THE WEB CONFERENCE | TAIPEI 2020

# Directional and Explainable Serendipity Recommendation


Xueqi Li[1], Wenjun Jiang[1], Weiguang Chen[1], Jie Wu[2] , Guojun Wang[3] and Kenli Li[1]

[1]Hunan University, [2]Temple University and [3]Guangzhou University

Presenter: Xueqi Li

# Outline

◊ Introduction

◊ DESR: The details

◊ Experiments

◊ Conclusions & Future Work

# Outline

◊ Introduction

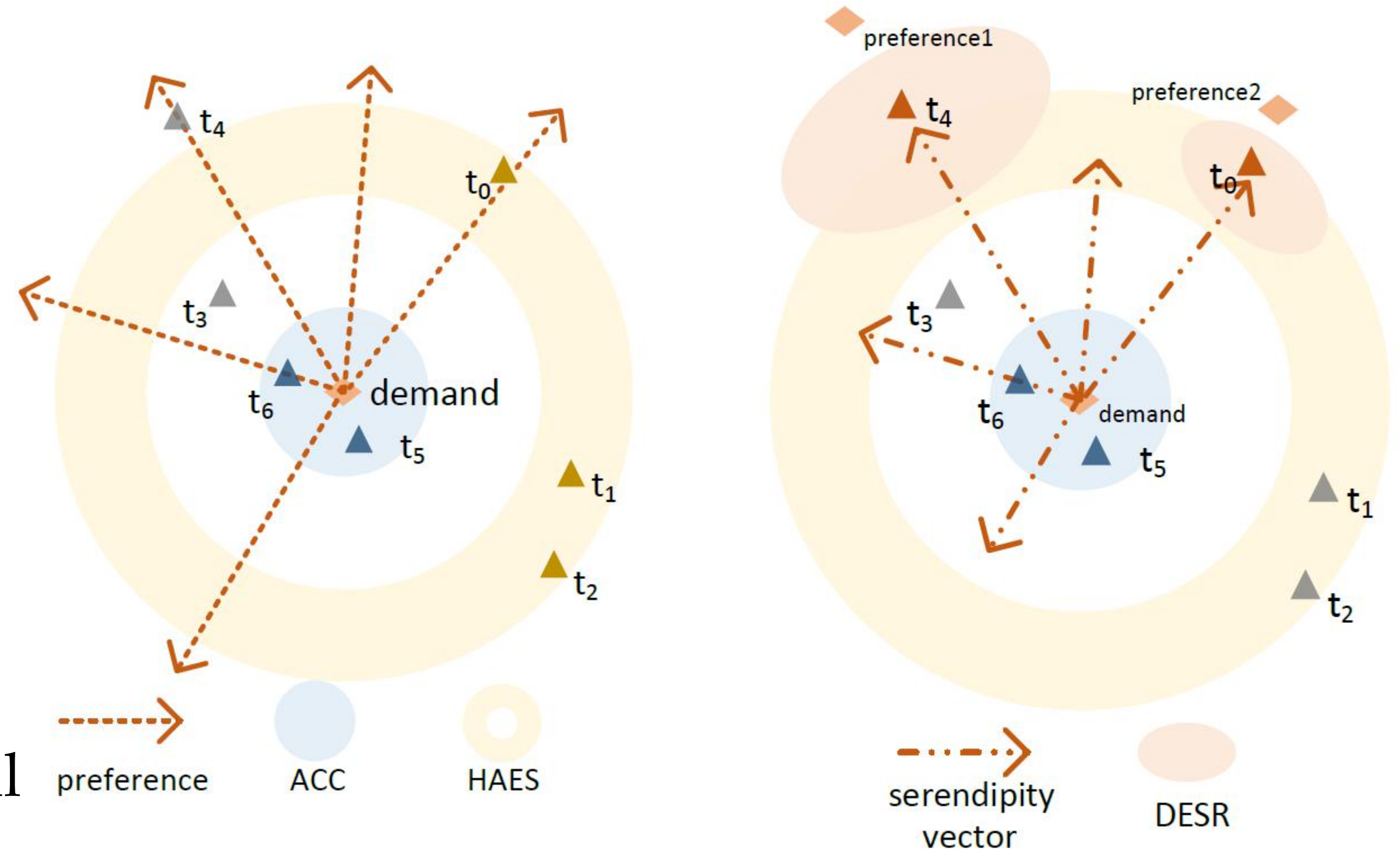
◊ DESR: The details

◊ Experiments

◊ Conclusions & Future Work

# An example

- ◊ Accuracy-oriented recommendation
- ◊ Serendipity-oriented recommendation
- ◊ Preference direction
- ◊ The necessity of directional recommendations




(a) (b)

