GRU4Rec, BERT4Rec, gSASRec
  - **MUR**: MIND, ComiRec, CAMI, PIMI, REMI
  - **DUR**: GPR4DUR (ours)
- Evaluation Metrics:
  - User side: (1) Interest-wise Coverage, (2) Interest-wise Relevance
  - Item side: (1) Exposure Deviation, (2) Tail Exposure Improvement

Mila

Google Research

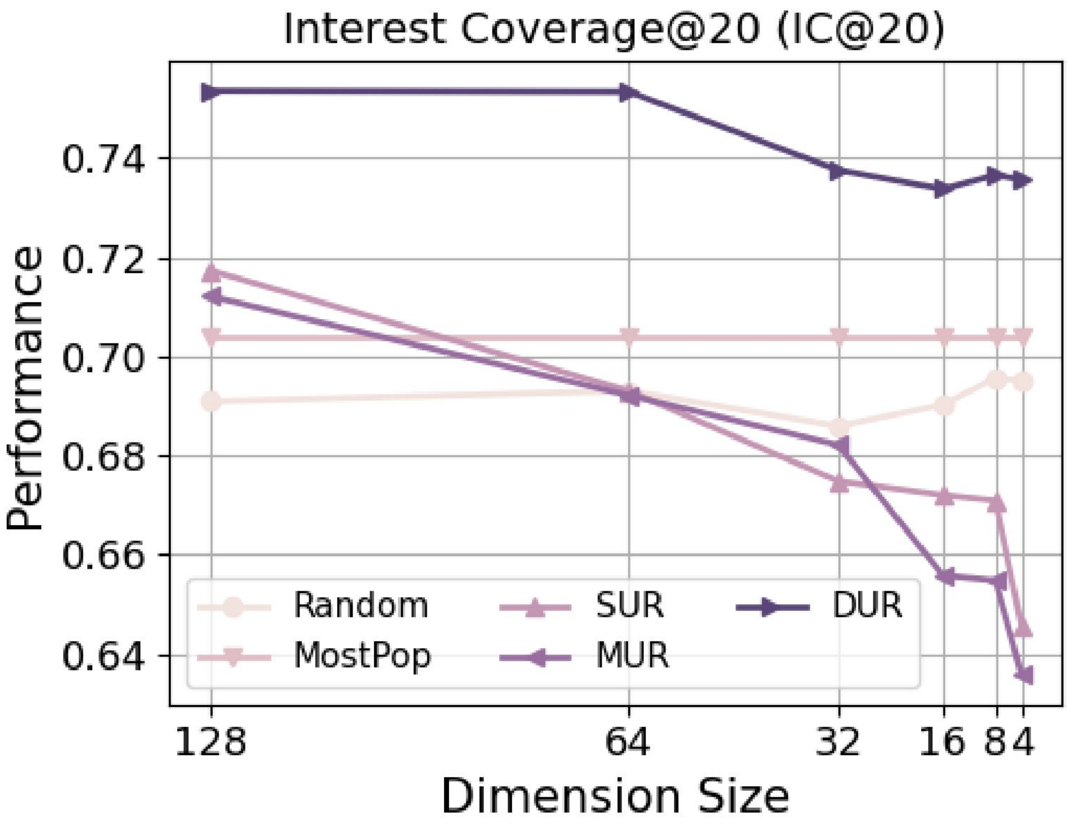
Performance across User Groups



(a) Users grouped by number of interactions

- ➔ *Improvement across all user groups*
- ➔ *Large improvement on multi-interest users*
- ➔ *Improved exposure to tail items for multi-interest users*

Robustness to Dimension Size



Summary

1. Understand limitations of *point-based user representation (SUR & MUR)*
2. A novel *density-based user representation (DUR)* using GPR
    a. improve on both retrieval and ranking
    b. largely **improve the interest coverage** and maintain **high relevance**
    c. reduce **exposure deviation** (*overall* + *niche interests*)
    d. **robust** to dimension size

Thanks for your attention!

Join us at Poster session 1 at Wed 11 Dec, 11 a.m. PST — 2 p.m. PST

Paper

Contact