From Adaptive Testing to Personalized Adaptive Testing:

Applications of Machine Learning Algorithms

Okan Bulut

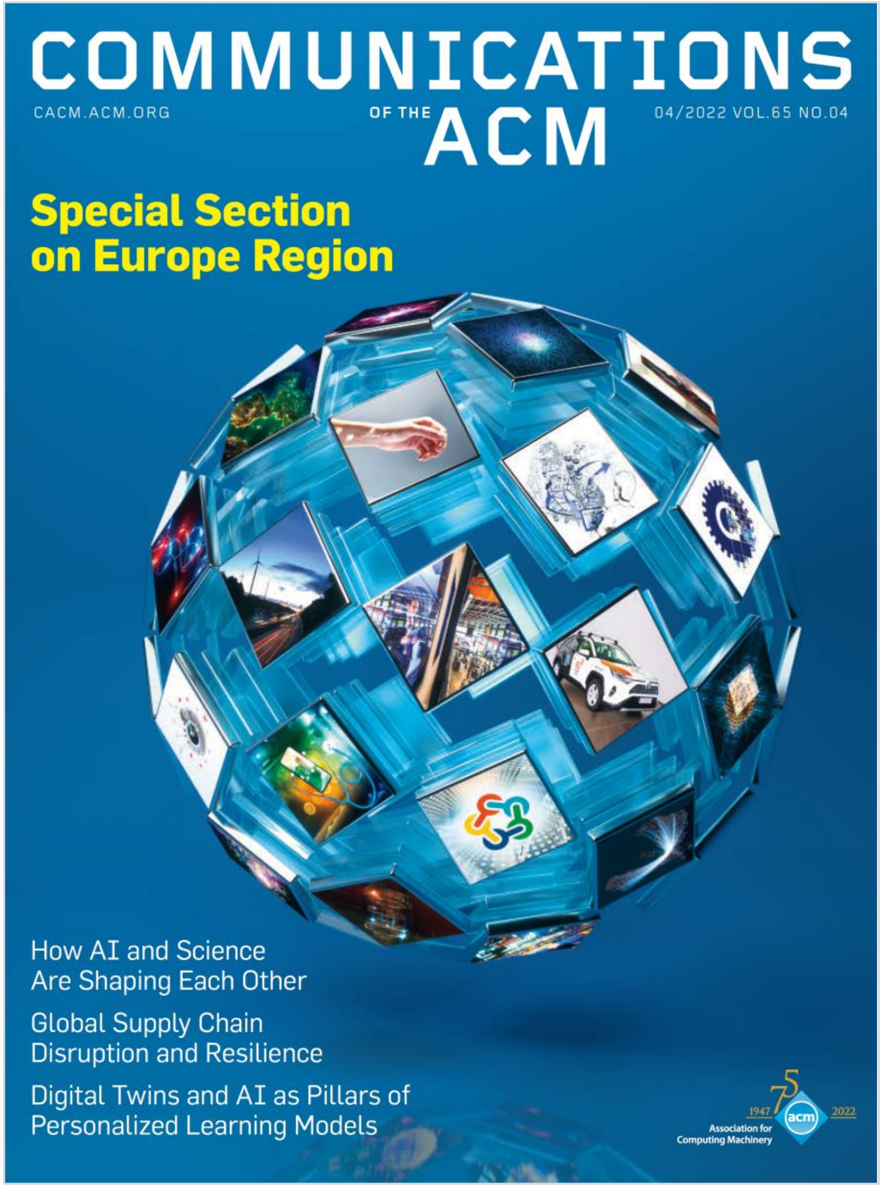
Measurement, Evaluation, and Data Science

University of Alberta

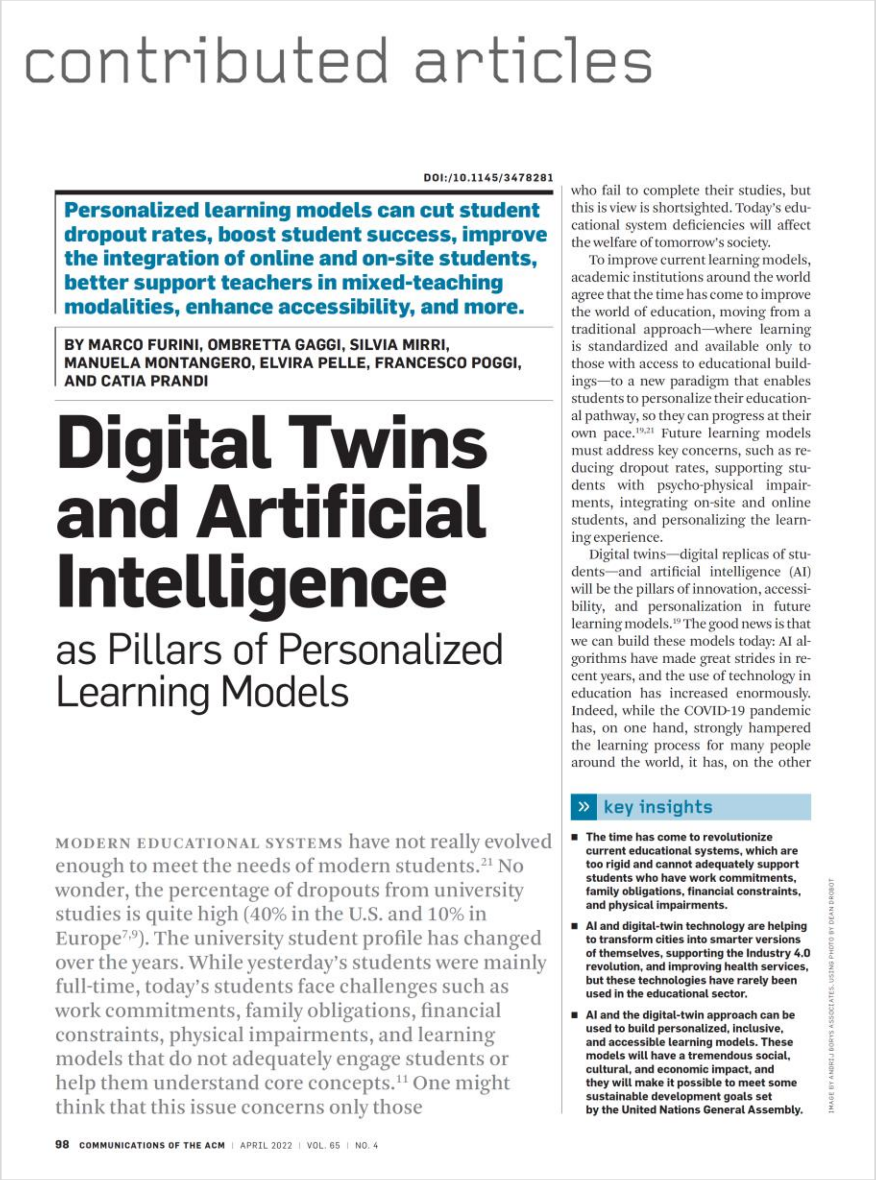
bulut@ualberta.ca www.okanbulut.com @drokanbulut

Outline

1. Beyond adaptive testing
2. Recommender systems (RS)
3. Two RS applications in adaptive testing



https://dl.acm.org/toc/cacm/2022/65/4



DOI:10.1145/3478281

Personalized learning models can cut student dropout rates, boost student success, improve the integration of online and on-site students, better support teachers in mixed-teaching modalities, enhance accessibility, and more.

BY MARCO FURINI, OMBRETTA GAGGI, SILVIA MIRRI, MANUELA MONTANGERO, ELVIRA PELLE, FRANCESCO POGGI, AND CATIA PRANDI

Digital Twins and Artificial Intelligence

as Pillars of Personalized Learning Models

MODERN EDUCATIONAL SYSTEMS have not really evolved enough to meet the needs of modern students.[21] No wonder, the percentage of dropouts from university studies is quite high (40% in the U.S. and 10% in Europe[7,9]). The university student profile has changed over the years. While yesterday's students were mainly full-time, today's students face challenges such as work commitments, family obligations, financial constraints, physical impairments, and learning models that do not adequately engage students or help them understand core concepts.[11] One might think that this issue concerns only those who fail to complete their studies, but this is view is shortsighted. Today's educational system deficiencies will affect the welfare of tomorrow's society.

To improve current learning models, academic institutions around the world agree that the time has come to improve the world of education, moving from a traditional approach—where learning is standardized and available only to those with access to educational buildings—to a new paradigm that enables students to personalize their educational pathway, so they can progress at their own pace.[19,21] Future learning models must address key concerns, such as reducing dropout rates, supporting students with psycho-physical impairments, integrating on-site and online students, and personalizing the learning experience.

Digital twins—digital replicas of students—and artificial intelligence (AI) will be the pillars of innovation, accessibility, and personalization in future learning models.[19] The good news is that we can build these models today: AI algorithms have made great strides in recent years, and the use of technology in education has increased enormously. Indeed, while the COVID-19 pandemic has, on one hand, strongly hampered the learning process for many people around the world, it has, on the other

» key insights

- The time has come to revolutionize current educational systems, which are too rigid and cannot adequately support students who have work commitments, family obligations, financial constraints, and physical impairments.
- AI and digital-twin technology are helping to transform cities into smarter versions of themselves, supporting the Industry 4.0 revolution, and improving health services, but these technologies have rarely been used in the educational sector.
- AI and the digital-twin approach can be used to build personalized, inclusive, and accessible learning models. These models will have a tremendous social, cultural, and economic impact, and they will make it possible to meet some sustainable development goals set by the United Nations General Assembly.

IMAGE BY ANDRIJ BORYS ASSOCIATES, USING PHOTO BY DEAN DROBOT

https://dl.acm.org/doi/10.1145/3478281

contributed articles

DOI:10.1145/3478281

Personalized learning models can cut student dropout rates, boost student success, improve the integration of online and on-site students, better support teachers in mixed-teaching modalities, enhance accessibility, and more.