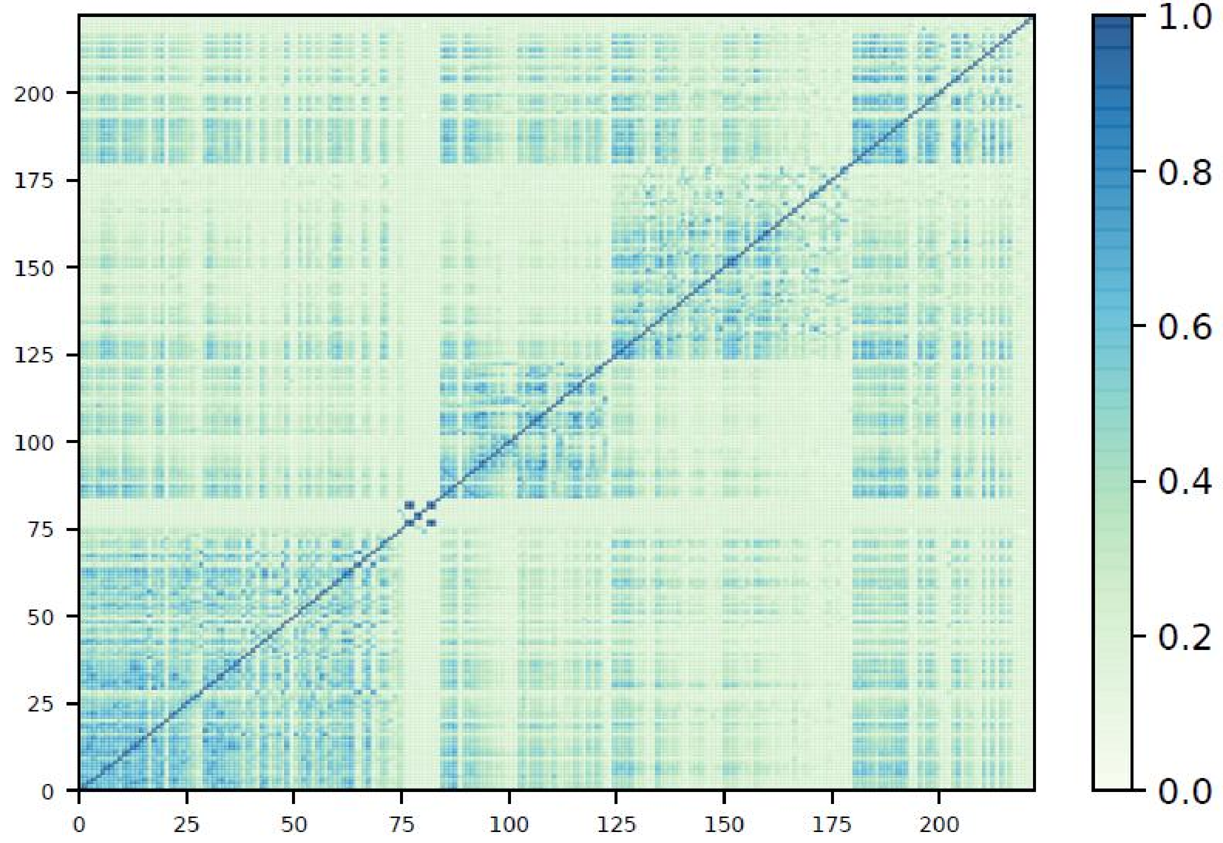
# An example

◊ Accuracy-oriented recommendation

◊ Serendipity-oriented recommendation

◊ Preference direction

◊ The necessity of directional recommendations




(c)

# Open Challenges

◊ What are the users' preference directions?

◊ How to generate serendipitous recommendations with user preference direction?

◊ Why the items are recommended?

# Outline

◊ Introduction

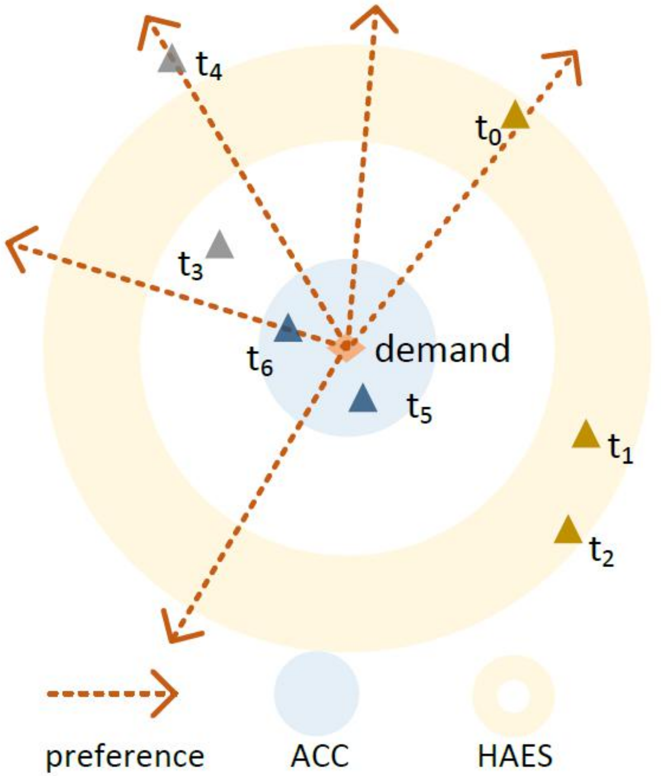
◊ DESR: The details

◊ Experiments

◊ Conclusions & Future Work

# The Concepts

- User Preference Direction
  - from a short-term demand to a long-term preference
- Serendipity Vector
  - Direction: the corresponding user preference direction
  - Magnitude: the range for serendipity recommendations



# Problem Definition

- Input:

  a user set U, an item set I, a rating matrix M and a target user $u_t$

- Output:

  a list of potential items $I_t$ and the corresponding explanations

- Objective:

  Maximize serendipity($I_t$, $u_t$)

