Dual Learning: Bridging Knowledge Between LLM and Recommender System

Pan Li
Assistant Professor, Georgia Institute of Technology
ISIR-eCom 2024

Georgia Tech
Scheller College of Business

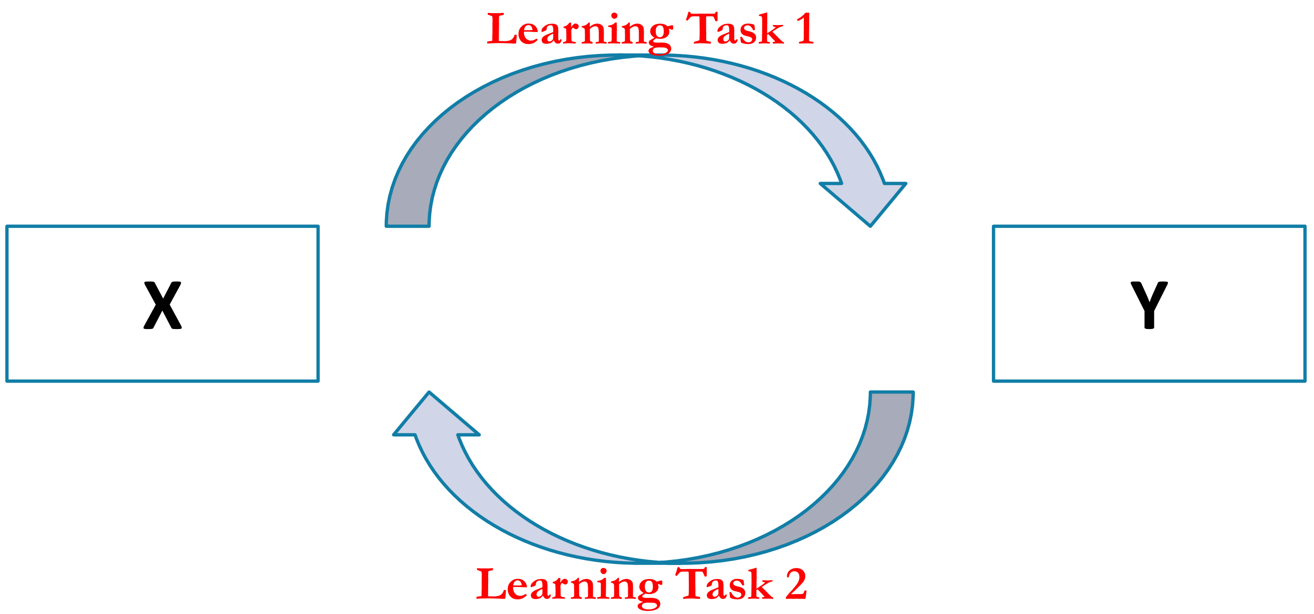
Agenda

- **Preliminaries in Dual Learning**
- **Dual Learning in Cross-Domain Recommendations**
- **Dual Learning, LLM, and Aspect-Based Recommendations**

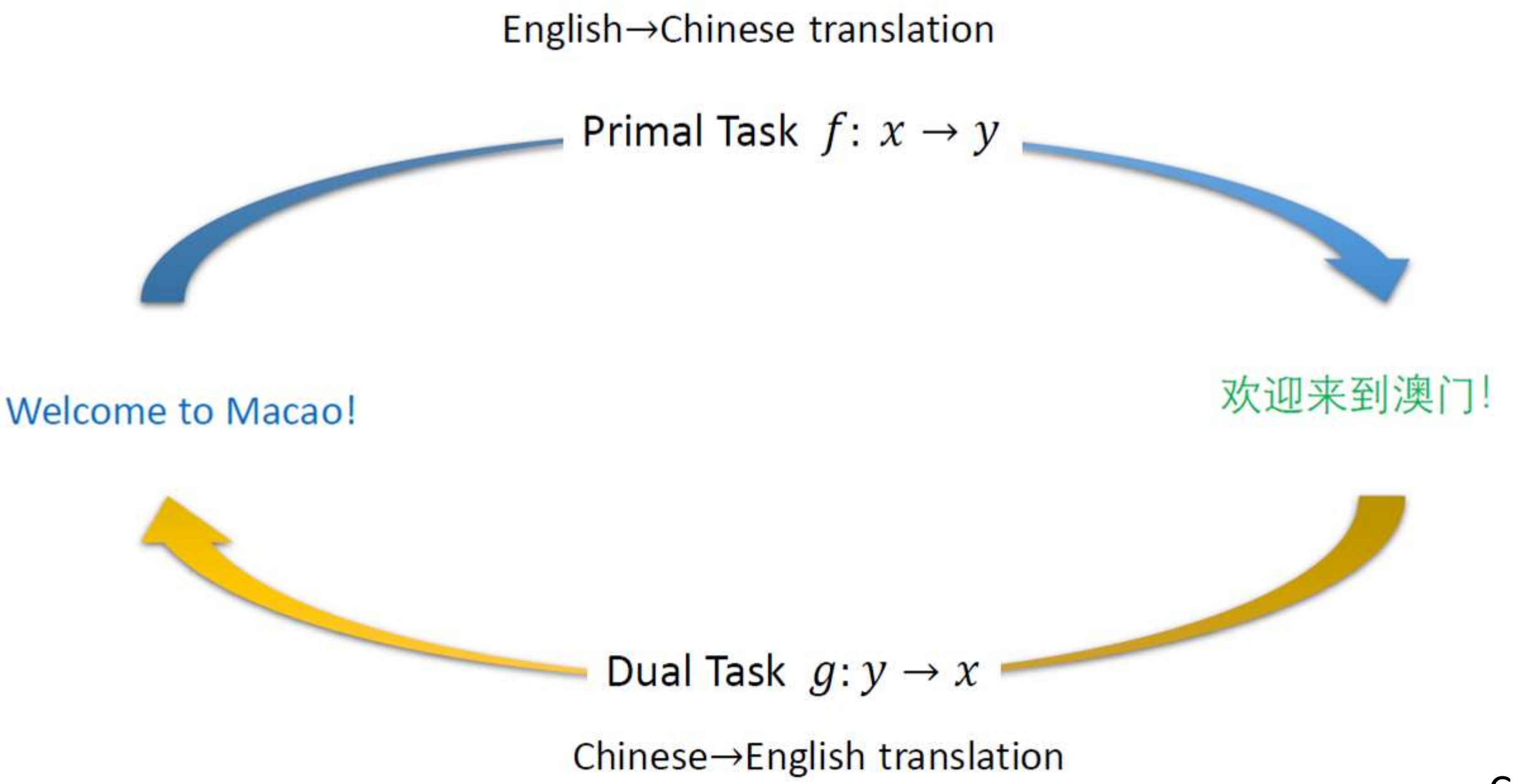
Motivation: Structural Duality

Definition

Two machine learning tasks are of structural duality if one learning task maps from space X to space Y, and the other learning task maps from space Y to space X.

