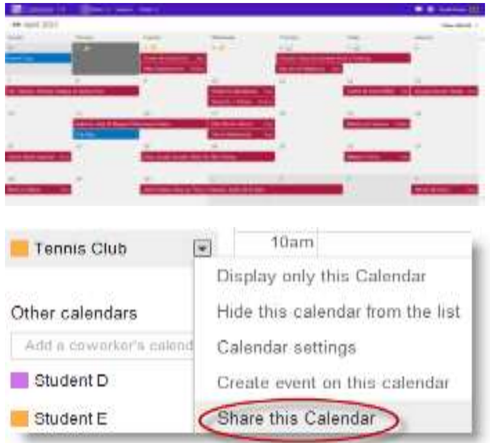
个性化推荐的未来

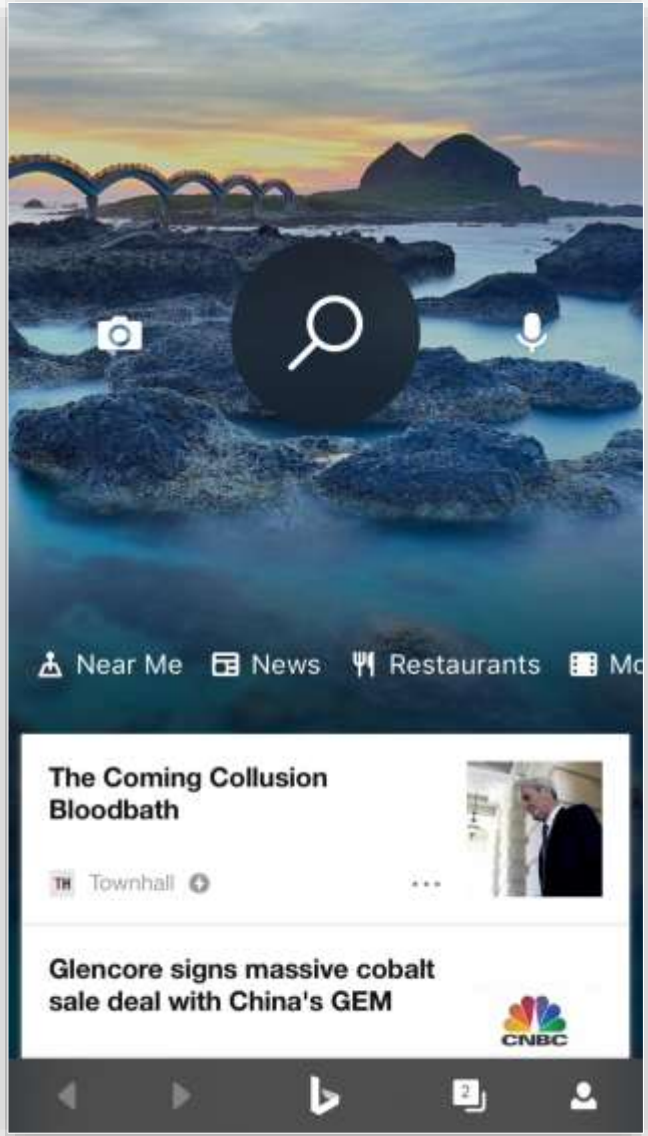
基于知识的推荐与可解释推荐

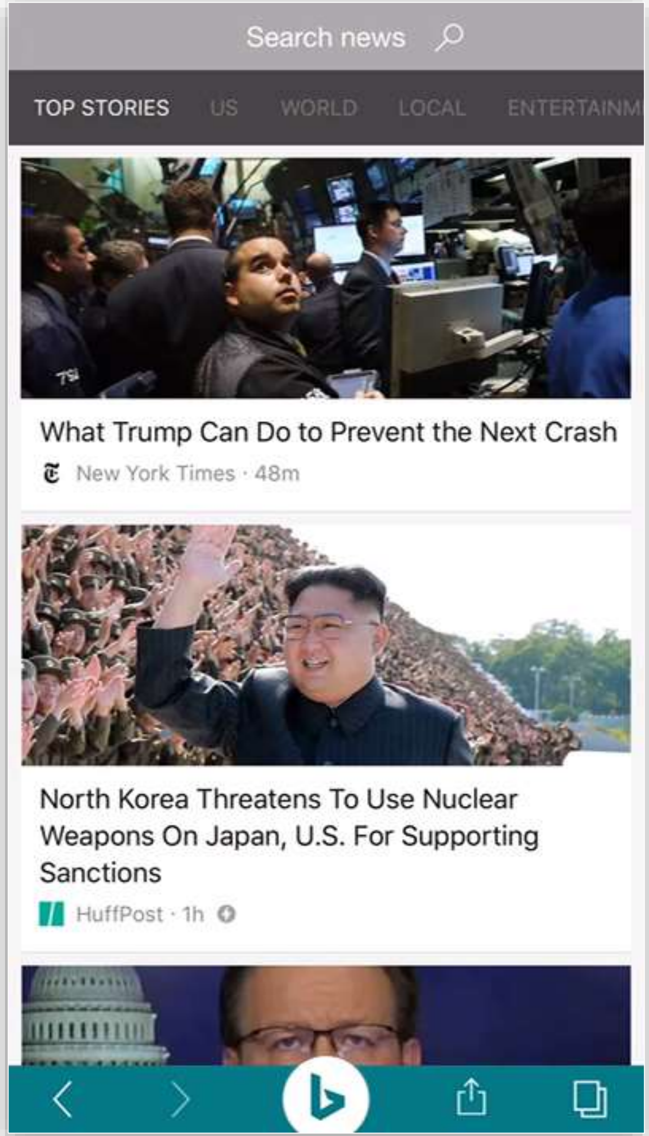
谢 幸，王希廷
微软亚洲研究院

User Behavioral Data

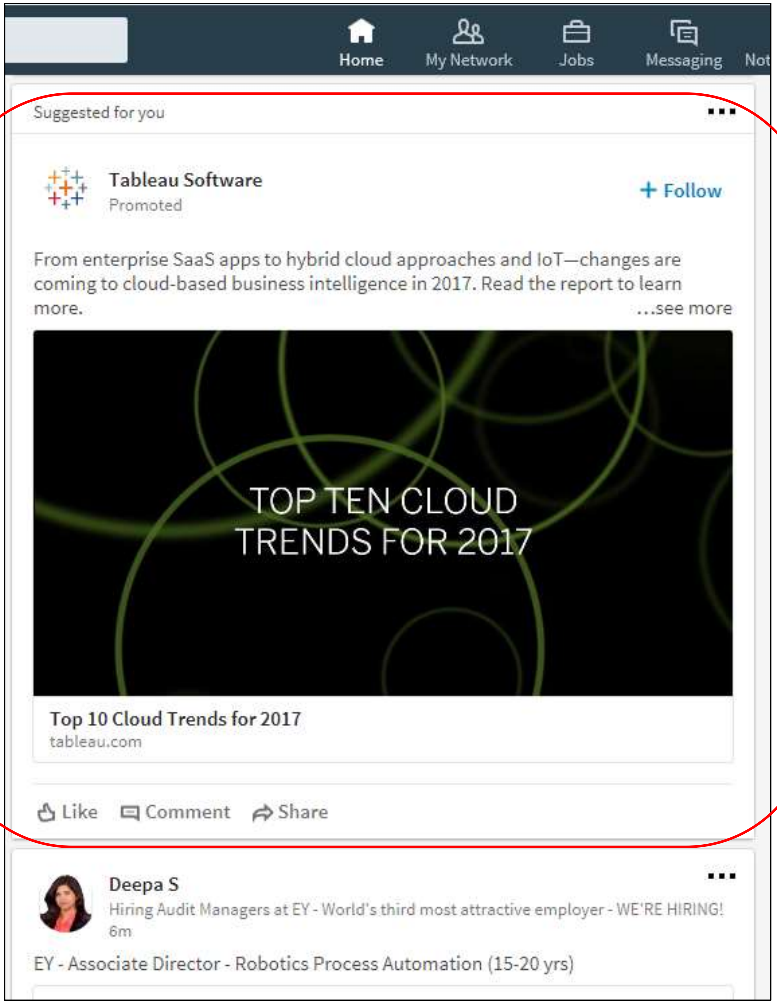


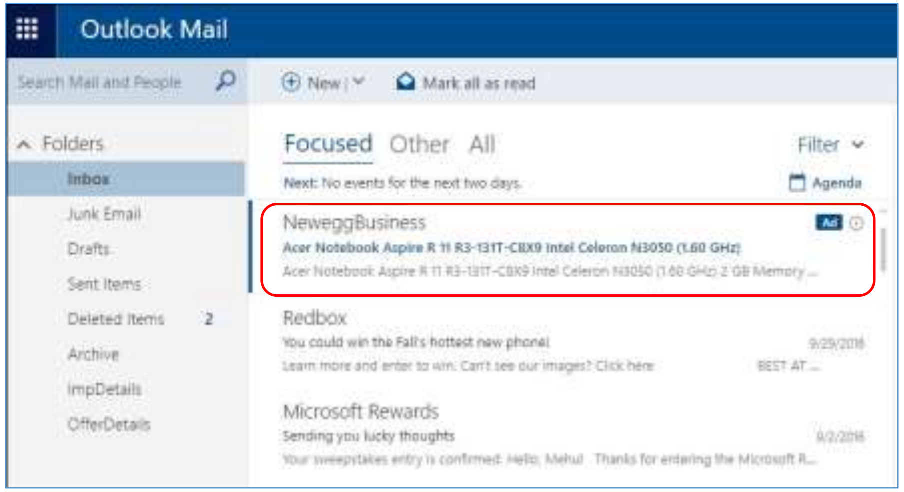
Personalized News Feed

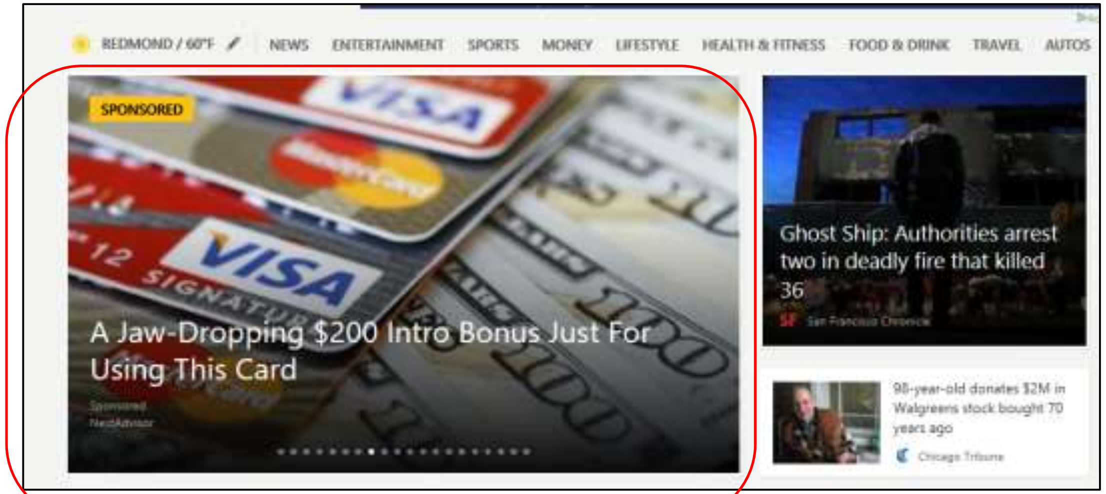




Online Advertising







Conversational Recommendaton

Every time I listen to this song, I will think of my first love.

What songs should I listen to when I can't sleep?

Can not sleep, listening to the song, recalling my story, and missing your hand.

Data Privacy

User Modeling

User Big Data

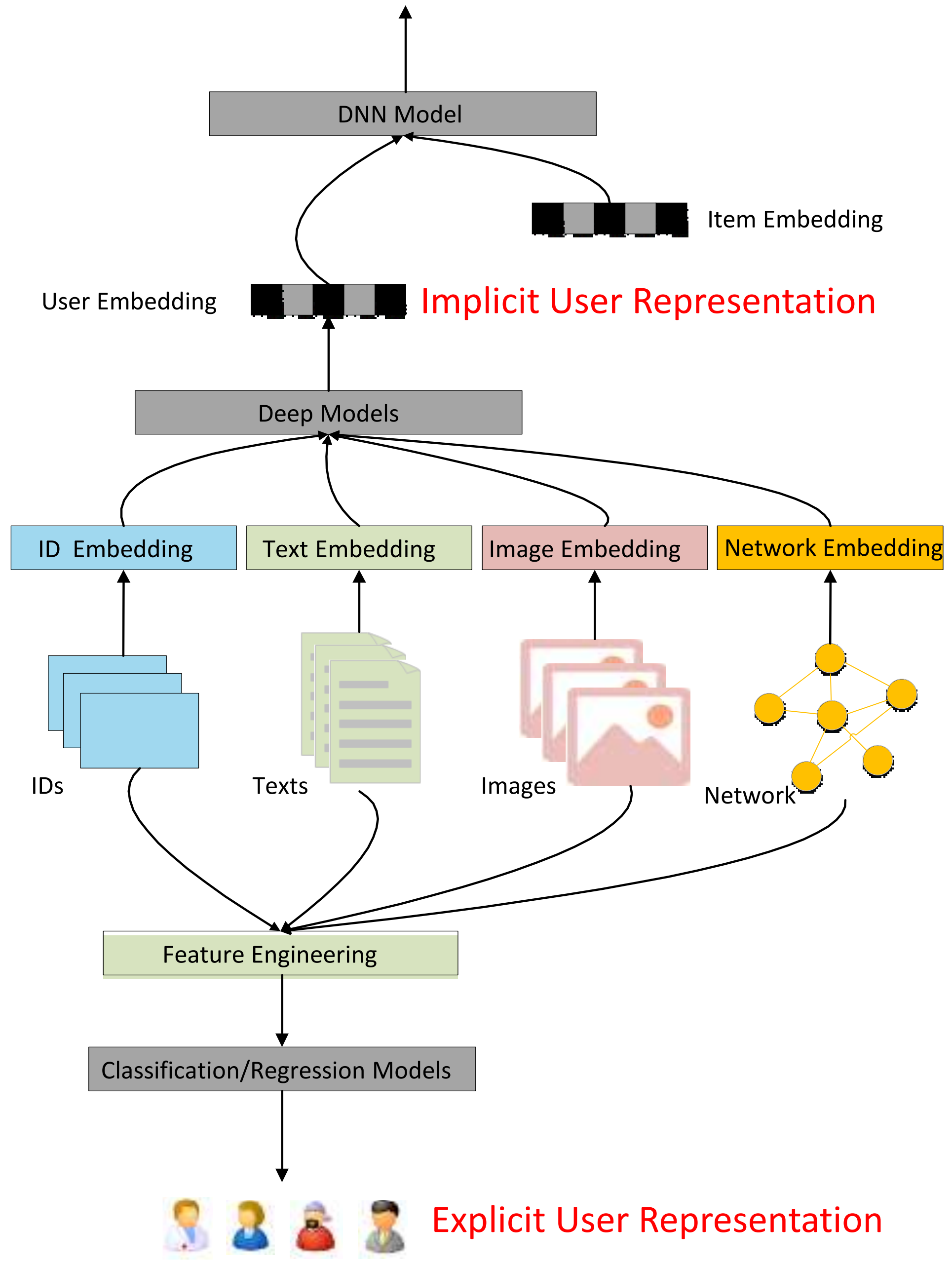
User Representation

Explicit User Representation

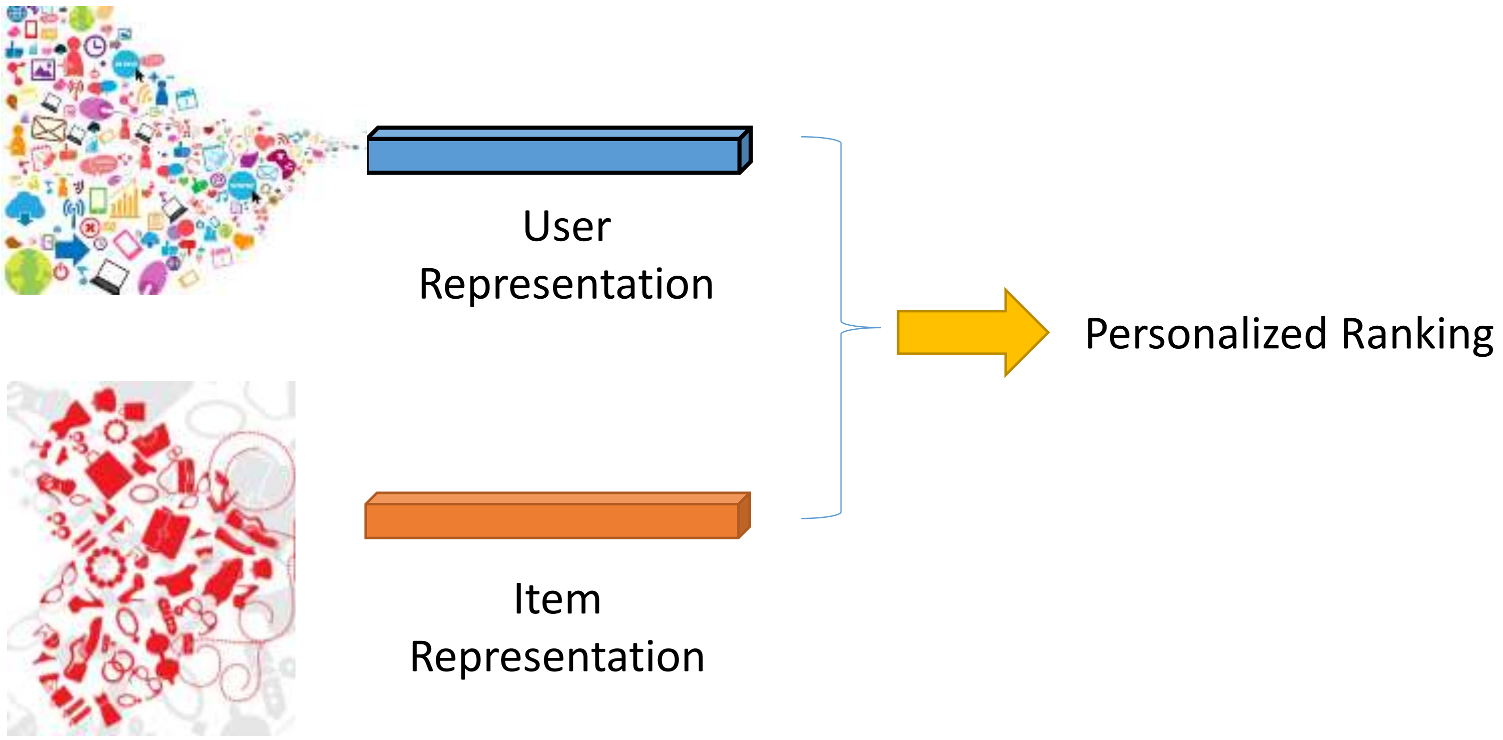
- Demographic
  - Age
  - Gender
  - Life stage
  - Marital status
  - Residence
  - Education
  - Vocation
- Personality
  - Openness
  - Conscientiousness
  - Extraversion
  - Agreeableness
  - Neuroticism
  - Impulsivity
  - Novelty-seeking
  - Indecisiveness
- Interests
  - Food
  - Book
  - Movie
  - Music
  - Sport
  - Restaurant
- Status
  - Emotion
  - Event
  - Health
  - Wealth
  - Device
- Social
  - Friend
  - Coworker
  - Spouse
  - Children
  - Other relatives
  - Tie strength
- Schedule
  - Task
  - Driving route
  - Metro/bus line
  - Appointment
  - Vacation

Explicit vs Implicit

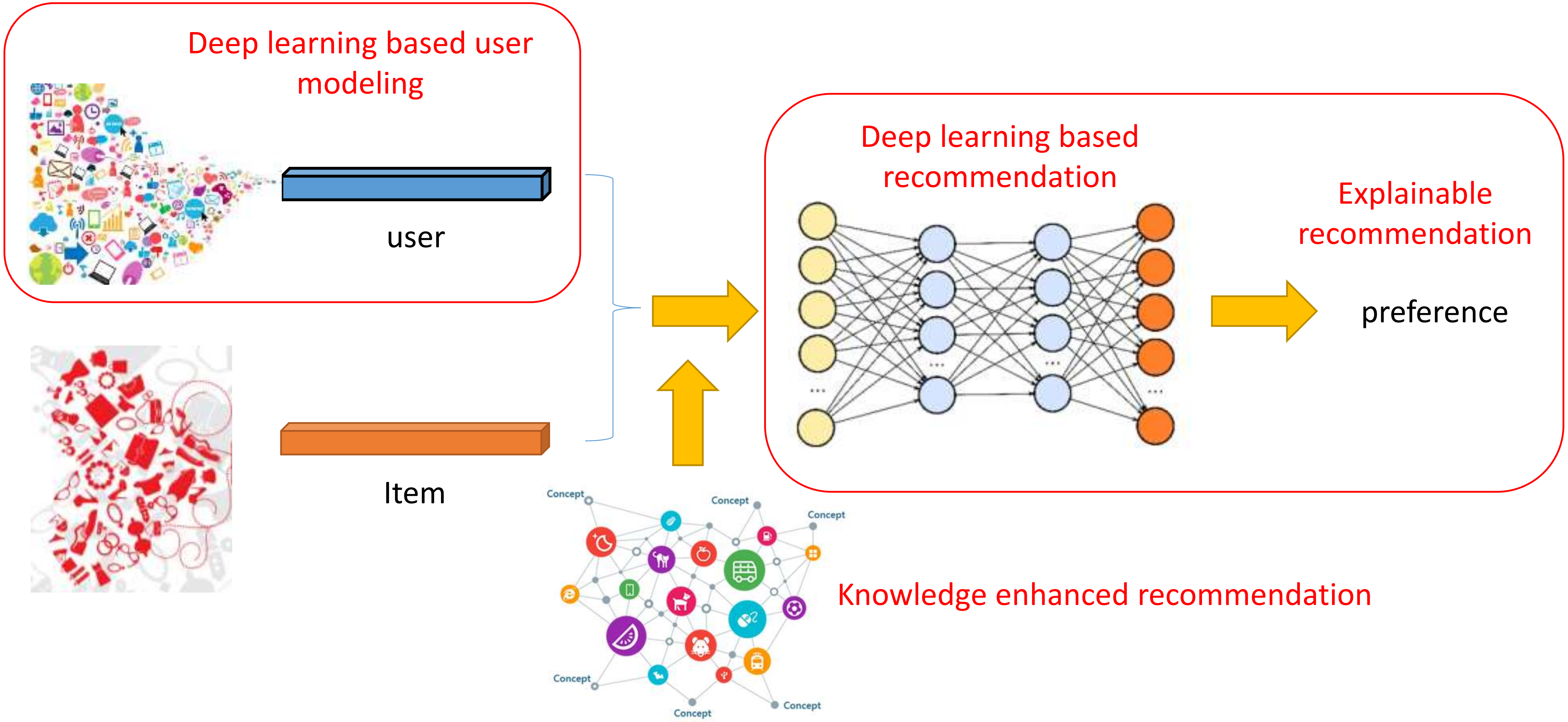
Representation	Pros	Cons
Explicit	• Easy to understand; • Can be directly bidden by advertisers	• Hard to obtain training data; • Difficult to satisfy complex and global needs;
Implicit	• Unified and heterogenous user representation; • End-to-end learning	• Difficult to explain; • Need to fine-tune in each task



Personalized Service

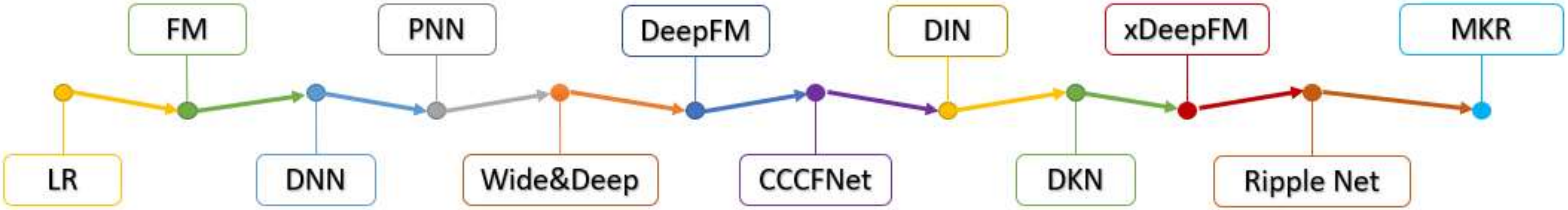


Our Research



DeepRec

- A collection of state-of-the-art deep learning based user representation and recommendation models. The systems are designed to be simple to use and easy to extend, while maintaining efficiency.



Recent Publication

- Xiting Wang, Yiru Chen, Jie Yang, etc. A Reinforcement Learning Framework for Explainable Recommendation, ICDM 2018
- Chanyoung Park, Donghyun Kim, Xing Xie, Hwanjo Yu, Collaborative Translational Metric Learning, ICDM 2018
- Zhigang Yuan, Fangzhao Wu, Junxin Liu, etc. Neural Sentence-level Sentiment Classification with Heterogeneous Supervision, ICDM 2018
- Hongwei Wang, etc. Ripple Network: Propagating User Preferences on the Knowledge Graph for Recommender Systems, CIKM 2018
- Jianxun Lian, Xiaohuan Zhou, etc., xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems, KDD 2018
- Zheng Liu, Xing Xie, Lei Chen, Context-aware Academic Collaborator Recommendation, KDD 2018
- Defu Lian, Kai Zheng, Vincent W. Zheng, etc. High-order Proximity Preserving Information Network Hashing, KDD 2018
- Jianxun Lian, etc. Towards Better Representation Learning for Personalized News Recommendation: a Multi-Channel Deep Fusion Approach, IJCAI 2018
- Haochao Ying, Fuzhen Zhuang, Fuzheng Zhang, etc. Sequential Recommender System based on Hierarchical Attention Networks, IJCAI 2018
- Yingzi Wang, Anastasios Noulas, Xiao Zhou, etc. Predicting the Spatio-Temporal Evolution of Chronic Diseases in Population with Human Mobility Data, IJCAI 2018
- Guanjie Zheng, Fuzheng Zhang, Zihan Zheng, etc. DRN: A Deep Reinforcement Learning Framework for News Recommendation, WWW 2018
- Hongwei Wang, Fuzheng Zhang, Xing Xie, Minyi Guo, DKN: Deep Knowledge-Aware Network for News Recommendation, WWW 2018
- Youngnam Lee, etc. How to Impute Missing Ratings? Claims, Solution, and Its Application to Collaborative Filtering, WWW 2018
- Hongwei Wang, etc. SHINE: Signed Heterogeneous Information Network Embedding for Sentiment Link Prediction, WSDM 2018

丛书简介

本套丛书是面向新形势下的大数据技术发展对人才培养提出的挑战以及知识更新的需求而策划组织的，旨在为学术研究和人才培养提供可供参考的“基石”。丛书内容涵盖大数据管理的理论、方法、技术等诸多方面，选题面向技术热点，弥补现有知识体系的漏洞和不足，力图为现有的数据管理知识查漏补缺，聚少成多，最终形成适应大数据技术发展和人才培养的知识体系和教材基础。丛书主编是中国人民大学孟小峰教授。

上架指导：计算机\大数据管理

定价：69.00元

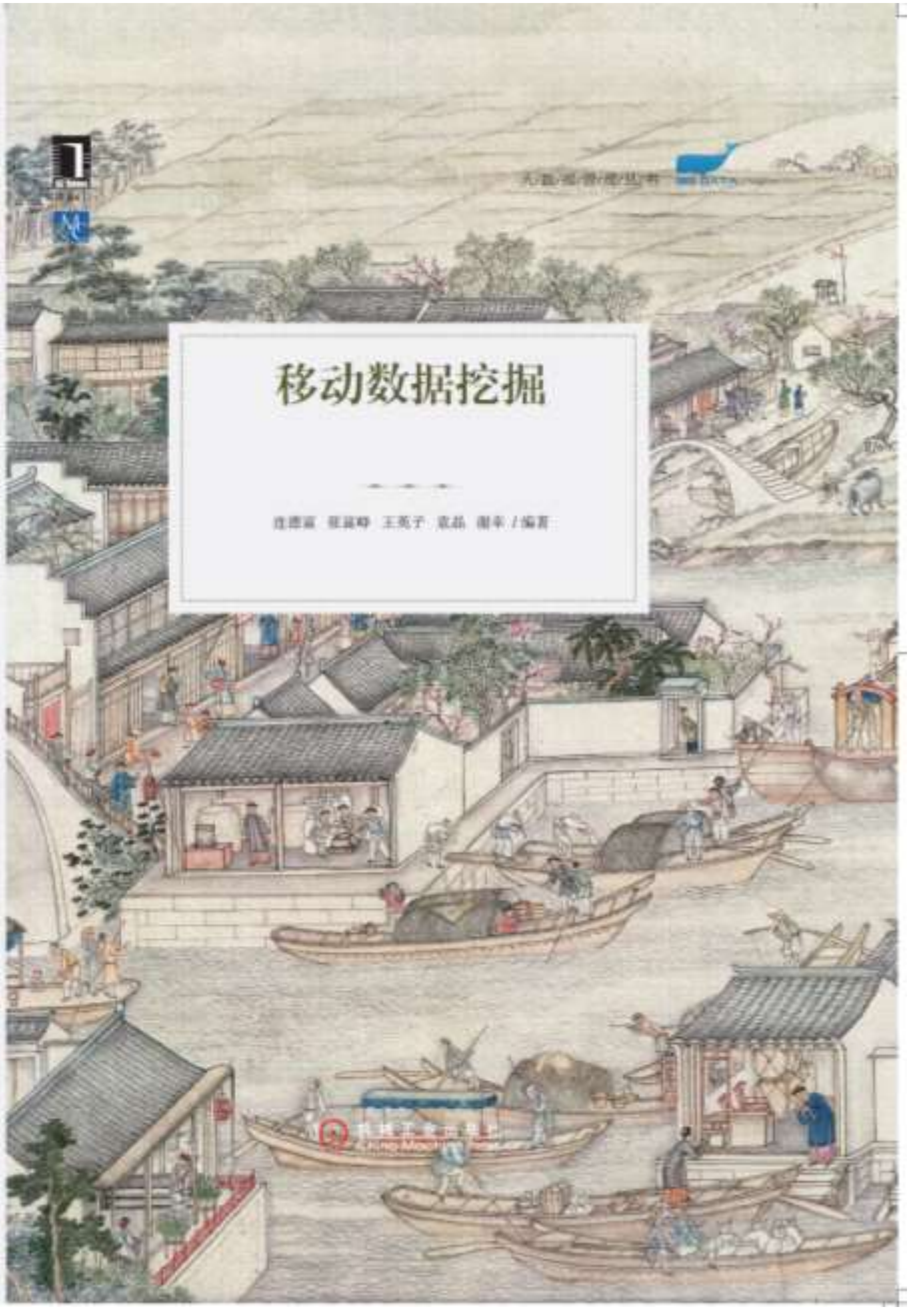
投稿热线：(010) 88379604
客服热线：(010) 88379426 88361066
购书热线：(010) 68326294 88379649 68995259