BY MARCO FURINI, OMBRETTA GAGGI, SILVIA MIRRI, MANUELA MONTANGERO, ELVIRA PELLE, FRANCESCO POGGI, AND CATIA PRANDI

Digital Twins and Artificial Intelligence

as Pillars of Personalized Learning Models

MODERN EDUCATIONAL SYSTEMS have not really evolved

who fail to complete their studies, but this is view is shortsighted. Today's educational system deficiencies will affect the welfare of tomorrow's society.

To improve current learning models, academic institutions around the world agree that the time has come to improve the world of education, moving from a traditional approach—where learning is standardized and available only to those with access to educational buildings—to a new paradigm that enables students to personalize their educational pathway, so they can progress at their own pace.[19,21] Future learning models must address key concerns, such as reducing dropout rates, supporting students with psycho-physical impairments, integrating on-site and online students, and personalizing the learning experience.

Digital twins—digital replicas of students—and artificial intelligence (AI) will be the pillars of innovation, accessibility, and personalization in future learning models.[19] The good news is that we can build these models today: AI algorithms have made great strides in recent years, and the use of technology in education has increased enormously. Indeed, while the COVID-19 pandemic has, on one hand, strongly hampered the learning process for many people around the world, it has, on the other

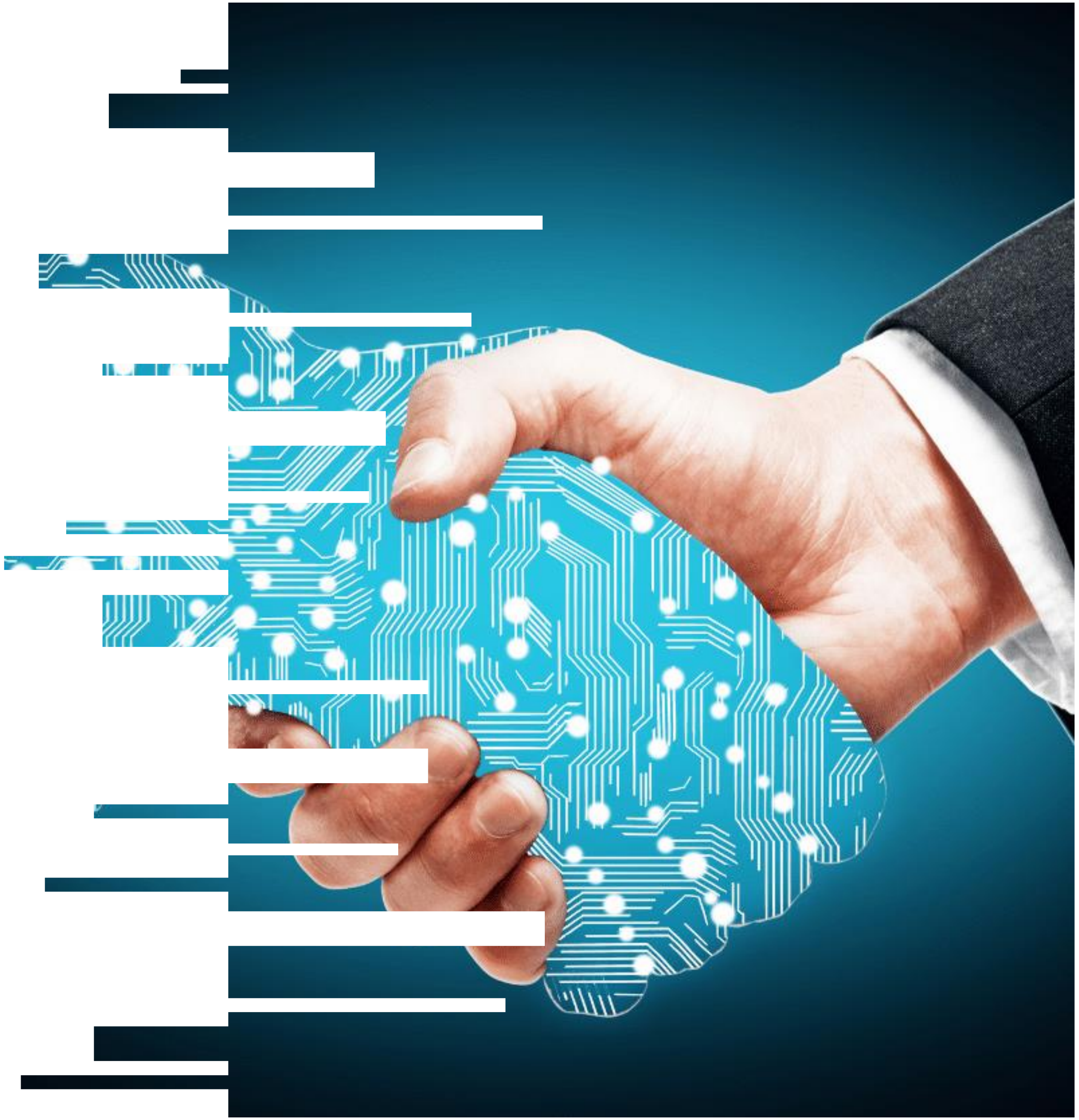
» key insights

- The time has come to revolutionize

Source: https://dl.acm.org/doi/10.1145/3478281

"A **digital twin** is a digital replica of a physical entity, and it is created by combining pieces of data from various sources."

Furini et al. (2022)



STUDENT

- Academic background
- Study habits
- Subject preferences
- Cognitive characteristics
- Learning behaviors
- Digital educational material consumption

DIGITAL TWIN

- Digital student records
- Online learning activities
- Digital learning behaviors
- Data from digital assessments
- Learner knowledge space
- Interactions with learning materials

On the Road to Adaptive Learning Systems

"Adaptive" Variables

Cognitive learning styles

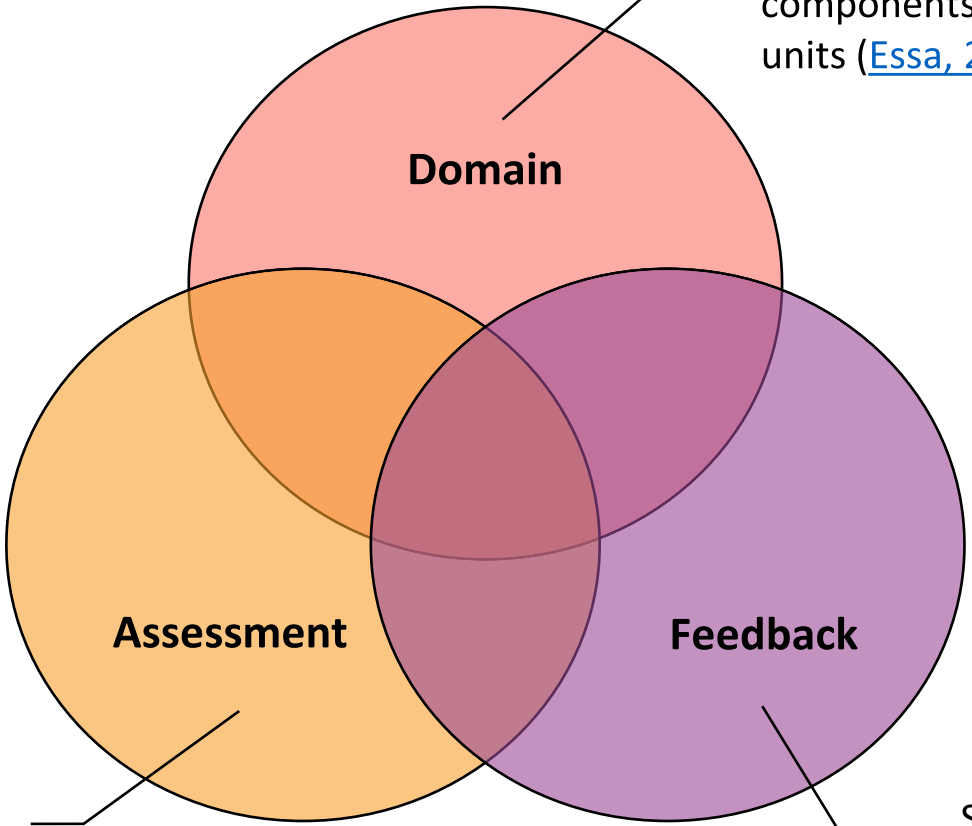
Preferences and interests

Learning progression

Demographic variables

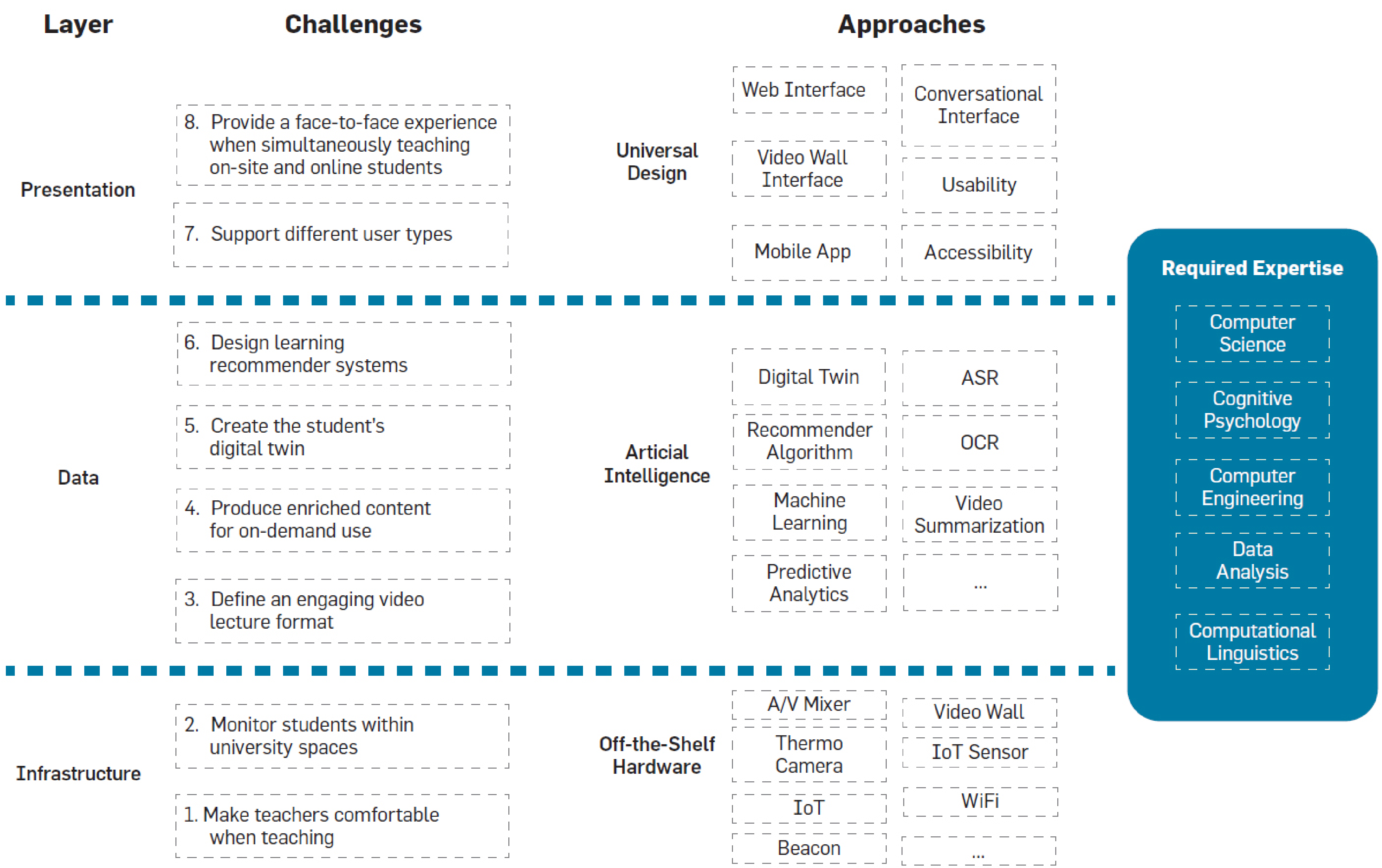
(Triantafillou et al., 2007)