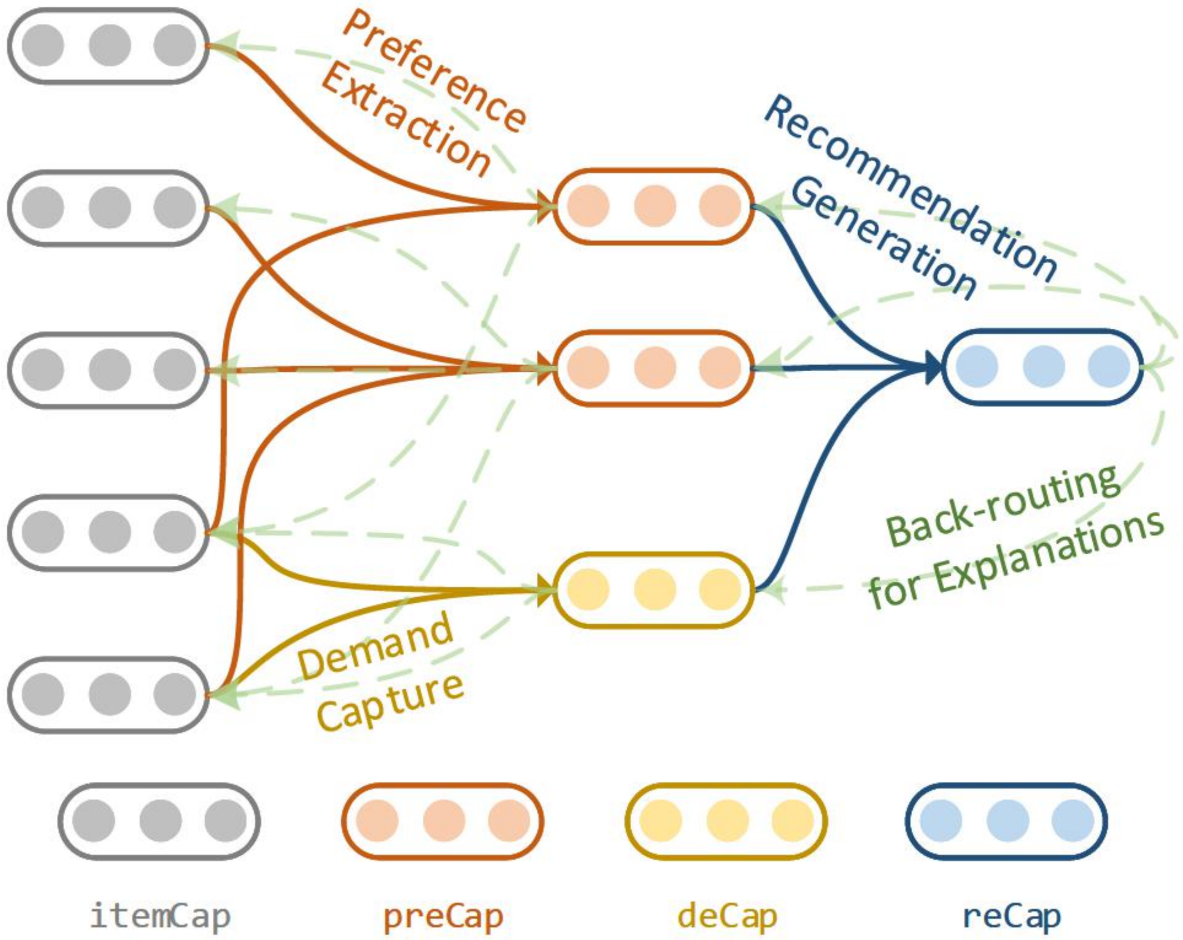
# Framework

◊ Long-term Preference Extraction

◊ Short-term Demand Capture

◊ Recommendation Generation

◊ Back-routing for Explanations



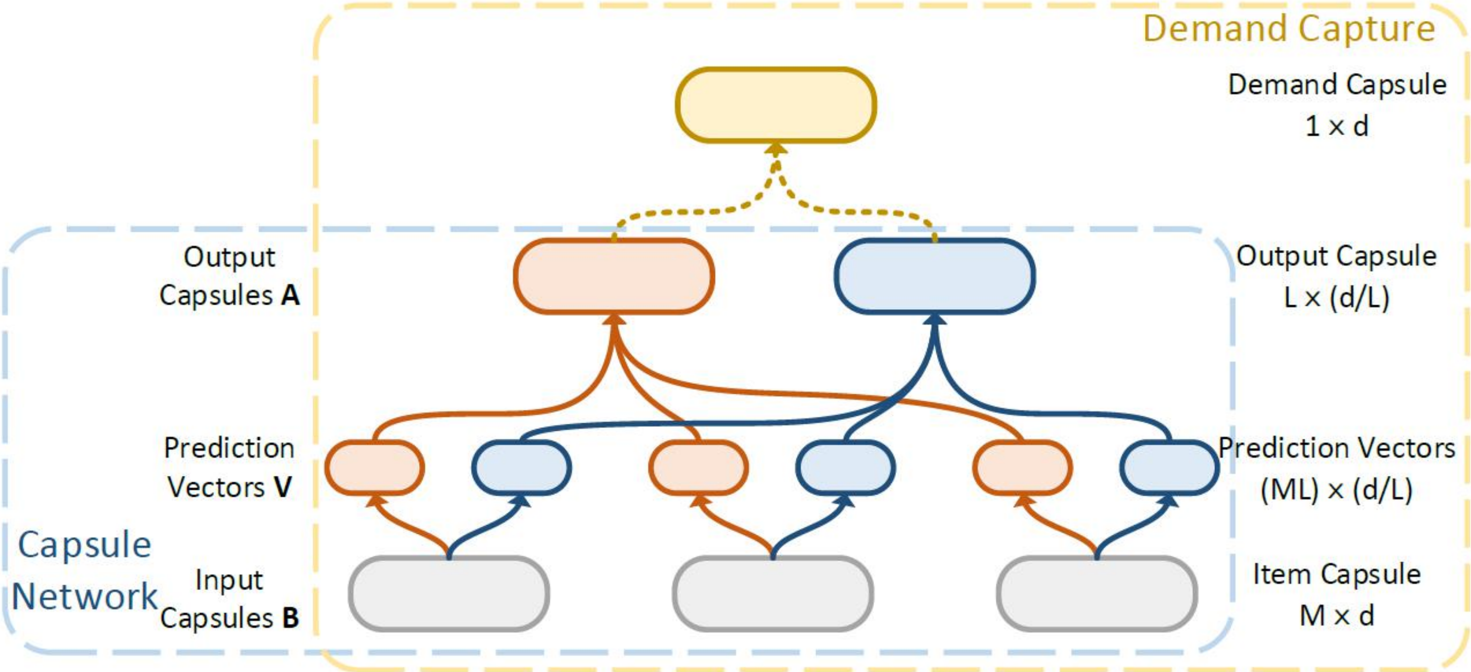
# Long-term Preference Extraction with GMM

$$P(y|\theta) = \sum_{k=1}^{K} \alpha_k \phi(y|\theta_k),$$

$$\phi(y|\theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp(-\frac{(y-\mu_k)^2}{2\sigma_k^2}),$$

| | Clustering | Preference Extraction |
|---|---|---|
| $\mu_k$ | k-th clustering center | k-th long-term preference |
| $\sigma_k$ | the corresponding clustering level | |