华章网站：www.hzbook.com
网上购书：www.china-pub.com
数字阅读：www.hzmedia.com.cn

移动数据挖掘

连德富 张富峥 王英子 袁晶 谢幸 /编著

机械工业出版社 China Machine Press



大数据管理丛书

移动数据挖掘

连德富 张富峥 王英子 袁晶 谢幸 /编著

机械工业出版社 China Machine Press

Query Log based User Modeling

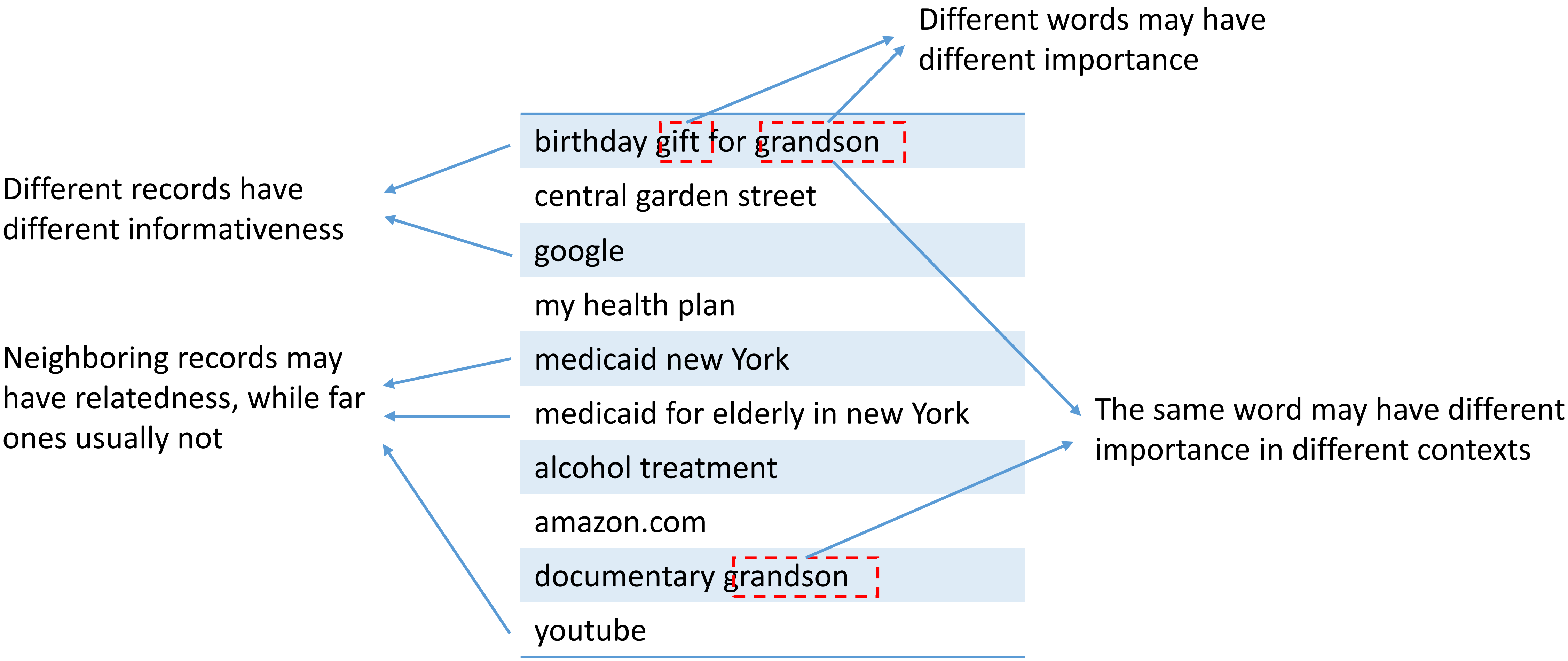
- gifts for classmates
- cool math games
- mickey mouse cartoon

- shower chair for elderly
- presbyopic glasses
- costco hearing aids

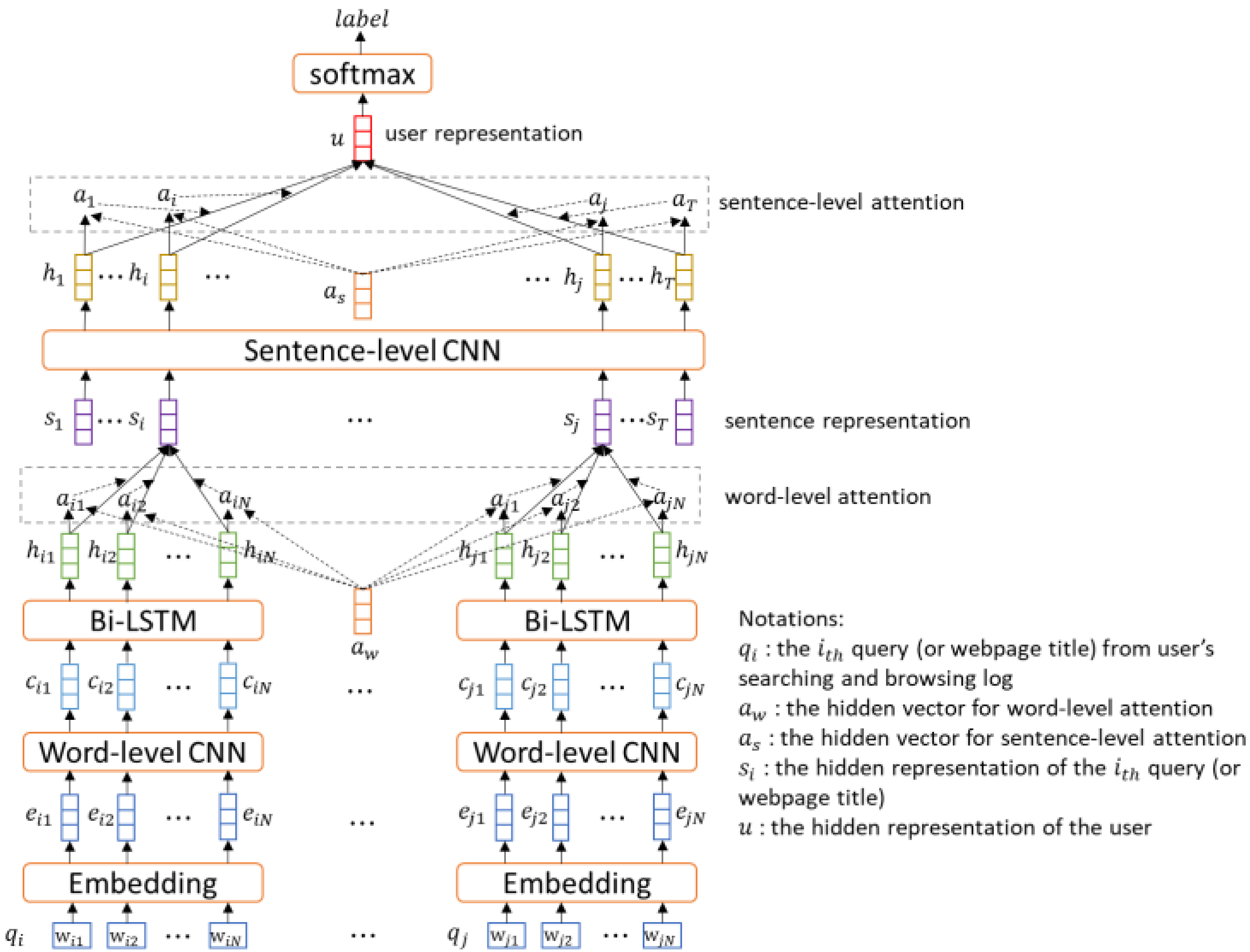
- groom to bride gifts
- tie clips
- philips shaver

- lipstick color chart
- womans ana blouse
- Dior Makeup

Query Log based User Modeling



Query Log based User Modeling

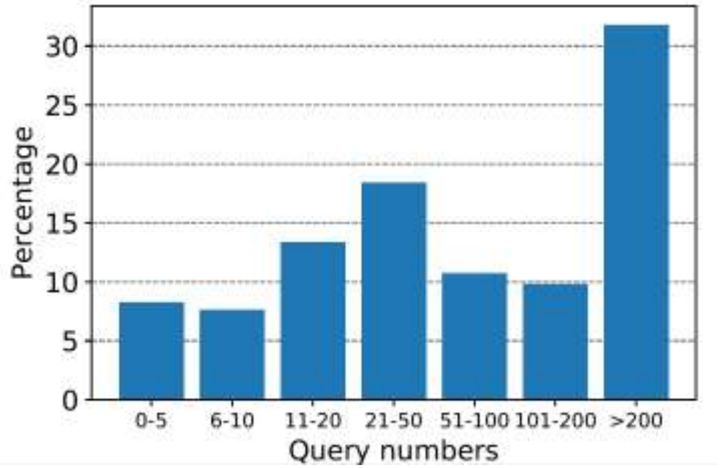


Experiments

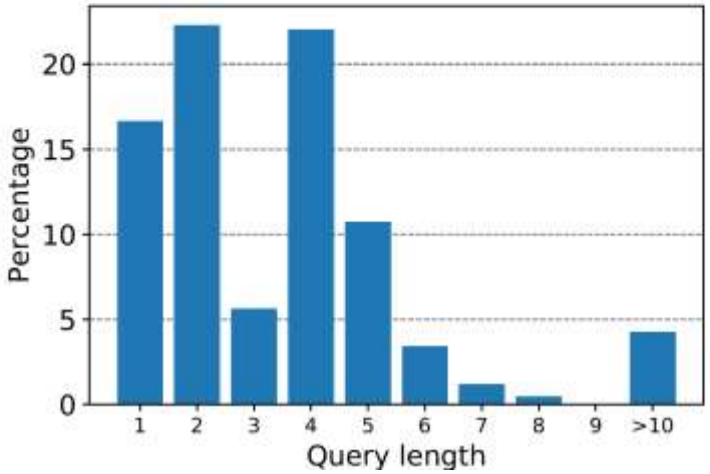
Mapping between age category and age range

Age category	1	2	3	4	5	6
Age range	< 18	$[18, 24]$	$[25, 34]$	$[35, 49]$	$[50, 64]$	> 64

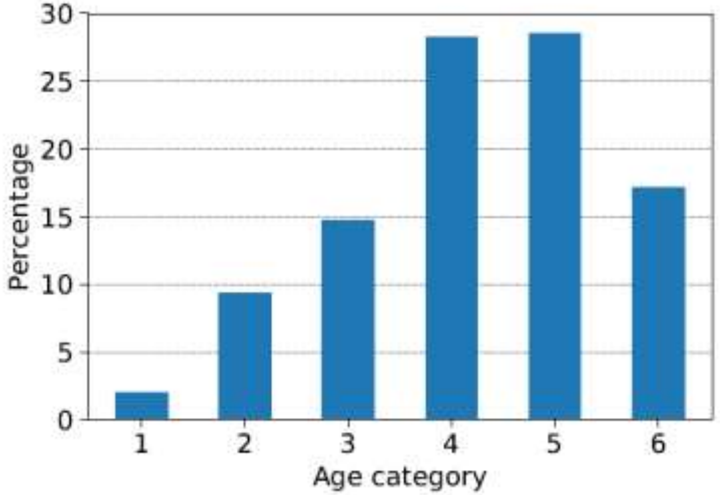
- Dataset:
  - 15,346,617 users in total with age category labels
    - Randomly sampled 10,000 users for experiments
    - Search queries posted from October 1, 2017 to March 31, 2018



Distribution of query number per user



Distribution of query length



Distribution of age category

Experiments

	10%		50%		100%	
	Accuracy	Fscore	Accuracy	Fscore	Accuracy	Fscore
SVM	31.97	21.96	34.20	26.32	34.53	27.44
LR	31.61	21.55	33.09	25.94	33.91	26.92
LinReg	27.12	17.38	29.64	22.48	30.34	23.52
FastText	28.65	21.09	30.40	23.55	30.90	24.01
CNN	30.08	19.66	35.58	26.17	37.31	26.96
LSTM	30.15	20.46	36.11	24.67	37.96	25.28
HAN	32.06	22.58	37.04	25.88	39.86	29.79
HURA	34.07	24.16	39.68	28.68	41.22	31.18

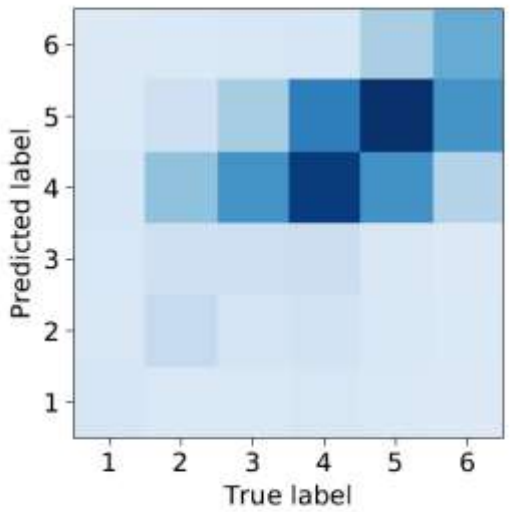
discrete feature, linear model

continuous feature, linear model

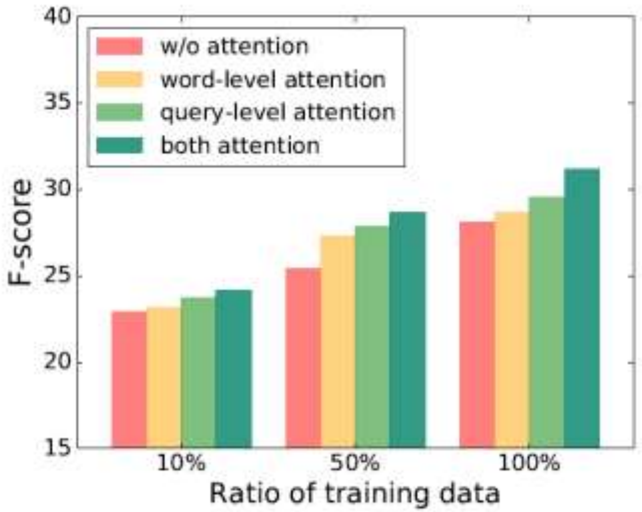
flat DNN models

hierarchical LSTM model

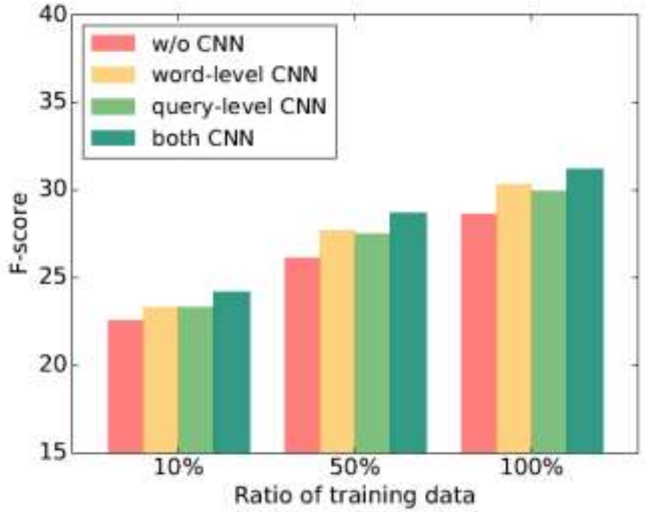
Experiments



Confusion matrix

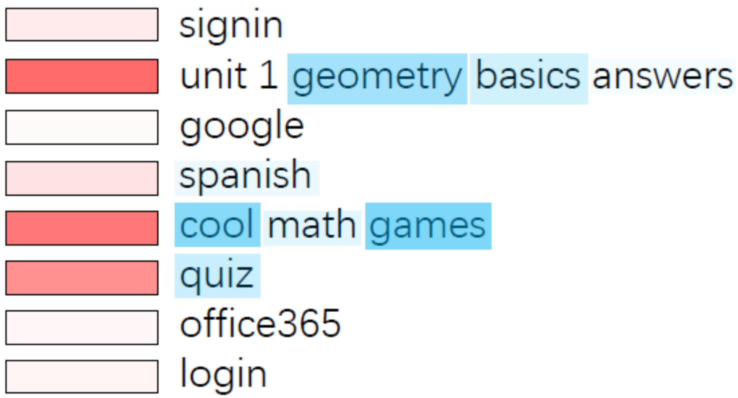


Word-level and query-level attention

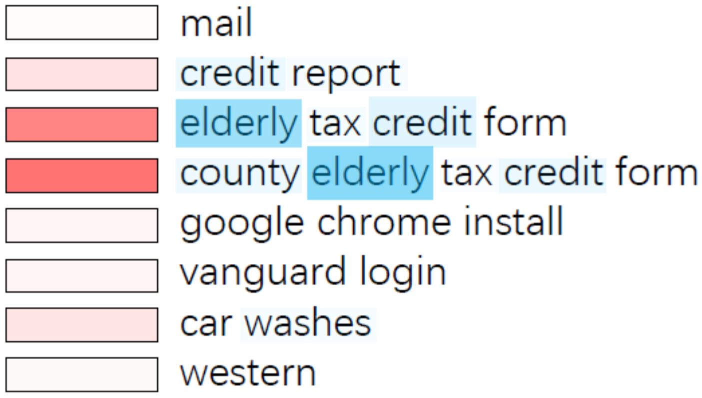


Word-level and query-level CNN networks

User Age Inference



Queries from a young user



Queries from an elder user

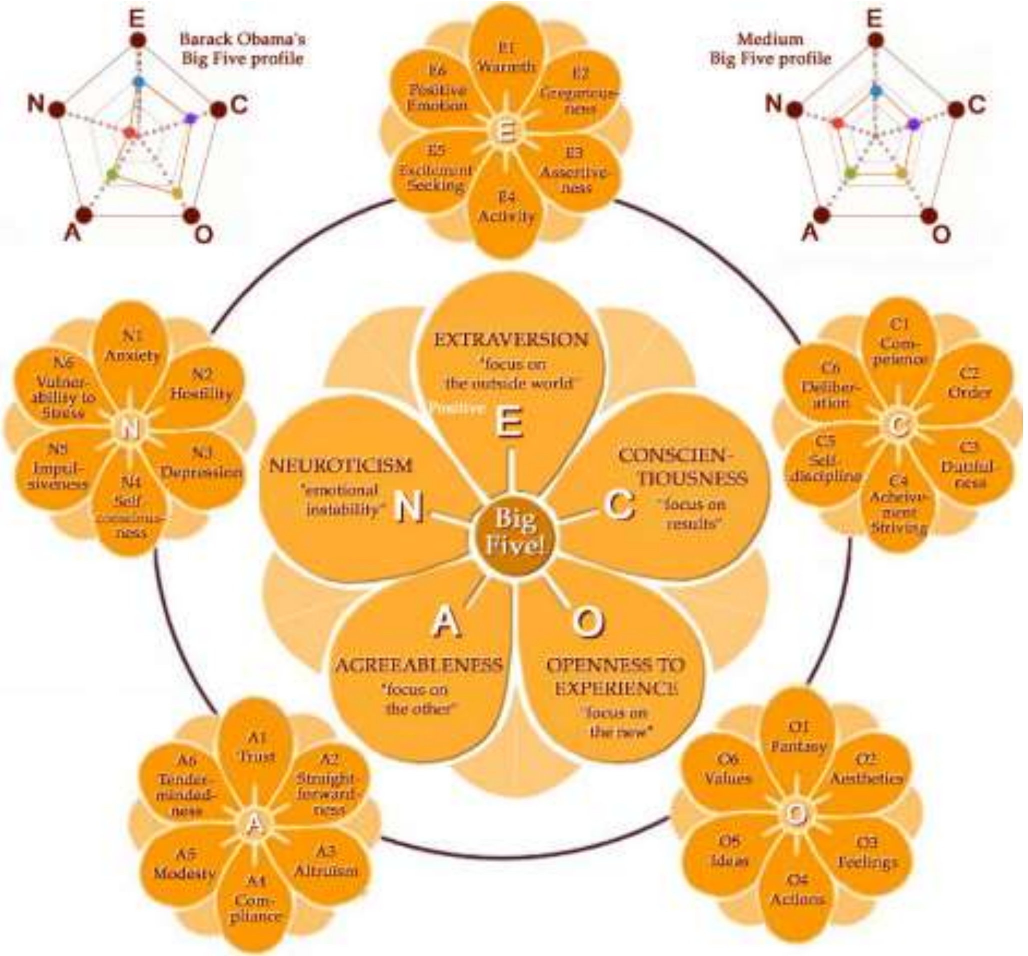
Car Segment

2018 mazda cx9 reliability
mathway math problem solver
open the dvd or cd drive in windows 10
lowes van & truck rental
facebook log in or sign up
buying high quality cars at a low price
plot summary imdb
how can i block a phone number from my home phone

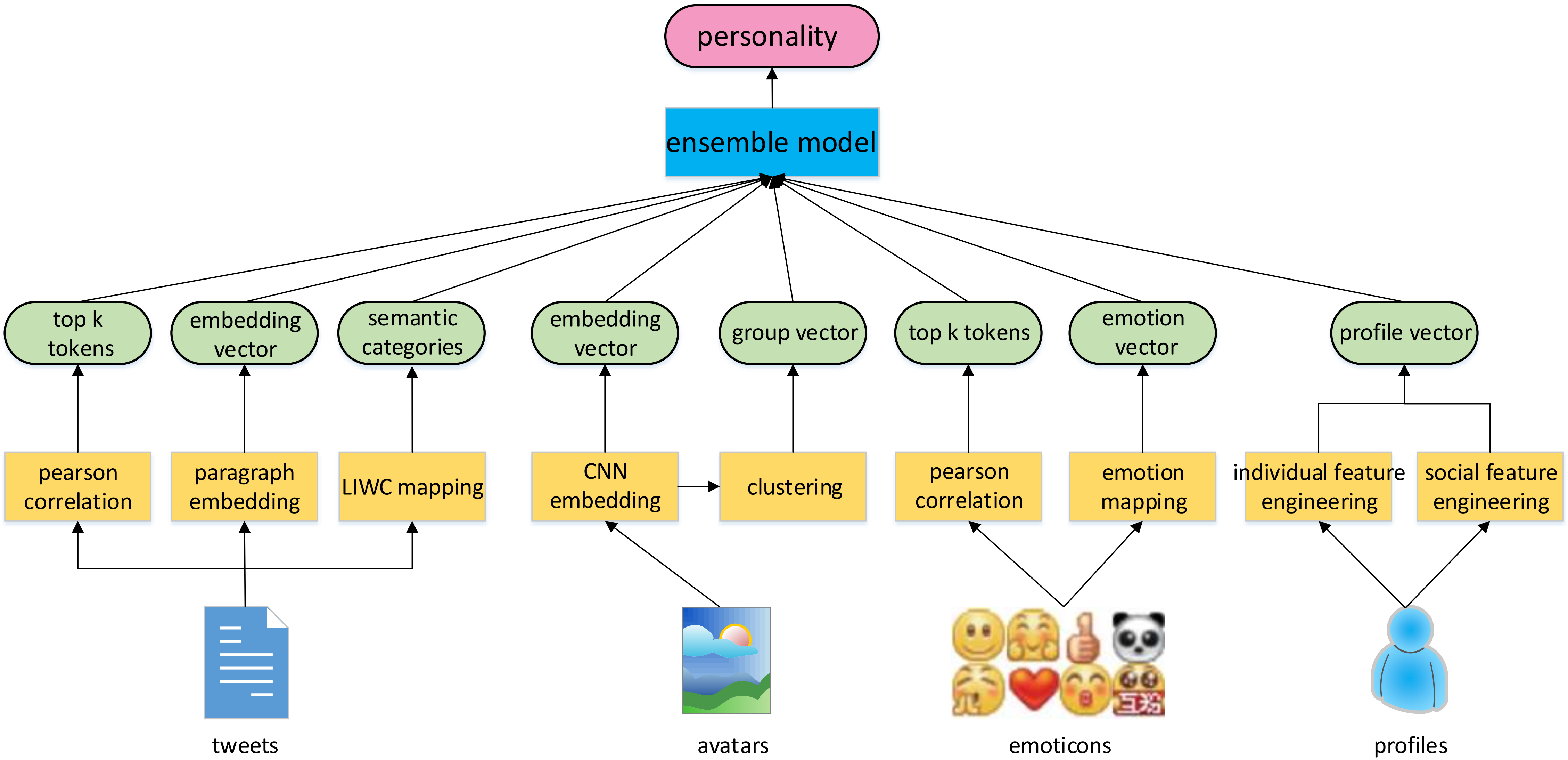
Pet Segment

dog food, cat food, and treats
the denver post official site
easybib: free bibliography generator
chords crowder guitar video
akc golden retriever pet adoption northern California
among large uk newspapers, which are considered
gmail email from google
heritage animal hospital care.com

Big Five


Barack Obama's Big Five profile
Medium Big Five profile
E
C
O
A
N
E1 Warmth
E2 Gregarious-ness
E3 Assertive-ness
E4 Activity
E5 Excitement Seeking
E6 Positive Emotion
N1 Anxiety
N2 Hostility
N3 Depression
N4 Self-conscious-ness
N5 Impul-siveness
N6 Vulner-ability to Stress
C1 Com-petence
C2 Order
C3 Dutiful-ness
C4 Acheive-ment Striving
C5 Self-discipline
C6 Deliber-ation
A1 Trust
A2 Straight-forward-ness
A3 Altruism
A4 Com-pliance
A5 Modesty
A6 Tender-minded-ness
O1 Fantasy
O2 Aesthetics
O3 Feelings
O4 Actions
O5 Ideas
O6 Values
EXTRAVERSION "focus on the outside world"
NEUROTICISM "emotional instability"
CONSCIEN-TIOUSNESS "focus on results"
AGREEABLENESS "focus on the other"
OPENNESS TO EXPERIENCE "focus on the new"
Big Five!