Concepts, knowledge
components, or knowledge
units (Essa, 2016)

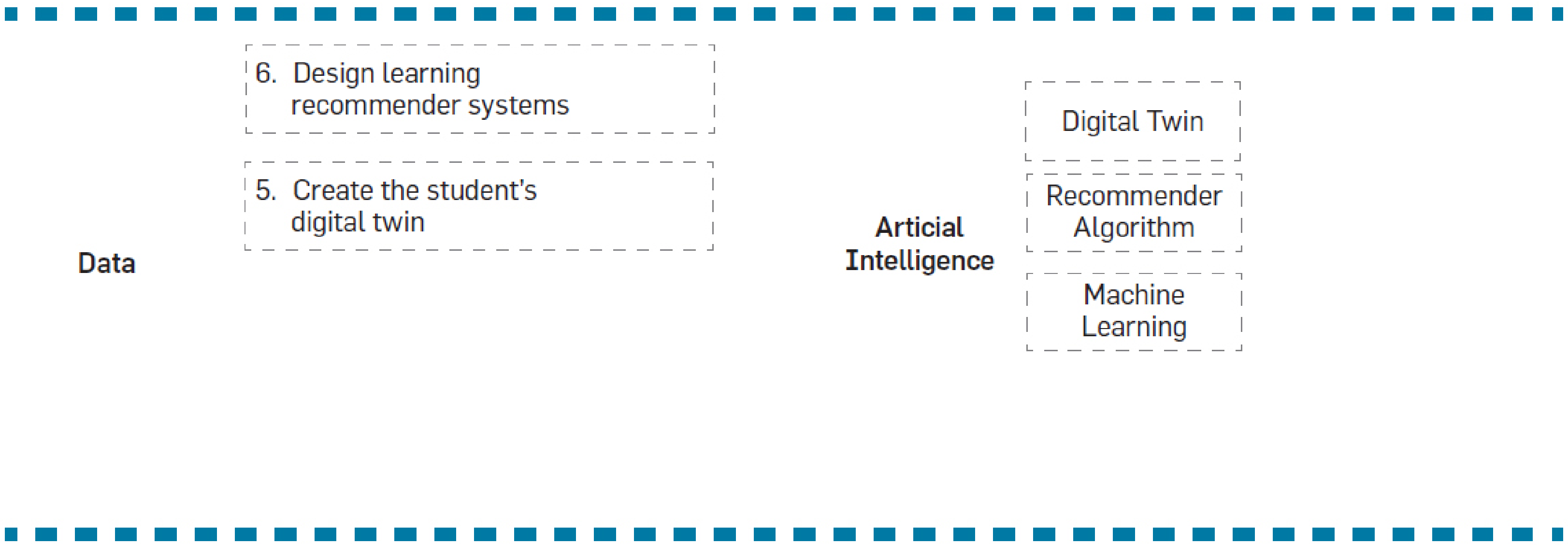


Difficulty
Time/frequency
Cognitive/metacognitive skills

Strengths/weaknesses
Actionable insights
What is next?



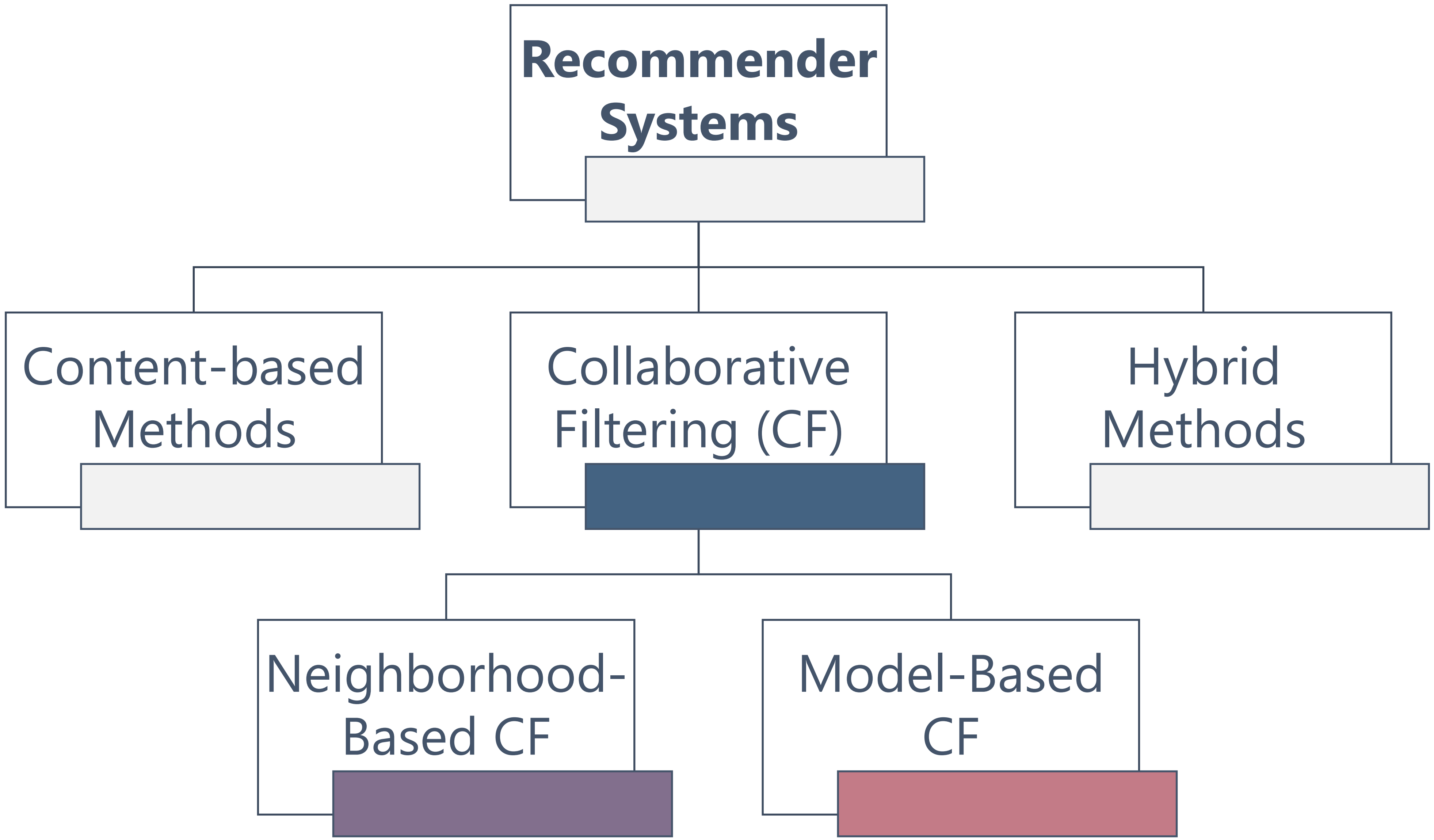
Source: https://dl.acm.org/doi/10.1145/3478281