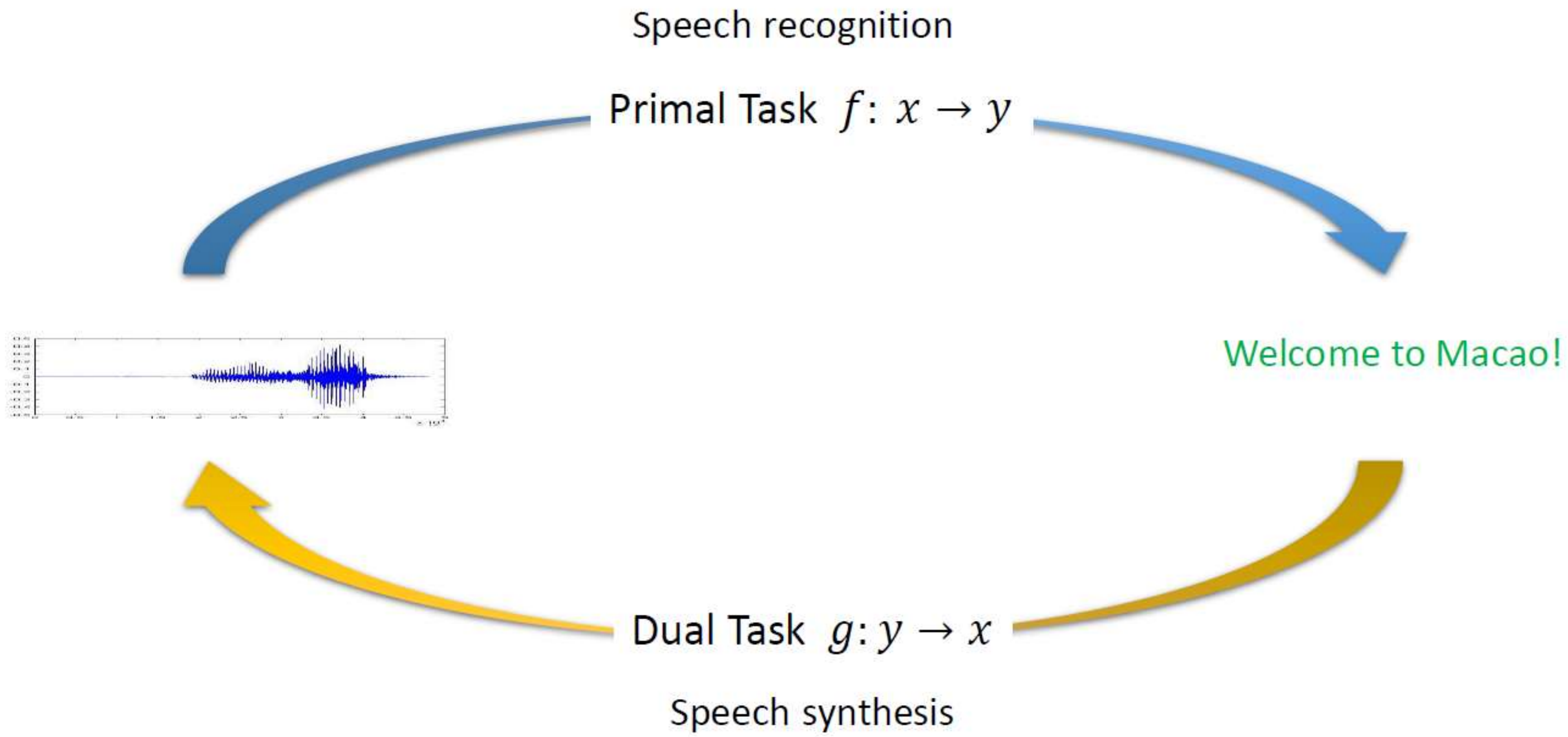


Example: Machine Translation



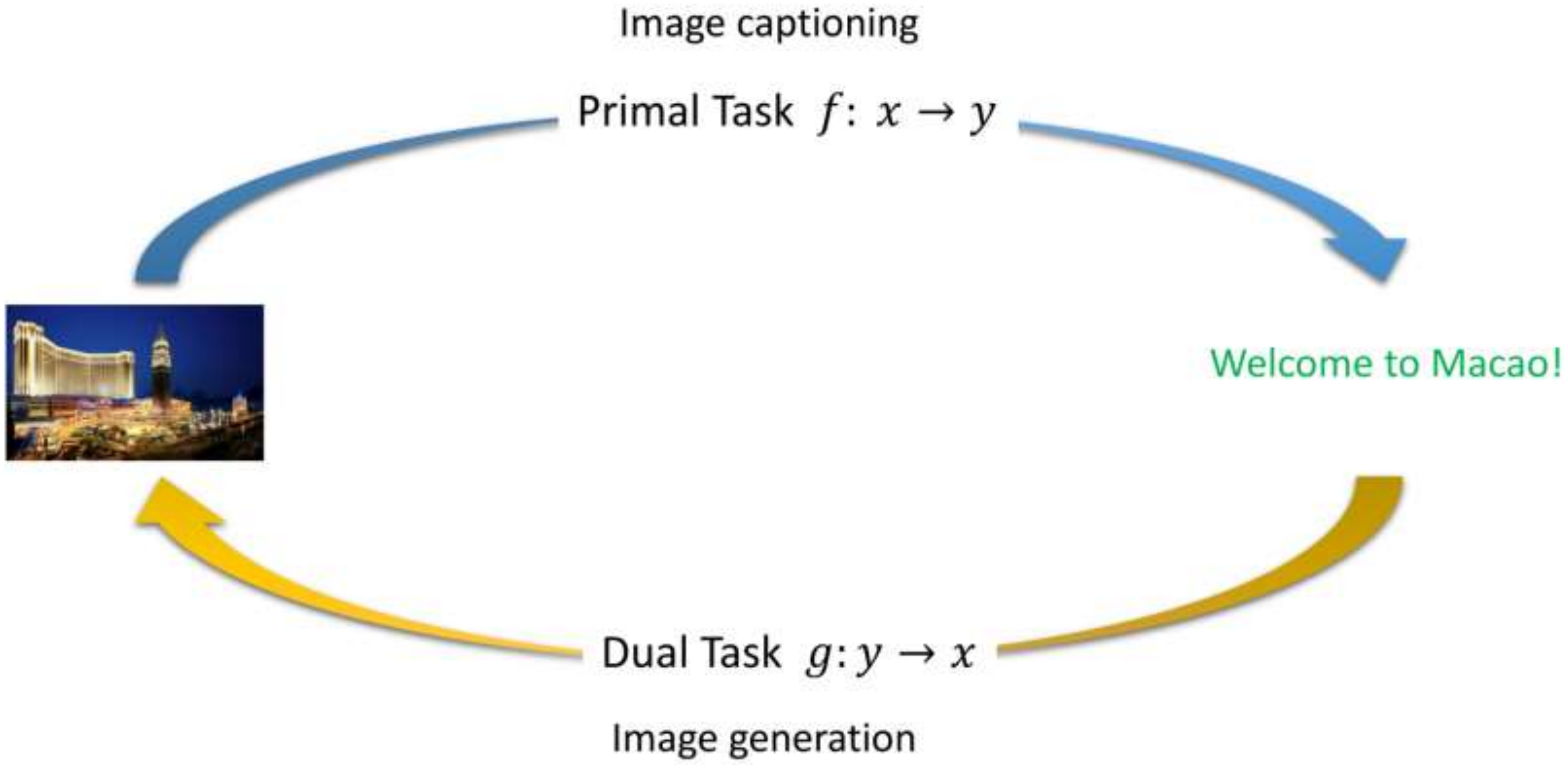
Credit: (Qin & Xia 2019)

Example: Speech Processing



Credit: (Qin & Xia 2019)

Example: Image Processing



Credit: (Qin & Xia 2019)

Structural Duality in AI

- **Structural Duality is very common in AI applications**

AI Application	X->Y	Y->X
Machine Translation	Translation from Language EN to CH	Translation from Language CH to EN
Speech Processing	Speech Recognition	Text-to-Speech
Image Processing	Image Captioning	Image Generation
Conversation	Question Answering	Question Generation
Search Engine	Query-Document Matching	Query/Keyword Suggestion

How Can We Exploit Structural Duality in AI Applications?