Personality Inference



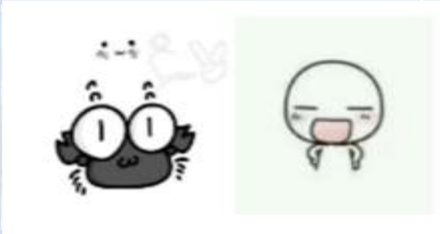
Data

- 3,162 users from a medical school
  - Major: nursing (524), clinical medicine (365) and pharmaceutics (342)
  - Region: Anhui, Zhejiang, and Jiangsu
  - Age: average 20.84
- Test Big Five Personality with a 44-item questionnaire

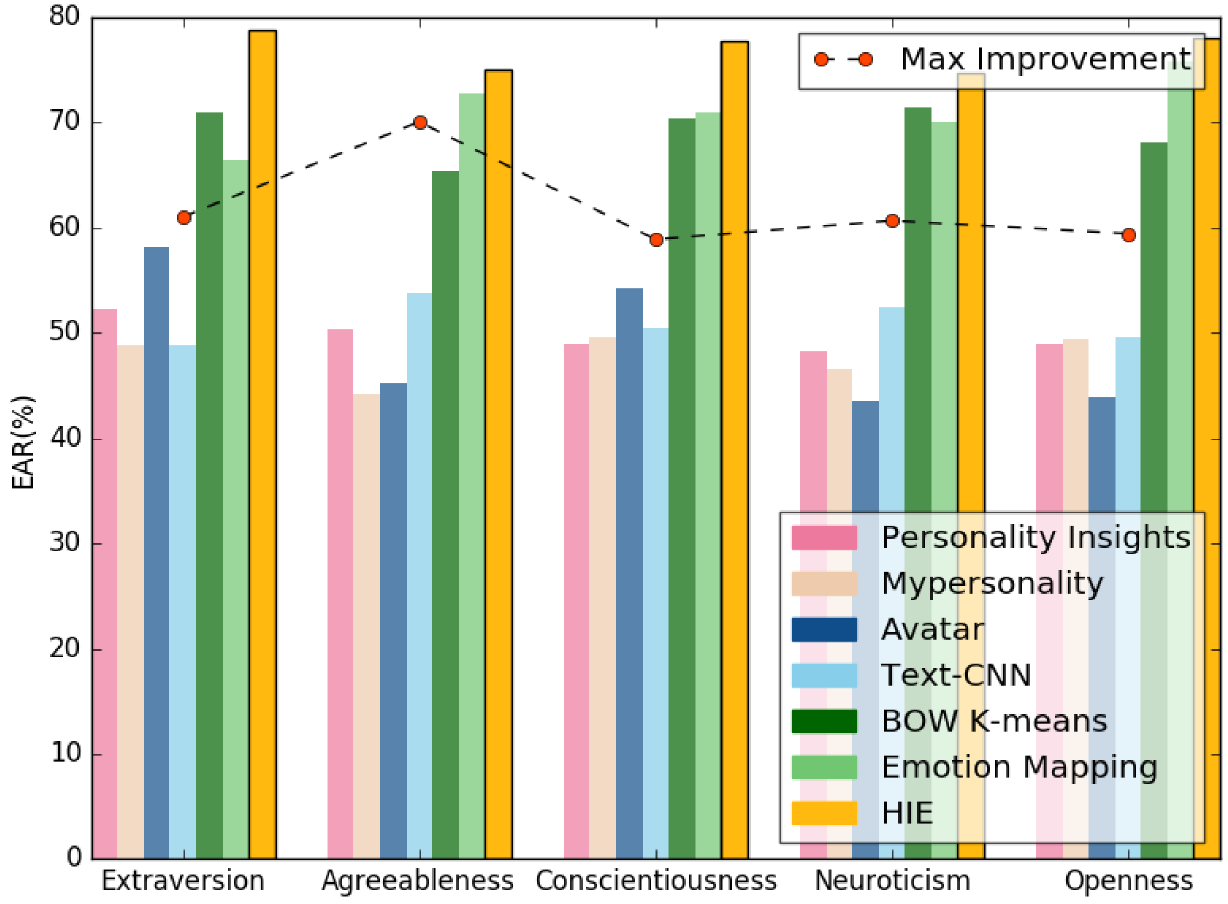
Correlation between Tweet and Personality

	Extraversion	Agreeableness	Neuroticism	Conscientiousness	Openness
Positive					
Negative					

Correlation between Avatar and Personality

	Extraversion	Agreeableness	Neuroticism	Conscientiousness	Openness
Positive					
Negative					

Experimental Results



Honghao Wei, Fuzheng Zhang, Nicholas Jing Yuan, Chuan Cao, etc. Beyond the Words: Predicting User Personality from Heterogeneous Information, WSDM 2017

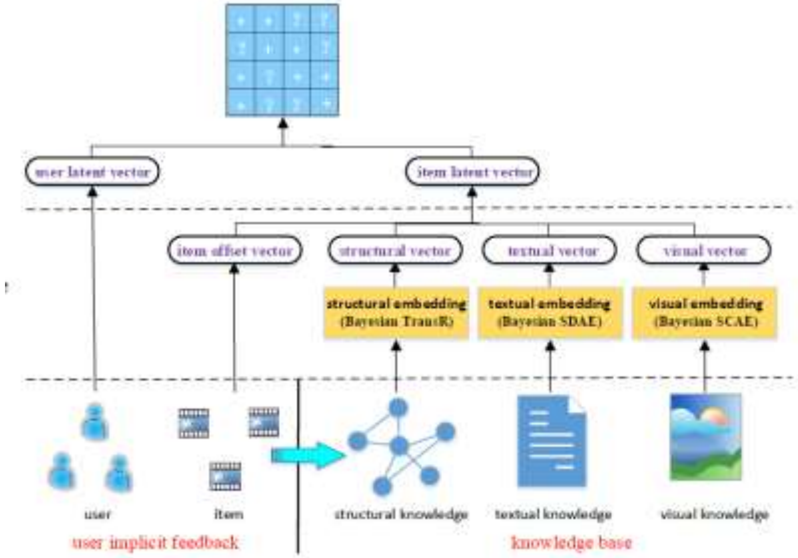
Personality in XiaoIce

Personality in Advertising

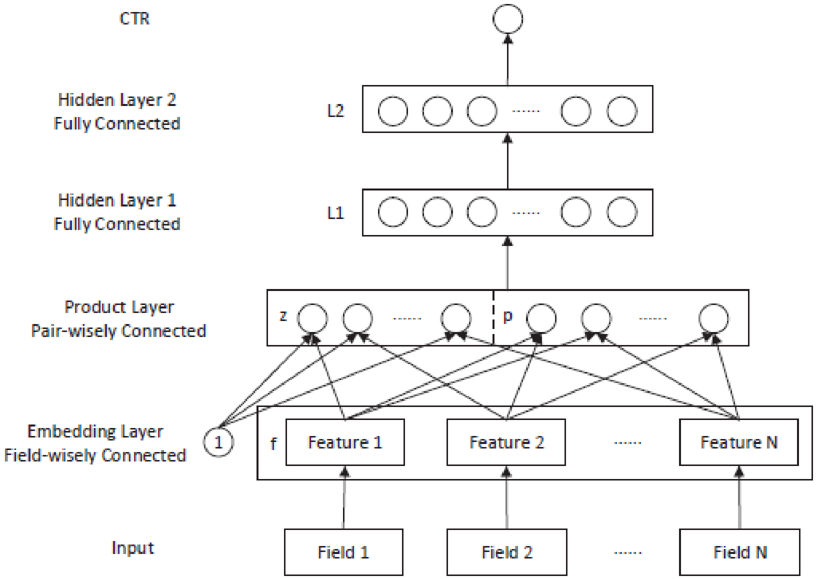
Tailoring messages to consumer personality increases effectiveness of digital advertising

Deep Learning Based Recommender System

Learning latent representations



Learning feature interactions



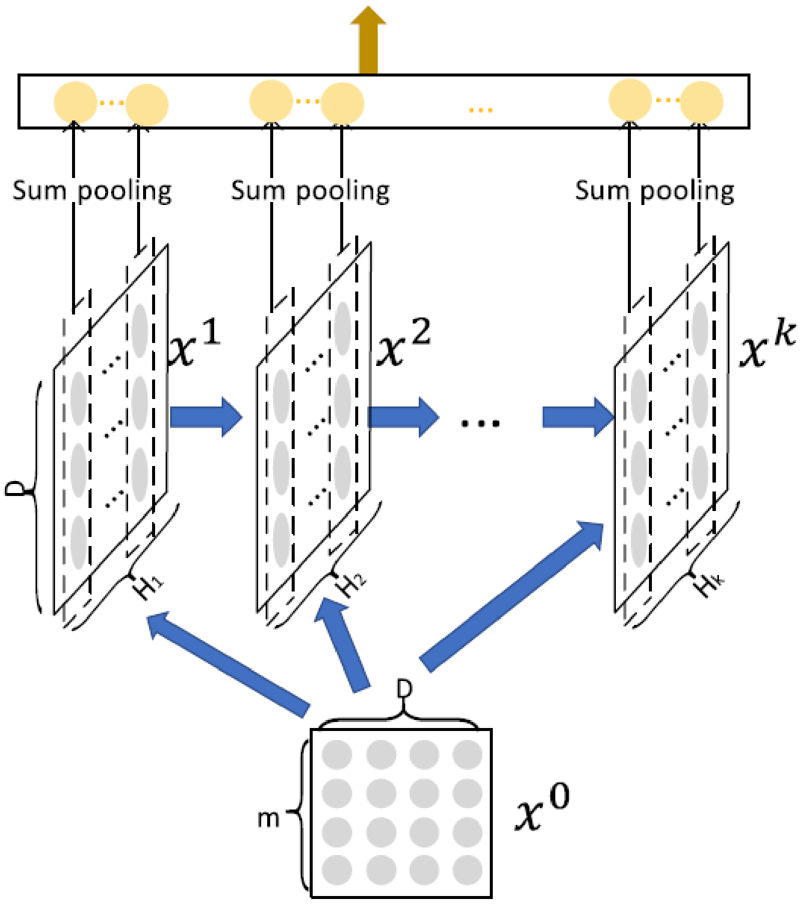
Motivations

- We try to design a new neural structure that
  - Automatically learns explicit high-order interactions
  - Vector-wise interaction, rather than bit-wise
    - $f(a_i, a_j) =< v_i, v_j > a_i a_j$
  - Different types of feature interactions can be combined easily
- Goals
  - Higher accuracy
  - Reducing manual feature engineering work

Compressed Interaction Network (CIN)

- Hidden units at the k-th layer:

$$\mathbf{X}^k_{h,*} = \sum_{i=1}^{H_{k-1}} \sum_{j=1}^{m} \mathbf{W}^{k,h}_{ij} (\mathbf{X}^{k-1}_{i,*} \circ \mathbf{X}^0_{j,*})$$

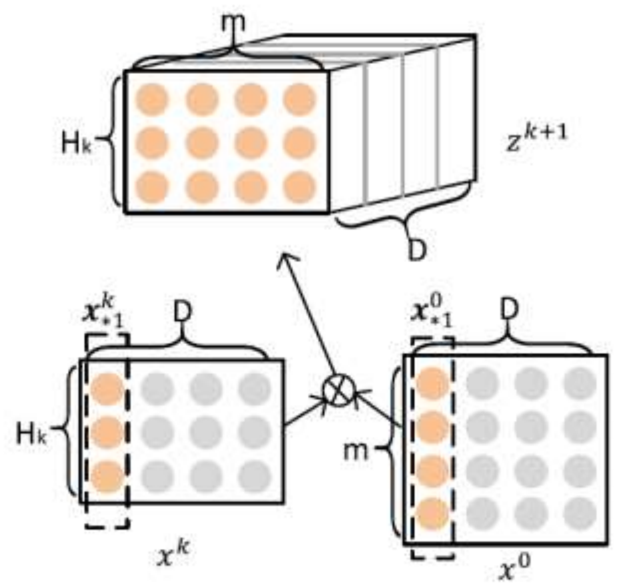


(c) An overview of the CIN architecture.

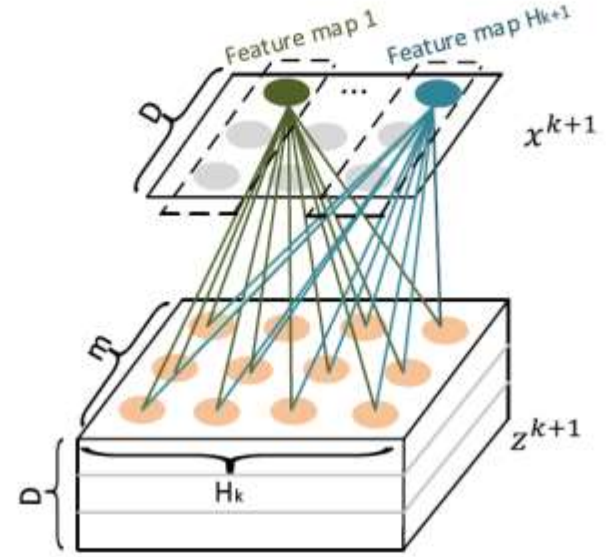
Compressed Interaction Network (CIN)

- Compression: interaction space from $O(m^2)$ down to $O(H_k)$
  - E.g., FM conduct the full pair-wise interaction, including necessary and unnecessary
- Keep the form of vectors
  - Hidden layers are matrices, rather than vectors
- Degree of feature interactions increases with the depth of layers (explicit)

Compressed Interaction Network (CIN)

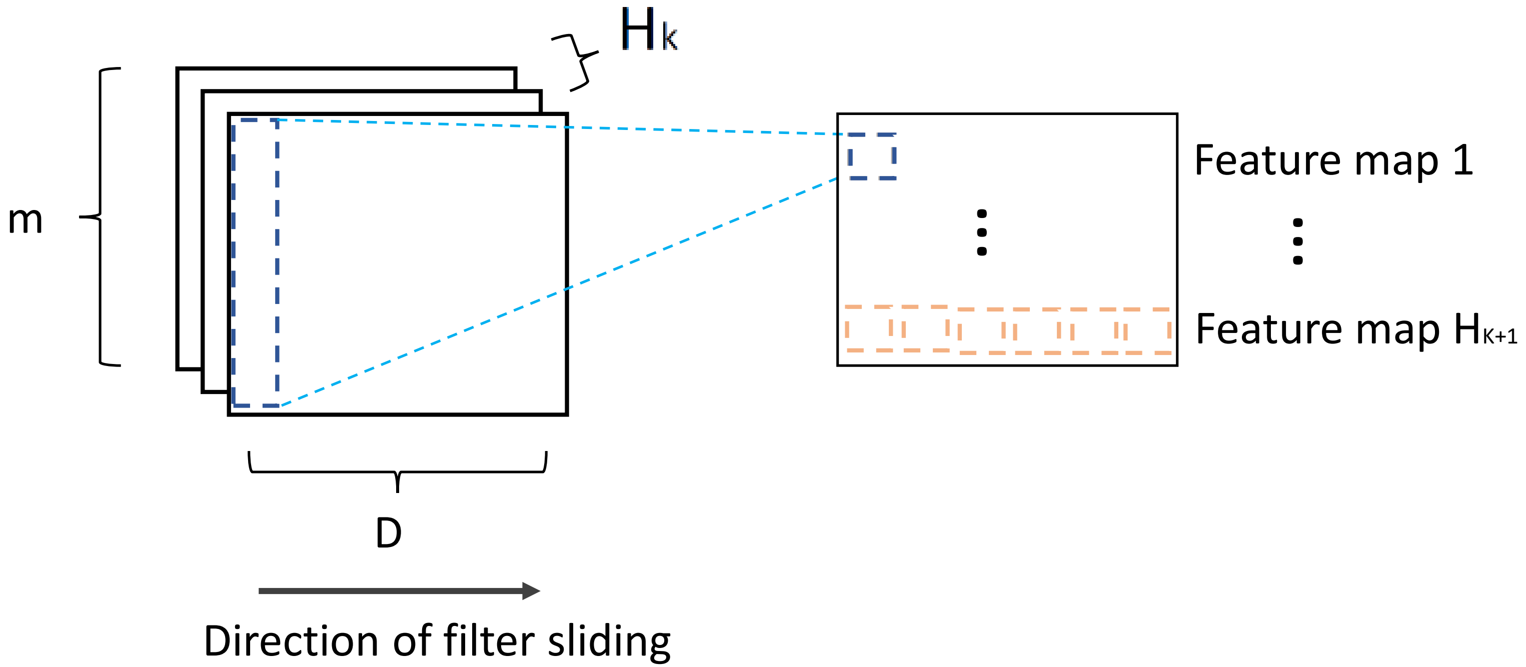


(a) Outer products along each dimension for feature interactions. The tensor Z^{k+1} is an intermediate result for further learning.

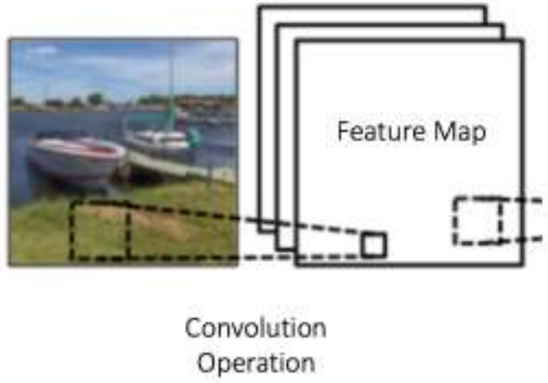


(b) The k-th layer of CIN. It compresses the intermediate tensor Z^{k+1} to H_{k+1} embedding vectors (aslo known as *feature maps*).

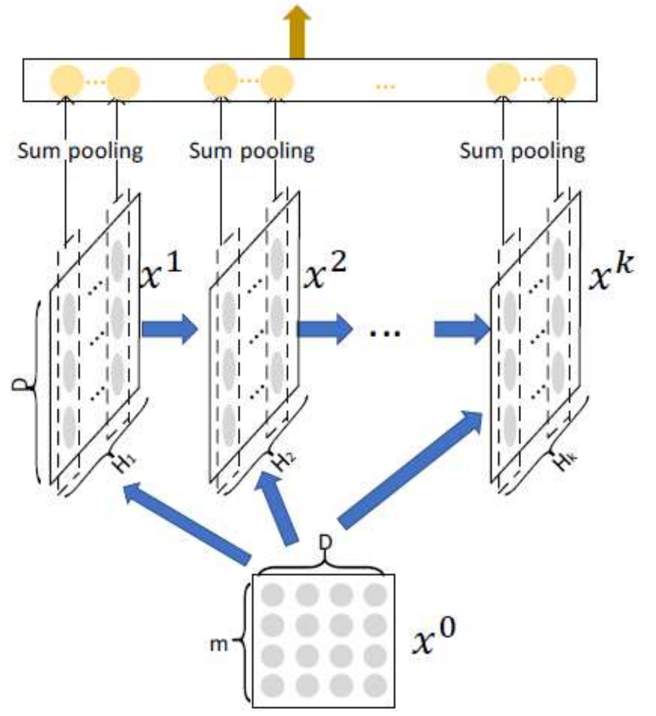
Relation with CNN



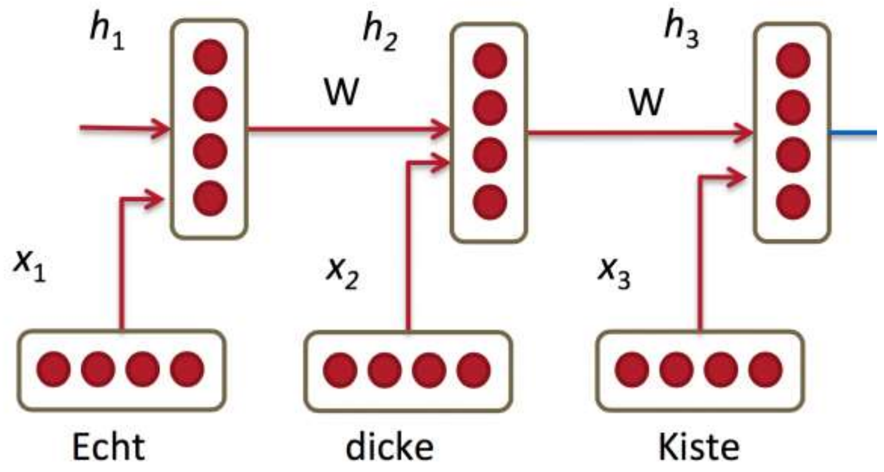
An example of image CNN



Relation with RNN



ours



RNN

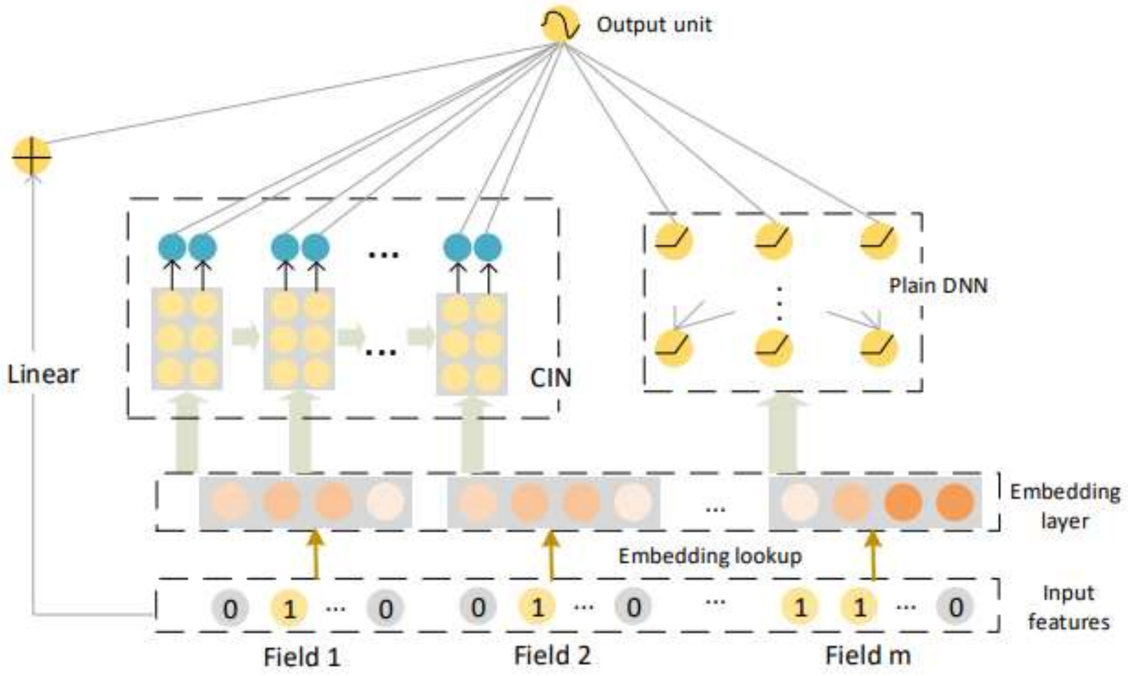
Extreme Deep Factorization Machine (xDeepFM)

- Combining explicit and implicit feature interaction network
- Integrate both memorization and generalization

$$\hat{y} = \sigma(\mathbf{w}_{linear}^T \mathbf{a} + \mathbf{w}_{dnn}^T \mathbf{x}_{dnn}^k + \mathbf{w}_{cin}^T \mathbf{p}^+ + b)$$

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} y_i log\hat{y}_i + (1 - y_i) log(1 - \hat{y}_i)$$

$$\mathcal{J} = \mathcal{L} + \lambda_* ||\Theta||$$



Experiments

- Three real-world datasets
  - Criteo: ads click-through-rate prediction
  - Dianping: restaurant recommendation
  - Bing News: news recommendation

Datasest	#instances	#fields	#features (sparse)
Criteo	45M	39	2.3M
Dianping	1.2M	18	230K
Bing News	5M	45	17K

- Evaluation metrics
  - AUC
  - Logloss

Experiments

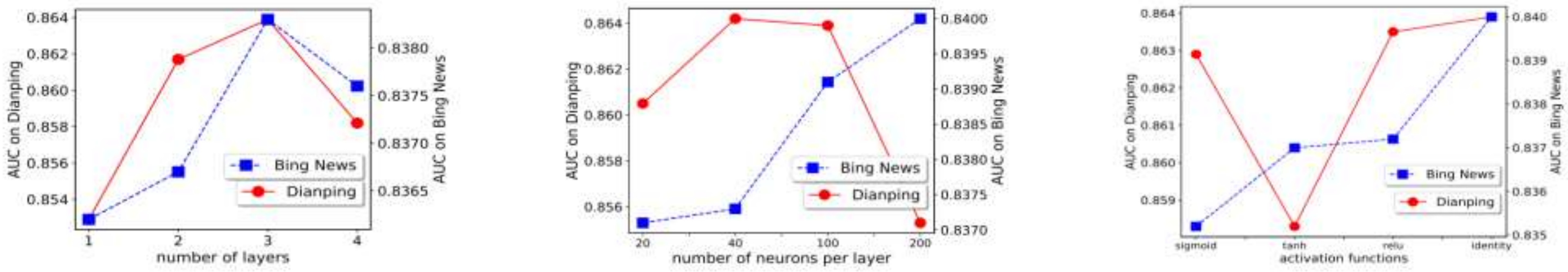
- High-order interactions are necessary
- Effectiveness of CIN

Model name	AUC	Logloss	Depth
Criteo			
FM	0.7900	0.4592	-
DNN	0.7993	0.4491	2
CrossNet	0.7961	0.4508	3
CIN	**0.8012**	0.4493	3
Dianping			
FM	0.8165	0.3558	-
DNN	0.8318	0.3382	3
CrossNet	0.8283	0.3404	2
CIN	**0.8576**	**0.3225**	2
Bing News			
FM	0.8223	0.2779	-
DNN	0.8366	0.273	2
CrossNet	0.8304	0.2765	6
CIN	**0.8377**	**0.2662**	5

Experiments

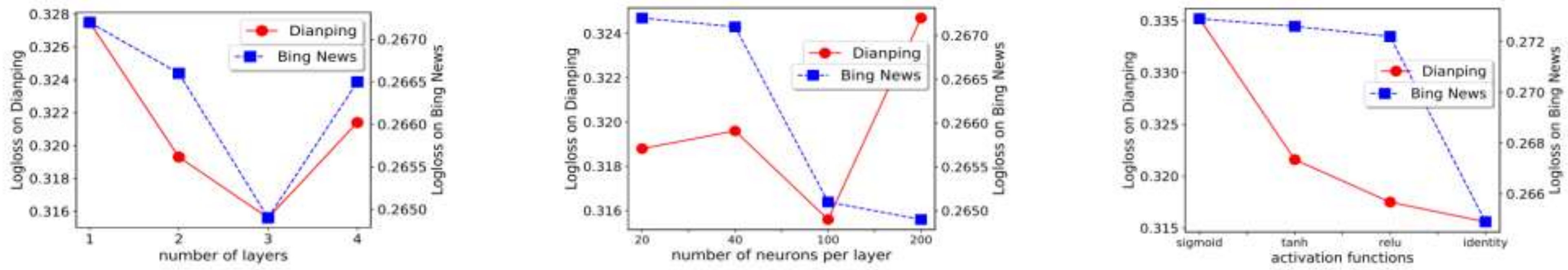
	Criteo			Dianping			Bing News		
Model name	AUC	Logloss	Depth	AUC	Logloss	Depth	AUC	Logloss	Depth
LR	0.7577	0.4854	-,-	0.8018	0.3608	-,-	0.7988	0.2950	-,-
FM	0.7900	0.4592	-,-	0.8165	0.3558	-,-	0.8223	0.2779	-,-
DNN	0.7993	0.4491	-,2	0.8318	0.3382	-,3	0.8366	0.2730	-,2
DCN	0.8026	0.4467	2,2	0.8391	0.3379	4,3	0.8379	0.2677	2,2
Wide&Deep	0.8000	0.4490	-,3	0.8361	0.3364	-,2	0.8377	0.2668	-,2
PNN	0.8038	0.4927	-,2	0.8445	0.3424	-,3	0.8321	0.2775	-,3
DeepFM	0.8025	0.4468	-,2	0.8481	0.3333	-,2	0.8376	0.2671	-,3
xDeepFM	**0.8052**	**0.4418**	3,2	**0.8639**	**0.3156**	3,3	**0.8400**	**0.2649**	3,2

Hyper-Parameter Sensitivity



(a) Number of layers. (b) Number of neurons per layer. (c) Activation functions

Figure 6: Impact of network hyper-parameters on AUC performance.



(a) Number of layers. (b) Number of neurons per layer. (c) Activation functions

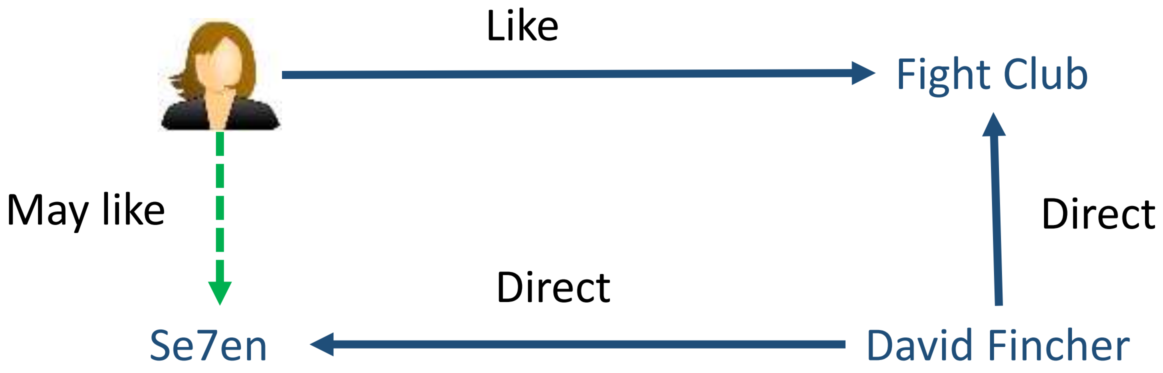
Figure 7: Impact of network hyper-parameters on Logloss performance.

Knowledge Graph

- A kind of semantic network, where node indicates entity or concept, edge indicates the semantic relation between entity/concept

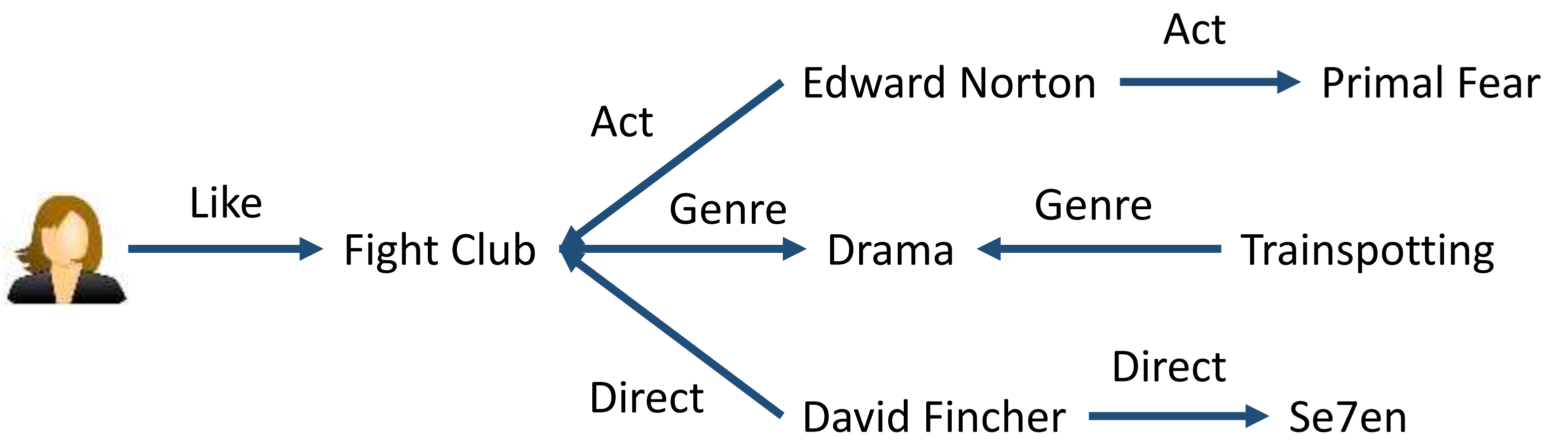
How Does Knowledge Graph Help?