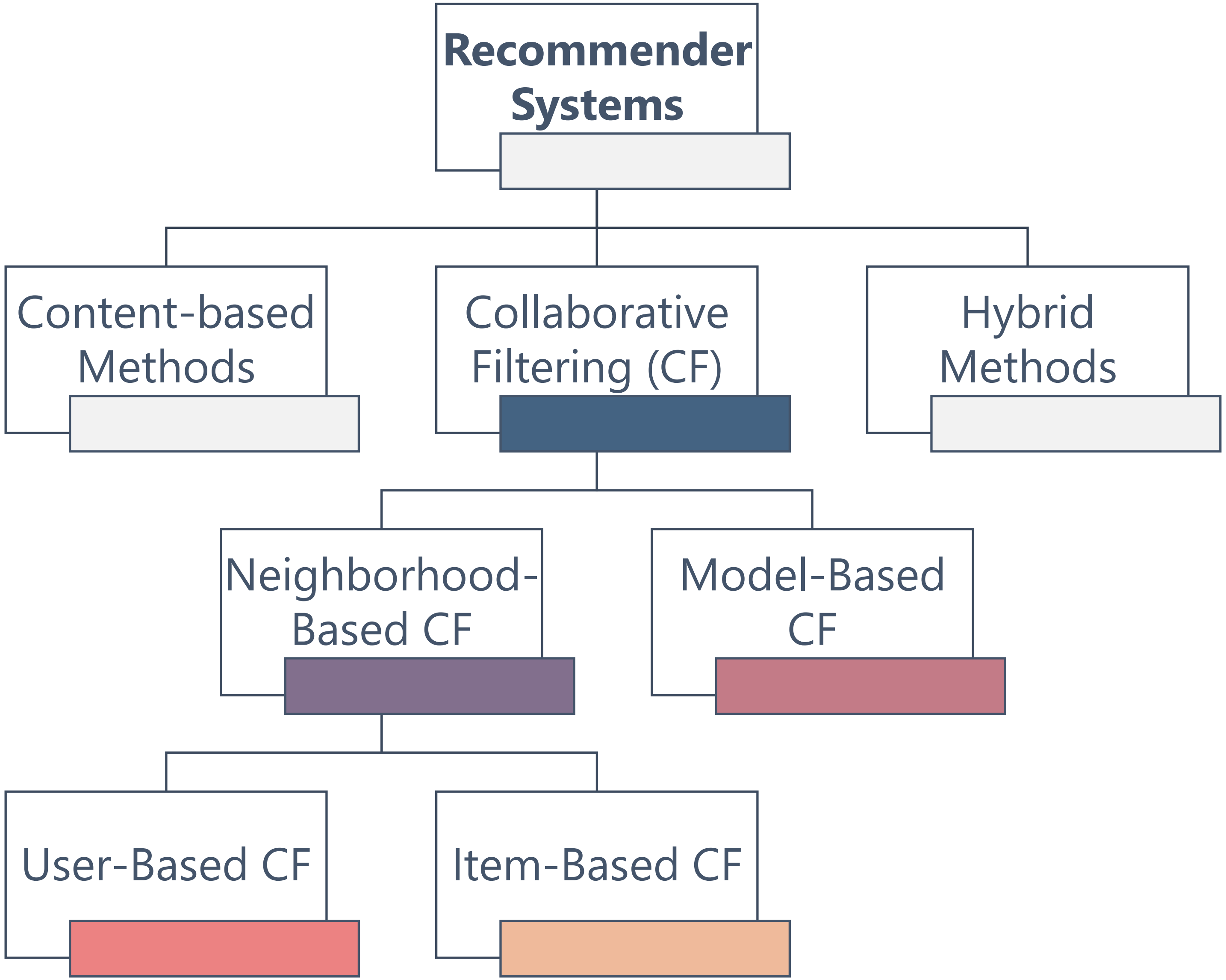


Source: https://dl.acm.org/doi/10.1145/3478281

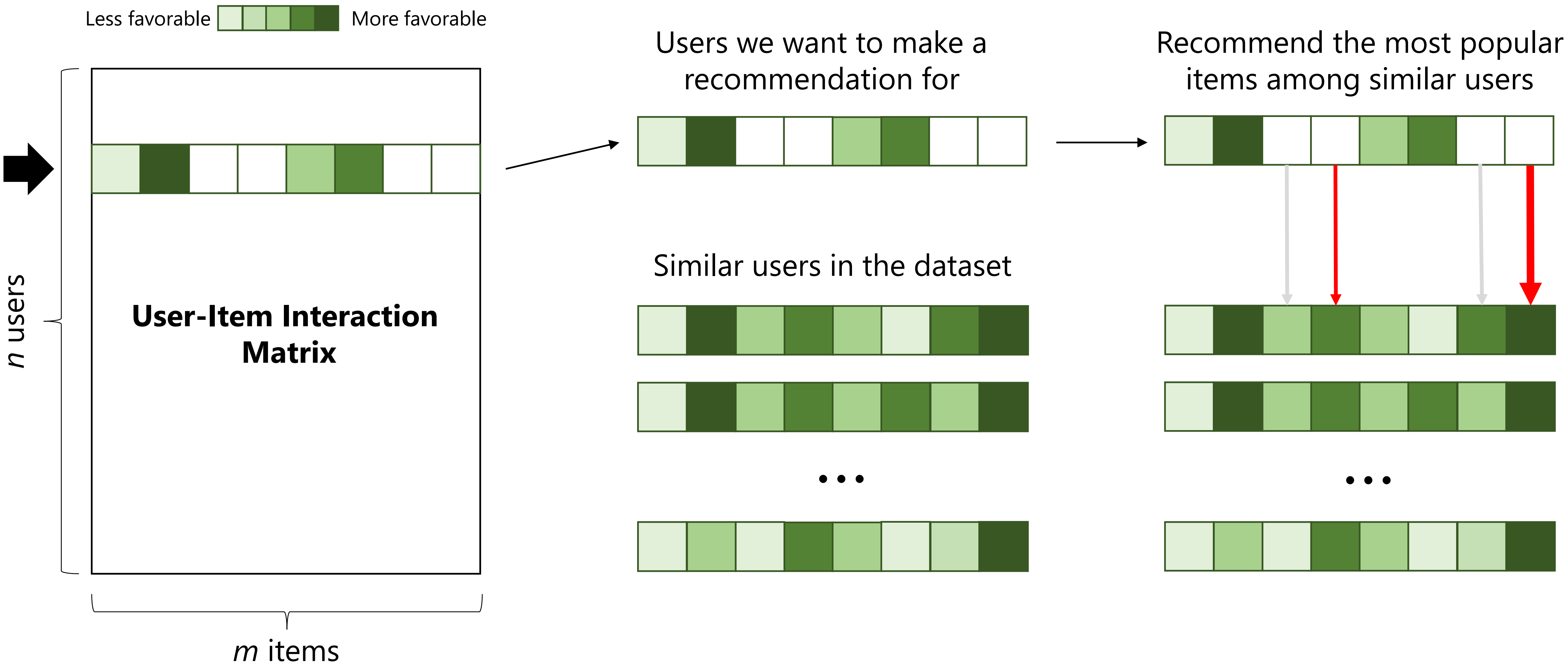
Recommender systems

"… personalized information agents that provide recommendations: suggestions for items likely to be of use to a user" (Burke, 2007)


Recommender Systems
Content-based Methods
Collaborative Filtering (CF)
Hybrid Methods
Neighborhood-Based CF
Model-Based CF


Recommender Systems
Content-based Methods
Collaborative Filtering (CF)
Hybrid Methods
Neighborhood-Based CF
Model-Based CF
User-Based CF
Item-Based CF

User-Based Collaborative Filtering



User-Based Collaborative Filtering

- **user** u_i, $i = 1, \dots, n$
- **item** p_j, $j = 1, \dots, m$
- **rating** r_{ij}

Step 1

$$sim(u_i, u_k) = \cos(u_i, u_k) = \frac{\sum_{j=1}^{m} r_{ij} r_{kj}}{\sqrt{\sum_{j=1}^{m} r_{ij}^2 \sum_{j=1}^{m} r_{kj}^2}}, \text{or}$$

$$sim(u_i, u_k) = cor(u_i, u_k) = \frac{\sum_{j=1} (r_{ij} - \overline{r_i})(r_{kj} - \overline{r_k})}{\sqrt{\sum_{j=1} (r_{ij} - \overline{r_i})^2 \sum_{j=1} (r_{kj} - \overline{r_k})^2}},$$

Step 2

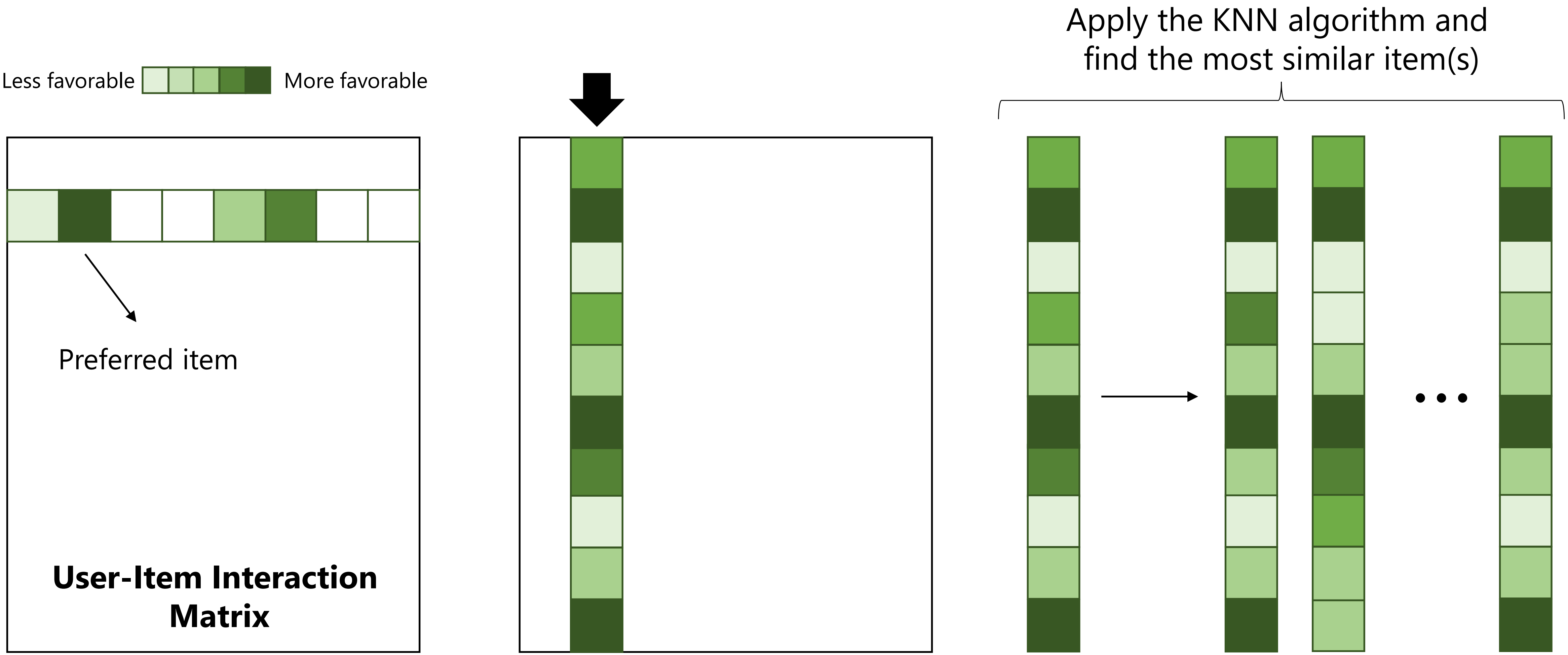
Perform *k*-nearest neighbors (KNN) to select the best neighbors of the target user (*alternatively, use a similarity threshold*)