Dual Learning

- Bidirectionally transfers information/knowledge/parameters between the primal task and the dual task.
- Optimizes simultaneously to achieve optimal performance for both tasks
- Bayes Theorem:

$$\boldsymbol{P}(\mathbf{x}, \mathbf{y}) = \boldsymbol{P}(\boldsymbol{x})\boldsymbol{P}(\mathbf{y}|\mathbf{x}; \mathbf{f}) = \boldsymbol{P}(\boldsymbol{y})\boldsymbol{P}(\boldsymbol{x}|\boldsymbol{y}; \boldsymbol{g})$$

- Dual Optimization:

$$\textbf{objective 1: } \min_{\boldsymbol{\theta}_{XY}} \frac{1}{|D|} \sum_{(x,y)\in D} L_1(f(x, \theta_{XY}), y)$$

$$\textbf{objective 2: } \min_{\boldsymbol{\theta}_{YX}} \frac{1}{|D|} \sum_{(x,y)\in D} L_2(g(y, \theta_{YX}), x)$$

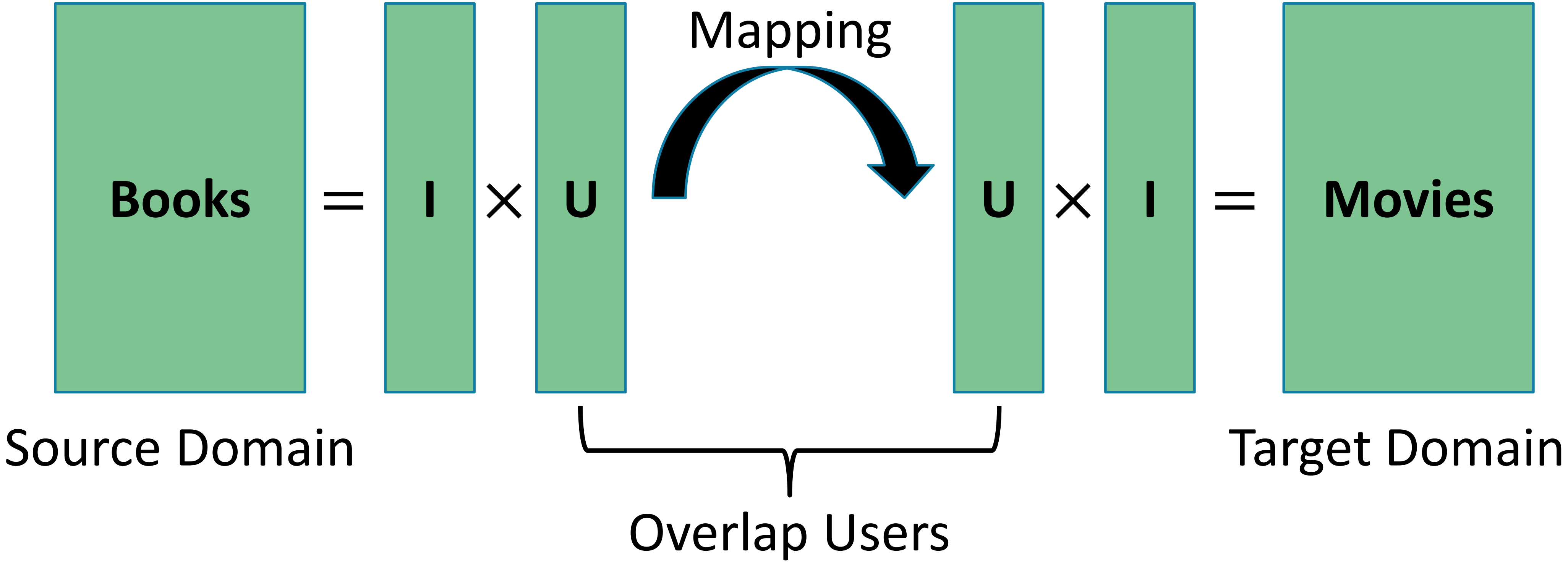
$$\textbf{s.t. } P(x, y) = P(x)P(y|x; f) = P(y)P(x|y; g), \ \forall\ (x, y) \in D$$

Application I: Cross-Domain RecSys

Suppose we know the user preferences in the book domain…

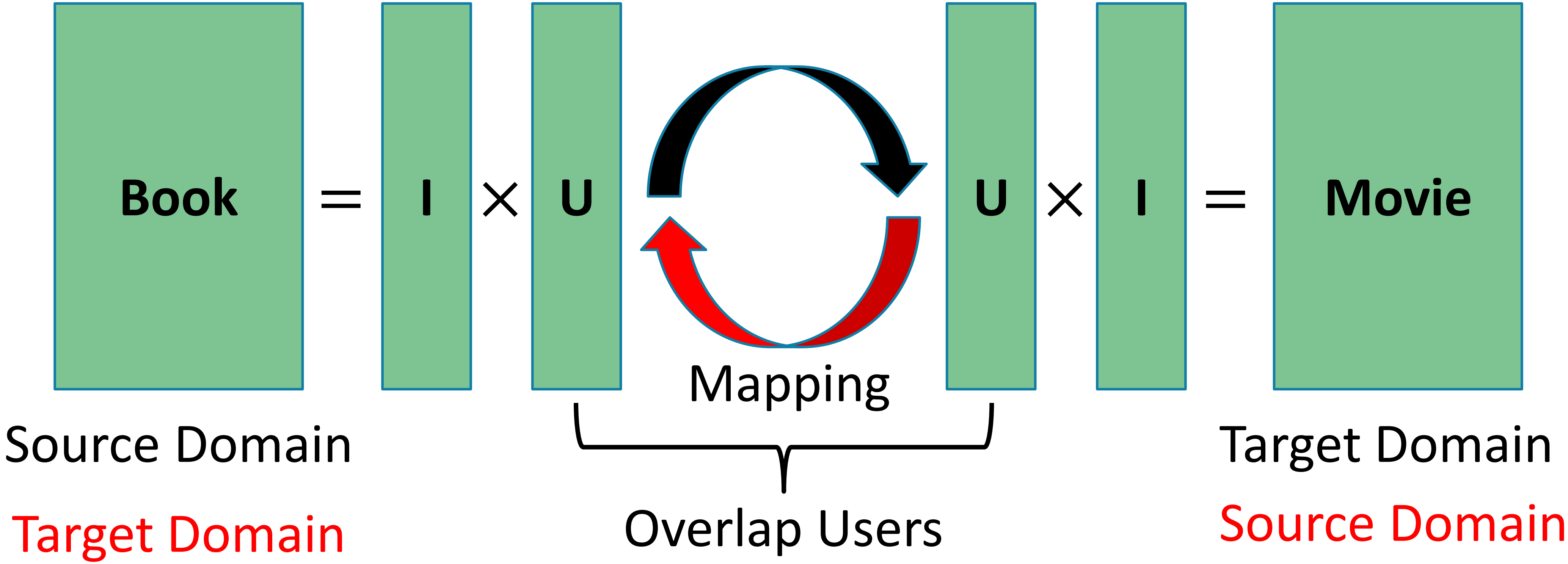
How to estimate the user preferences in the movie domain?