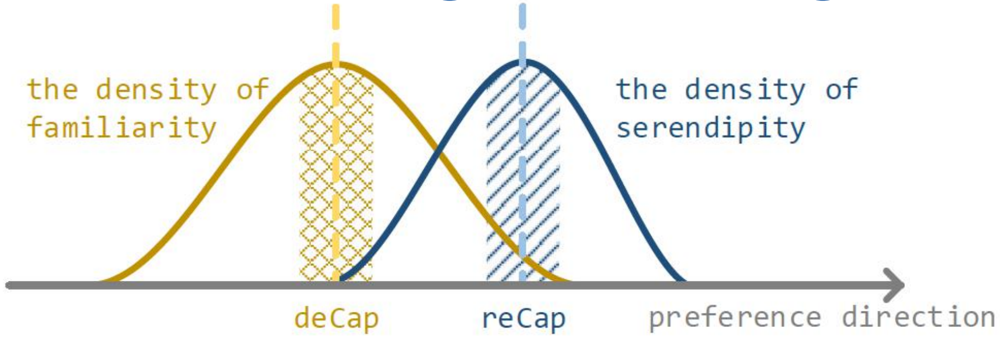
# Short-term Demand Capture with Capsule Network




$$loss = \frac{1}{n}\sum_{i=1}^{n}(deCap_i - de\hat{C}ap_i)^2$$

# Recommendation Generation

## Determining the Magnitude




$S(u_{tar})$ : the scope of the long-term preferences

$F(preCap)$: the familiarity for preCap

$$||\overrightarrow{serendipity\ vector}|| = m_{base}(1 + S(u_{tar}))(1 + F(preCap))$$

# Recommendation Generation

## Setting the direction

$$\frac{\overrightarrow{serendipity\ vector}}{||\overrightarrow{serendipity\ vector}||} = \frac{\overrightarrow{preCap - deCap}}{||\overrightarrow{preCap - deCap}||}$$

## Recommendation Generation

$s_i = \frac{T_i}{\sum_{i=1}^{K} T_i}$ , $T_i$ is the number of items related to the corresponding preference

# Back-routing for Explanations

$$expCap = Minimize\ dis(cap, reCap_{tar})$$

$$explanation = \begin{cases} explanation_1 & \text{expCap} \in \text{preCaps} \\ explanation_2 & \text{expCap} \in \text{deCaps} \end{cases}$$

"The item is similar to the items, {$t_1$, $t_2$, ... , $t_E$ }, which you watched for a long time.“

"The item is similar to the items, {$t_1$, $t_2$, ... , $t_E$ }, which you recently watched."

# Outline

◊ Introduction

◊ DESR: The details

◊ Experiments

◊ Conclusions & Future Work

# Experiments

◊ Experiment Settings

◊ Effects of Parameters

◊ Overall Comparison

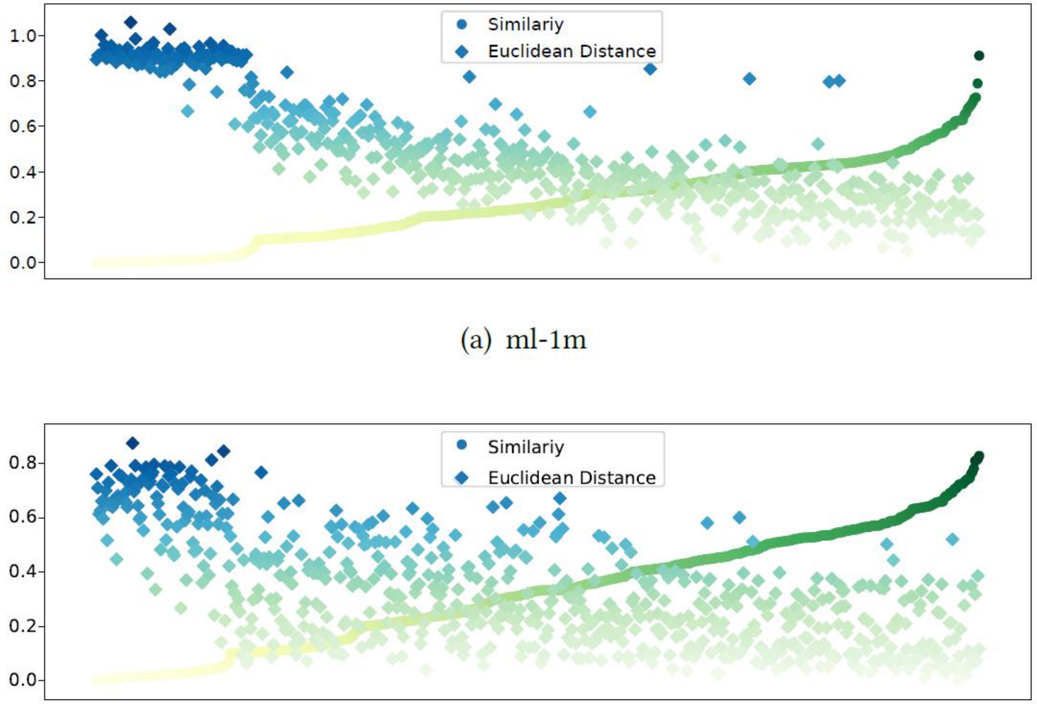
◊ Verification on Components

◊ A Case Study

# Experiment Settings -- Data Preprocessing

**Table 2: Statistics of datasets.**

| Item | Statistic | |
|---|---|---|
| | MovieLens-1m | Amazon-Kindle-Store |
| # users | 6040 | 3061 |
| # items | 3260 | 6073 |
| # ratings | 998539 | 132594 |
| density [5] | 5.07% | 0.71% |