- Precision
  - More semantic content about items
  - Deep user interest



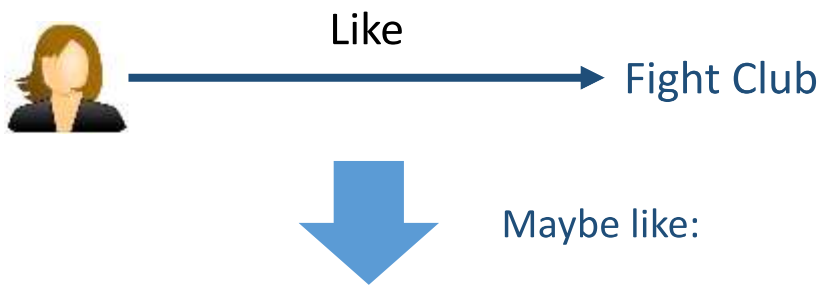
How Does Knowledge Graph Help?

- Diversity
  - Different types of relations in knowledge graph
  - Extend user's interest in different paths



How Does Knowledge Graph Help?

- Explanation ability
  - Connect user interest and recommendation results
  - Improve user satisfaction, boost user trust



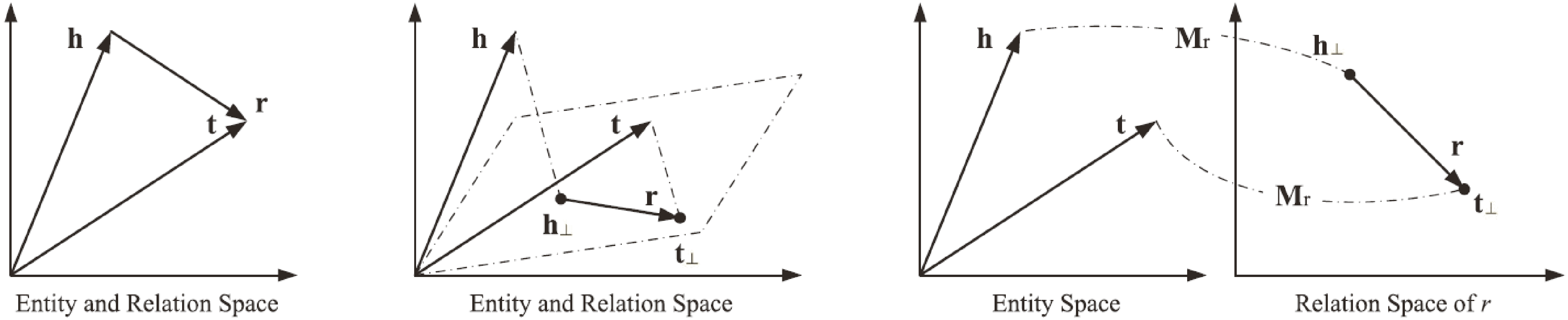
Primal Fear, because they share the same actor
Trainspotting, because they share the same genre
Se7en, because they share the same director

Knowledge Graph Embedding

- Learns a low-dimensional vector for each entity and relation in KG, which can keep the structural and semantic knowledge

Distance-based Models

- ❑ Apply distance-based score function to estimate the triple probability
- ❑ TransE, TransH, TransR, etc.

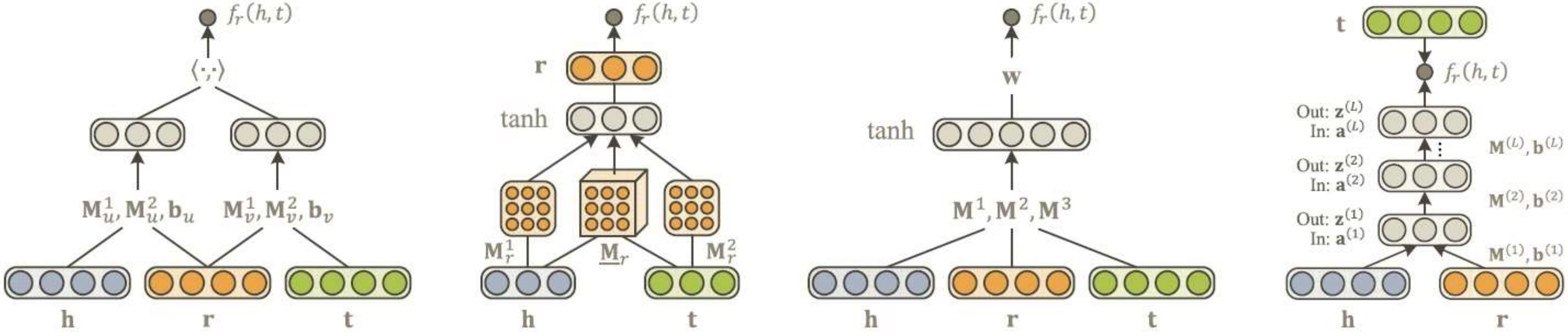


(a) TransE. (b) TransH. (c) TransR.

Knowledge Graph Embedding

Matching-based Models

- Apply similarity-based score function to estimate the triple probability
- SME, NTN, MLP, NAM, etc.



(a) SME. (b) NTN. (c) MLP. (d) NAM.

Knowledge Graph Embedding (KGE)

- Learns a low-dimensional vector for each entity and relation, which can keep the structural and semantic knowledge

KG, RS $\xrightarrow[\text{Learning}]{\text{KGE}}$ Entity Vector, Relation Vector / User Vector, Item Vector (Joint Training)

KG $\xrightarrow{\text{KGE}}$ Entity Vector / Relation Vector $\xrightarrow{\text{Feed into}}$ RS Task $\xrightarrow{\text{Learning}}$ User Vector / Item Vector (Successive Training)

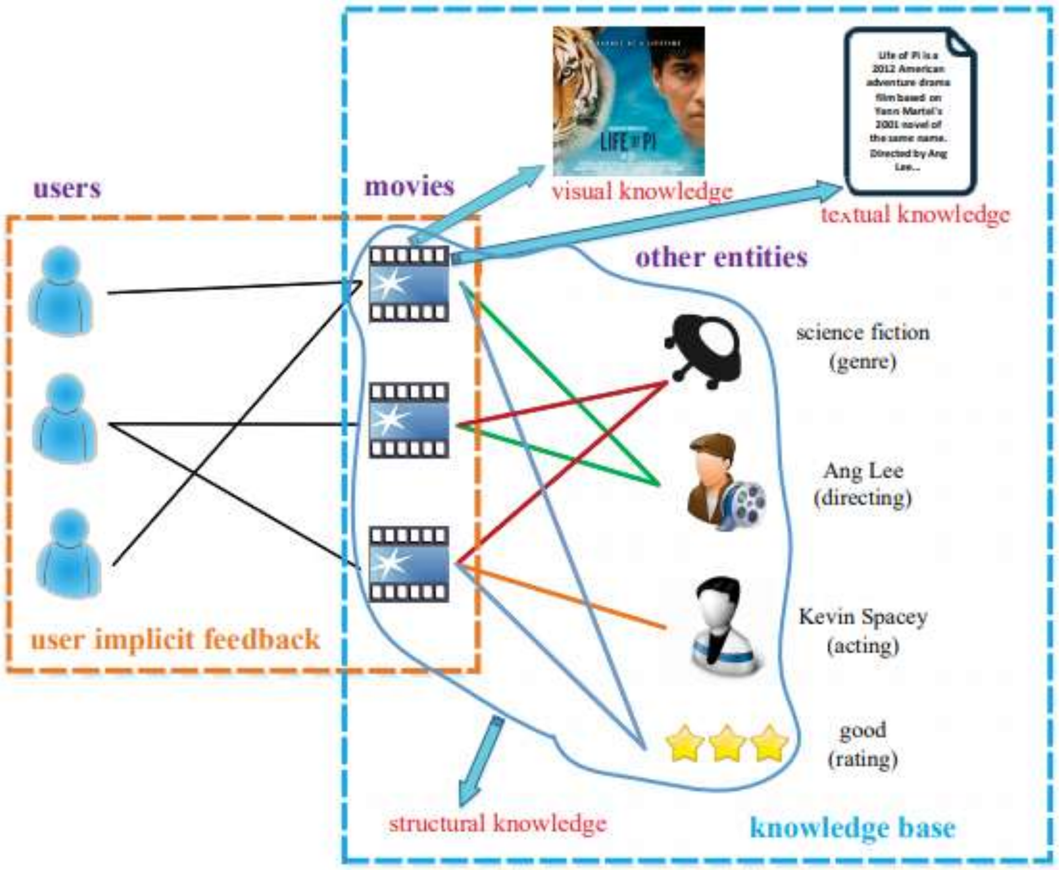
KG $\xrightarrow{\text{KGE}}$ Entity Vector, Relation Vector

RS $\xrightarrow{\text{Learning}}$ User Vector, Item Vector

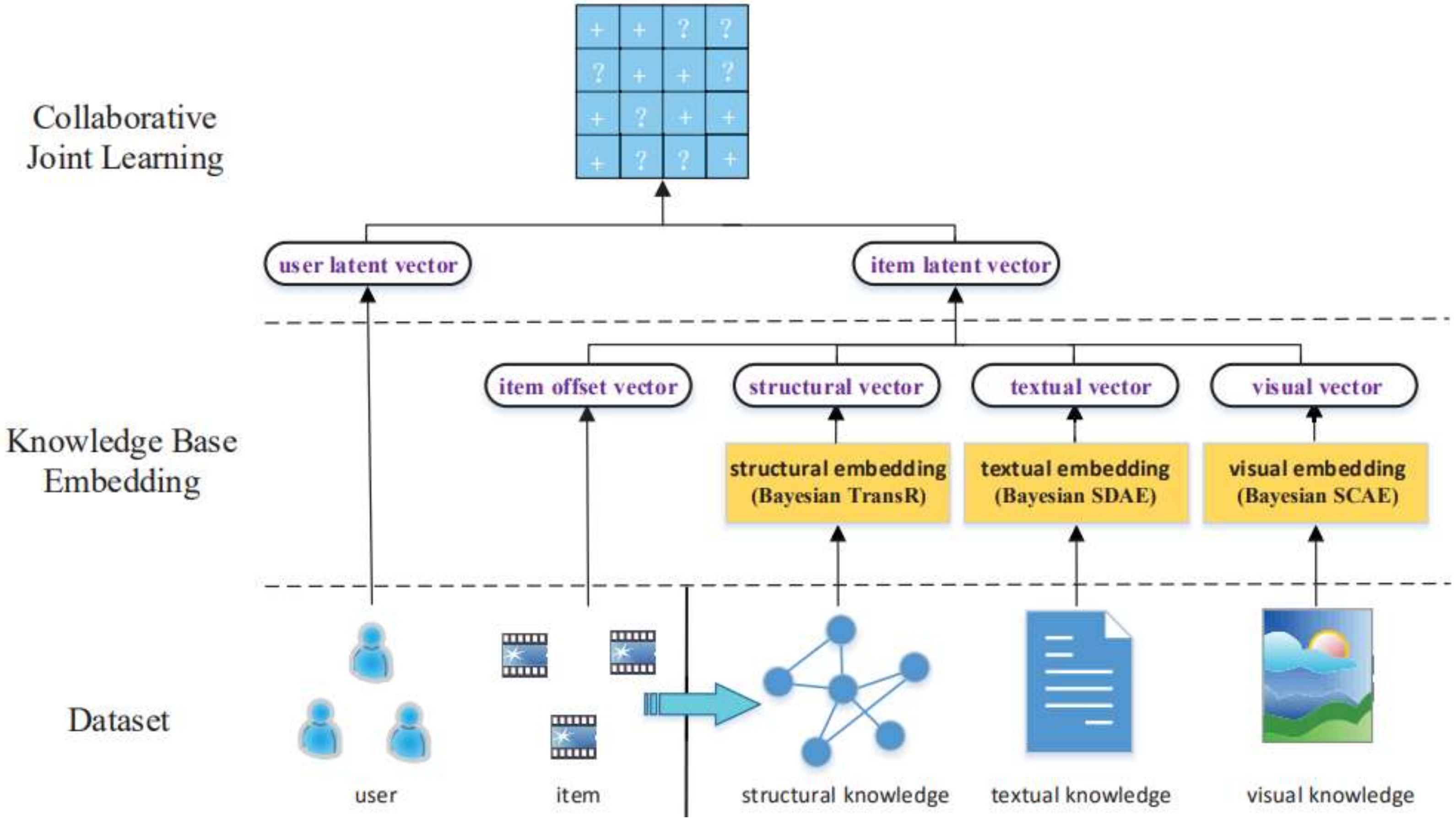
(Alternate Training)

Collaborative Knowledge Embedding (Joint Training)

- Structural knowledge
  - Direct, act, etc.
- Visual knowledge
  - Movie poster, book cover image, etc.
- Textual knowledge
  - Movie description, reviews, etc.



Collaborative Knowledge Embedding



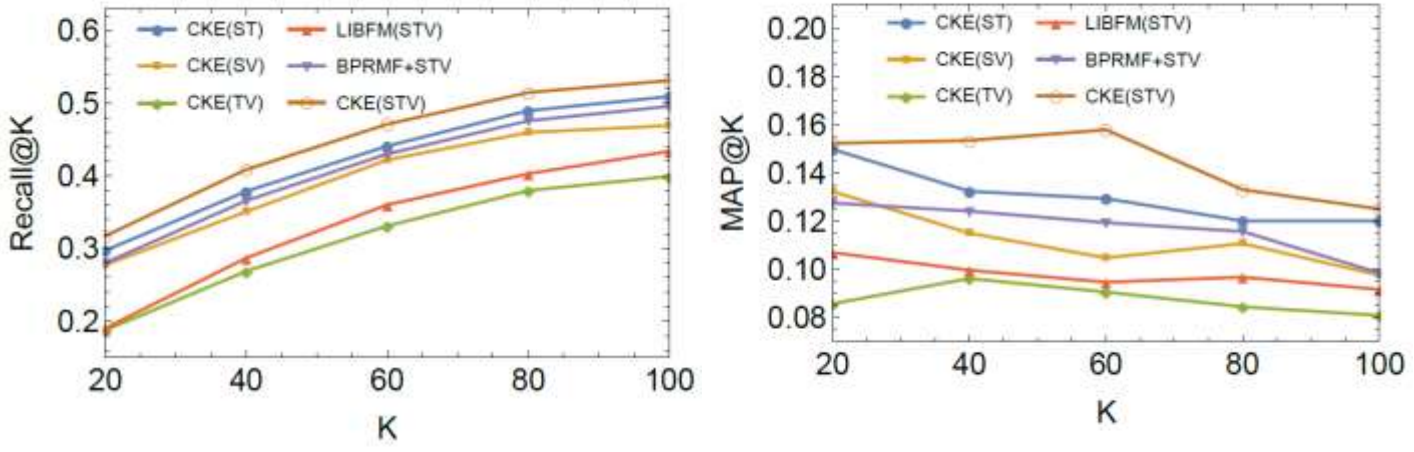
Data

- MovieLens-1M
  - 1-step subgraph includes category, director, writer, actors, language, country, production date, rating, nominated awards, and received awards
- IntentBooks
  - 9-month Bing query logs, apply entity linking to find out book entity
  - 1-step subgraph includes category, author, publish date, belonged series, language, and rating

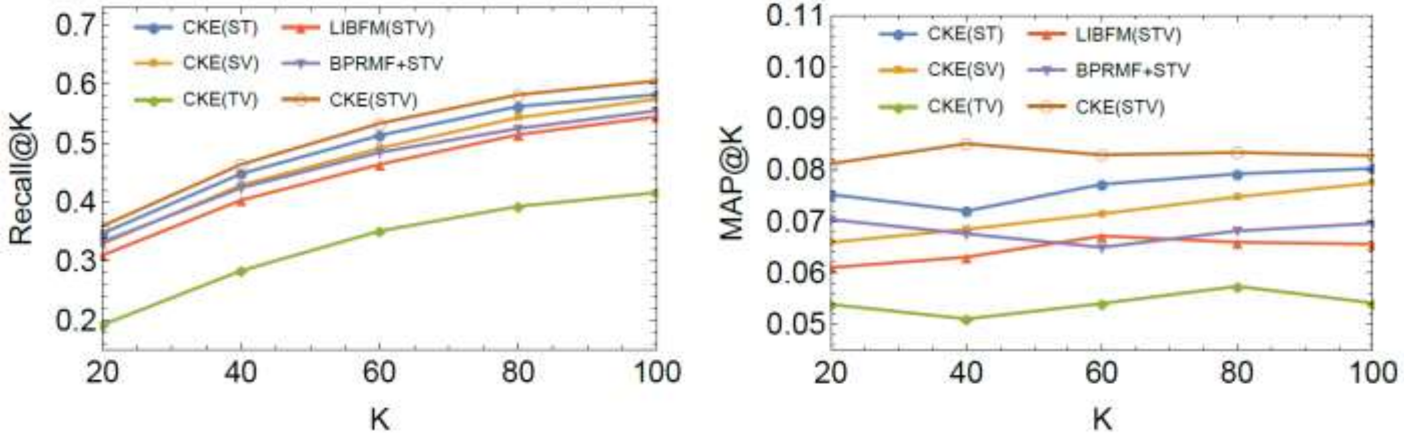
	MovieLens-1M	IntentBooks
#user	5,883	92,564
#item	3,230	18,475
#interactions	226,101	897,871
#sk nodes	84,011	26,337
#sk edges	169,368	57,408
#sk edge types	10	6
#tk items	2,752	17,331
#vk items	2,958	16,719

Results

- Baselines
  - CKE(ST), CKE(SV), CKE(TV): only two types of knowledge
  - LIBFM(STV): all knowledge as raw features
  - BPRMF+STV: not joint-learning



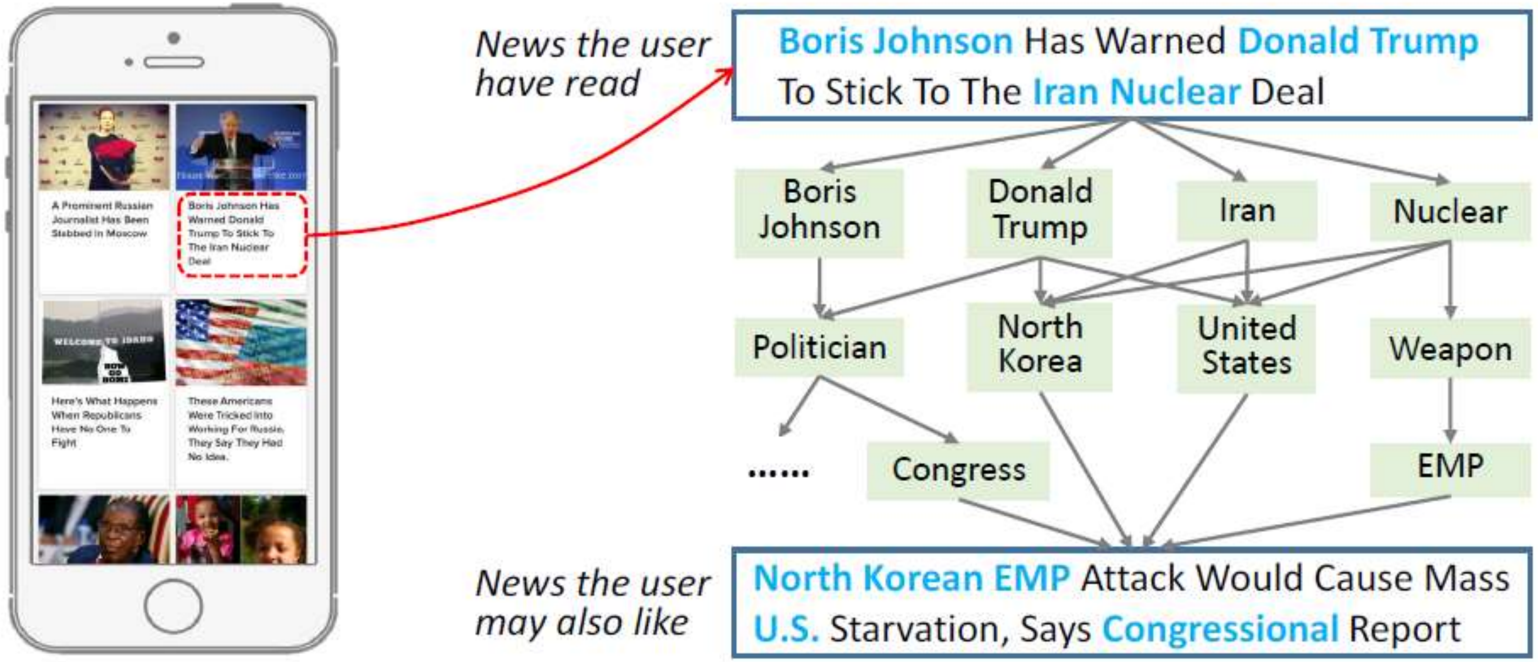
MovieLens-1M



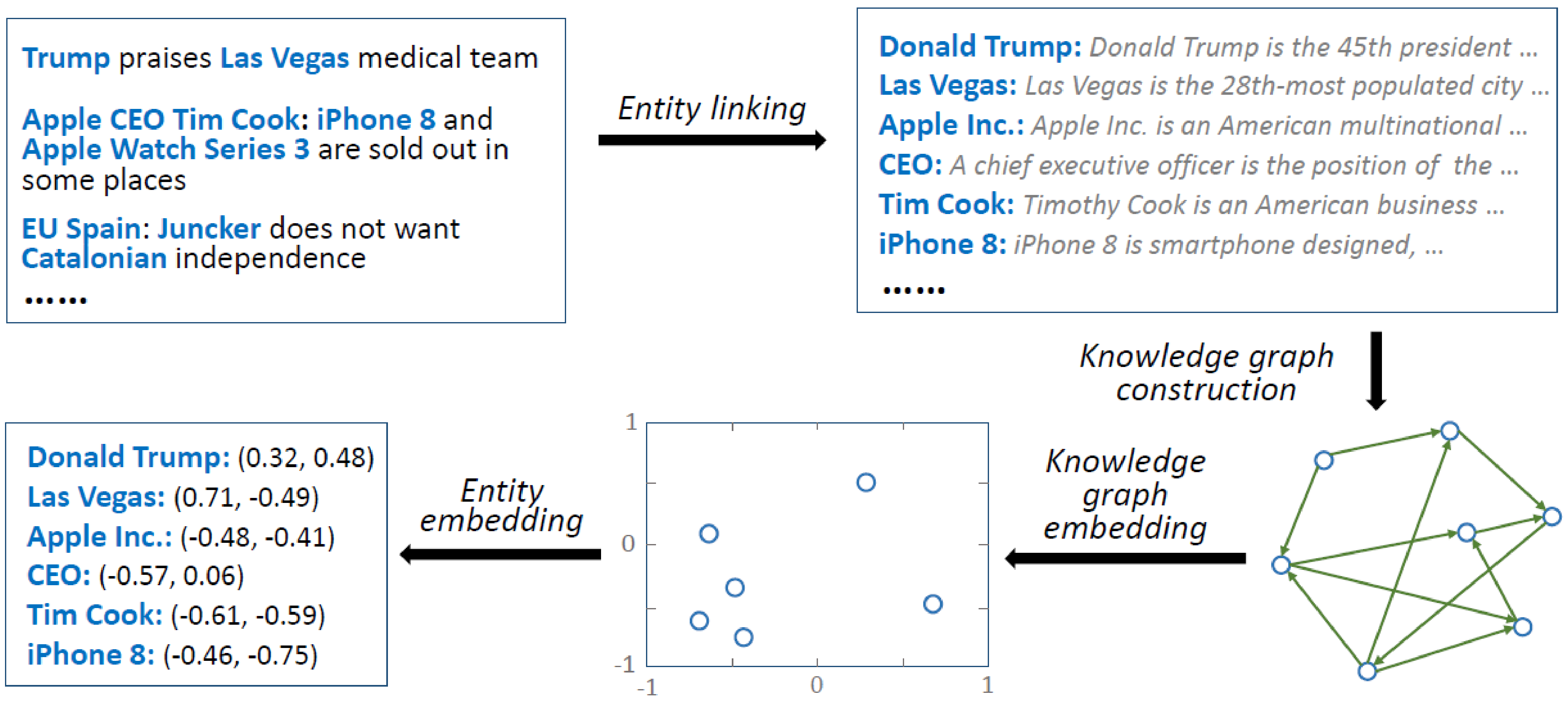
IntentBooks

Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, Wei-Ying Ma, Collaborative Knowledge Base Embedding for Recommender Systems, KDD 2016

Deep Knowledge-aware Network (Successive Training)

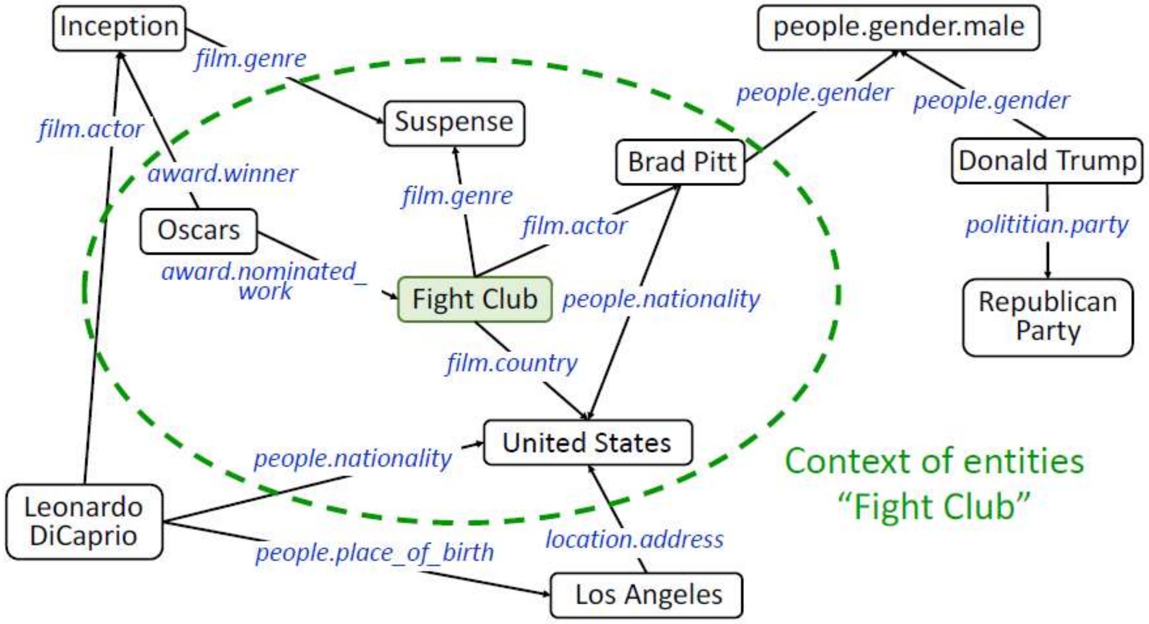


Deep Knowledge-aware Network

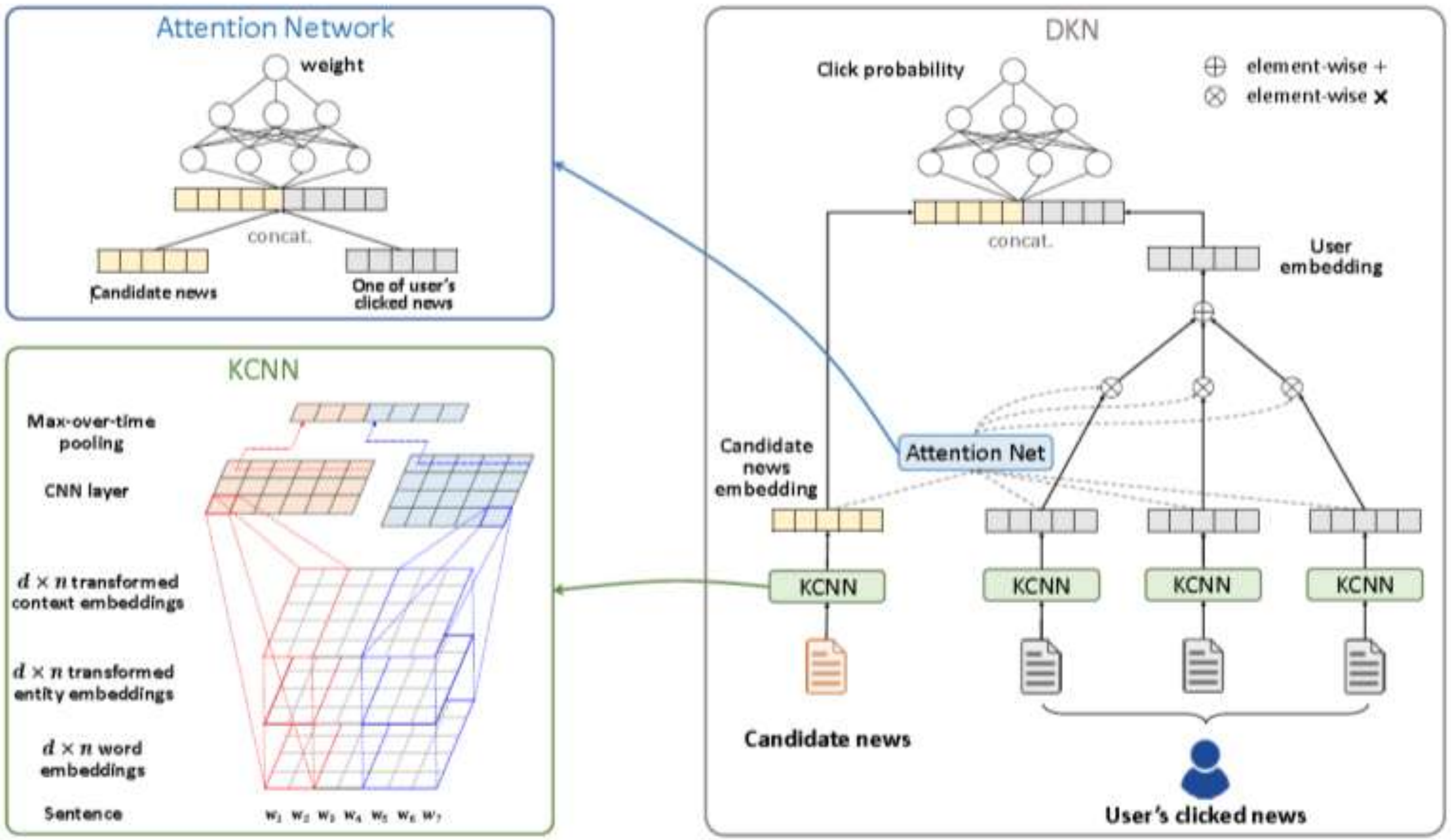


Extract Knowledge Representations

- Additionally use contextual entity embeddings to include structural information
- Context implies one-step neighbor