Transfer Learning for CDR



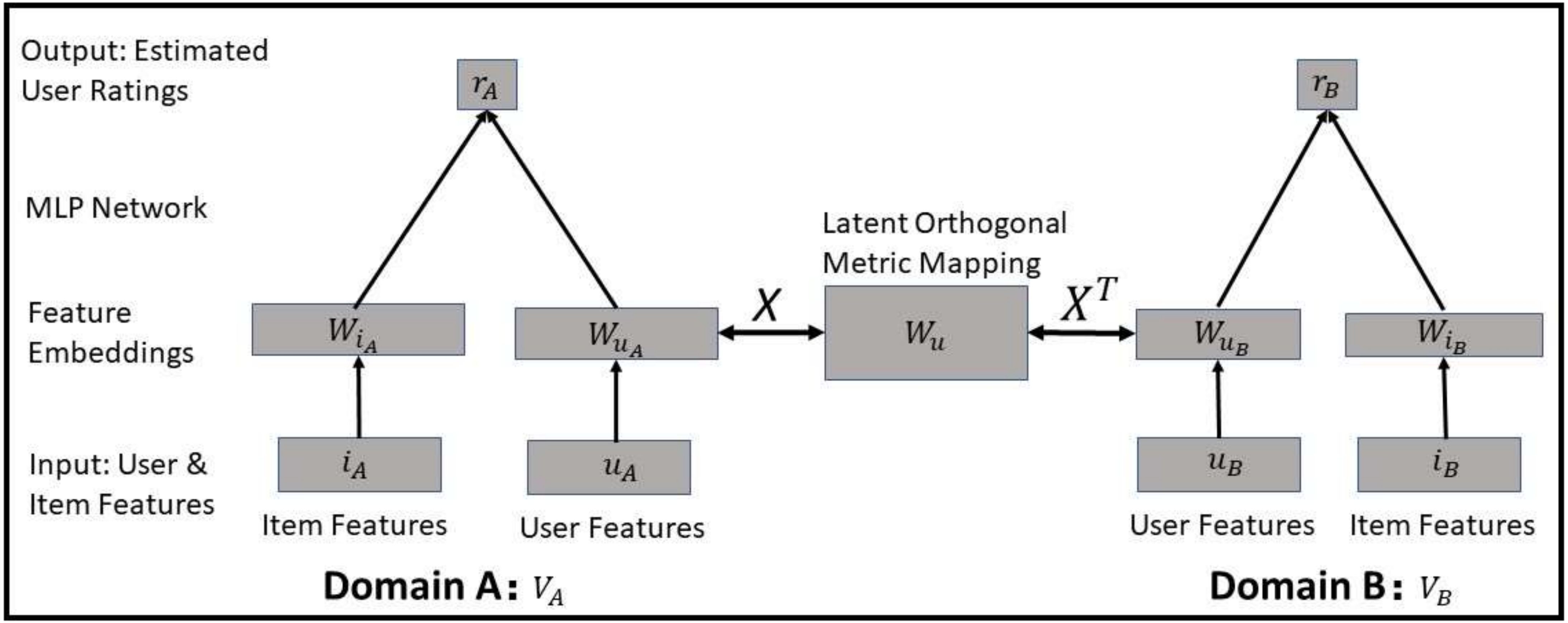
<u>Key Idea: apply dual learning to cross-domain recommendations!</u>

Solution: Dual Learning



Improving recommendation performance in one domain would also lead to improvement in the other domain!

Model [WSDM 2020, TKDE 2021]



Latent Orthogonal Metric Mapping

- Learns the bidirectional orthogonal mapping (X, X^T) between user embeddings across different domains.
- Minimize the Euclidean distance in the latent space

$$L_{o_A} = argmin_X \sum_{\{W_{ou_A}, W_{ou_B}\} \in \{ou_A, ou_B\}} \left| X W_{ou_A} - W_{ou_B} \right|^2 \quad L_{o_B} = argmin_X \sum_{\{W_{ou_A}, W_{ou_B}\} \in \{ou_A, ou_B\}} \left| W_{ou_A} - X^T W_{ou_B} \right|^2$$

Orthogonality is important because it

- preserves similarities between user embeddings across different latent spaces.
- automatically derives the inverse mapping function.

Experiments: Data

Dataset: collected from an online recommendation service for books, movies, music

Contains rich information of user features and item features:

- User (Gender, Age, Movie Taste, Residence, Preference, Usage, Marital Status, Personality)
- Book (Category, Title, Author, Publisher, Language, Country, Price, Date)
- Movie (Genre, Title, Director, Writer, Runtime, Country, Rating, Votes)
- Music (Listener, PlayCount, Artist, Album, Tag, Release, Duration, Title)

Domain	Book	Movie	Music
# of Users	804,825	959,502	45,962
# of Items	182,653	79,866	183,114
# of Ratings	223,007,805	51,269,130	2,536,273
Sparsity	0.0157%	0.0669%	0.0301%

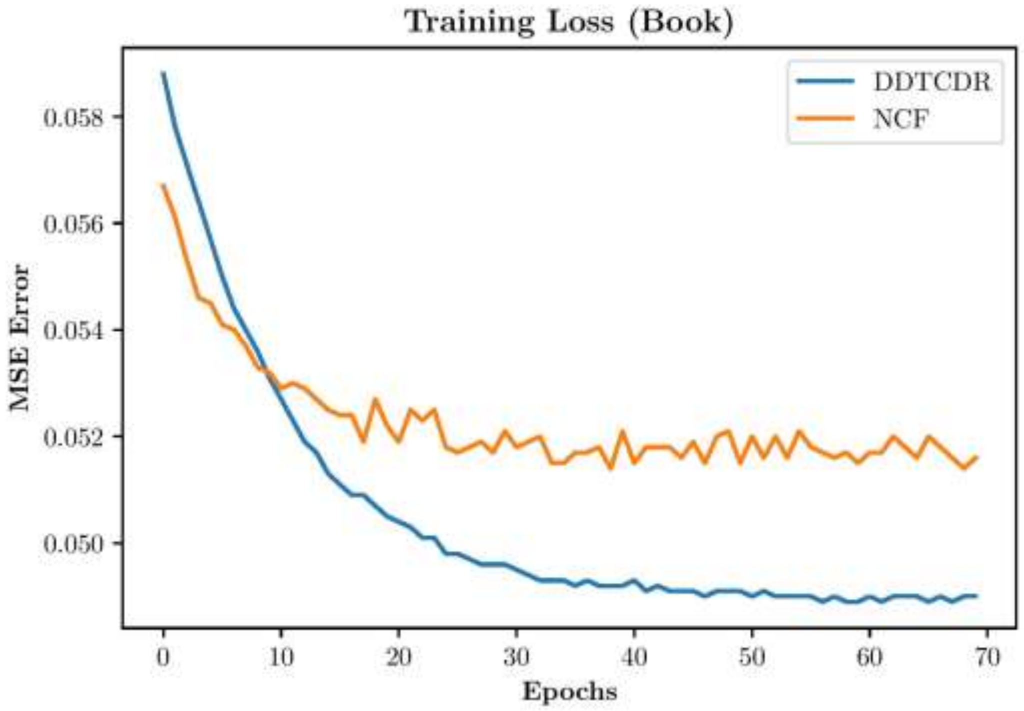
Experiments: Baselines

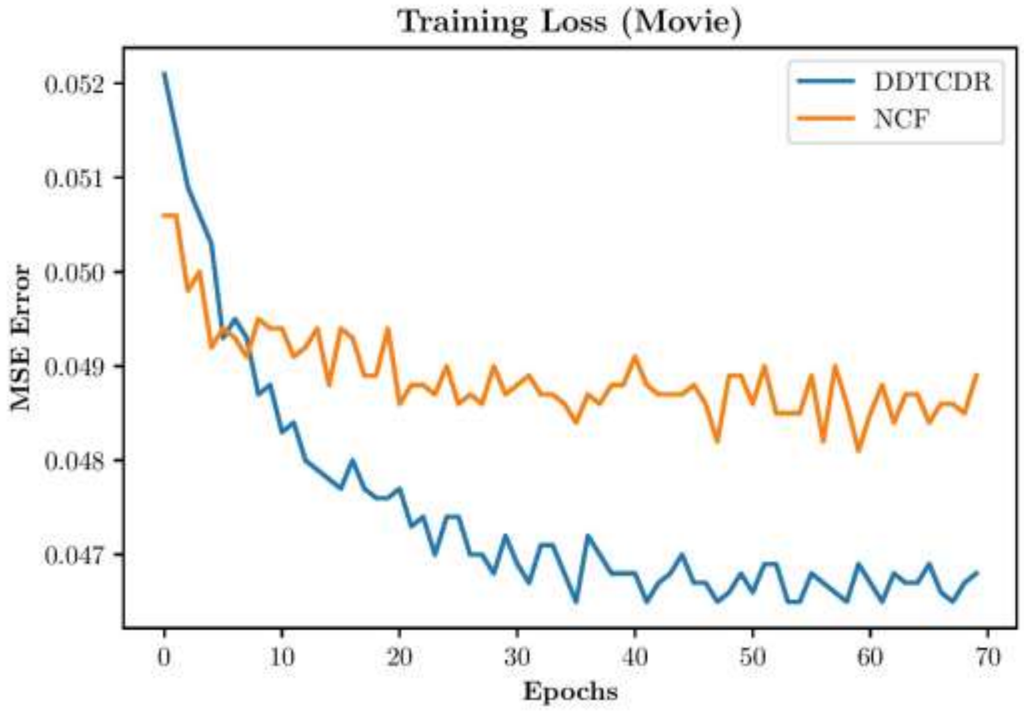
- Baseline Methods:
  - **CCCFNet**: Cross-domain Content-boosted Collaborative Filtering neural NETwork (Lian et al. 2017)
  - **CDFM**: Cross Domain Factorization Machine (Loni et al. 2014)
  - **CoNet**: Collaborative Cross Network (Hu et al. 2018)
  - **CMF**: Collective Matrix Factorization (Singh & Gordon, 2008)
  - **NCF**: Neural Collaborative Filtering (He et al. 2018)
- Conducted record-stratified 5-fold cross validation
- Evaluated performance using RMSE, MAE, Precision and Recall metrics