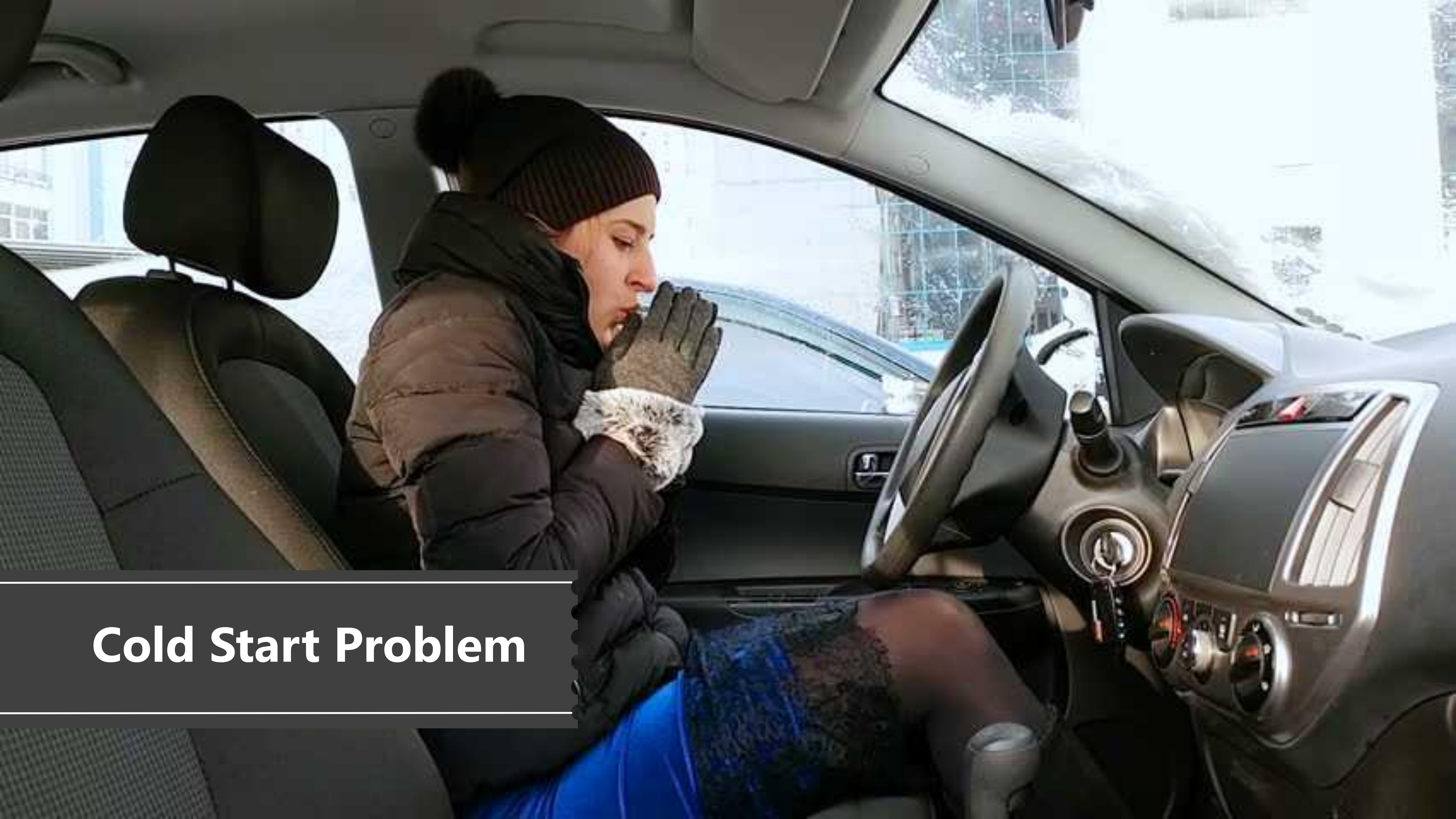
Step 3

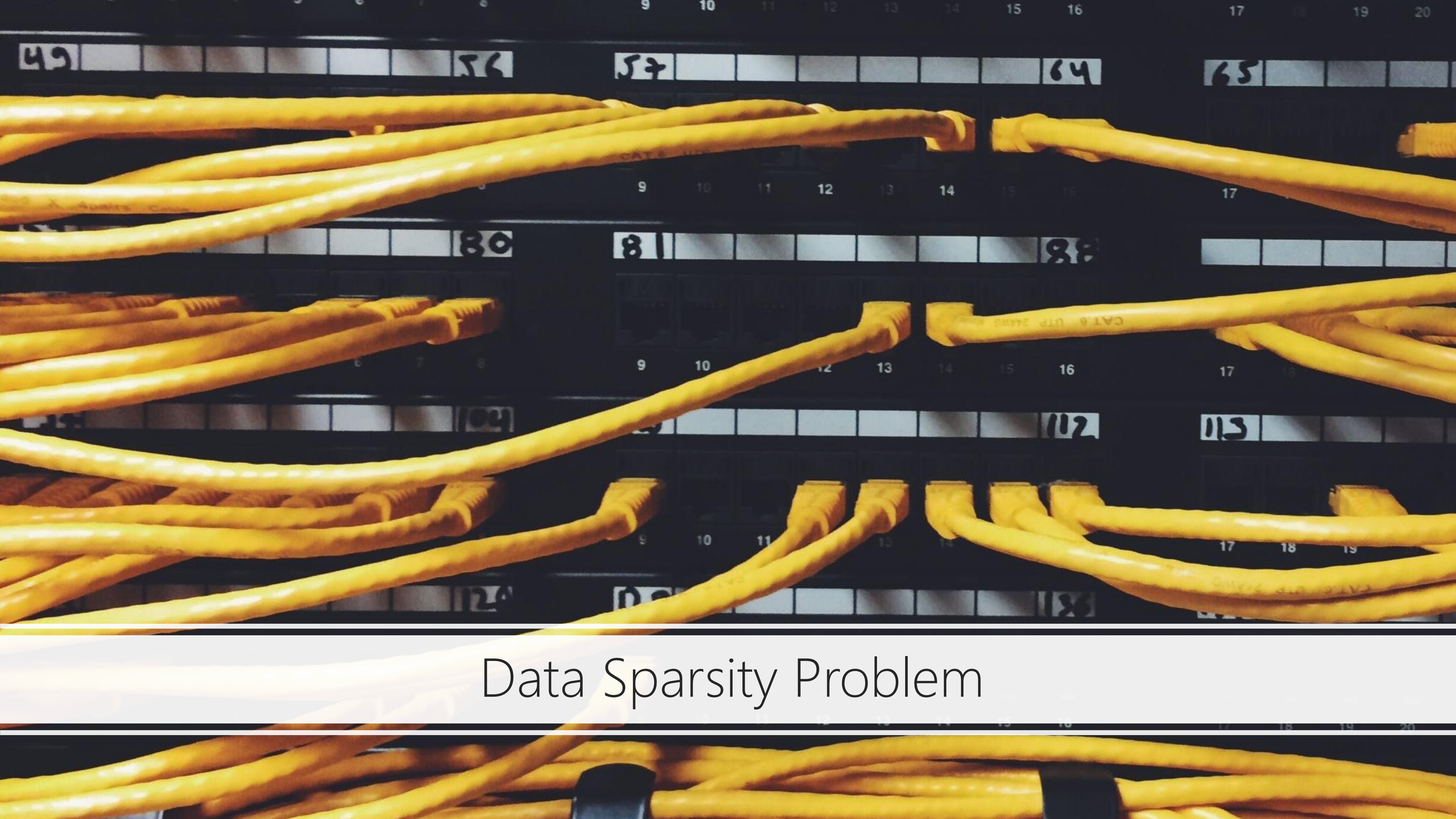
Predict an unknown rating for the target user based on the best neighbors identified in Step 2.

$$\hat{r}_{ij} = \frac{\sum_k sim(u_i, u_k) r_{kj}}{\#\ of\ ratings} \quad \text{or} \quad \hat{r}_{ij} = \overline{r_i} + \frac{\sum_k sim(u_i, u_k)(r_{kj} - \overline{r_k})}{\#\ of\ ratings}$$