Deep Knowledge-aware Network

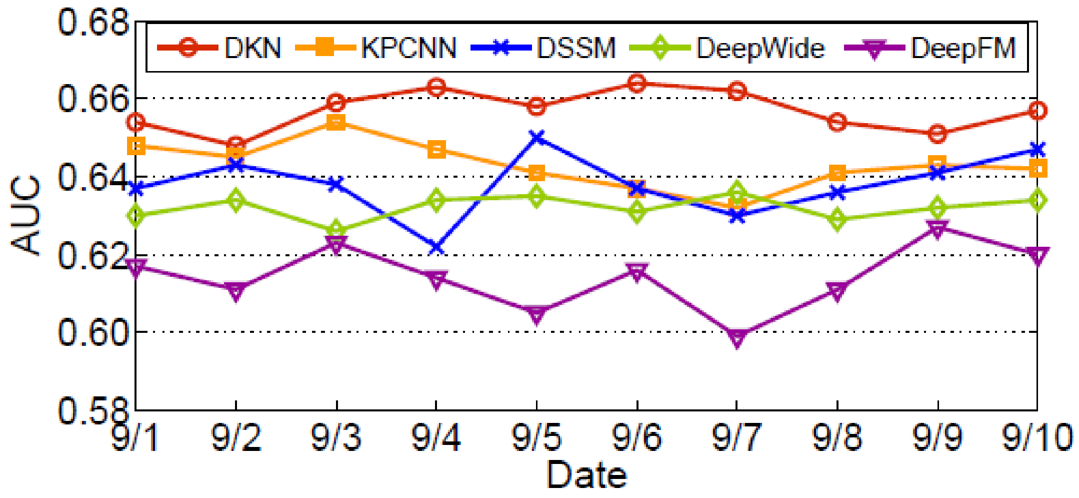


Experiments

Models*	F1	AUC	p-value**
DKN	**68.9 ± 1.5**	**65.9 ± 1.2**	–
LibFM	61.8 ± 2.1 (-10.3%)	59.7 ± 1.8 (-9.4%)	$< 10^{-3}$
LibFM(-)	61.1 ± 1.9 (-11.3%)	58.9 ± 1.7 (-10.6%)	$< 10^{-3}$
KPCNN	67.0 ± 1.6 (-2.8%)	64.2 ± 1.4 (-2.6%)	0.098
KPCNN(-)	65.8 ± 1.4 (-4.5%)	63.1 ± 1.5 (-4.2%)	0.036
DSSM	66.7 ± 1.8 (-3.2%)	63.6 ± 2.0 (-3.5%)	0.063
DSSM(-)	66.1 ± 1.6 (-4.1%)	63.2 ± 1.8 (-4.1%)	0.045
DeepWide	66.0 ±1.2 (-4.2%)	63.3 ± 1.5 (-3.9%)	0.039
DeepWide(-)	63.7 ± 0.9 (-7.5%)	61.5 ± 1.1 (-6.7%)	0.004
DeepFM	63.8 ± 1.5 (-7.4%)	61.2 ± 2.3 (-7.1%)	0.014
DeepFM(-)	64.0 ± 1.9 (-7.1%)	61.1 ± 1.8 (-7.3%)	0.007
YouTubeNet	65.5 ± 1.2 (-4.9%)	63.0 ± 1.4 (-4.4%)	0.025
YouTubeNet(-)	65.1 ± 0.7 (-5.5%)	62.1 ± 1.3 (-5.8%)	0.011
DMF	57.2 ± 1.2 (-17.0%)	55.3 ± 1.0 (-16.1%)	$< 10^{-3}$

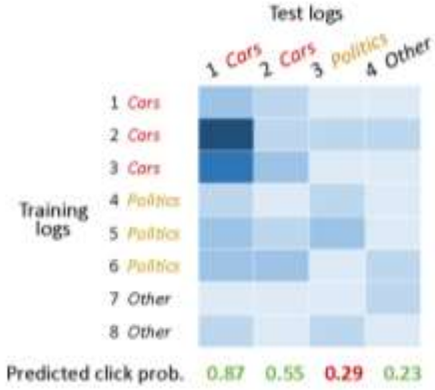
* "(-)" denotes "without input of entity embeddings".

** p-value is the probability of no significant difference with DKN on AUC by t-test.

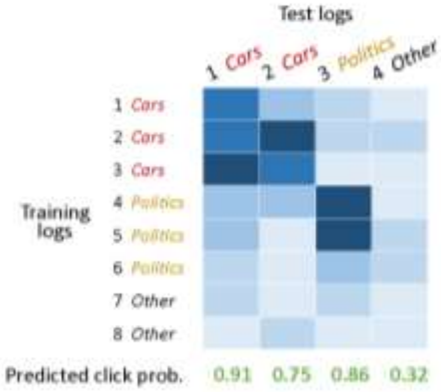


Examples

	No.	Date	News title	Entities	Label	Category
training	1	12/25/2016	Elon Musk teases huge upgrades for Tesla's supercharger network	Elon Musk; Tesla Inc.	1	Cars
	2	03/25/2017	Elon Musk offers Tesla Model 3 sneak peek	Elon Musk; Tesla Model 3	1	Cars
	3	12/14/2016	Google fumbles while Tesla sprints toward a driverless future	Google Inc.; Tesla Inc.	1	Cars
	4	12/15/2016	Trump pledges aid to Silicon Valley during tech meeting	Donald Trump; Silicon Valley	1	Politics
	5	03/26/2017	Donald Trump is a big reason why the GOP kept the Montana House seat	Donald Trump; GOP; Montana	1	Politics
	6	05/03/2017	North Korea threat: Kim could use nuclear weapons as "blackmail"	North Korea; Kim Jong-un	1	Politics
	7	12/22/2016	Microsoft sells out of unlocked Lumia 950 and Lumia 950 XL in the US	Microsoft; Lumia; United States	1	Other
	8	12/08/2017	6.5 magnitude earthquake recorded off the coast of California	earthquake; California	1	Other
			......			
test	1	07/08/2017	Tesla makes its first Model 3	Tesla Inc; Tesla Model 3	1	Cars
	2	08/13/2017	General Motors is ramping up its self-driving car: Ford should be nervous	General Motors; Ford Inc.	1	Cars
	3	06/21/2017	Jeh Johnson testifies on Russian interference in 2016 election	Jeh Johnson; Russian	1	Politics
	4	07/16/2017	"Game of Thrones" season 7 premiere: how you can watch	Game of Thrones	0	Other



(a) without knowledge graph

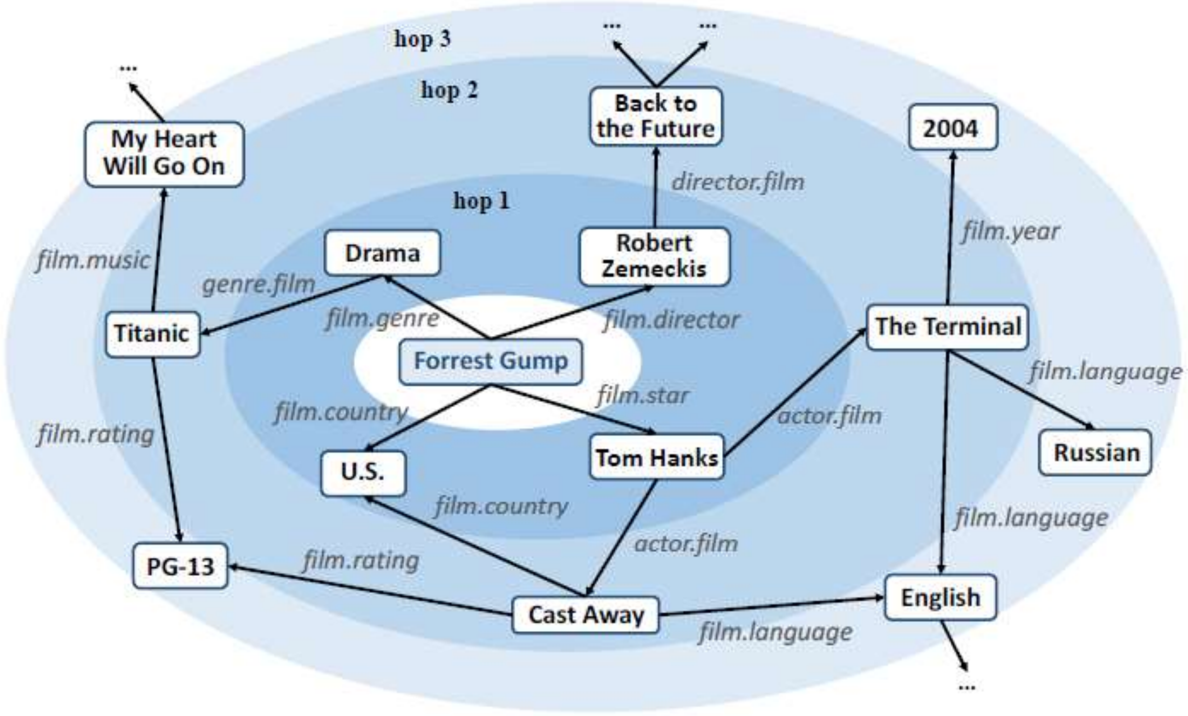


(b) with knowledge graph

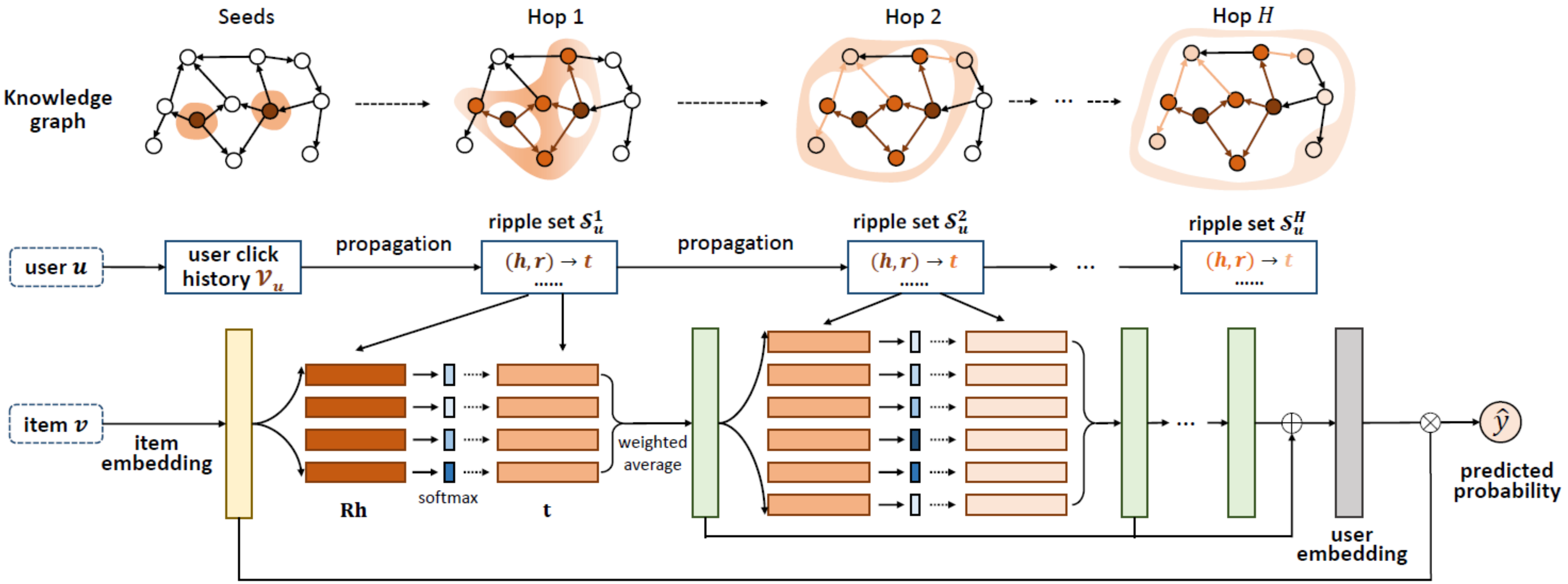
Hongwei Wang, Fuzheng Zhang, Xing Xie, Minyi Guo, DKN: Deep Knowledge-Aware Network for News Recommendation, WWW 2018

Ripple Network (Joint Training)

- Users interests as seed entity, propagates in the graph step by step
- Decay in the propagating process



Ripple Network

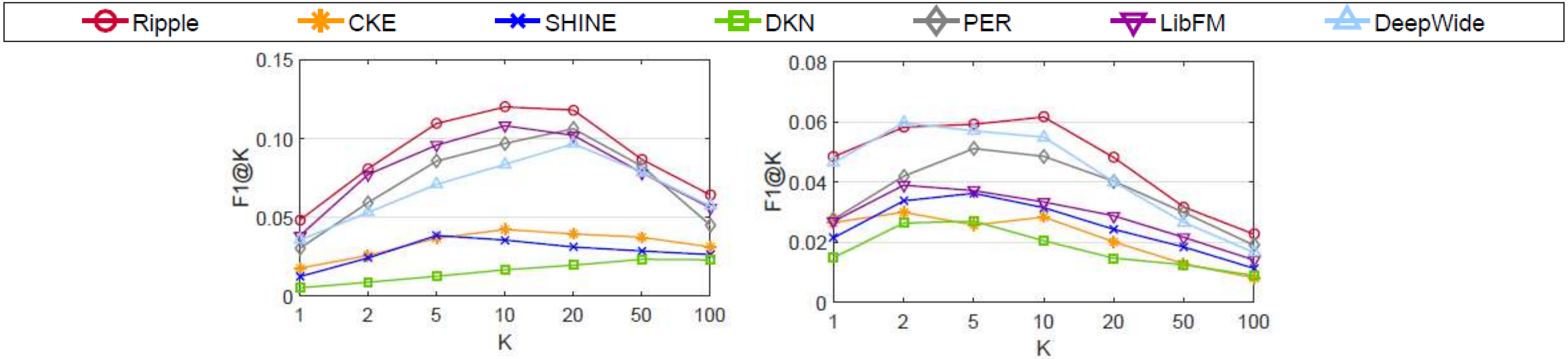


$$
\begin{aligned}
\min \mathcal{L} &= -\log\left(p(\mathrm{Y}|\Theta,\mathcal{G}) \cdot p(\mathcal{G}|\Theta) \cdot p(\Theta)\right) \\
&= \sum_{(u,v)\in \mathrm{Y}} -\Big(y_{uv}\log\sigma(\mathbf{u}^{\mathrm{T}}\mathbf{v}) + (1-y_{uv})\log\left(1-\sigma(\mathbf{u}^{\mathrm{T}}\mathbf{v})\right)\Big) \\
&\quad + \frac{\lambda_2}{2}\sum_{r\in\mathcal{R}} \|\mathbf{I}_r - \mathbf{E}^{\mathrm{T}}\mathbf{R}\mathbf{E}\|_2^2 + \frac{\lambda_1}{2}\Big(\|\mathbf{V}\|_2^2 + \|\mathbf{E}\|_2^2 + \sum_{r\in\mathcal{R}}\|\mathbf{R}\|_2^2\Big)
\end{aligned}
$$

Experiments

Model	MovieLens-1M		Book-Crossing		Bing-News	
	AUC	*ACC*	*AUC*	*ACC*	*AUC*	*ACC*
Ripple*	**0.913**	**0.835**	**0.840**	**0.775**	**0.778**	**0.732**
CKE	0.796	0.739	0.634	0.606	0.660	0.617
SHINE	0.778	0.732	0.668	0.636	0.614	0.587
DKN	0.655	0.589	0.621	0.598	0.761	0.704
PER	0.901	0.826	0.814	0.735	-	-
LibFM	0.892	0.812	0.763	0.705	0.744	0.688
DeepWide	0.903	0.822	0.806	0.731	0.754	0.695

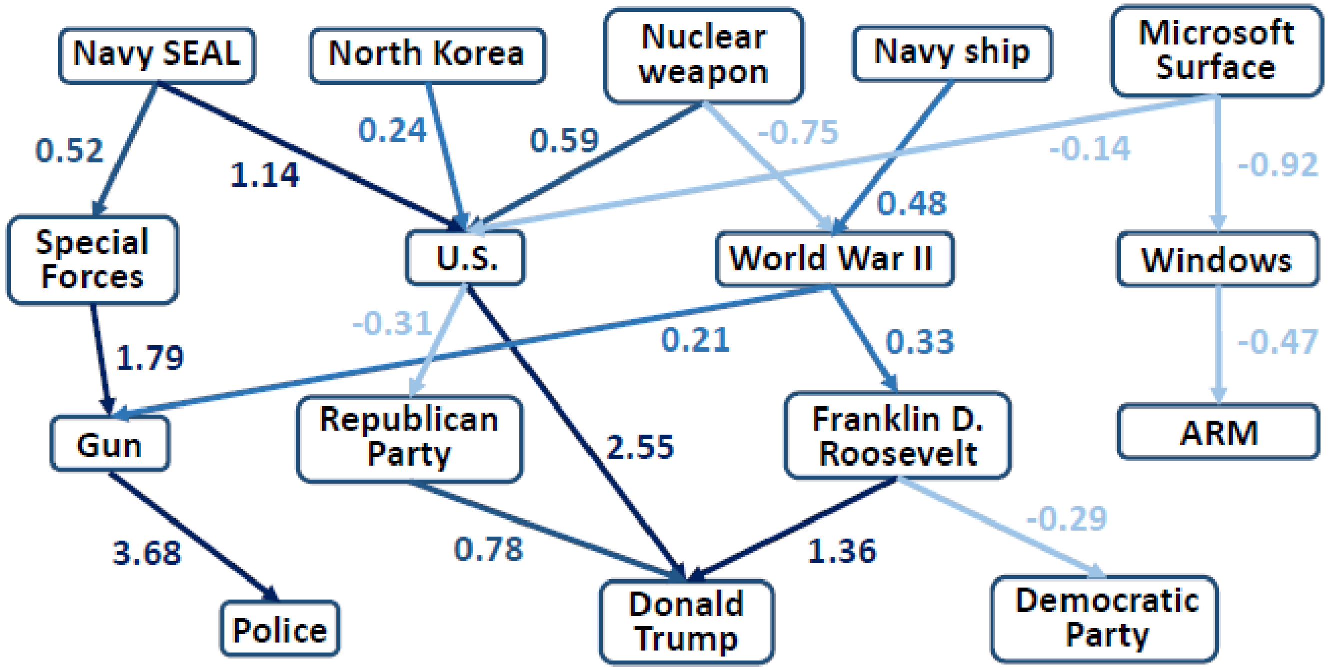
* Statistically significant improvements by t-test.



Example

Click history:

1. *Family of **Navy SEAL** Trainee Who Died During Pool Exercise Plans to Take Legal Action*
2. ***North Korea** Vows to Strengthen **Nuclear Weapons***
3. ***North Korea** Threatens 'Toughest Counteraction' After **U.S.** Moves **Navy Ships***
4. *Consumer Reports Pulls Recommendation for **Microsoft Surface** Laptops*



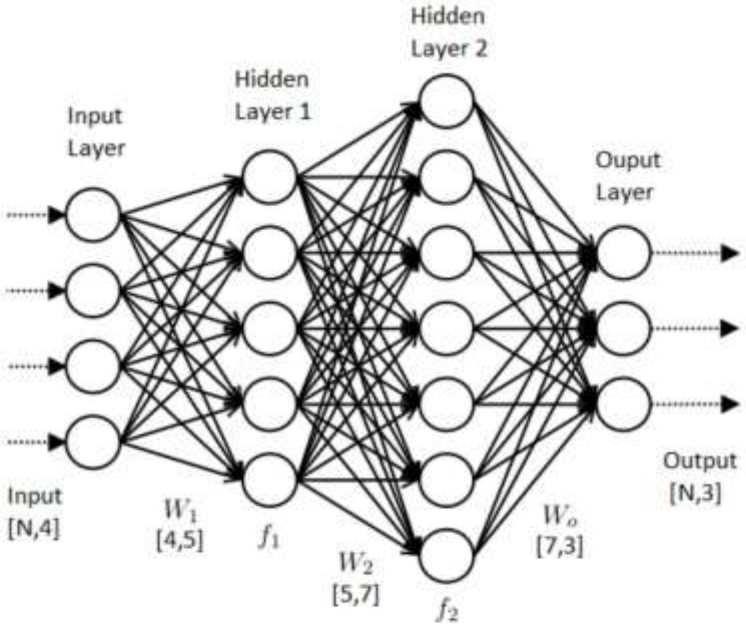
Candidate news: ***Trump** Announces Gunman Dead, Credits 'Heroic Actions' of **Police***

可解释推荐

Explainable AI

Attention from

- Government
- Industry
- Academia

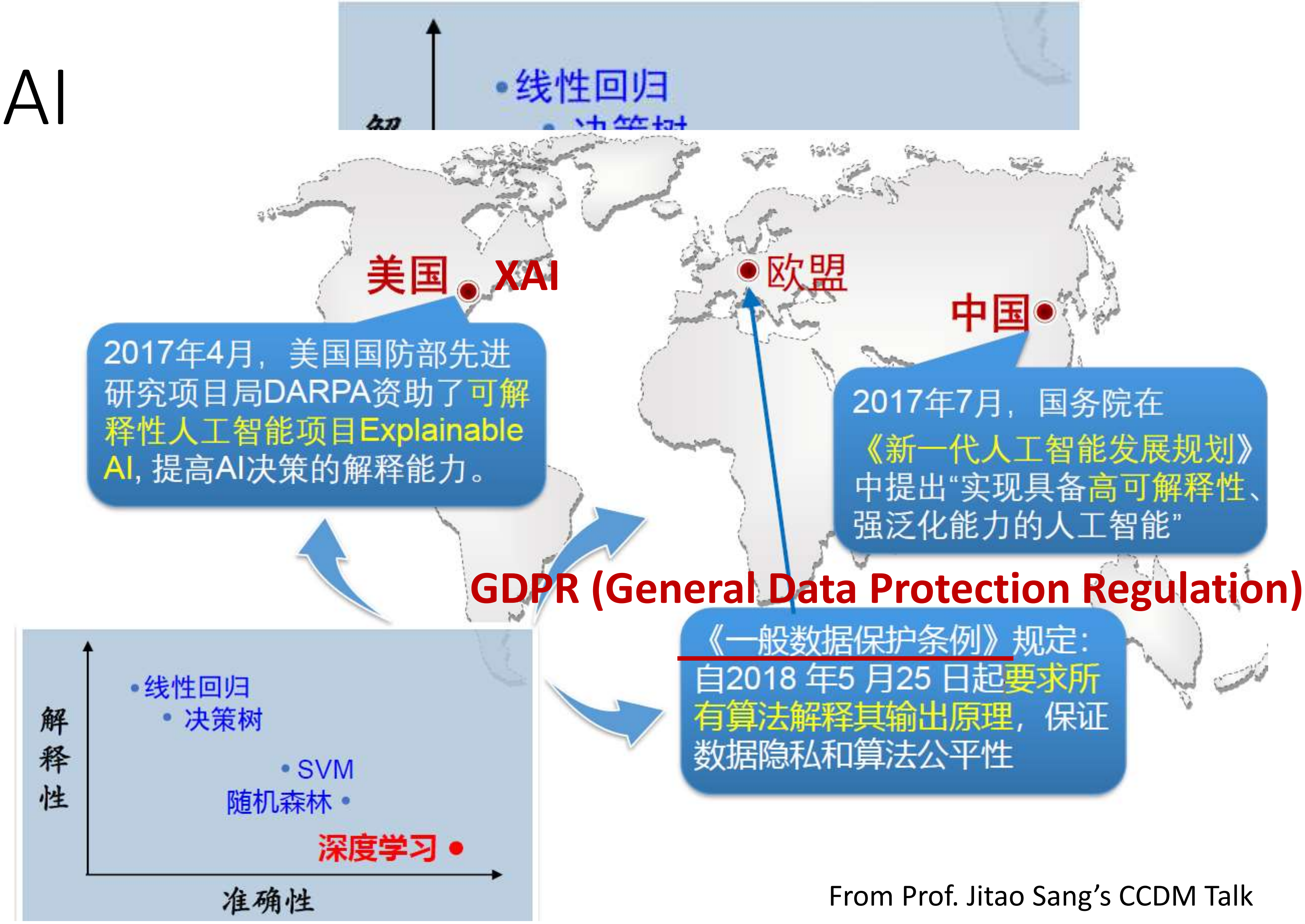


? =

Explainable AI

Attention from

- **Government**
- Industry
- Academia



From Prof. Jitao Sang’s CCDM Talk

Explainable AI

Attention from

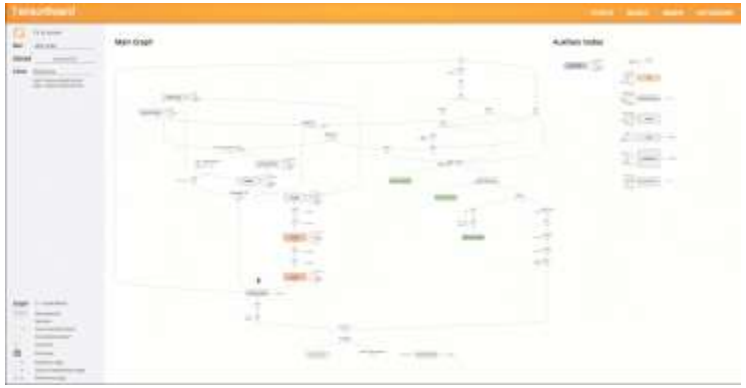
- Government
- **Industry**
- Academia

Microsoft: Invests 1,600 engineers to support GDPR compliance

facebook: Moves more than 1.5 billion users out of reach of European privacy law

Google:

TensorBoard: Graph Visualization



Explainable AI

Attention from
- Government
- Industry
- **Academia**

ICML 2017 Awards

Best Paper Award

Understanding Black-box Predictions via Influence Functions

Pang Wei Koh, Percy Liang

NIPS | 2017 **Best paper awards:**

A Linear-Time Kernel Goodness-of-Fit Test.

- Wittawat Jitkrittum, Wenkai Xu, Zoltan Szabo, Kenji Fukumizu, Arthur Gretton.