Results: Book/Movie Domains

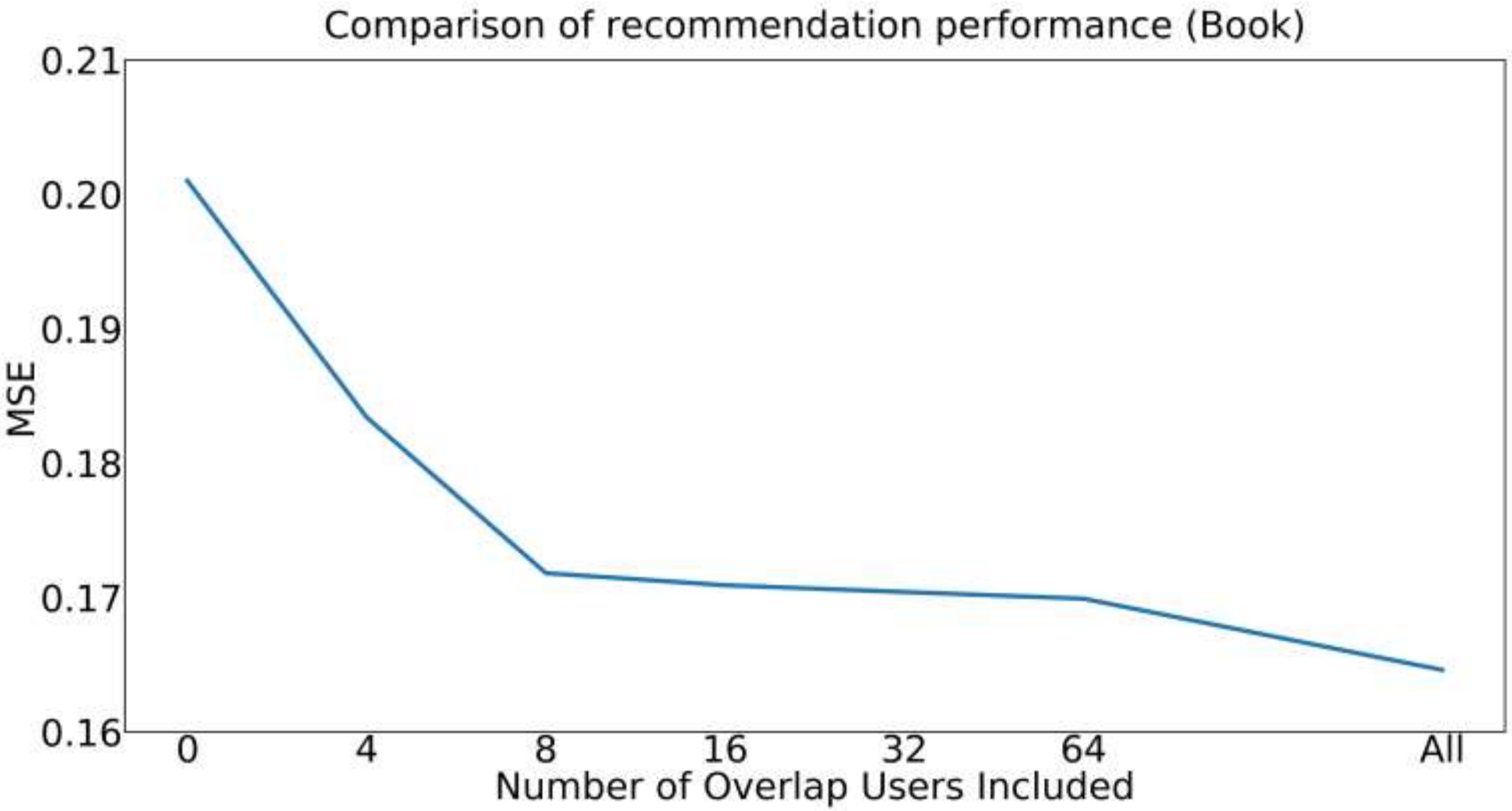
Algorithm	Book				Movie			
	RMSE	MAE	Precision@5	Recall@5	RMSE	MAE	Precision@5	Recall@5
DDTCDR	**0.2213***	**0.1708***	**0.8595***	**0.9594***	**0.2213***	**0.1714***	**0.8925***	**0.9871***
Improved %	(+3.98%)	(+9.54%)	(+2.77%)	(+6.30%)	(+2.44%)	(+9.80%)	(+2.75%)	(+2.74%)
NCF	0.2315	0.1887	0.8357	0.8924	0.2276	0.1895	0.8644	0.9589
CCFNet	0.2639	0.1841	0.8102	0.8872	0.2476	0.1939	0.8545	0.9300
CDFM	0.2494	0.2165	0.7978	0.8610	0.2289	0.1901	0.8498	0.9312
CMF	0.2921	0.2478	0.7972	0.8523	0.2738	0.2293	0.8324	0.9012
CoNet	0.2305	0.1892	0.8328	0.8990	0.2298	0.1903	0.8680	0.9601