Item-Based Collaborative Filtering

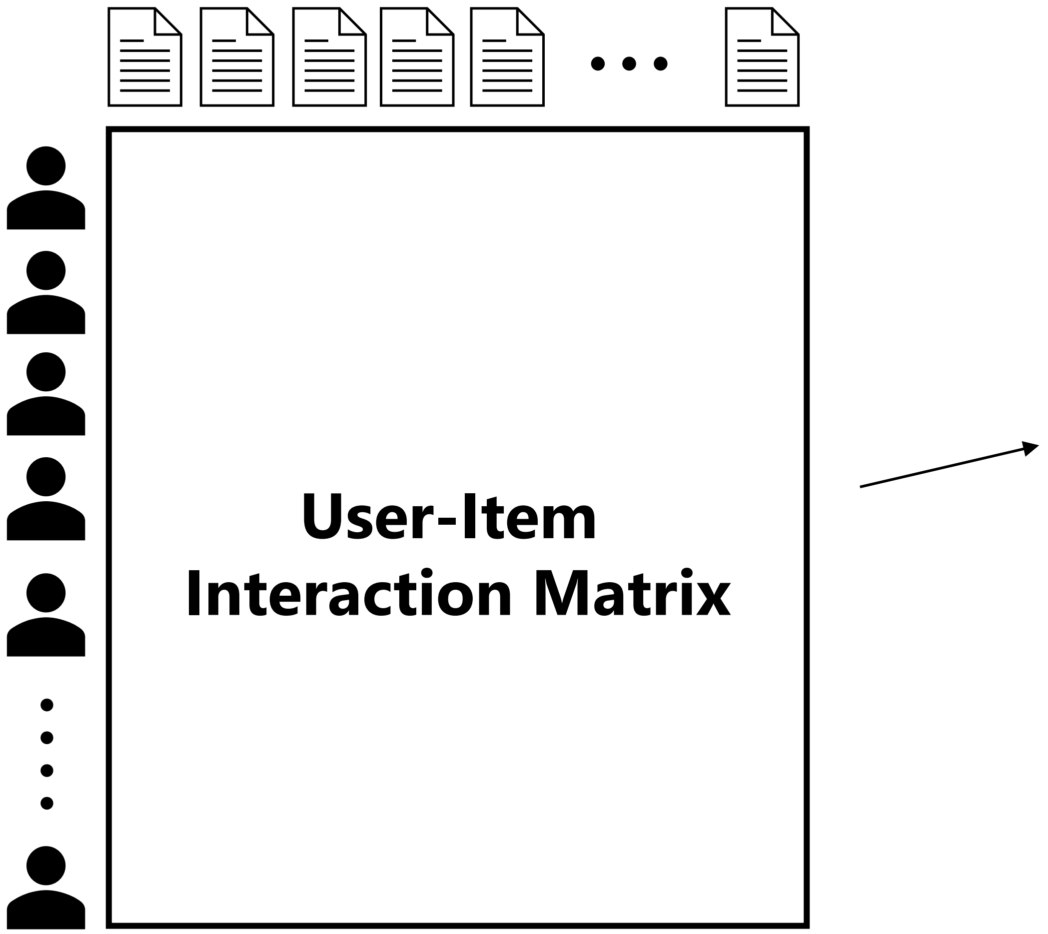



Cold Start Problem



Data Sparsity Problem

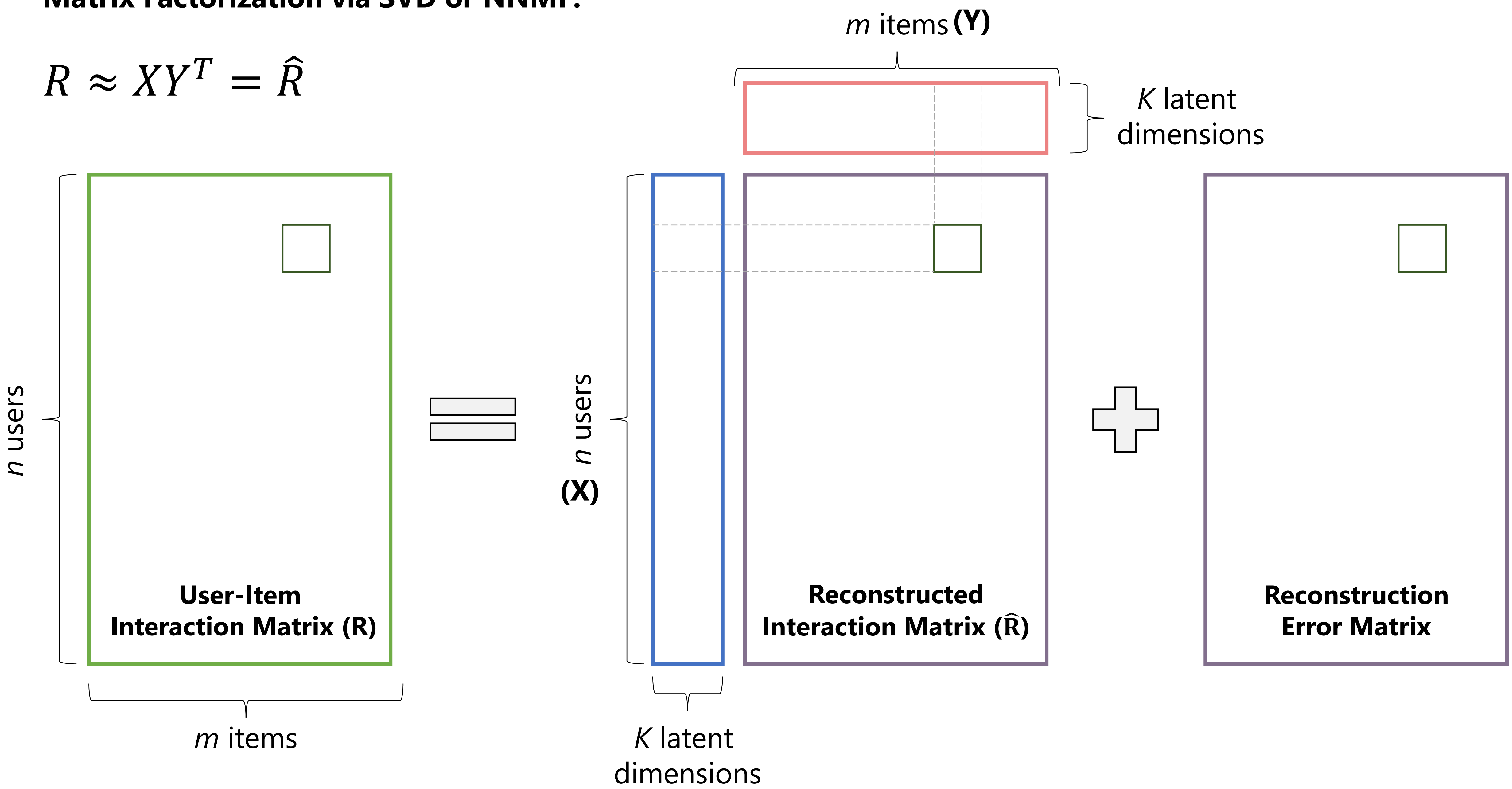
Model-Based Collaborative Filtering

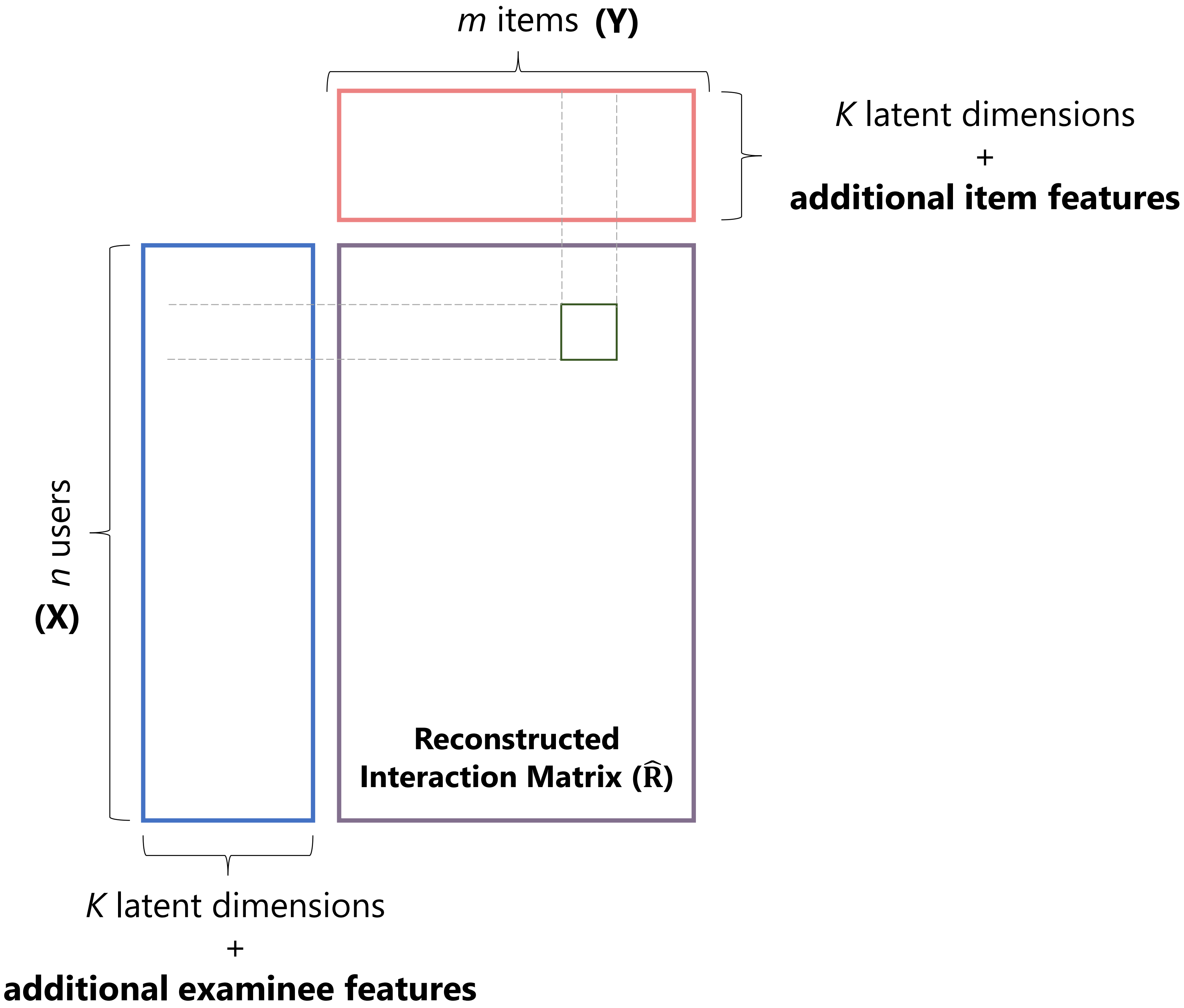


An underlying **generative** model that explains the user-item interactions.