(a) ml-1m

(b) Kindle

# Experiment Settings -- Baselines

◊ RAND  a random-based method

◊ $ACC_{LSTM}$  an accuracy-oriented method

◊ KFN  a serendipity-oriented method

◊ HAES  a serendipity-oriented method

# Experiment Settings -- Metrics

| Metric | Abbreviation | Description |
|---|---|---|
| $accuracy$ (Eq. 16) | - | a metric to measure the recommendation accuracy |
| $accuracy\_preference$ (Eq. 17) | $acc\_pre$ | a metric to measure the similarity of $reCaps$ and $preCaps$ |
| $accuracy\_demand$ (Eq. 18) | $acc\_de$ | a metric to measure the similarity of $reCaps$ and $deCaps$ |
| $accuracy\_all$ (Eq. 19) | $acc$ | a metric to summarize $accuracy$, $acc\_pre$ and $acc\_de$ |
| $diversity$ (Eq. 20) | $div$ | a metric to measure the diversity of recommendations |
| $difference$ (Eq. 21) | - | a metric to measure the difference between recommendations and $u_{tar}$'s history |
| $difference\_all$ (Eq. 22) | $dif$ | a metric to summarize $diversity$ and $difference$ |
| $AD$ (Eq. 23) | - | a metric to comprehensively measure $acc$ and $dif$ |

$$AD = \frac{acc * dif}{acc + dif}$$

An overall metric for recommendation serendipity.

More details refer to the paper.

# Effects of Parameters -- in metrics

◊ Varying parameters exerts different impacts for most approaches, especially $\mathrm{ACC}_{\mathrm{LSTM}}$ and DESR.

◊ A bigger weight of acc_pre (i.e., $(\eta, \theta) = (1, 0.5)$) makes DESR outperform $\mathrm{ACC}_{\mathrm{LSTM}}$ on acc.

$$acc = \frac{1}{1+\eta+\theta}(accuracy + \eta acc\_pre + \theta acc\_de)$$

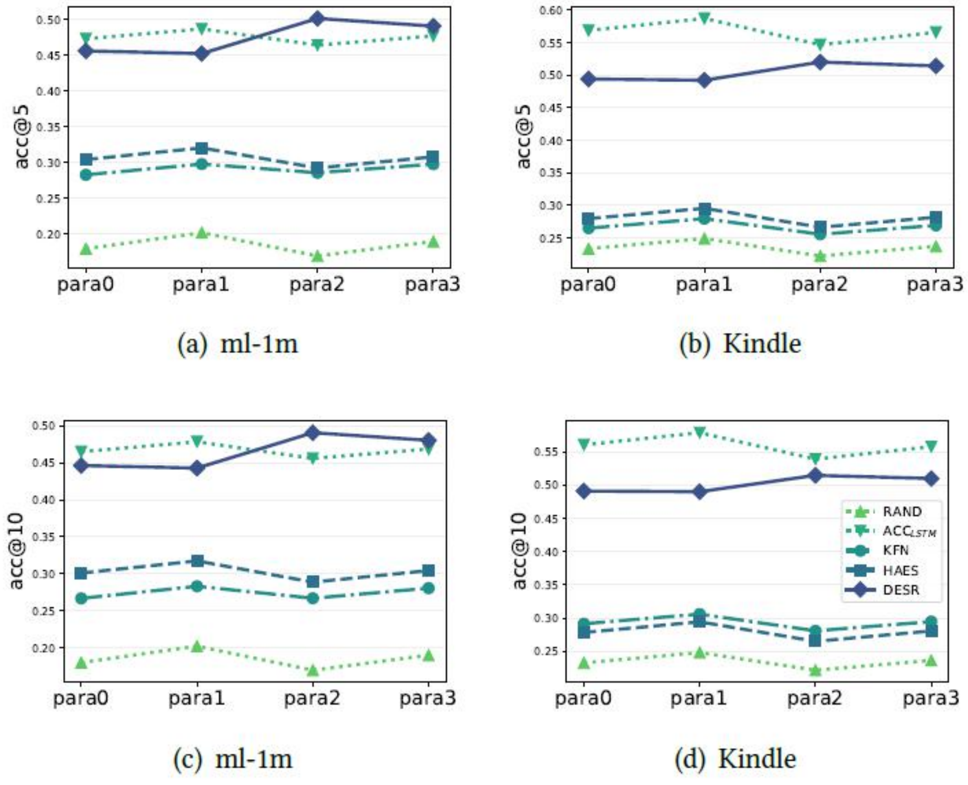



**Figure 6: Effects of parameters on acc. (para0:** $\eta = 0.5, \theta = 0.5$; **para1:** $\eta = 0.5, \theta = 1$; **para2:** $\eta = 1, \theta = 0.5$; **para3:** $\eta = 1, \theta = 1$.)

# Effects of Parameters -- in metrics

- Serendipity-oriented methods (i.e., KFN, HAES, DESR) perform better on dif when a bigger weight is assigned to difference.

- DESR performs poorer than HAES on dif.