Results: Convergence





Results: Number of Overlap Users

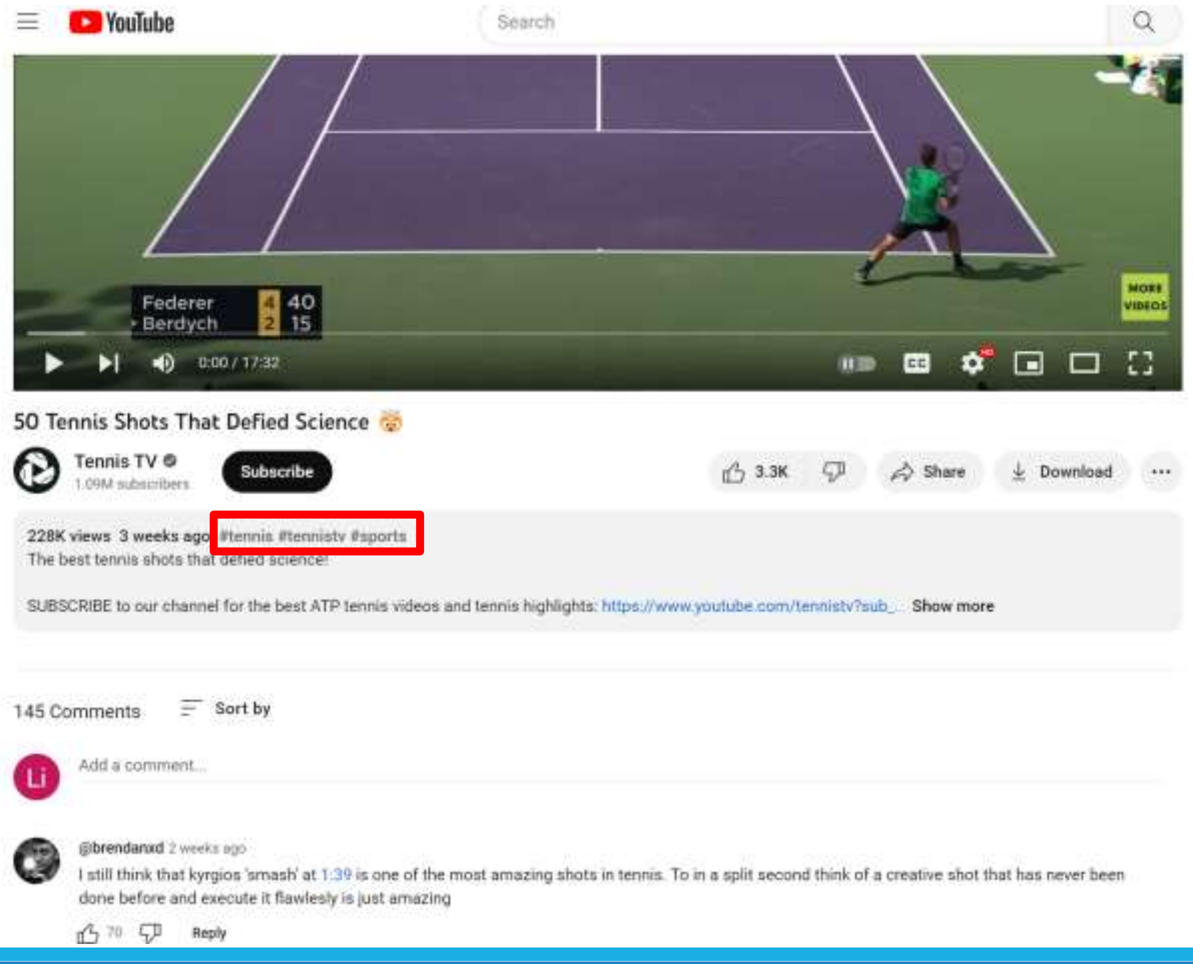


Online A/B Test

We conduct an online A/B test at Alibaba-Youku, one of the leading video streaming platforms in China.

- Test Period: January 2021
- User Sample: over 1 million
- Applications: TV Shows/Short Videos
- Business Metrics Improvements: +7.07% in total video views

Application II: LLM and Aspects in RecSys

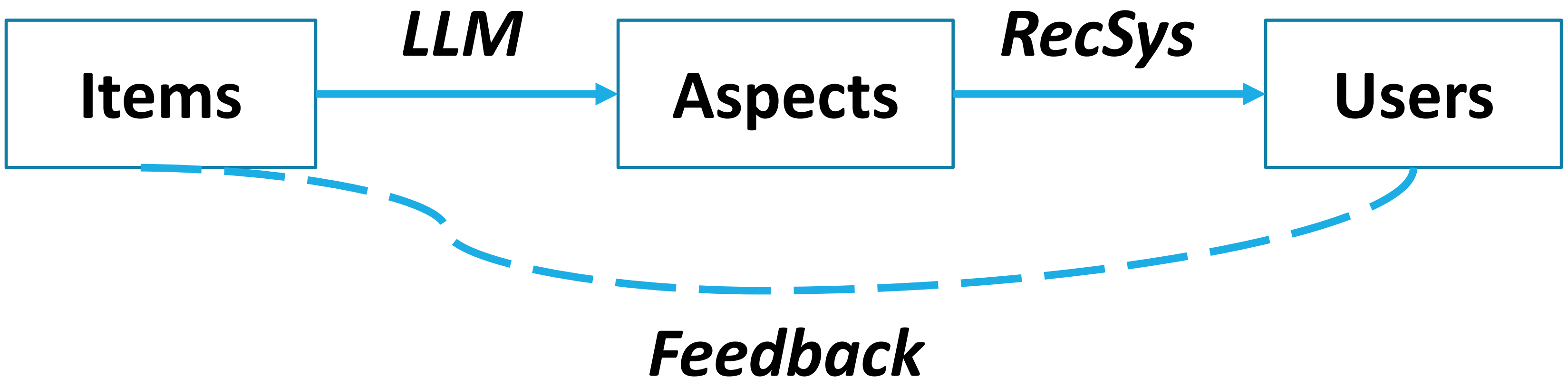


Question 1: How can we better explain the recommended product based on its aspect?

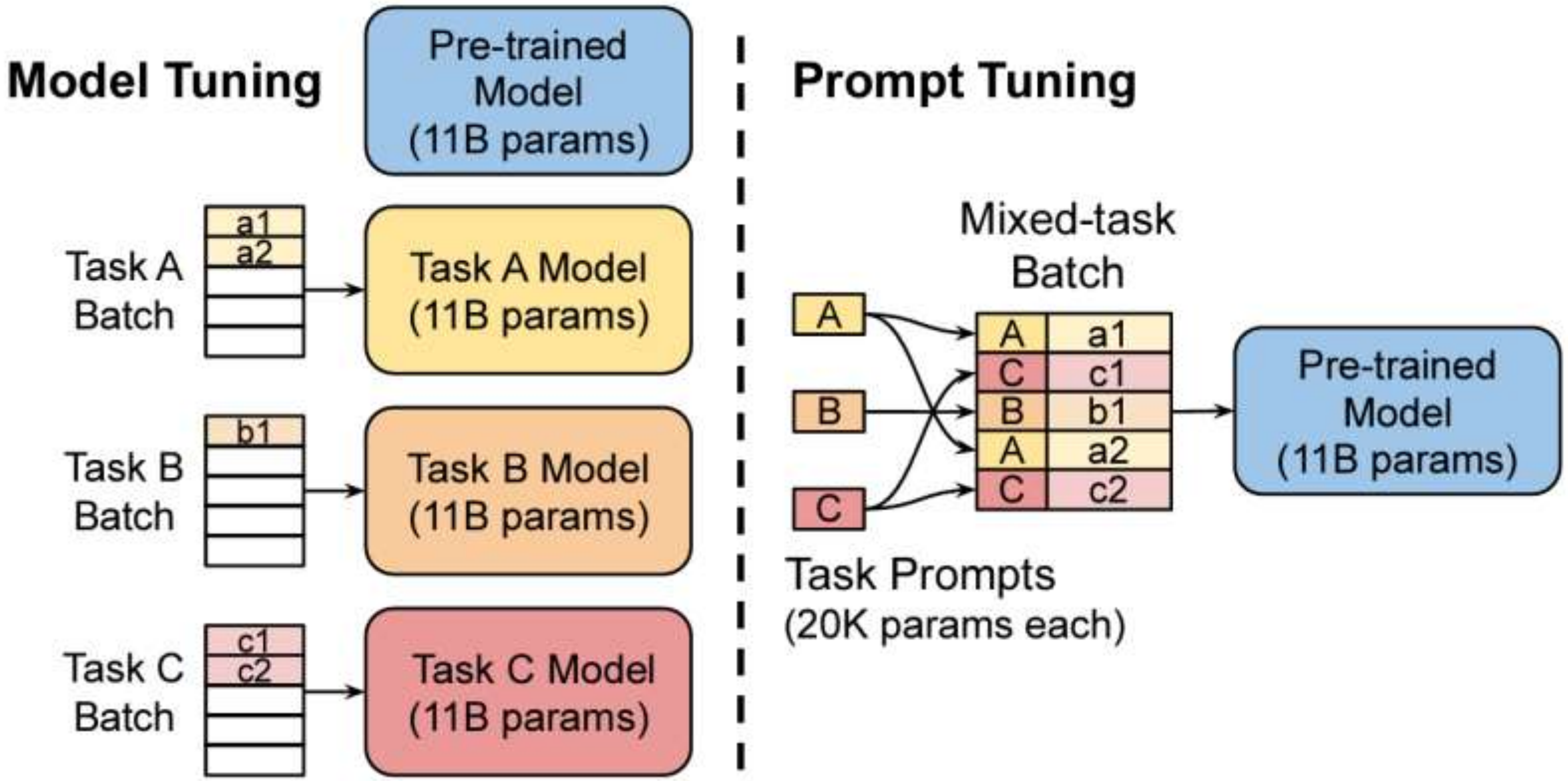
Question 2: How can we provide better recommendations based on user preference over different aspects?