Sam Charringtion from TWiML&AI, the authors of the NIPS 2017 best paper said at 14:10 in the following video that " … *explainability was one of the reasons that the paper was given the award* …"

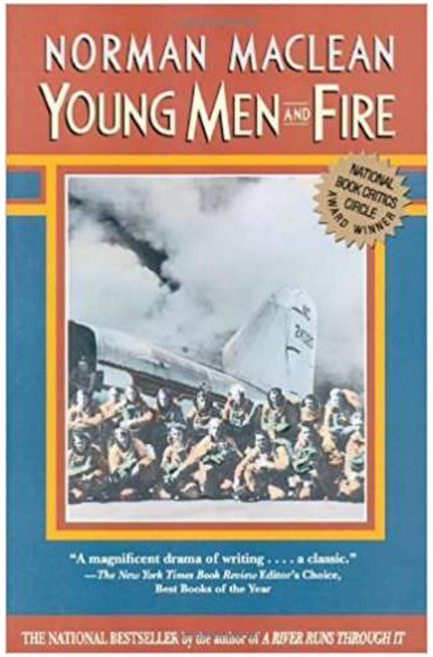
KDD2018

11 accepted papers mentioned interpretation/explanation in the title

Traditional vs. Explainable Recommendation

- Traditional recommendation
  - What, Who, When, Where
- Explainable recommendation
  - **Why**

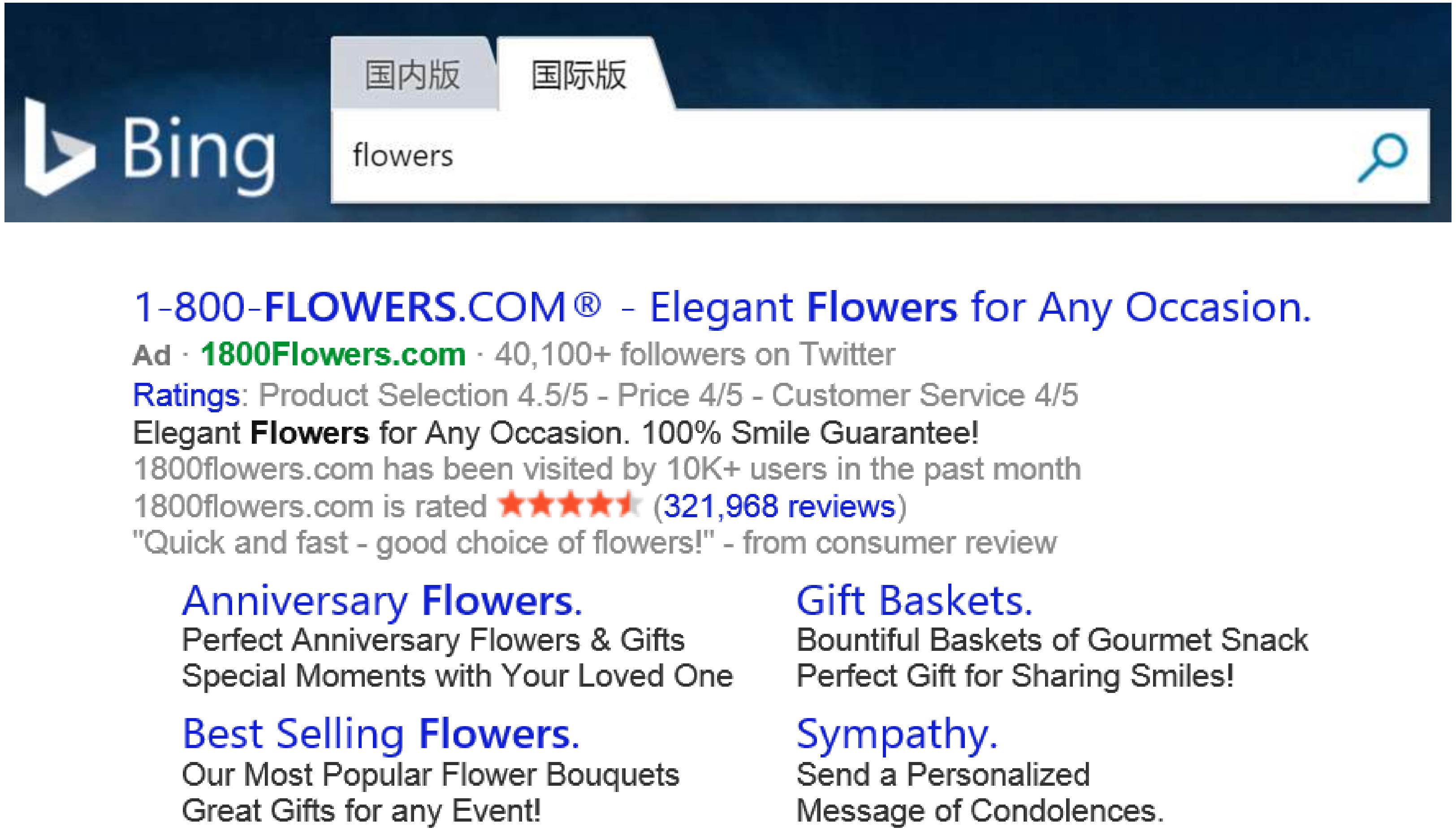
Connect the item with the user: <u>persuasiveness</u>, <u>trust</u>, <u>satisfaction</u>



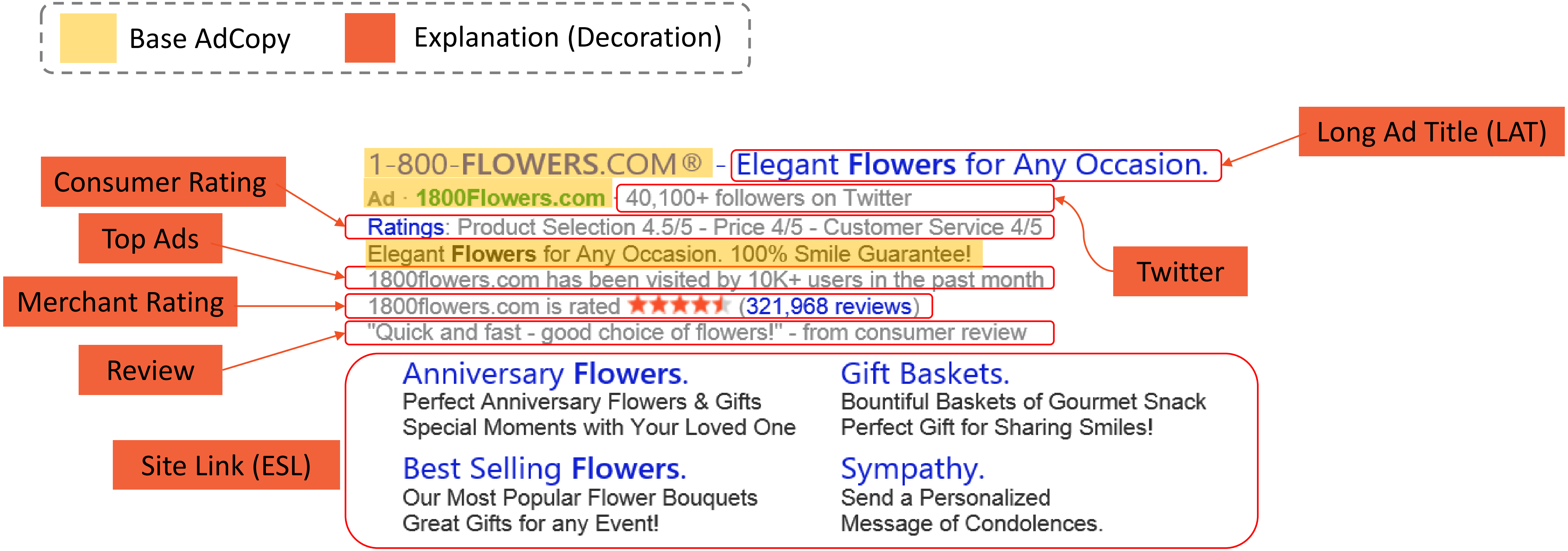
It impacts how Satya thinks about leadership

It may help **you** better understand some major decisions of Satya

Explainable Recommendation for Ads

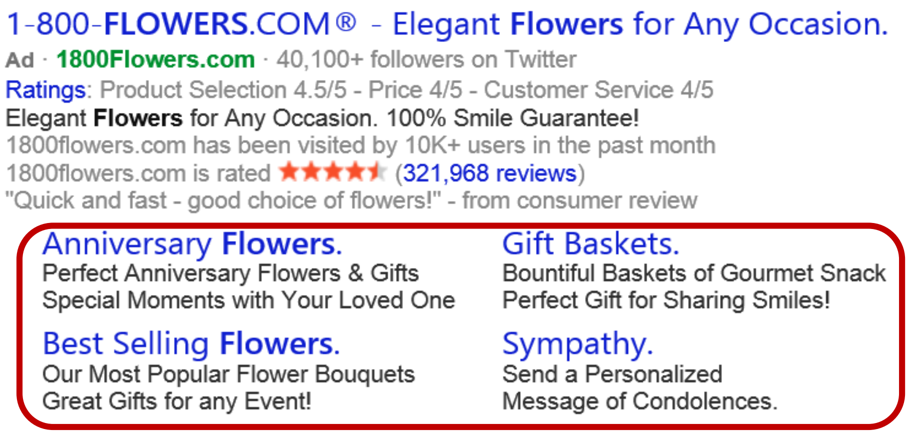


Explainable Recommendation for Ads

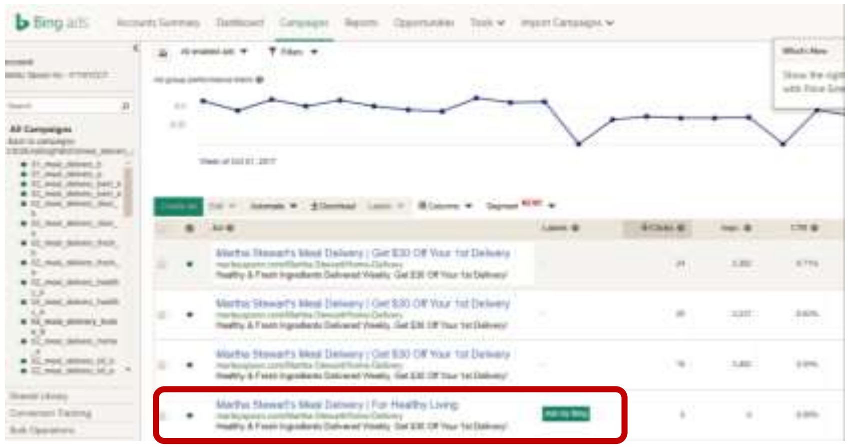


Application Scenarios In Ads

Search Ads

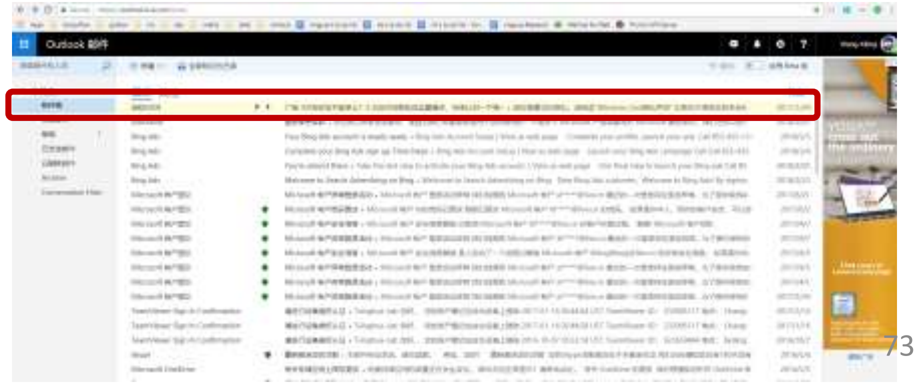


Bing Ads Platform



Native Ads on MSN

Native Ads on outlook.com



Outline

- Definition and goals
- Forms of explanations
- Explainable recommendation pipelines

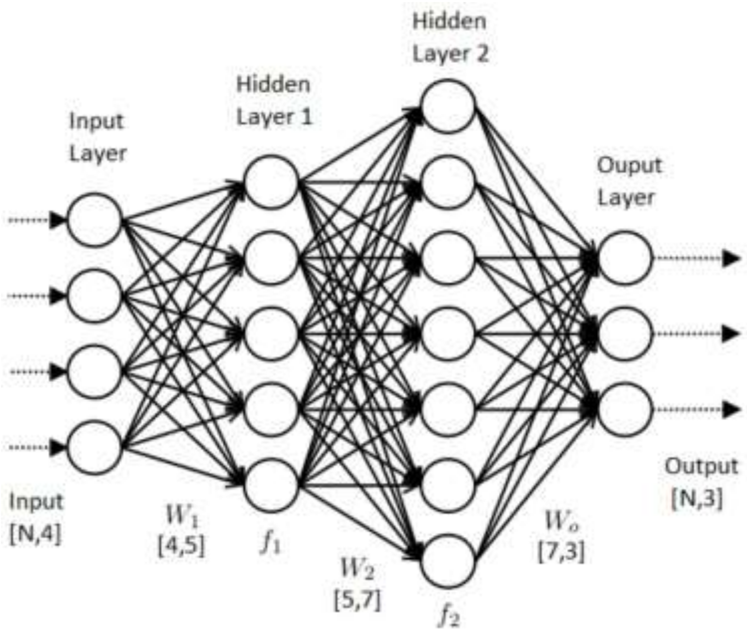
Outline

- **Definition and goals**
- Forms of explanations
- Explainable recommendation pipelines

Goals of Explainable AI

Model explainability

- Transparency — Open/gray black boxes
- Debugging — Debug failed models
- Trust — Understand why some models work; Important for high-stakes applications such as healthcare and finance



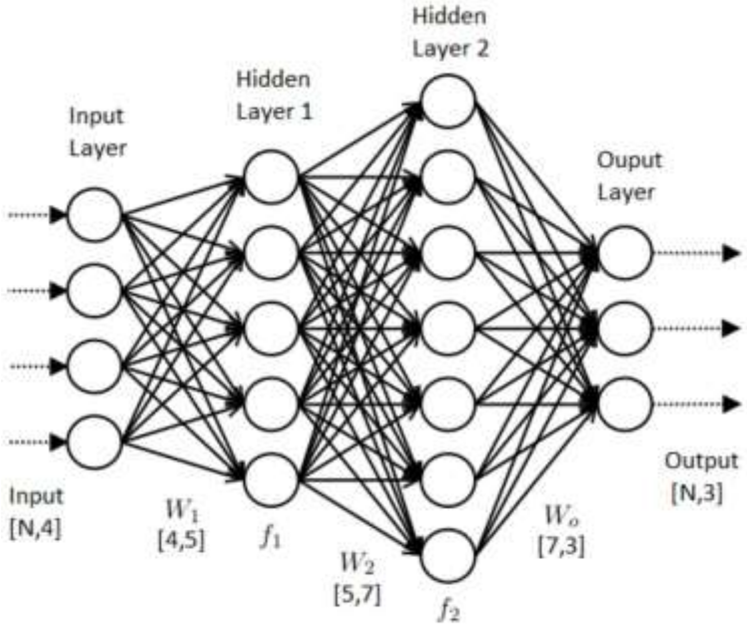
Goals of Explainable Recommendation

Model explainability
- Transparency
- Debugging
- Trust

Persuasiveness

Readability
Effectiveness
Satisfaction

Presentation quality



Advertisers

Researchers
Algorithm developers

Research beginners

Ordinary Users

Goals of Explainable Recommendation

Model explainability: Transparency, Debugging, Trust

Persuasiveness

Presentation quality: Readability, Effectiveness, Satisfaction

- Understanding their relationships
  - Correlated
  - Trade-off

Relationships between the Goals: Correlated

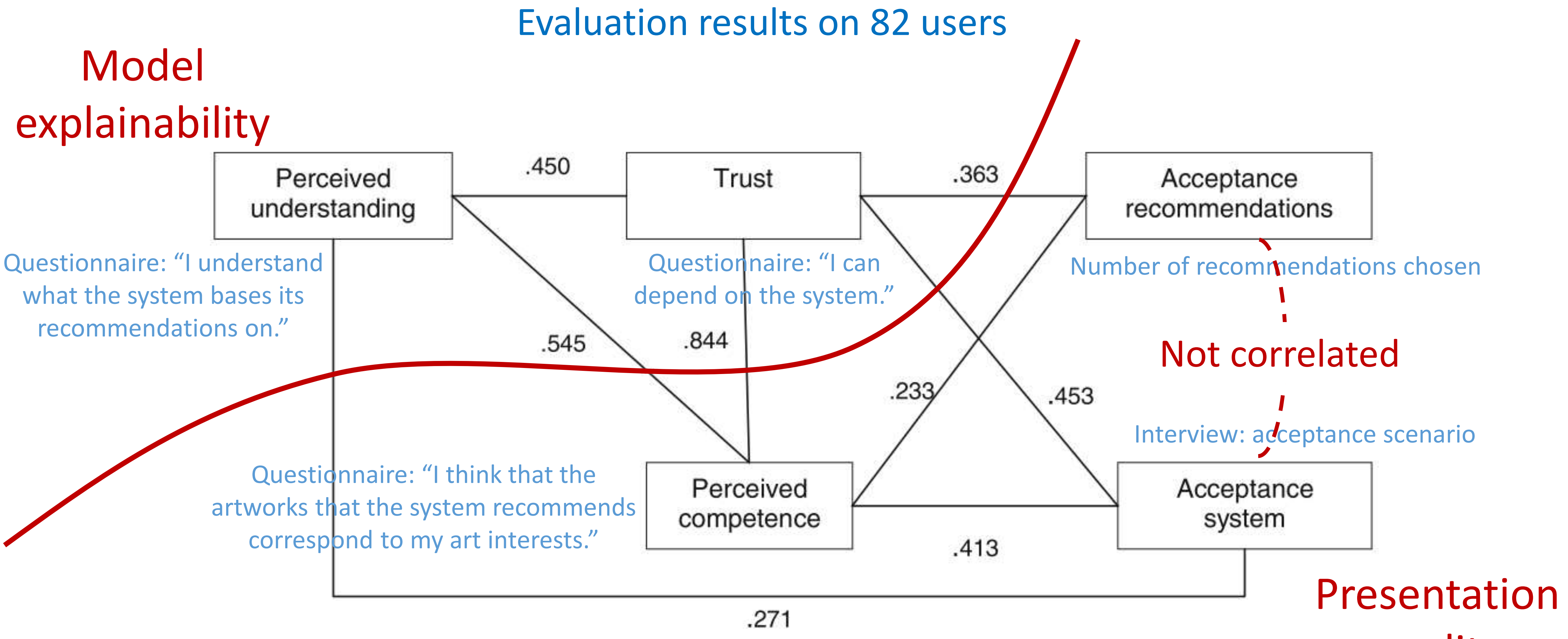


Fig. 5 Main significant correlations, Spearman's rho, $p(1 - \text{tailed}) < .05$

[UMUAI2008]

Relationships between the Goals: Trade-Off

Model explainability

Aim	Definition
Transparency (Tra.)	Explain how the system works
Scrutability (Scr.)	Allow users to tell the system it is wrong
Trust	Increase users' confidence in the system
Effectiveness (Efk.)	Help users make good decisions
Persuasiveness (Pers.)	Convince users to try or buy
Efficiency (Efc.)	Help users make decisions faster
Satisfaction (Sat.)	Increase the ease of usability or enjoyment

Trade-off

Trade-off

Trade-off

Presentation quality

[UMUAI2008]

Goals of Explainable Recommendation

Model explainability: Transparency, Debugging, Trust

Persuasiveness

Presentation quality: Readability, Effectiveness, Satisfaction

- Understanding their relationships
  - Correlated
  - Trade-off

Goals of Explainable Recommendation

Model explainability: Transparency, Debugging, Trust

Persuasiveness

Presentation quality: Readability, Effectiveness, Satisfaction

- Understanding their relationships
  - Correlated
  - Trade-off

- Most existing methods consider both criteria
  - Model explainability: 9 out of 10 papers
  - Presentation quality: all papers

Definition of Explainable Recommendation

- In general, give statements that support the recommendations [WWW2013]

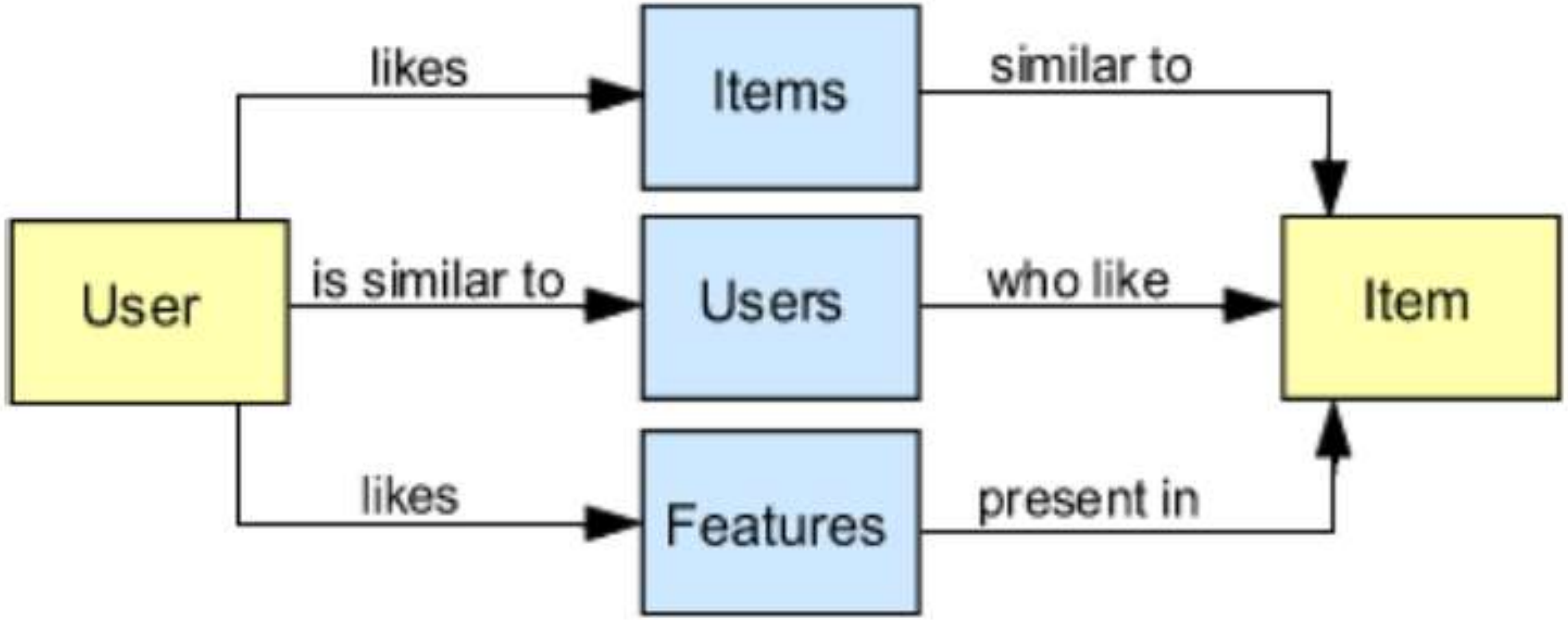
- Application specific
  - [Model-explainability] help users understand the system behavior [CHI2012]
  - [Presentation quality-Effectiveness] help users make more accurate decisions [IUI2015]
  - [Presentation quality-Persuasiveness] convincing users to adopt recommendations [IUI2015, IUI2009]

Outline

- Definition and goals
- **Forms of explanations**
- Explainable recommendation pipelines

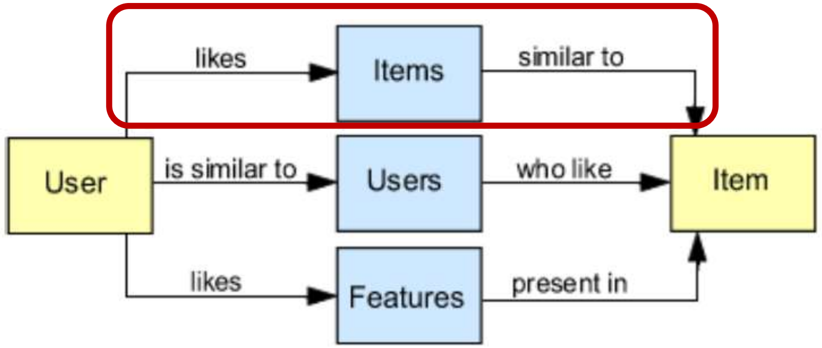
Forms of Explanations

- Three basic forms



[IUI2009]

Item-Based Explanations



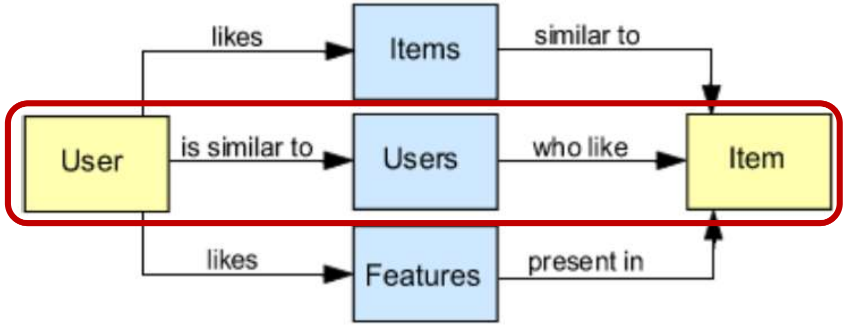
- “You may like the item because it is similar to items you previously like”

Amazon

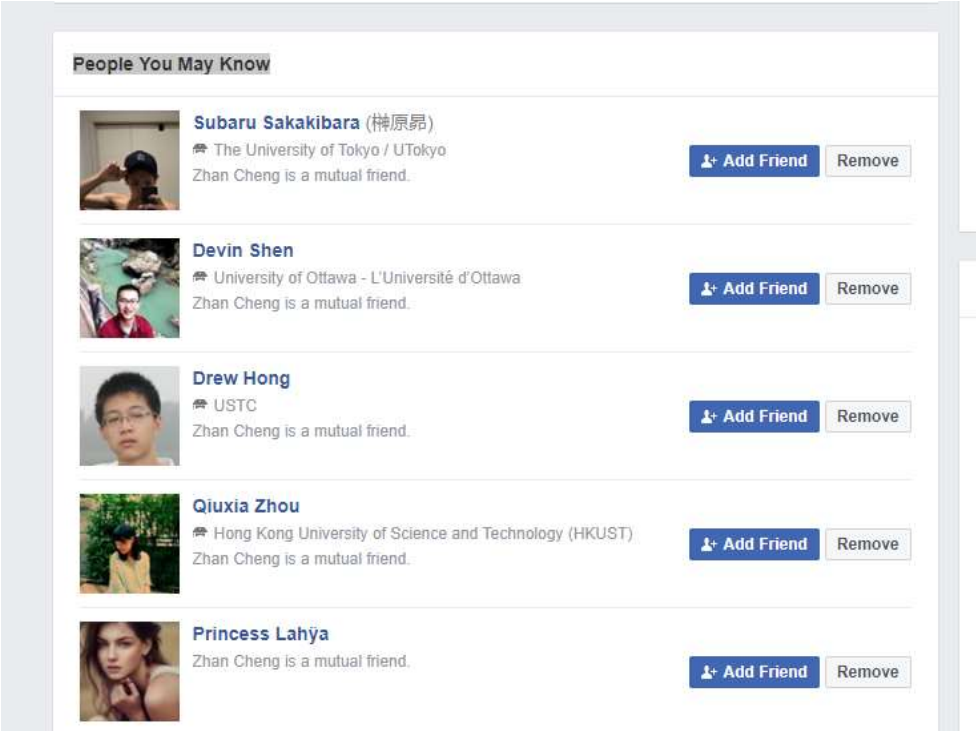
BOOK	YOUR RATING Out of 5	INFLUENCE Out of 100
Of Mice and Men	4	54
1984	4	50
Till We Have Faces : A Myth Retold	5	50
Crime and Punishment	4	46
The Gambler	5	11

[IUI2005]

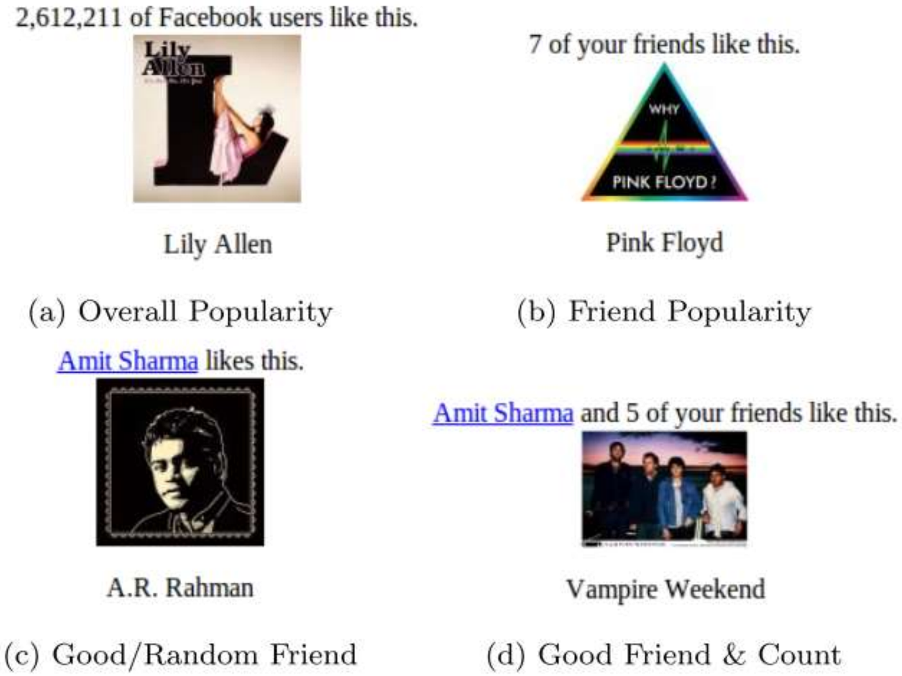
User-Based Explanations



- “You may like the item because a user similar to you like this item”

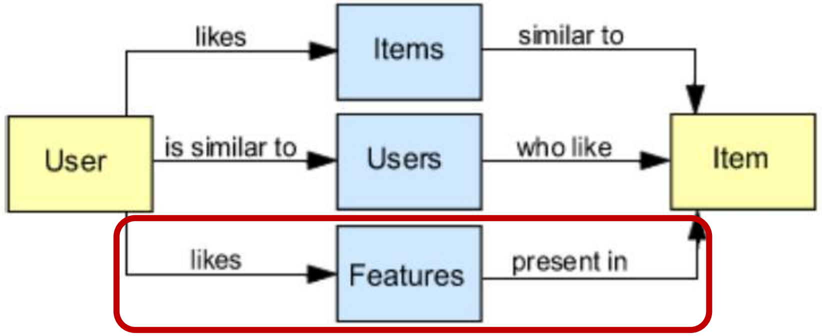


Facebook



[WWW2013]

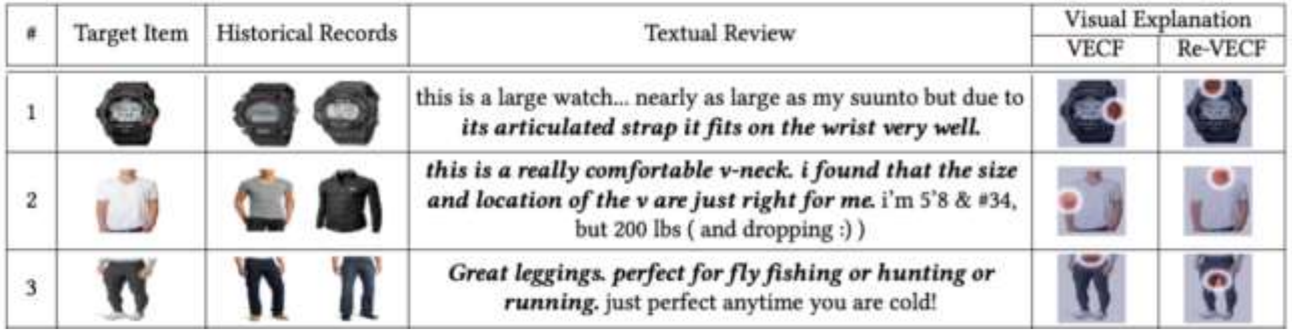
Feature-Based Explanations



- "You may like the item because this item contains features you like"

Slot	Word	Count	Strength	Explain
DESCRIPTION	HEART	2	94.14	Explain
DESCRIPTION	BEAUTIFUL	1	17.07	Explain
DESCRIPTION	MOTHER	3	11.55	Explain
DESCRIPTION	READ	14	10.63	Explain
DESCRIPTION	STORY	16	9.12	Explain

[IUI2005]



#	Target Item	Historical Records	Textual Review	Visual Explanation VECF	Re-VECF
1			this is a large watch... nearly as large as my suunto but due to ***its articulated strap it fits on the wrist very well.***		
2			***this is a really comfortable v-neck. i found that the size and location of the v are just right for me.*** i'm 5'8 & #34, but 200 lbs (and dropping :))		
3			***Great leggings. perfect for fly fishing or hunting or running.*** just perfect anytime you are cold!		