$$dif = \frac{1}{\lambda + \gamma}(\lambda div + \gamma difference)$$

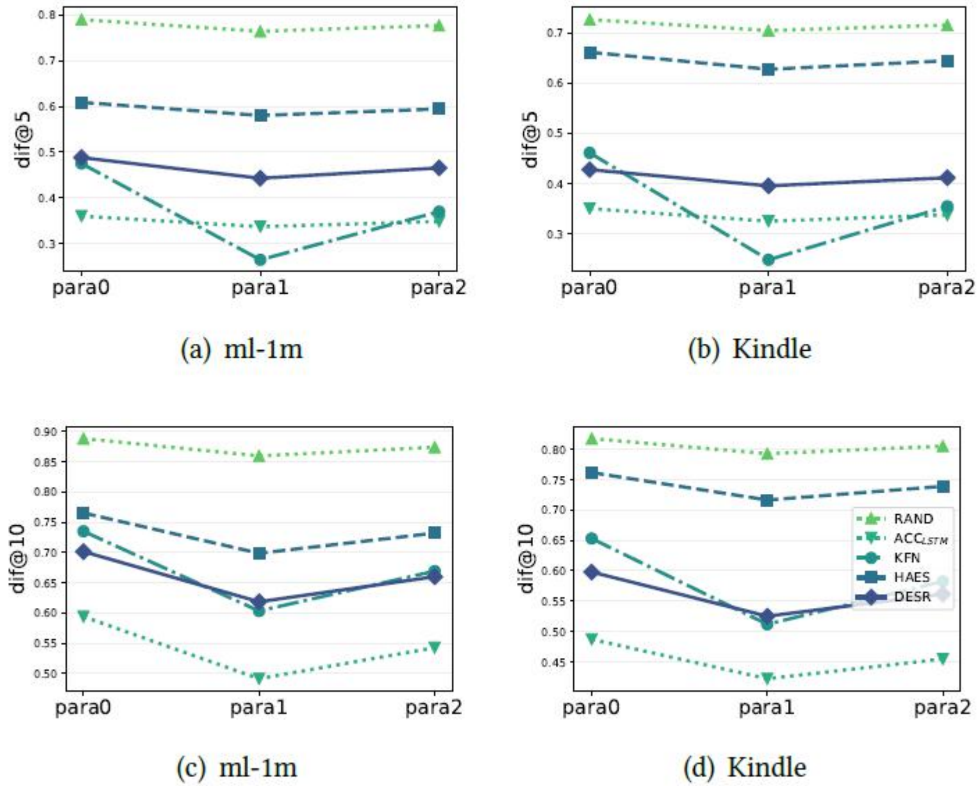
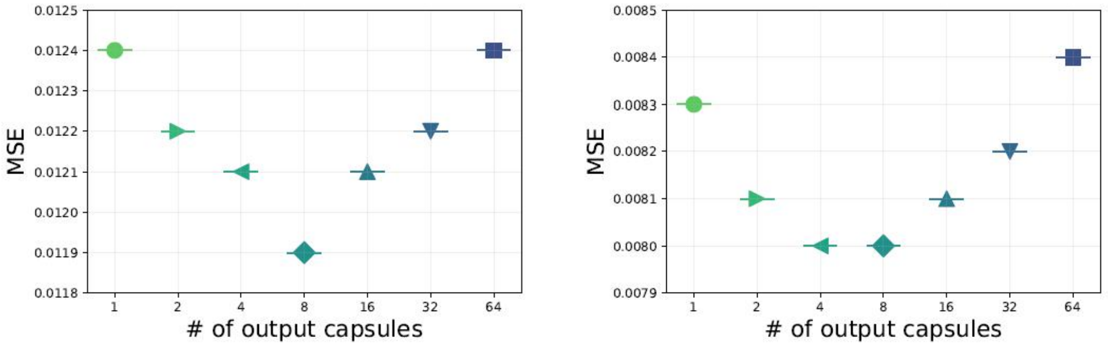



**Figure 7: Effects of parameters on dif. (para0:** $\lambda = 0.5, \gamma = 1$; **para1:** $\lambda = 1, \gamma = 0.5$; **para2:** $\lambda = 1, \gamma = 1$.)

# Effects of Parameters -- in DESR

◊ It is hardly possible to capture accurate short-term demands when L is too big or small.

◊ The optimal value for L is 8.




(a) ml-1m (b) Kindle

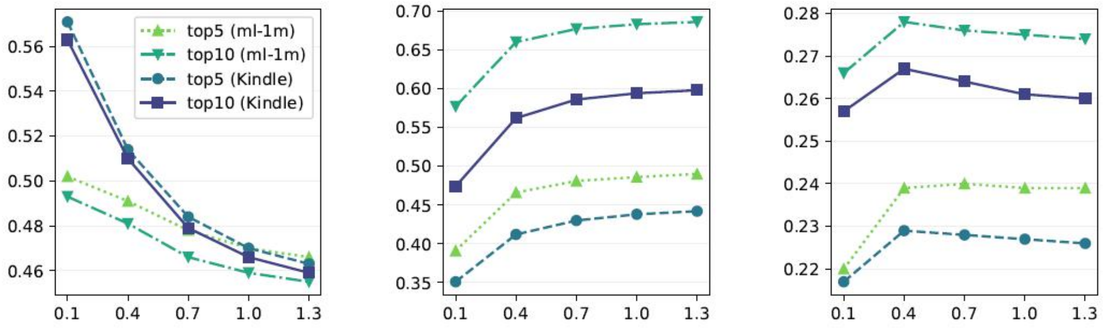
**Figure 8: Effects of the number of output capsules.**

# Effects of Parameters -- in DESR

◊ By increasing of $m_{base}$ , the performance on dif becomes better and that on acc becomes worse.

◊ When $m_{base}$ = 0.4, DESR reaches the best performance on AD.

$$\|\overrightarrow{serendipity\ vector}\| = m_{base}(1 + S(u_{tar}))(1 + F(preCap)).$$




(a) $acc$ (b) $dif$ (c) $AD$

**Figure 9: Effects of $m_{base}$**

# Overall Comparison -- on Serendipity

◊ DESR achieves the best performance on both datasets.

◊ DESR performs better on MovieLens-1m.