Matrix Factorization via SVD or NNMF:

$$R \approx XY^T = \hat{R}$$



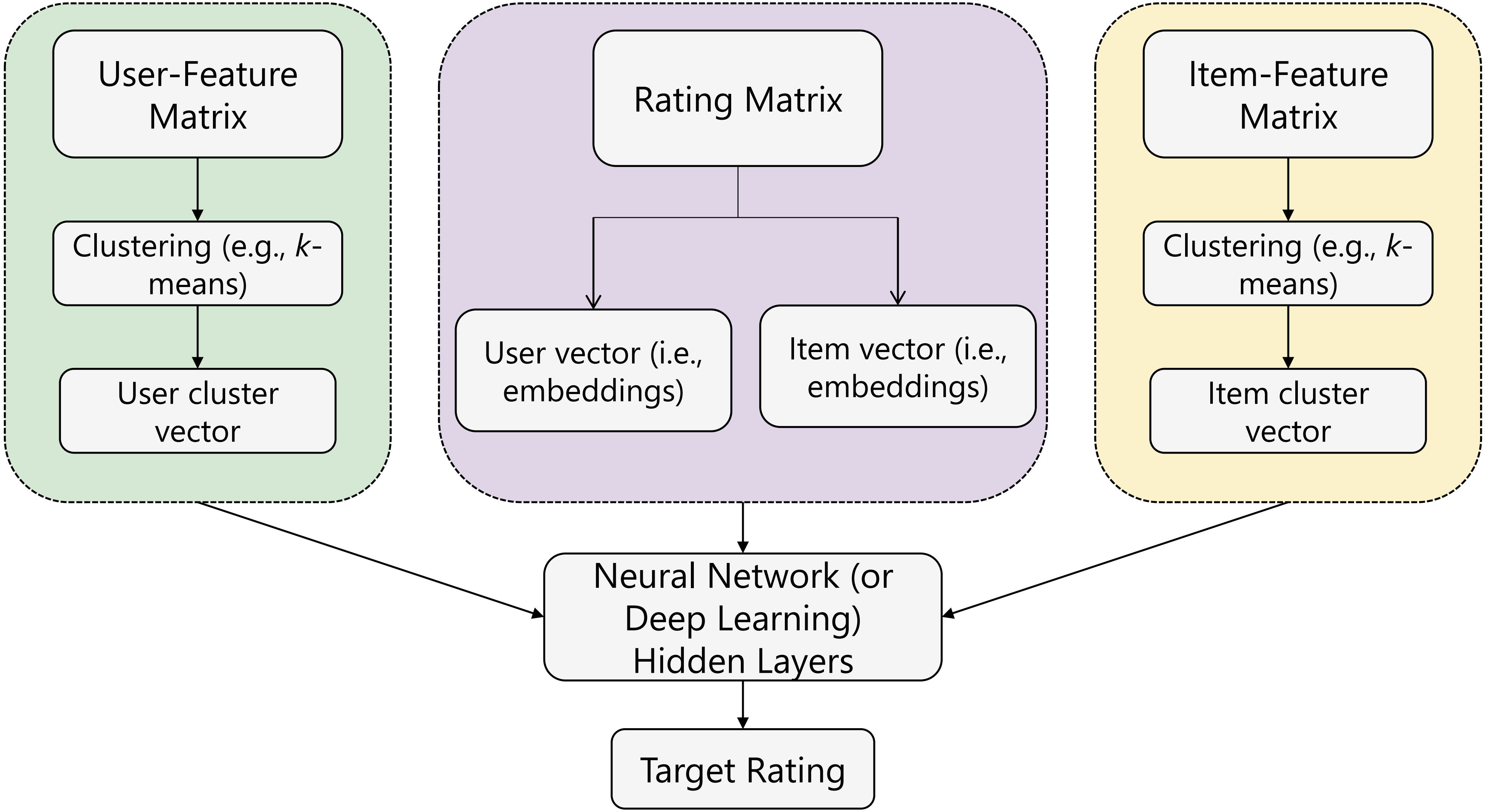


m items **(Y)**
K latent dimensions
+
additional item features
(X) *n* users
Reconstructed
Interaction Matrix ($\hat{\mathrm{R}}$)
K latent dimensions
+
additional examinee features

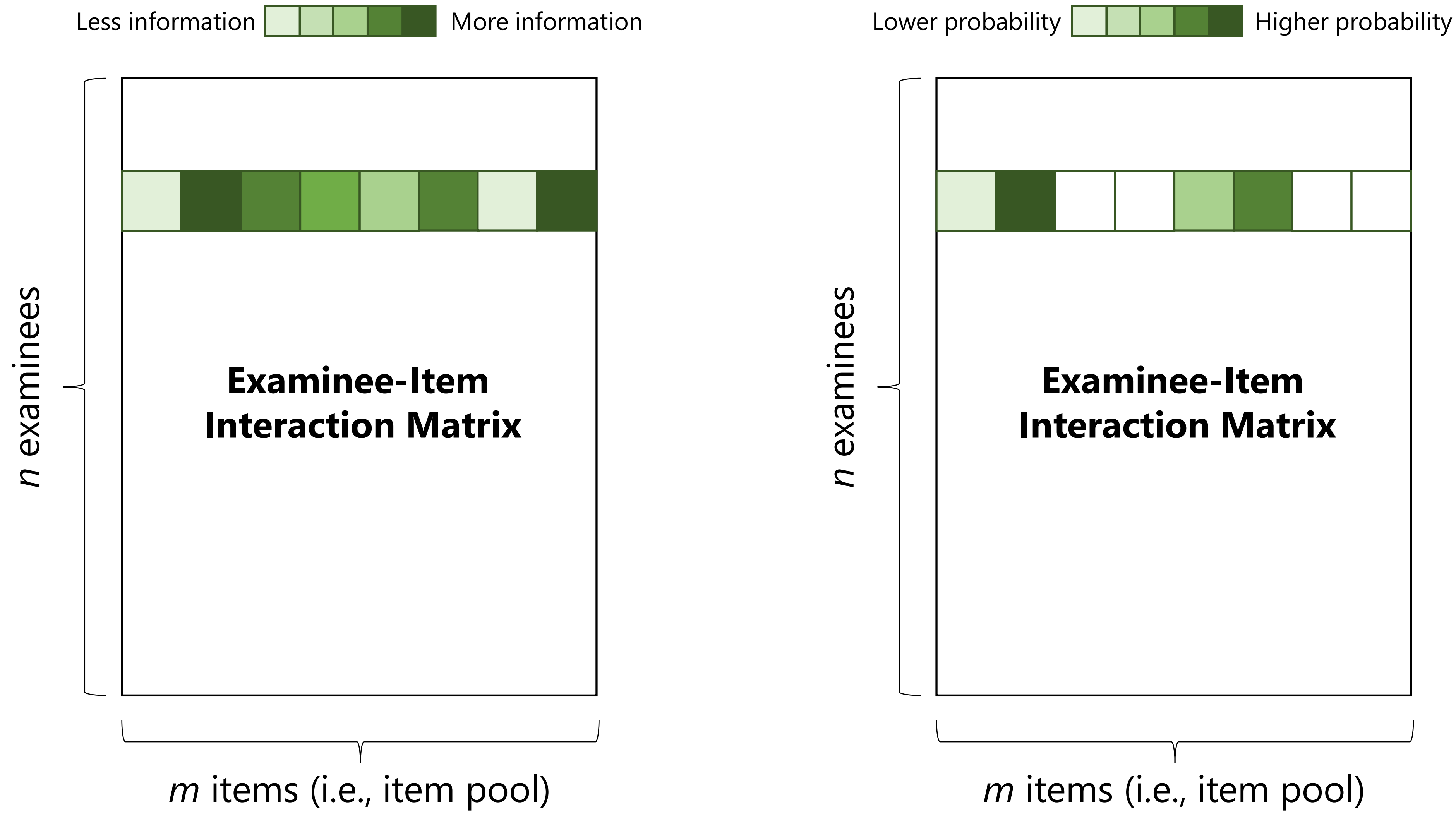
Hybrid Recommender Systems

User x Item x Context

$$\hat{R} = X \times Y \times C$$



Adaptive Testing via Recommender Systems (RS)



Item Selection With Collaborative Filtering in On-The-Fly Multistage Adaptive Testing

Applied Psychological Measurement
2022, Vol. 46(8) 690–704

DOI: 10.1177/01466216221124089
journals.sagepub.com/home/apm
SAGE

Jiaying Xiao[1] and Okan Bulut[2]

Abstract
An important design feature in the implementation of both computerized adaptive testing and multistage adaptive testing is the use of an appropriate method for item selection. The item selection method is expected to select the most optimal items depending on the examinees' ability level while considering other design features (e.g., item exposure and item bank utilization). This study introduced collaborative filtering (CF) as a new method for item selection in the *on-the-fly assembled multistage adaptive testing* framework. The user-based CF (UBCF) and item-based CF (IBCF) methods were compared to the maximum Fisher information method based on the accuracy of ability estimation, item exposure rates, and item bank utilization under different test conditions (e.g., item bank size, test length, and the sparseness of training data). The simulation results indicated that the UBCF method outperformed the traditional item selection methods regarding measurement accuracy. Also, the IBCF method showed the most superior performance in terms of item bank utilization. Limitations of the current study and the directions for future research are discussed.