Why LLM?

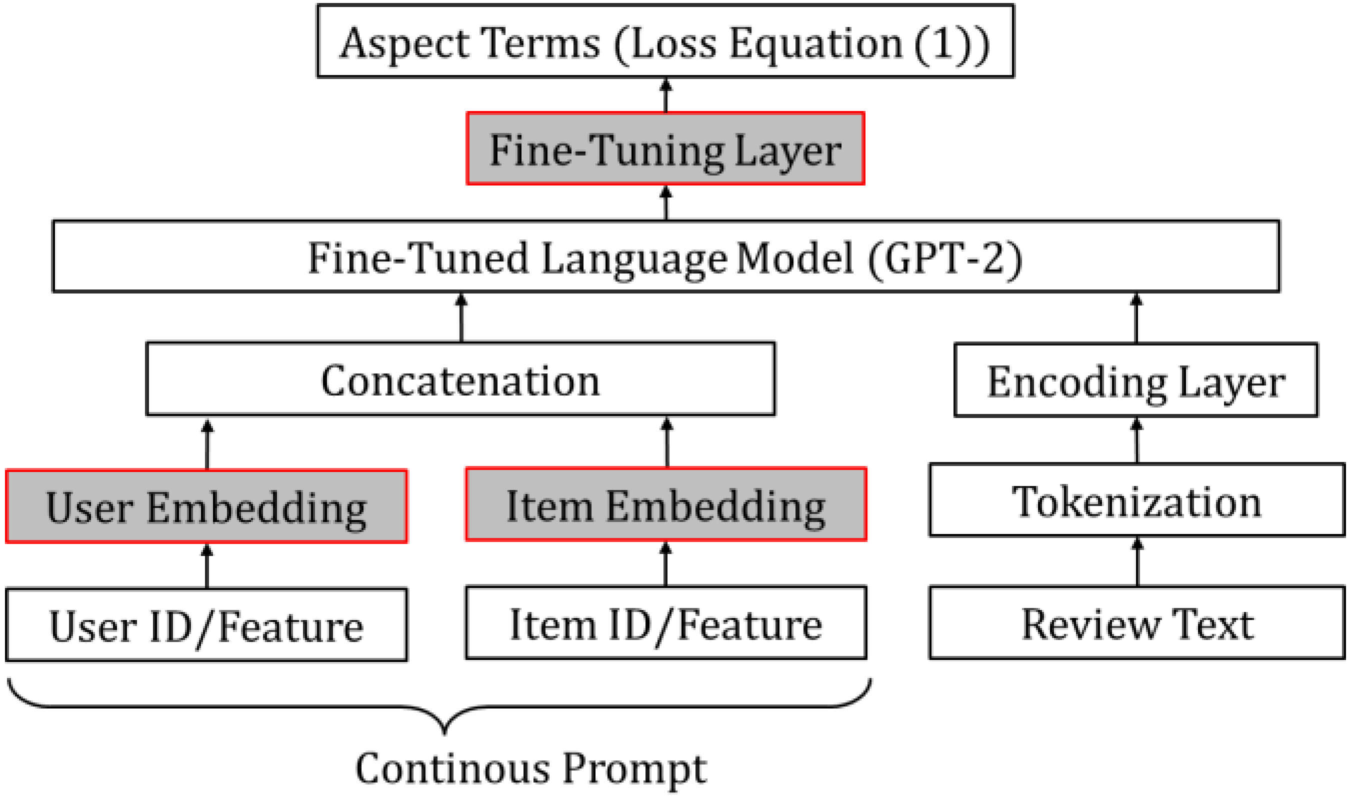
- LLMs have shown excellent capabilities in common-sense reasoning and utilizing background knowledge in a variety of tasks (e.g., aspect extraction).
- LLMs transfer the rich world knowledge from the universe of web textual data to better understand users' behavior and preferences.



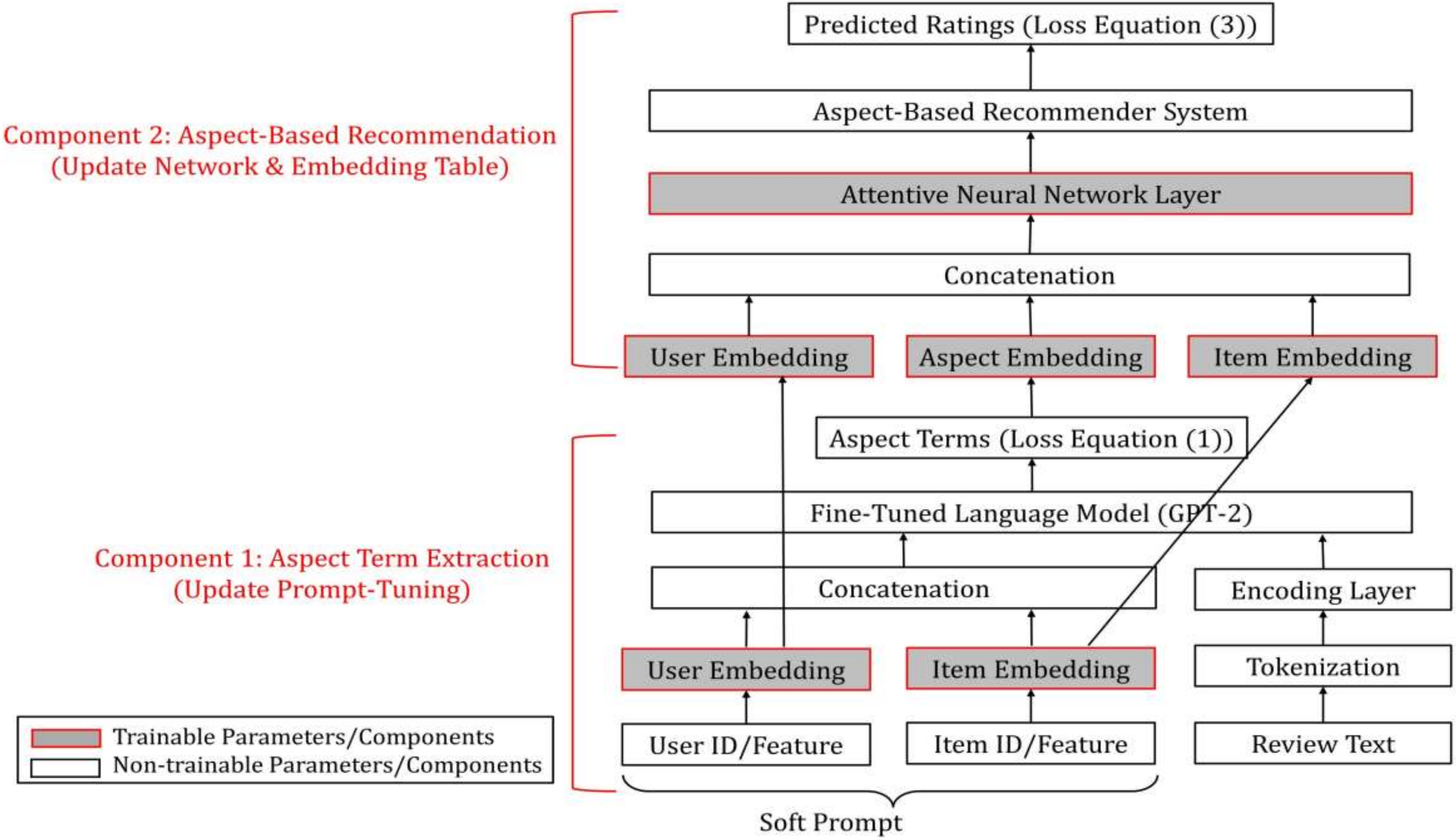
Prompt Tuning in LLM



Continuous Prompt Tuning: Benefits



- **Effectively incorporate User/Item ID information & features into LLM**
- **Can be easily concatenated with the review text to identify the most important aspect terms**
- **Can be dynamically updated based on user preference learned from the downstream recommendation task**


Predicted Ratings (Loss Equation (3))
Aspect-Based Recommender System
Attentive Neural Network Layer
Concatenation
User Embedding
Aspect Embedding
Item Embedding
Aspect Terms (Loss Equation (1))
Fine-Tuned Language Model (GPT-2)
Concatenation
Encoding Layer
User Embedding
Item Embedding
Tokenization
User ID/Feature
Item ID/Feature
Review Text
Soft Prompt
Component 2: Aspect-Based Recommendation
(Update Network & Embedding Table)
Component 1: Aspect Term Extraction
(Update Prompt-Tuning)
Trainable Parameters/Components
Non-trainable Parameters/Components

Experiments: Data

Datasets: Collected from TripAdvisor (hotel), Amazon (movies), and Yelp (restaurant).