Table 6: Comparison on overall recommendation (ml-1m).

| method | acc@5 | acc@10 | dif@5 | dif@10 | AD@5 | AD@10 |
|---|---|---|---|---|---|---|
| **RAND** | 0.19 | 0.19 | **0.777** | **0.874** | 0.152 | 0.156 |
| **ACC$_{LSTM}$** | 0.477 | 0.469 | 0.348 | 0.543 | 0.201 | 0.252 |
| **KFN** | 0.298 | 0.28 | 0.37 | 0.669 | 0.165 | 0.198 |
| **HAES** | 0.308 | 0.304 | 0.595 | 0.732 | 0.203 | 0.215 |
| **DESR** | **0.491** | **0.48** | 0.466 | 0.66 | **0.239** | **0.278** |

Table 7: Comparison on overall recommendation (Kindle).

| method | acc@5 | acc@10 | dif@5 | dif@10 | AD@5 | AD@10 |
|---|---|---|---|---|---|---|
| **RAND** | 0.237 | 0.236 | **0.715** | **0.805** | 0.178 | 0.183 |
| **ACC$_{LSTM}$** | **0.566** | **0.558** | 0.338 | 0.455 | 0.212 | 0.25 |
| **KFN** | 0.269 | 0.295 | 0.355 | 0.582 | 0.153 | 0.196 |
| **HAES** | 0.282 | 0.281 | 0.644 | 0.739 | 0.196 | 0.203 |
| **DESR** | 0.514 | 0.51 | 0.412 | 0.562 | **0.229** | **0.267** |

# Overall Comparison -- on Diversity

Table 4: Comparison on original metrics. (ml-1m)

| method | accuracy@5 | accuracy@10 | acc_pre@5 | acc_pre@10 | acc_de@5 | acc_de@10 | div@5 | div@10 | difference@5 | difference@10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **RAND** | 0.15 | 0.15 | 0.128 | 0.129 | 0.291 | 0.291 | **0.739** | **0.831** | **0.819** | **0.918** |
| **ACC$_{LSTM}$** | **0.461** | **0.454** | 0.428 | 0.42 | **0.541** | **0.533** | 0.314 | 0.389 | 0.383 | 0.697 |
| **KFN** | 0.238 | 0.225 | 0.297 | 0.268 | 0.358 | 0.349 | 0.054 | 0.472 | 0.686 | 0.867 |
| **HAES** | 0.293 | 0.291 | 0.245 | 0.239 | 0.386 | 0.383 | 0.551 | 0.632 | 0.638 | 0.832 |
| **DESR** | 0.352 | 0.344 | **0.683** | **0.668** | 0.437 | 0.429 | 0.398 | 0.536 | 0.533 | 0.784 |

Table 5: Comparison on original metrics. (Kindle)

| method | accuracy@5 | accuracy@10 | acc_pre@5 | acc_pre@10 | acc_de@5 | acc_de@10 | div@5 | div@10 | difference@5 | difference@10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **RAND** | 0.222 | 0.221 | 0.177 | 0.177 | 0.311 | 0.31 | **0.683** | **0.768** | **0.748** | **0.843** |
| **ACC$_{LSTM}$** | **0.577** | **0.569** | 0.459 | 0.453 | **0.661** | **0.652** | 0.3 | 0.357 | 0.375 | 0.552 |
| **KFN** | 0.252 | 0.282 | 0.218 | 0.238 | 0.338 | 0.364 | 0.036 | 0.371 | 0.674 | 0.794 |
| **HAES** | 0.273 | 0.272 | 0.213 | 0.212 | 0.358 | 0.358 | 0.593 | 0.671 | 0.695 | 0.807 |
| **DESR** | 0.433 | 0.435 | **0.624** | **0.609** | 0.485 | 0.486 | 0.363 | 0.452 | 0.46 | 0.671 |

◊ The more accurate the recommendations are, the less diverse they become.

◊ DESR maximizes diversity under the premise of ensuring recommendation accuracy.

# Verification on Components -- Preference Extraction

Table 4: Comparison on original metrics. (ml-1m)

| method | accuracy@5 | accuracy@10 | acc_pre@5 | acc_pre@10 | acc_de@5 | acc_de@10 | div@5 | div@10 | difference@5 | difference@10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **RAND** | 0.15 | 0.15 | 0.128 | 0.129 | 0.291 | 0.291 | **0.739** | **0.831** | **0.819** | **0.918** |
| **$ACC_{LSTM}$** | **0.461** | **0.454** | 0.428 | 0.42 | **0.541** | **0.533** | 0.314 | 0.389 | 0.383 | 0.697 |
| **KFN** | 0.238 | 0.225 | 0.297 | 0.268 | 0.358 | 0.349 | 0.054 | 0.472 | 0.686 | 0.867 |
| **HAES** | 0.293 | 0.291 | 0.245 | 0.239 | 0.386 | 0.383 | 0.551 | 0.632 | 0.638 | 0.832 |
| **DESR** | 0.352 | 0.344 | **0.683** | **0.668** | 0.437 | 0.429 | 0.398 | 0.536 | 0.533 | 0.784 |

Table 5: Comparison on original metrics. (Kindle)

| method | accuracy@5 | accuracy@10 | acc_pre@5 | acc_pre@10 | acc_de@5 | acc_de@10 | div@5 | div@10 | difference@5 | difference@10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **RAND** | 0.222 | 0.221 | 0.177 | 0.177 | 0.311 | 0.31 | **0.683** | **0.768** | **0.748** | **0.843** |
| **$ACC_{LSTM}$** | **0.577** | **0.569** | 0.459 | 0.453 | **0.661** | **0.652** | 0.3 | 0.357 | 0.375 | 0.552 |
| **KFN** | 0.252 | 0.282 | 0.218 | 0.238 | 0.338 | 0.364 | 0.036 | 0.371 | 0.674 | 0.794 |
| **HAES** | 0.273 | 0.272 | 0.213 | 0.212 | 0.358 | 0.358 | 0.593 | 0.671 | 0.695 | 0.807 |
| **DESR** | 0.433 | 0.435 | **0.624** | **0.609** | 0.485 | 0.486 | 0.363 | 0.452 | 0.46 | 0.671 |

◊ DESR increases *acc_pre* by at least 33.33% even when compared to $ACC_{LSTM}$.

◊ DESR performs even better than other serendipity-oriented methods on *accuracy*.