[Arxiv2018]

[SIGIR2014]

Dialog-Based Explanations

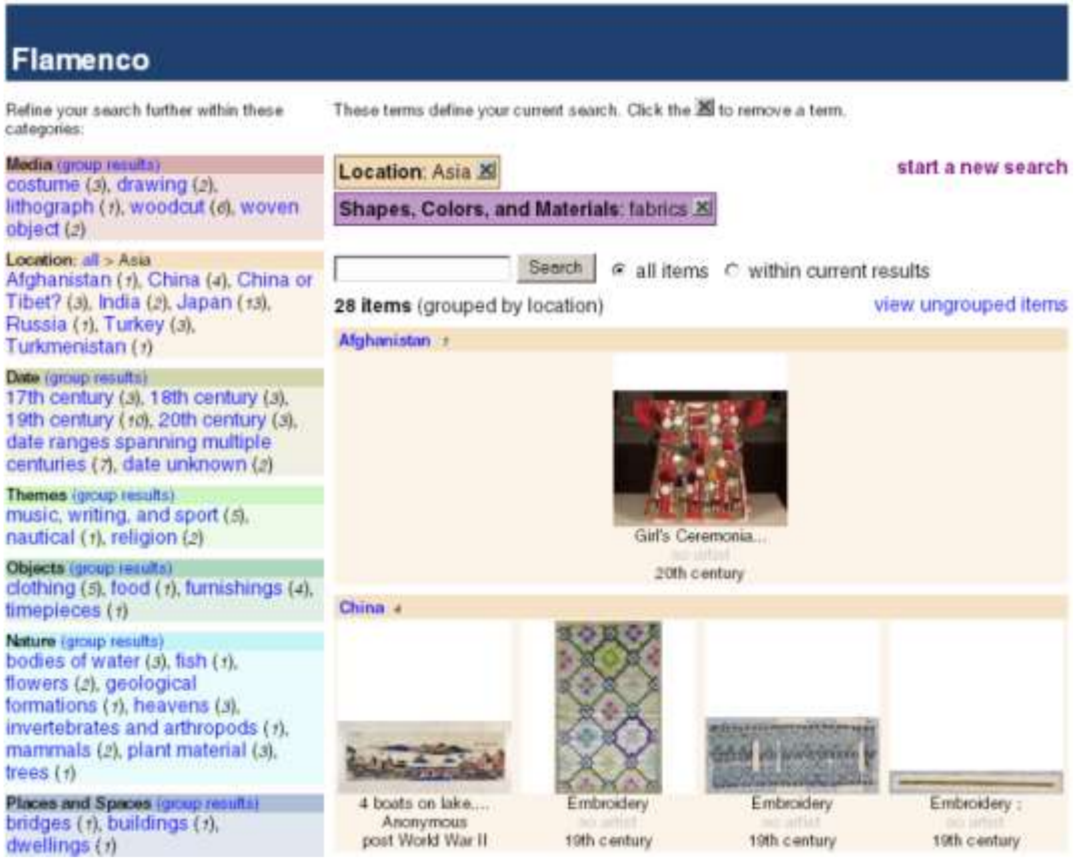
Microsoft XiaoIce (小冰)

1. Pick a model (Absolute/Pairwise) (Sec. 4.2) and preference elicitation mechanism: `Abs` (Sec. 4.3) / `Abs Pos` / `Abs Pos & Neg` / `Pairwise` (Sec. 4.4).
2. Initialize model parameters using offline data.
3. A new user arrives. Now iterate for a few questions[3]:
   (a) Mechanism selects a question to ask
   (b) User answers the question
   (c) All model parameters are updated
   (d) Remove the question from the allowed questions
4. System presents the final recommended list

[KDD2016]

Structured Overview Explanations

NewsMap

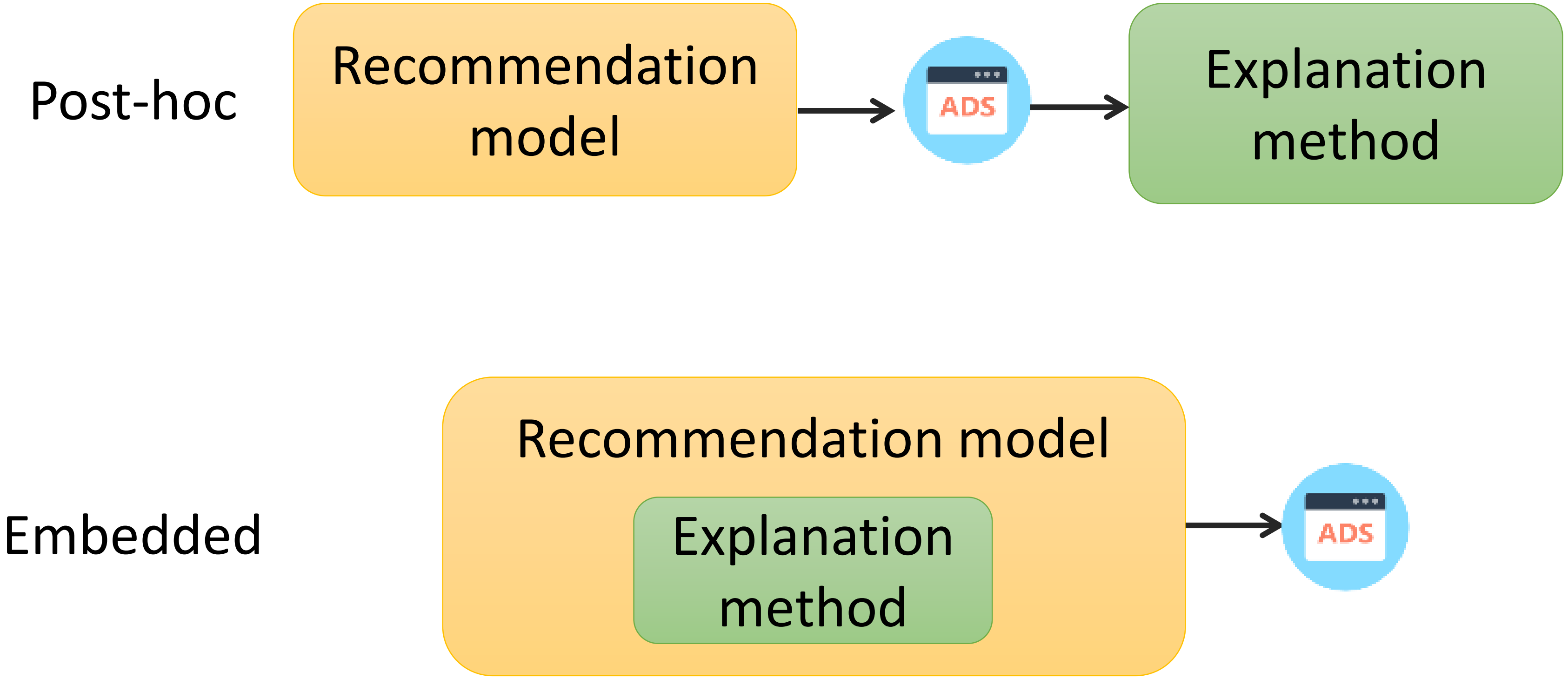


[CHI2003]

Outline

- Definition and goals
- Forms of explanations
- **Explainable recommendation pipelines**

Existing Pipelines



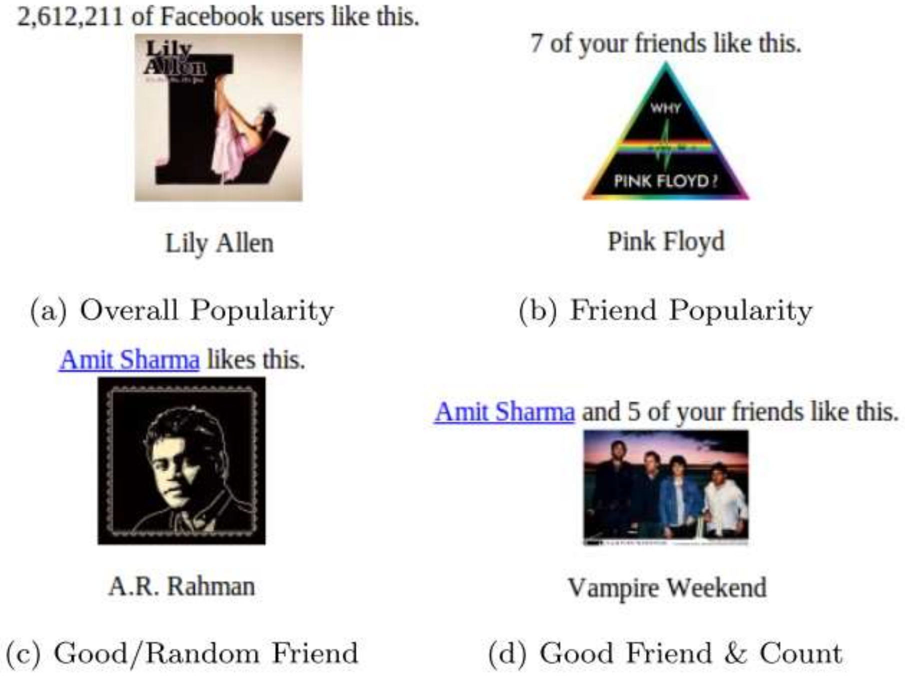
Model agnostic

Presentation quality

Model explainability

Post-Hoc Methods

- Rule-based



Fraction of likelihood ratings above 5 (neutral rating) for each explanation strategy.

Explanation	**Fraction > 5**
FriendPop	0.137
RandFriend	0.141
OverallPop	0.175
GoodFriend	0.200
GoodFrCount	0.239

Friend with maximum tie strength: maximum number of interactions (likes, comments, wall posts)

[WWW2013]

Post-Hoc Methods

- Rule-based
- Retrieval-based

Scenario: book recommendation

Feature-based recommendation

Slot	Word	Count	Strength	Explain
DESCRIPTION	HEART	2	94.14	Explain
DESCRIPTION	BEAUTIFUL	1	17.07	Explain
DESCRIPTION	MOTHER	3	11.55	Explain
DESCRIPTION	READ	14	10.63	Explain
DESCRIPTION	STORY	16	9.12	Explain

Ranking score: $c * strength(t)$

$$strength(t) = \frac{P(t|c_l, s)}{P(t|c_d, s)}$$

c : number of times t appears

c_l : the category of likes

c_d : the category of dislikes

[IUI2005]

Post-Hoc Methods

- Rule-based
- Retrieval-based

Scenario: book recommendation

Item-based recommendation

BOOK	YOUR RATING Out of 5	INFLUENCE Out of 100
Of Mice and Men	4	54
1984	4	50
Till We Have Faces : A Myth Retold	5	50
Crime and Punishment	4	46
The Gambler	5	11

Influence score of item j on i:

$$p(i|u,S^+,S^-) - p(i|u,S^+\backslash j,S^-)$$

Similarity score between item j and i:

$$Pearson(i,j)$$

[IUI2005]

Post-Hoc Methods

- Rule-based
- Retrieval-based
- Generative

Scenario: explanation generation for music recommendation

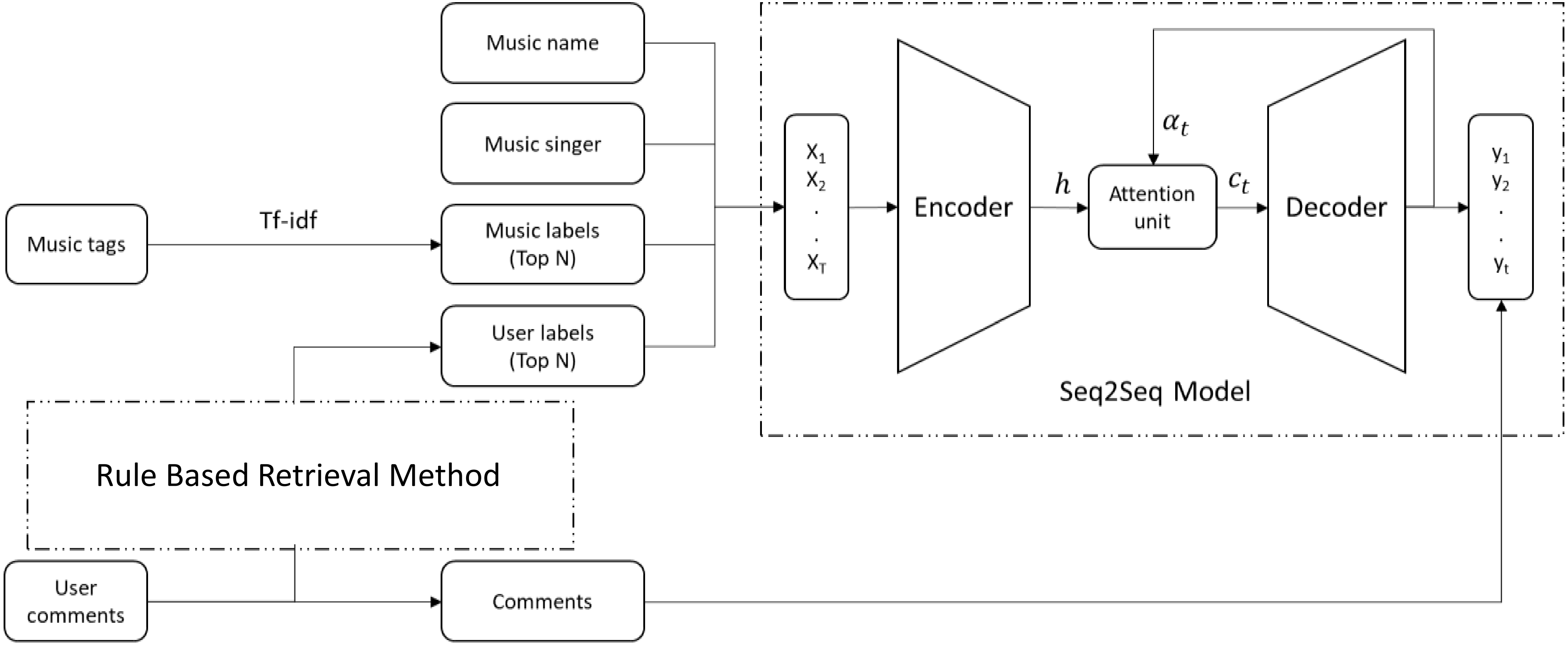
Data Preparation

- 163 music data
  - Song + Singer + Album + Lyric + Music tags + Comments
- User tags
  - XiaoIce tag
  - Weibo tag

Requirements

- User Profile Related Reasons 
  - Age and Gender
  - User tags
- Song Related Reasons 
  - Lyric

Our Framework



Examples-User Profiles Related

- 民谣大学时光校园时光光阴摇滚朴树我爱你再见
  - 学生
    - 以前学校每天中午都会放这首歌
    - 校园十佳歌手，我就听了这首歌。[可爱]
  - 电音
    - 这首歌真是越听越带感
    - 每次听到这首歌都会热血沸腾
  - 民谣
    - 每次听这首歌都会有一种很安静的感觉
    - 很喜欢这首歌，很喜欢民谣
  - 失恋
    - 今天分手了，听到这首歌，心都碎了下来
    - 我失恋了，听着这首歌，感觉自己也是醉了
  - 晚睡
    - 每天晚上睡觉前听这首歌，越听越有感觉，越听越有感觉，
    - 这首歌是我最喜欢的一首歌，晚安

- Music Tags
- Singer
- Song Name
- User Profiles

Examples-Song Related

- 正面指南无非自拍拍拍有趣欢笑悲哀姿态女孩李荣浩自拍
  - 李荣浩八个专辑里的歌最喜欢的就是这首了
  - 昨天今天真的好喜欢这首歌好想听现场版啊[亲亲][亲亲][亲亲]
  - 听李荣浩的歌不会分享给朋友们[可爱][可爱]
  - 看到李荣浩的歌啊[可爱][可爱][可爱][可爱][可爱]人我爱你一辈子~
  - 李荣浩还行！这首歌真的有味道？？？？？？
- 迷迭香甜味喜好发酵味道性感无可救药讯号飘扬优雅周杰伦迷迭香
  - 这首歌千万千万别火呀[发怒][发怒][发怒][发怒][发怒]
  - 我的阿珍什么时候来听这首歌
  - 这首歌好骚，喜欢[色]
  - 所以这首歌是我伦唱的最正常的[大哭]
  - 周杰伦慵懒的嗓音能驾驭这首歌
- 火星人地球猿人起火难过心脏小孩空袭话音舍利薛之谦火星人来过
  - 薛之谦唱的这首歌啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊啊
  - 这首歌是越听越好听
  - 因为薛之谦，我喜欢上了薛之谦的歌
  - 薛，我喜欢你，希望你的歌能给我带来好运[爱心]
  - 这首歌简直不要太酷，太喜欢这首了[色]

- Music Lyrics
- Singer
- Song Name

Evaluation on User Profiles

Music:　朴树 我爱你再见
User profile:　失恋
Music tags:　民谣大学时光校园时光光阴摇滚

Generated Reasons	Fluency	Personalization	Relevance	Overall
每次听这首歌都会想到初恋	3	2	2	2
这首歌是我初恋最喜欢的歌	3	1	2	2
今天分手了，听到这首歌，心都碎了下来	3	3	3	3
初恋女友最喜欢的歌	3	1	1	1
失恋了，听着这首歌，感觉自己也是醉了	3	3	3	2
单身狗听这首歌真的是真的好吗	3	3	2	2
分手后，听着这首歌，感觉自己也是醉了	3	2	2	2
我失恋了，听着这首歌，感觉自己也是醉了	3	2	2	2
单身狗听这首歌真的是真的好吗	3	3	2	2
听了这首歌，我就知道我失恋了	3	3	3	3

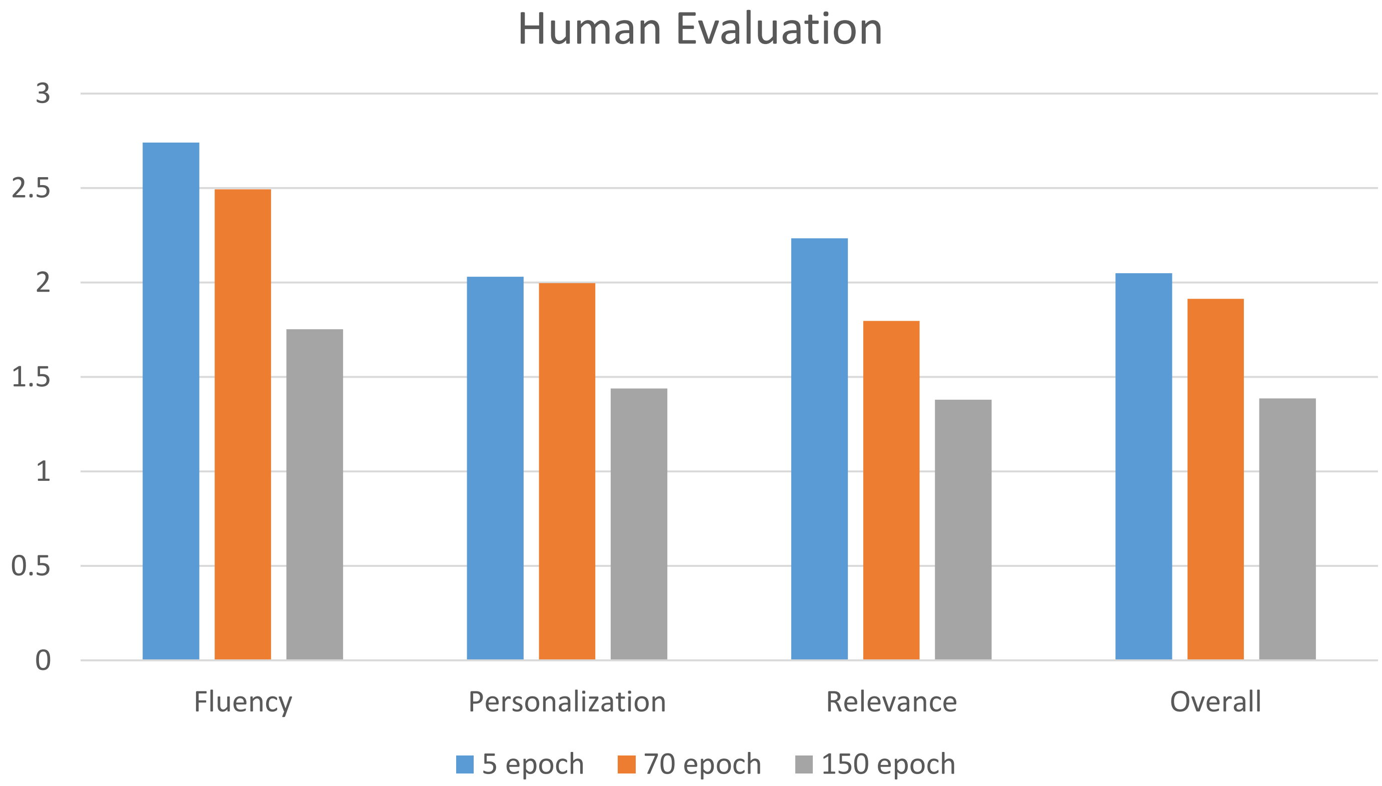
1: Bad

2: Medium

3: Good

Evaluation on User Profiles

- 5 songs
- 20 user profiles
- 300 reasons



Impact of Epoch

Music: 李宇春 下个路口见

User profile: 学生

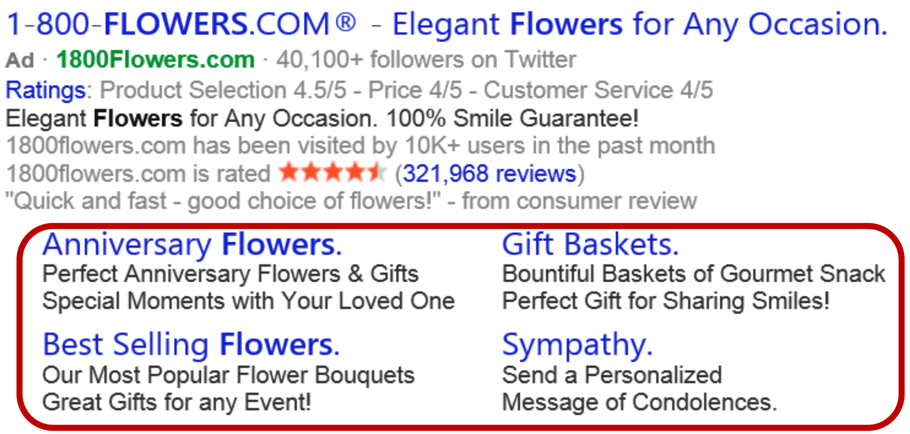
Music tags: 国语回忆校园时光记忆怀念

5 epochs (*General*)	**70 epochs (*Border*)**	**150 epochs (*Specific*)**
我们学校每天中午都放这首歌	学校每天都放这首歌	每次听这首歌，都想哭。那年的同学，你的未来，不知我
学校每天中午都放这首歌	刚刚学校广播站放的就是这首歌	刚刚听这首歌，在学校广播听到，爱上了这首歌
以前学校每天中午都会放这首歌	唉，学校每天都放这首歌，每次听到都好想哭，说好的	学校每天都是这首歌
记得以前学校每天中午都会放这首歌	今天学校广播站放了这首歌，好想回家路上	中学时代喜欢陈奕迅的歌，我知道你明天会来，这也是音乐老师的
今天学校放了这首歌，我就知道这首歌了	听着这首歌，想着你，想着你，想着你，想着你	坐在公交上，学校放的这首歌，当时没感觉，眼神中看着我
这首歌是我们学校每天中午放学的铃声	在学校的广播听到这首歌，感觉自己像赵小雷[吐舌][吐舌]	很多年前在学校听过的男生唱的最好的歌
当时学校广播放了这首歌，当时就觉得好听	每次听到这首歌都会想到以前校园生活的味道～	听着这首歌，想回到校园时代
我同学说这首歌是我最喜欢的一首歌，每次听都会觉得很	这首歌是我的上课铃声[奸笑]	在听这首歌是因为是因为学校生活开始认识你，记得还记得那一天
曾经在学校的广播里听到这首歌，当时觉得好幸福	校园广播听到了这首歌，很好听	刚听到这首歌，在学校广播里听到，才反应过来是这首歌…忘
当年学校放这首歌，当时觉得好幸福	同桌的你，听这首歌让我想哭	唉，听着这首歌写着曾经的我们曾经学校的周一，现在还有祖国送给

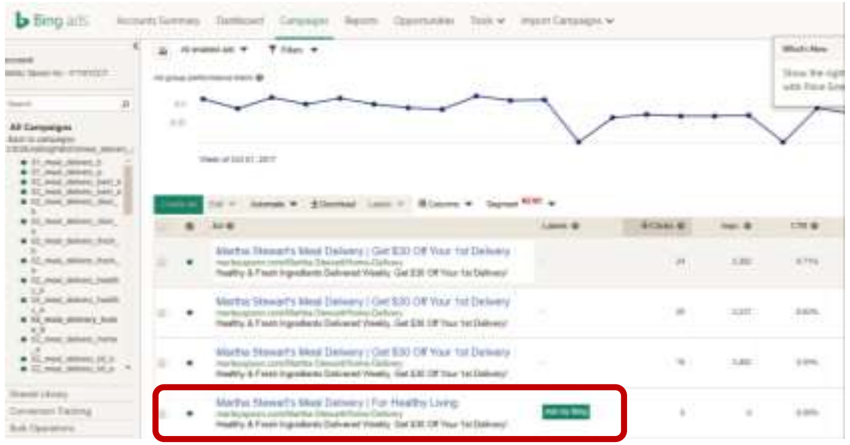
Post-Hoc Methods

- Rule-based
- Retrieval-based
- Generative

Search Ads



Bing Ads Platform

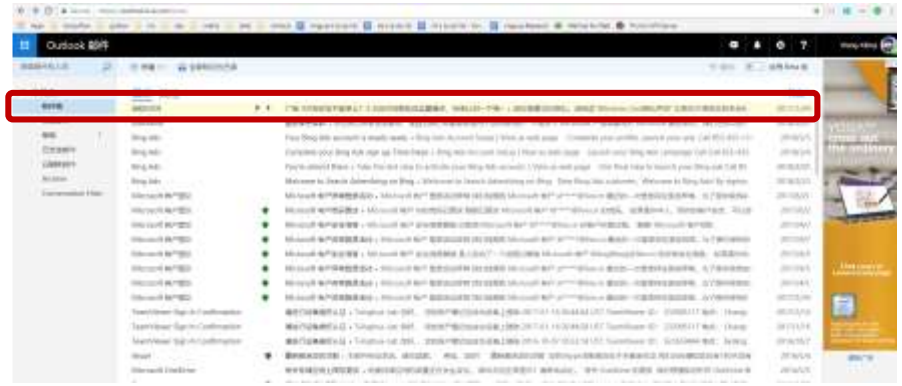


Post-Hoc Methods

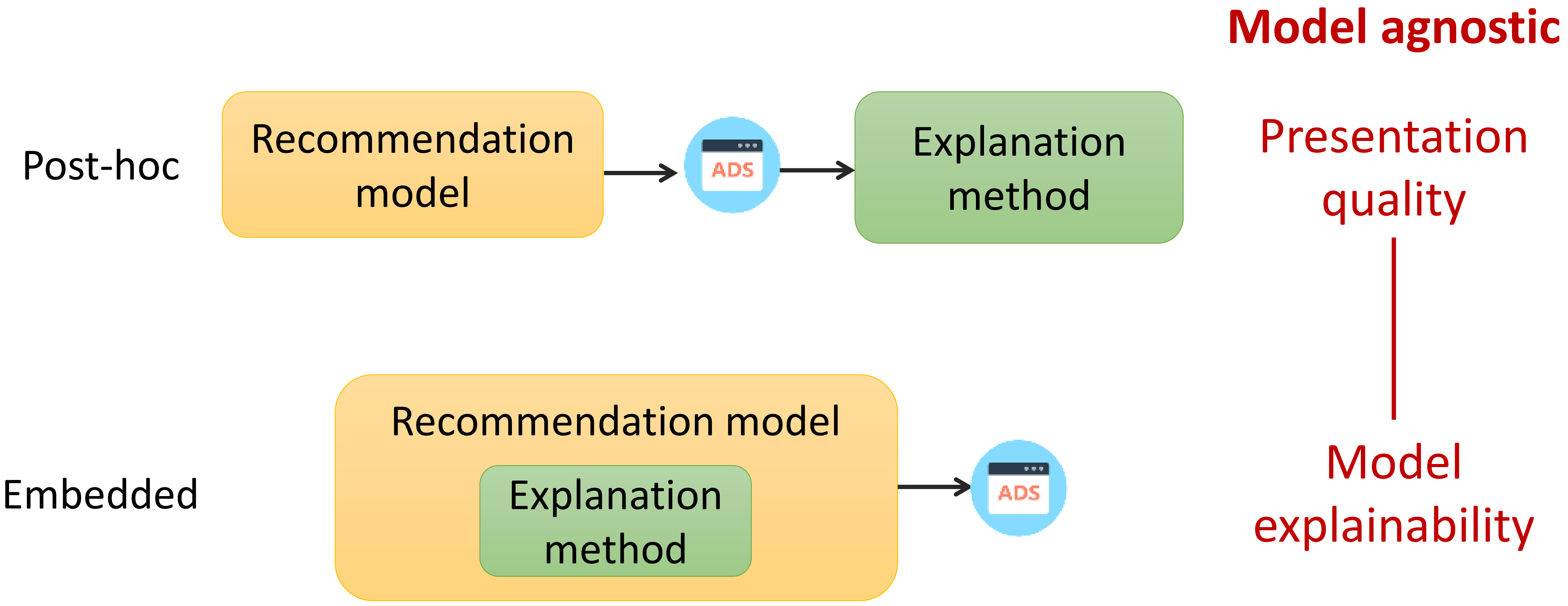
- Rule-based
- Retrieval-based
- Generative