Each Dataset Contains:

- User/Item IDs
- Ratings (Scale 1-5)
- User Reviews,
- Aspect Terms (Ground Truth)

Domain	TripAdvisor	Amazon	Yelp
# of Users	9,765	7,506	27,147
# of Items	6,280	7,360	20,266
# of Ratings	320,023	441,783	1,293,247
Sparsity	0.522%	0.800%	0.235%

Experiments: Baselines

- Baseline Methods (Aspect Extraction)
  - **DE-CNN**
  - **LCFS**
  - **ABAE**
  - **BERT**
  - **IMN**
  - **JASA**
- Baseline Methods (Aspect RecSys)
  - **A3NCF**
  - **SULM**
  - **AARM**
  - **MMALFM**
  - **ANR**
  - **MTER**
- Conducted record-stratified 5-fold cross validation
- Evaluated performance using RMSE, MAE, Precision and Recall metrics

Aspect-Term Extraction Performance

Dataset	Amazon			Yelp			TripAdvisor		
Algorithm	Precision@3	Recall@3	F1-Score	Precision@3	Recall@3	F1-Score	Precision@3	Recall@3	F1-Score
Our Model	**0.2533***	**0.2846***	**0.2680***	**0.2431***	**0.2568***	**0.2498***	**0.2755***	**0.2519***	**0.2632***
	(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0012)	(0.0011)	(0.0011)
(Improvement %)	+2.57%	+5.51%	+3.96%	+2.59%	+2.73%	+2.68%	+0.98%	+0.83%	+0.91%
DE-CNN	<u>0.2468</u>	<u>0.2689</u>	<u>0.2574</u>	<u>0.2368</u>	<u>0.2498</u>	<u>0.2431</u>	0.2723	0.2496	0.2605
LCFS	0.2449	0.2677	0.2558	0.2362	0.2496	0.2427	0.2705	0.2488	0.2592
ABAE	0.2416	0.2650	0.2528	0.2350	0.2491	0.2418	0.2688	0.2471	0.2575
BERT	0.2449	0.2681	0.2560	0.2359	0.2496	0.2426	<u>0.2728</u>	<u>0.2498</u>	<u>0.2608</u>
IMN	0.2430	0.2634	0.2528	0.2347	0.2481	0.2412	0.2715	0.2493	0.2599
JASA	0.2408	0.2634	0.2516	0.2343	0.2481	0.2410	0.2691	0.2487	0.2585
Ablation 1	0.2420	0.2641	0.2526	0.2359	0.2498	0.2427	0.2688	0.2480	0.2580
Ablation 2	0.2485	0.2739	0.2606	0.2381	0.2515	0.2446	0.2726	0.2501	0.2609
Ablation 3	0.2428	0.2667	0.2542	0.2346	0.2498	0.2420	0.2680	0.2468	0.2570
Ablation 4	0.2428	0.2661	0.2539	0.2346	0.2491	0.2416	0.2685	0.2472	0.2574
Ablation 5	0.2496	0.2780	0.2631	0.2393	0.2538	0.2463	0.2736	0.2510	0.2618
Ablation 6	0.2498	0.2786	0.2634	0.2397	0.2541	0.2467	0.2738	0.2510	0.2619

Table 2: Aspect term extraction performance in three datasets. '*' represents statistical significance with confidence level = 0.95. Improvement percentages are computed over the performance of the best baseline model for each metric.

Aspect-Based RecSys Performance

Dataset	Amazon			Yelp			TripAdvisor		
Algorithm	RMSE	MAE	AUC	RMSE	MAE	AUC	RMSE	MAE	AUC
Our Model	**0.2083***	**0.1757***	**0.7243***	**0.2413***	**0.2053***	**0.6991***	**0.1975***	**0.1709***	**0.7071***
	(0.0011)	(0.0009)	(0.0017)	(0.0011)	(0.0009)	(0.0016)	(0.0011)	(0.0009)	(0.0017)
(Improvement %)	+4.08%	+4.89%	+2.91%	+6.80%	+4.43%	+2.59%	+5.62%	+5.38%	+2.39%
A3NCF	0.2246	0.1895	0.6964	0.2611	0.2176	0.6780	0.2108	0.1814	0.6875
SULM	0.2478	0.1977	0.6851	0.2825	0.2255	0.6612	0.2199	0.1874	0.6733
AARM	0.2168	0.1843	0.7032	0.2589	0.2159	0.6799	0.2089	0.1805	0.6898
MMALFM	0.2305	0.1930	0.6928	0.2596	0.2163	0.6801	0.2120	0.1822	0.6892
ANR	0.2277	0.1915	0.6958	0.2577	0.2144	0.6810	0.2086	0.1801	0.6902
MTER	0.2286	0.1903	0.6964	0.2621	0.2163	0.6801	0.2101	0.1827	0.6885
Ablation 1	0.2250	0.1900	0.6980	0.2568	0.2141	0.6825	0.2081	0.1801	0.6933
Ablation 2	0.2142	0.1799	0.7197	0.2440	0.2090	0.6962	0.2001	0.1741	0.7045
Ablation 3	0.2398	0.1942	0.6903	0.2677	0.2189	0.6784	0.2144	0.1886	0.6855
Ablation 4	0,2375	0.1926	0.6915	0.2661	0.2180	0.6776	0.2140	0.1867	0.6877
Ablation 5	0.2298	0.1917	0.6966	0.2581	0.2152	0.6801	0.2095	0.1844	0.6898
Ablation 6	0.2196	0.1820	0.7158	0.2479	0.2117	0.6844	0.2059	0.1770	0.6967

Table 3: Aspect-based recommendation performance in three datasets. '*' represents statistical significance with confidence level = 0.95. Improvement percentages are computed over the performance of the best baseline model for each metric.

A Few Examples

Original Review 1	"It is a great collection version of star wars original episodes and worth purchasing through amazon if you are a fan."
Ground Truth	Star Wars, Original, Worth
Our Model	Star Wars, Original, Worth Purchasing
DE-CNN	Collection, Star Wars, Episode
LCFS	Star Wars, Worth, Amazon
ABAE	Collection, Episode, Worth
Original Review 2	"This movie is still a wonderful adventure which stands up well to the test of time."
Ground Truth	Wonderful, Adventure, Test of Time
Our Model	Wonderful, Adventure, Test of Time
DE-CNN	Movie, Wonderful, Well
LCFS	Movie, Wonderful, Adventure
ABAE	Movie, Wonderful, Time
Original Review 3	"The bathroom looked a little dated and the water pressure was on the low end."
Ground Truth	Bathroom, Dated, Low End
Our Model	Bathroom, Dated, Water Pressure
DE-CNN	Bathroom, Dated, Low
LCFS	Bathroom, Little, Water
ABAE	Bathroom, Little, Dated

Table 4: Case study of the aspect term extraction task

Thank you!

Pan Li

Assistant Professor@ITM, Georgia Tech

Email: pan.li@scheller.gatech.edu

Website: lpworld.github.io