# Verification on Components -- Demand Capture

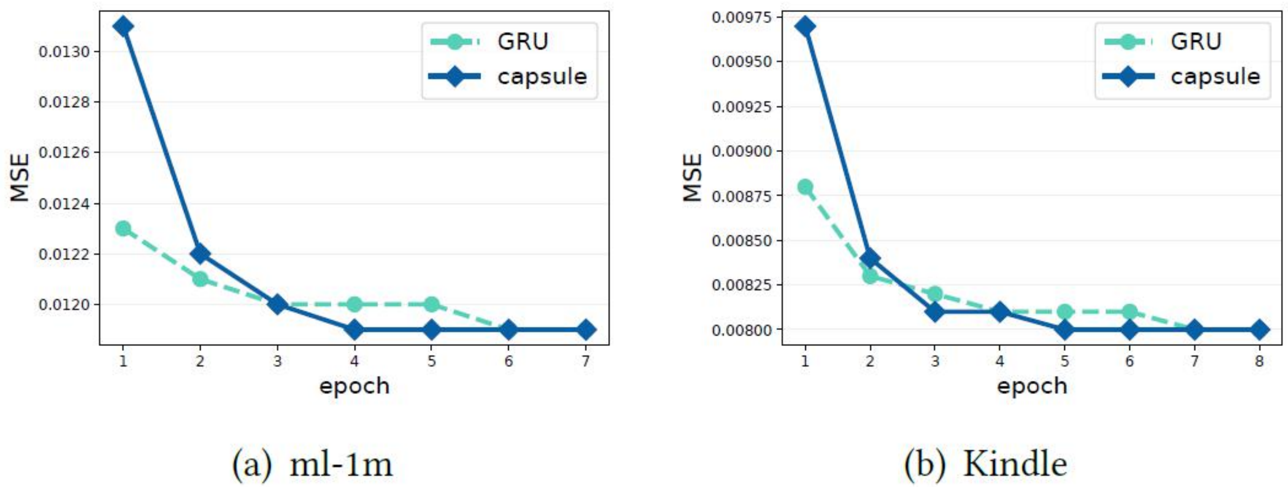



**Figure 11: Comparison between the capsule network and GRU on capturing users' short-term demands.**

◊ Capsule network and GRU reach the same optimal MSE on both datasets.

◊ Capsule network has a faster convergence and outperforms GRU on explainability.

# A Case Study – on Explanation Generation

For reCap0 (Almost Famous), "The movie is similar to the movies, Forrest Gump, Life Is Beautiful and American Beauty, which you watched for a long time."

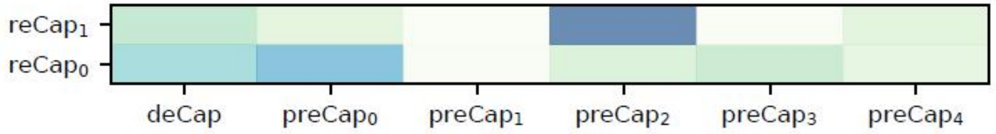



Figure 14: An illustration to show the similarity between recommendations and preferences, demands.

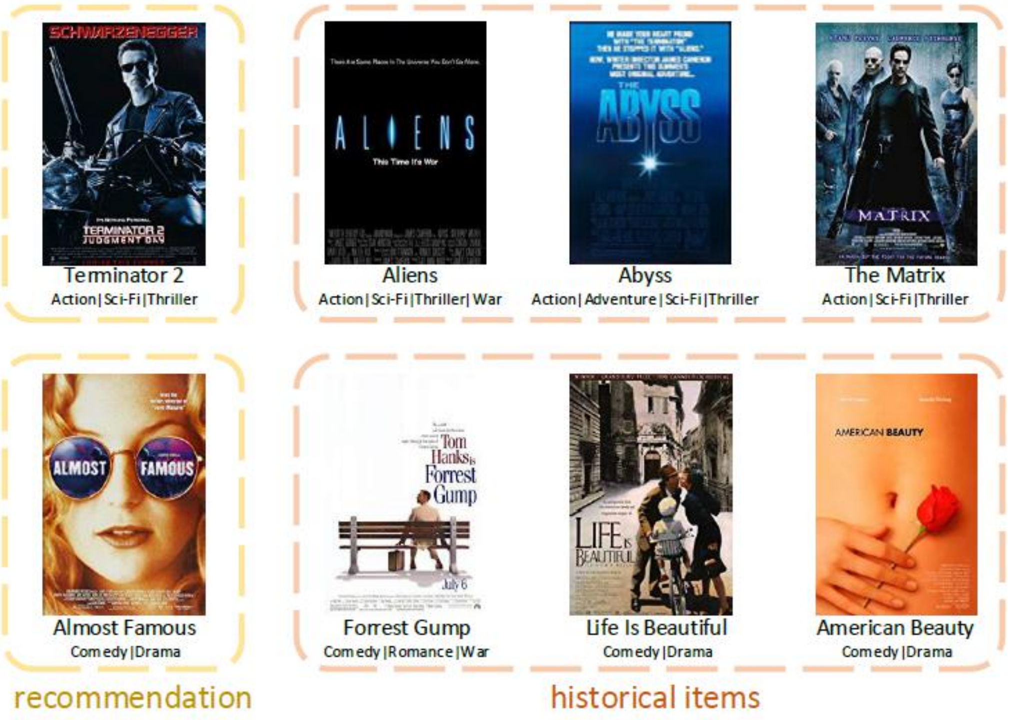



Figure 15: Selection of historical items.

# Summary of Experiments

- ◊ GMM outperforms RNN on developing a comprehensive representation on long-term preferences.
- ◊ Capsule network has great potential in sequence processing.
- ◊ Keeping accuracy on users' long-term preferences would improve the recommendation serendipity.
- ◊ In general, DESR achieves a better performance on *AD*.

# Outline

◊ Introduction

◊ DESR: The details

◊ Experiments

◊ Conclusions & Future Work

# Conclusions

◊ the reinforcement of user preference direction and explainability in serendipity recommendation

◊ Proposal for novel fine-grained metrics for serendipity

# Future work

◊ provide more user-friendly explanations in the serendipity recommendation

Thanks!