Keywords
collaborative filtering, multistage adaptive testing, item selection, measurement accuracy

Item Selection for On-the-Fly Multi-Stage Adaptive Testing

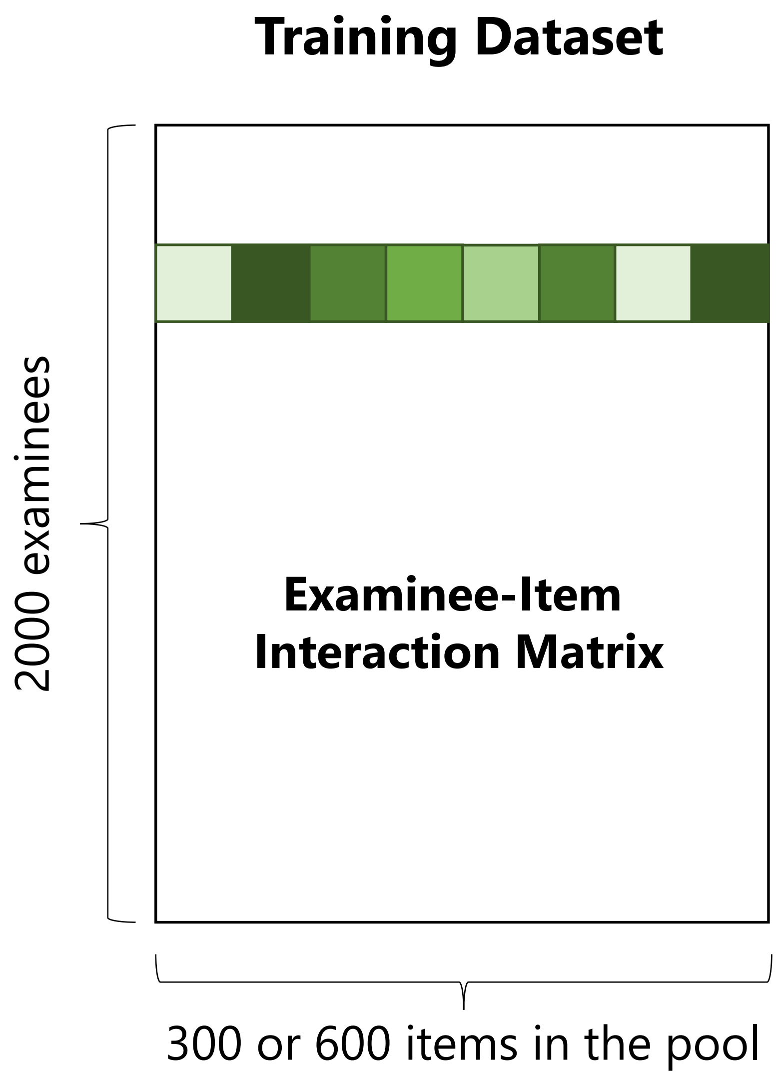
Stage 1: A pre-assembled module

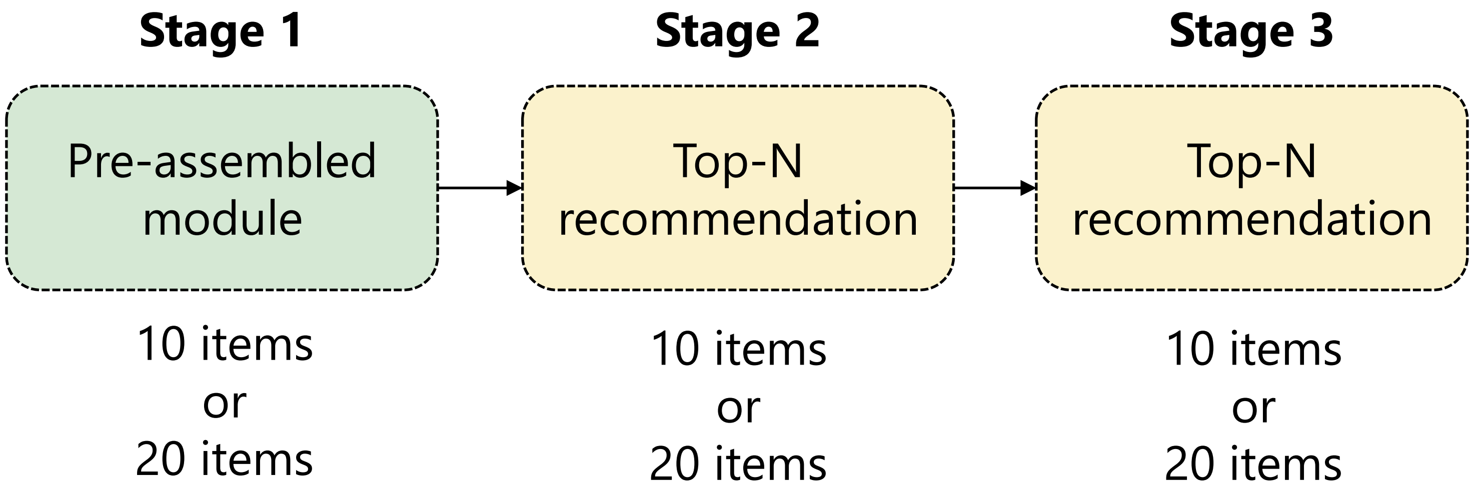
Stages 2 & 3: On-the-fly assembled modules via user-based and item-based collaborative filtering

No additional item or user feature used

Item selection using the recommenderlab package in R

https://doi.org/10.1177/01466216221124089





On-the-fly module assembly using:

- User-based CF (UBCF)
- Item-based CF (IBCF)
- Maximum Fisher information (MFI)

$\theta = [-3, -2.6, \ldots, 2.6, 3] \rightarrow$ 500 simulated examinees per ability

Item Bank Size	Method	30-item design				60-item design			
		Bias	RMSE	Reliability	Proportion of unused items	Bias	RMSE	Reliability	Proportion of unused items
300 items	UBCF	-0.019	0.362	0.970	57%	-0.013	0.277	0.981	31%
	IBCF	-0.007	0.428	0.962	38%	0.002	0.341	0.974	27%
	MFI	-0.016	0.369	0.969	59%	-0.013	0.279	0.981	32%
600 items	UBCF	0.044	0.464	0.957	75%	0.015	0.335	0.975	57%
	IBCF	-0.010	0.341	0.973	66%	-0.010	0.256	0.984	51%
	MFI	-0.012	0.365	0.970	76%	-0.011	0.270	0.982	59%

Personalized Scheduling for Adaptive Tests

What is the optimal test schedule for each student based on their learning progress?

Progress monitoring with Renaissance's Star Reading and Star Math adaptive tests for K-12

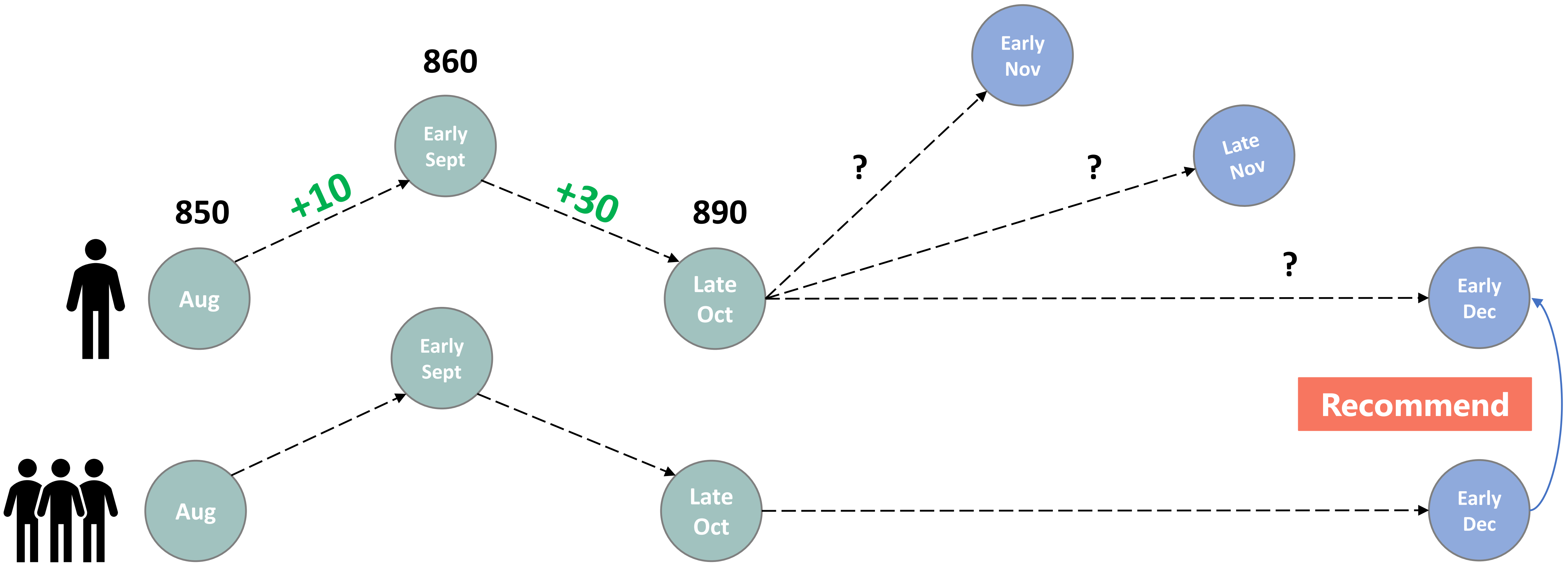
Grade 2 (n = 668,324) and Grade 4 (n = 727,147)

2 to 18 test administrations per student

(Bulut, Shin, & Cormier, 2022; Shin & Bulut, 2022; Bulut, Cormier, & Shin, 2020)

User-Based Collaborative Filtering with Dijkstra's Shortest Path First Algorithm

- **Maximize** the positive and absolute score change between test administrations
- **Minimize** the number of test administrations



- Find similar students (with max score change + fewest test administrations)
- Select the most similar students based on Euclidean distance and recommend their schedule

Standard Practice = Schedules Determined by Teachers **RS** = Recommender System

Evaluation Criteria	Grade 2			Grade 4		
	Standard Practice		RS	Standard Practice		RS
Average number of tests administered	5.42	→	**3.51**	5.37	→	**3.84**
Average score change between tests	8.32	→	**12.25**	3.49	→	**4.63**
Range of tests required	(1, 18)	→	(1, 5)	(1, 17)	→	(1, 6)
Non-recommendable cases	-		0.05%			0.10%

Concluding Remarks

- Recommender systems can help us take a more holistic approach to designing adaptive learning systems.
  - Shifting the focus from “examinees” to “learners”
  - ~~Less psychometrics & more AI~~; an amalgamation of psychometrics and AI

- Using the auxiliary information about learners as “adaptive variables”
  - Enhanced adaptivity and precision (especially when there is no prior information on learners)

  - Prioritizing the text-taker experience (TTX) in decision-making ([Duolingo, 2021](#))

  - Driving innovation in the cycle of domain, assessment, and feedback

Future Directions

- Recommender systems can involve real-time process data (e.g., response time) to consider test-taking engagement in adaptive testing.

- Recommender systems can be used with other psychometric models such as Bayesian Knowledge Tracing to measure mastery of content domain.

- Recommender systems utilizing deep learning algorithms can model both responses and sequential action data in adaptive learning environments.
  - Chen et al. (2019)'s Behavior Sequence Transformer Model
  - Wu et al. (2017)'s Recurrent Recommender Networks

Thank You!

For questions/comments:

bulut@ualberta.ca