Native Ads on MSN

Native Ads on outlook.com



Existing Pipelines



Recommendation model | Explanation method

Embedded Methods

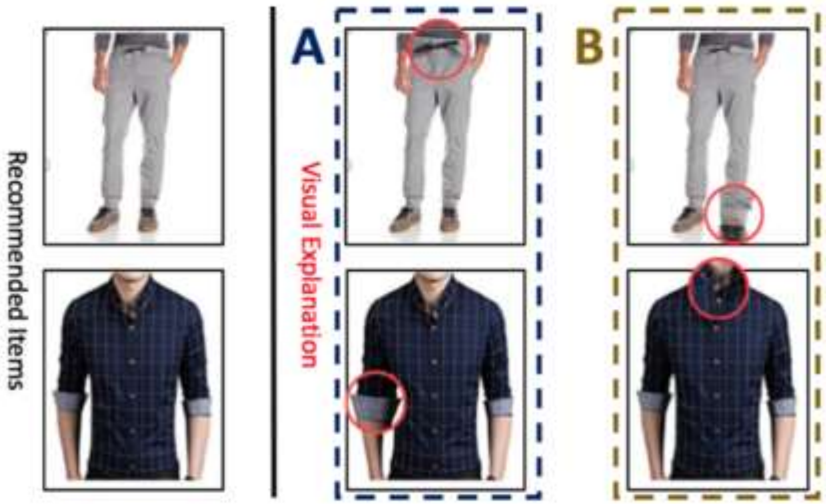
- Most embedded methods are feature-based
  - Features are usually parts from the auxiliary information (review, images)
  - It fits well with existing recommendation models (can even improve accuracy)
- Types of features
  - Phrases
  - Sentences
  - Images

EFM: Phrase-level explanation

Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis

Yongfeng Zhang[†], Guokun Lai[†], Min Zhang[†], Yi Zhang[‡], Yiqun Liu[†], Shaoping Ma[†]
[†]State Key Laboratory of Intelligent Technology and Systems
[†]Department of Computer Science & Technology, Tsinghua University, Beijing, 100084, China
[‡]School of Engineering, University of California, Santa Cruz, CA 95060, USA
{zhangyf07,laiguokun}@gmail.com,{z-m,yiqunliu,msp}@tsinghua.edu.cn,yiz@soe.ucsc.edu

[SIGIR2014]

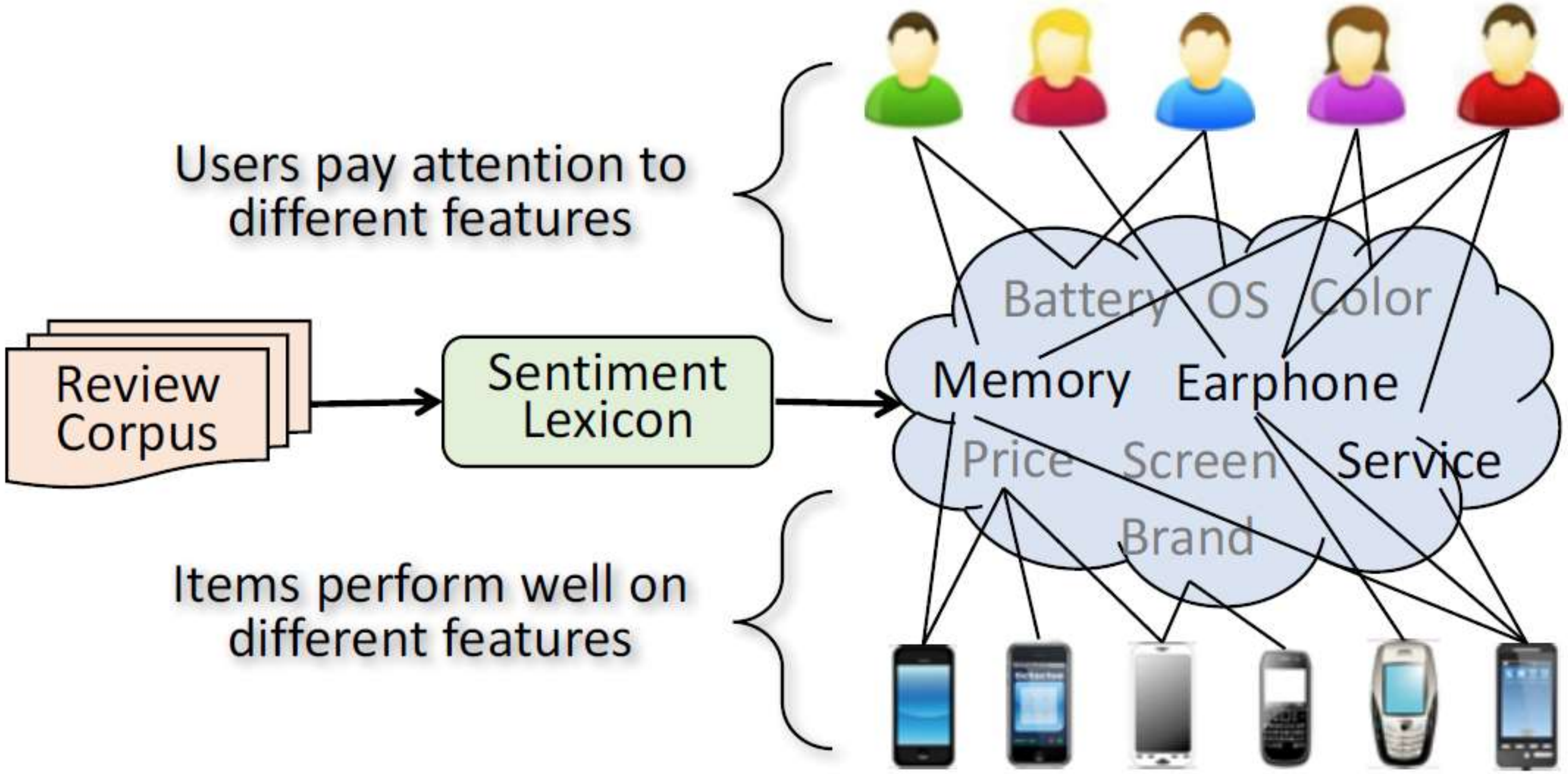


Recommendation model

Explanation method

Intuition

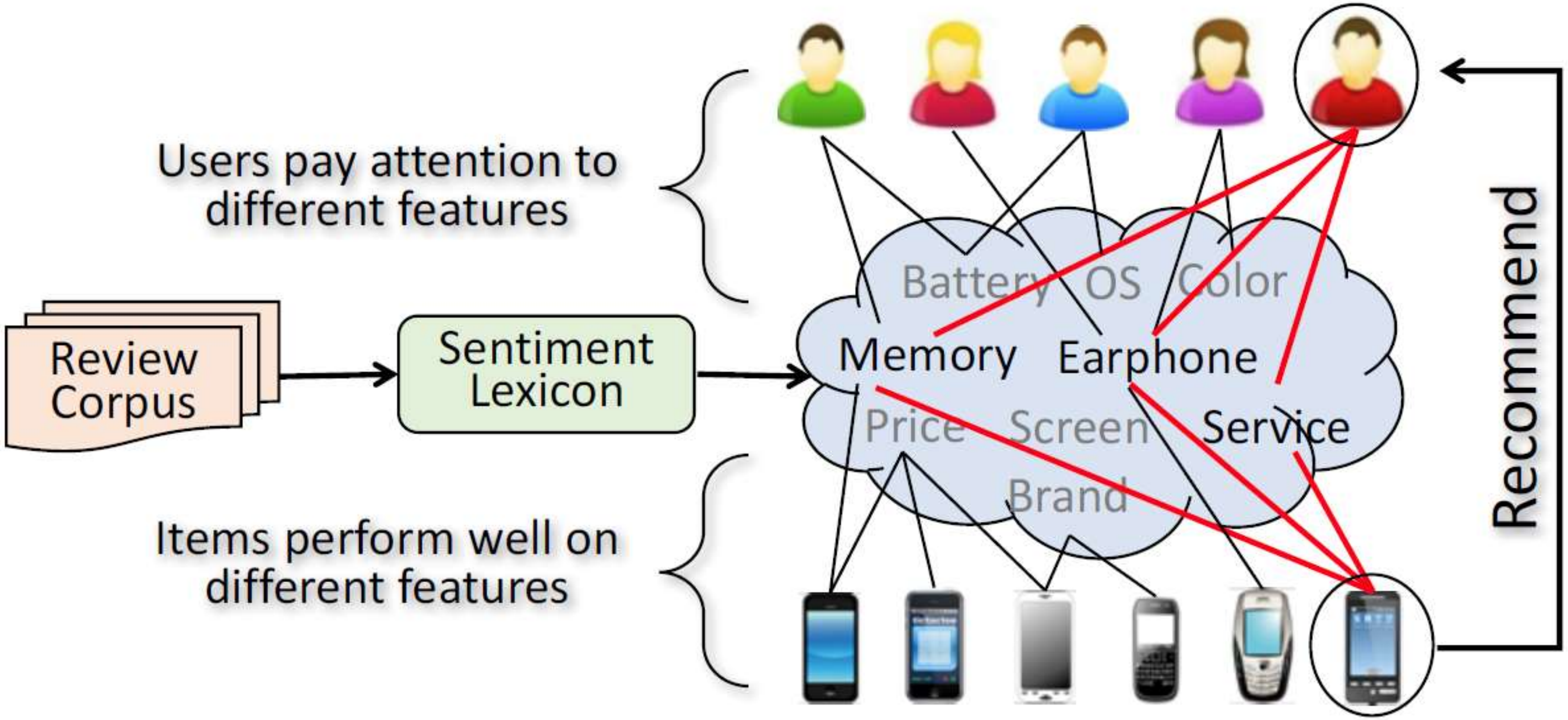
- To recommend a product that performs well on the features that a user concerns



Recommendation model | Explanation method

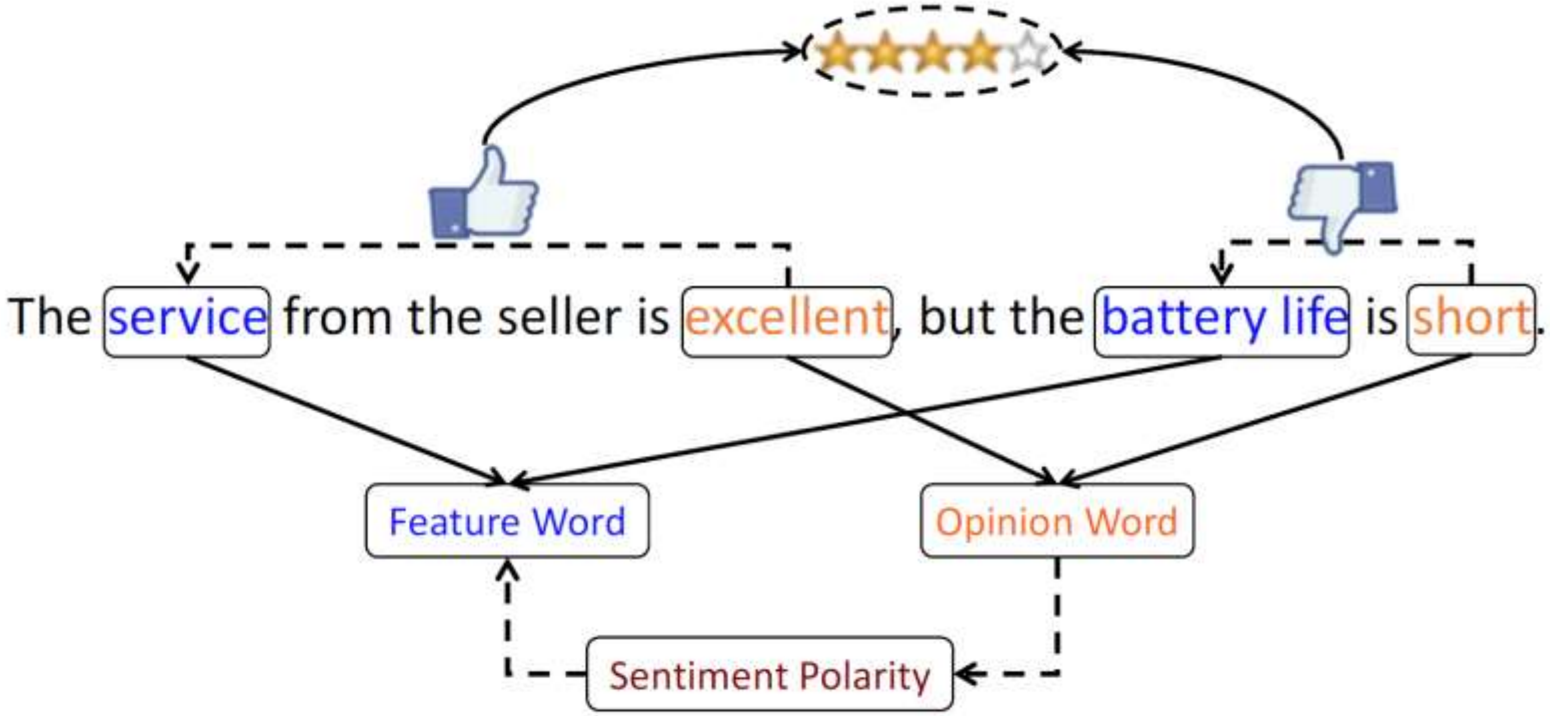
Intuition

- To recommend a product that performs well on the features that a user concerns



Aspect-level Sentiment Analysis

- To extract and organize features and opinions in unstructured reviews

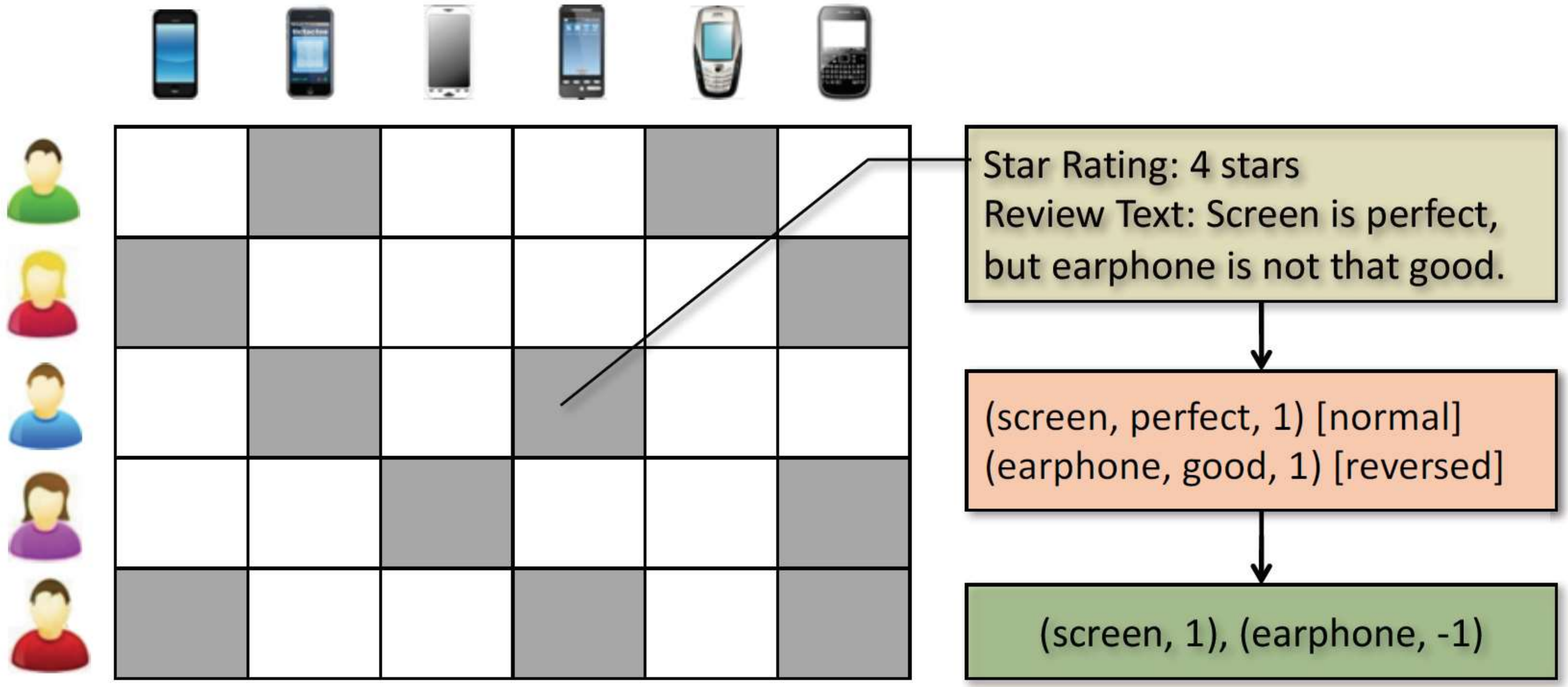


Hu and Liu, Mining and Summarizing Customer Reviews, KDD 2004
Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers.

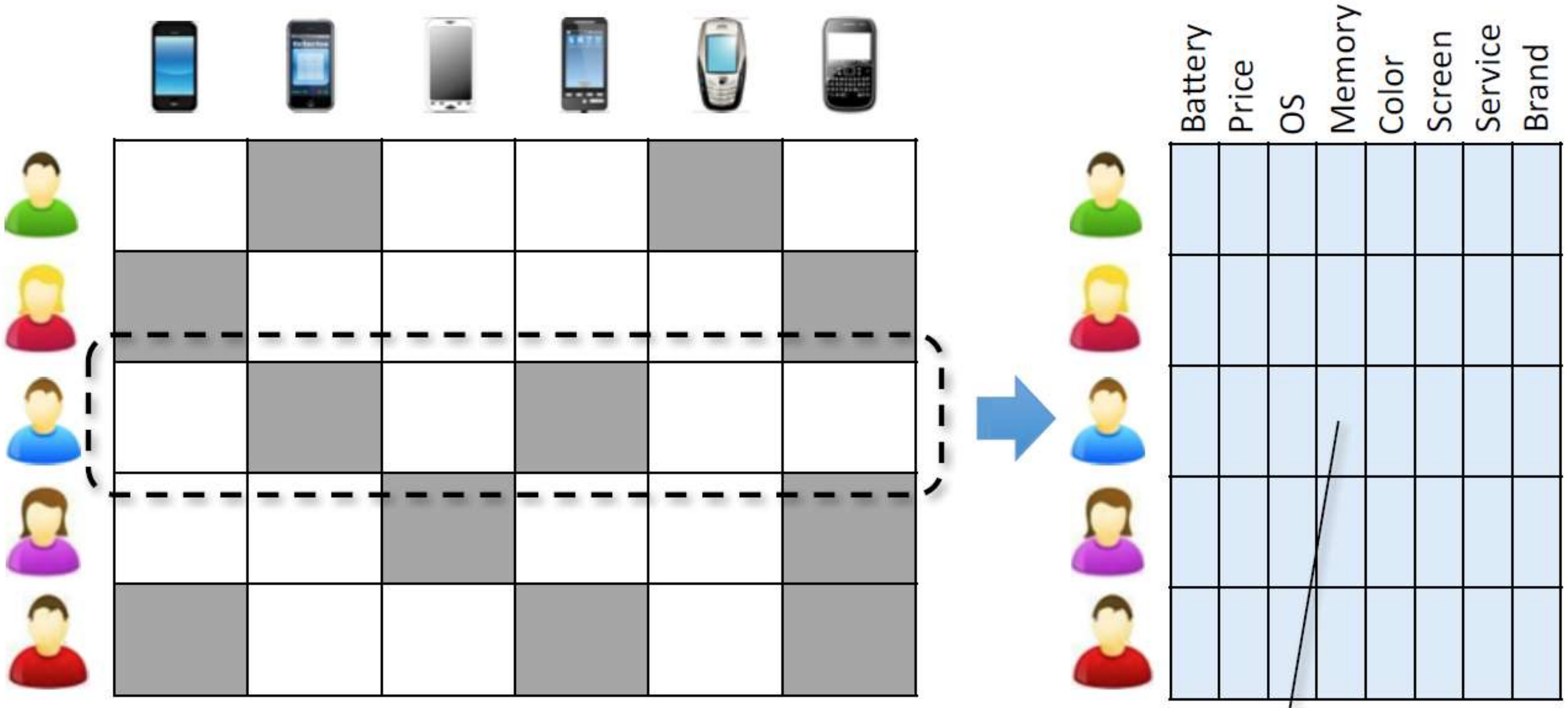
Recommendation model

Explanation method

Structure the Textual Reviews



User-Feature Attention Matrix



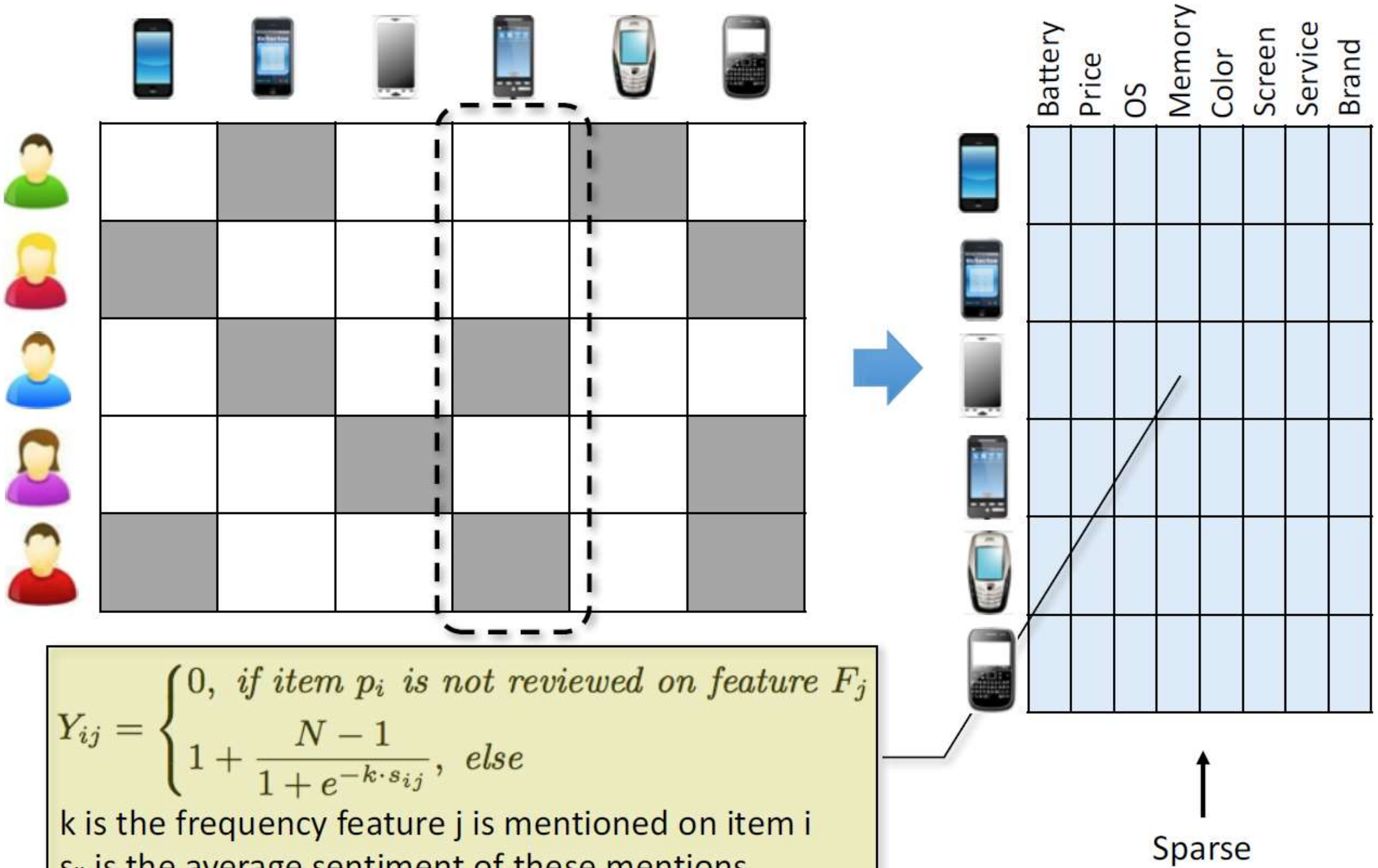
Sparse

$$X_{ij} = \begin{cases} 0, & \textit{if user } u_i \textit{ did not mention feature } F_j \\ 1 + (N-1)\left(\frac{2}{1+e^{-t_{ij}}} - 1\right), & \textit{else} \end{cases}$$

t_{ij} is the frequency that user i mentions feature j

Recommendation model | Explanation method

Item-Feature Attention Matrix



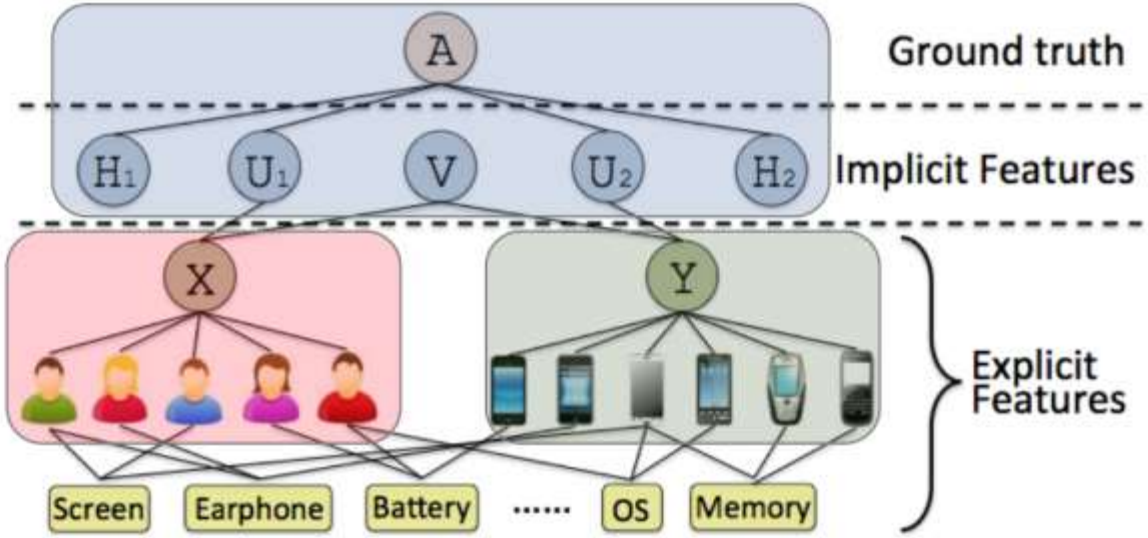
$$Y_{ij} = \begin{cases} 0, & \textit{if item } p_i \textit{ is not reviewed on feature } F_j \\ 1 + \dfrac{N-1}{1+e^{-k \cdot s_{ij}}}, & \textit{else} \end{cases}$$

k is the frequency feature j is mentioned on item i

s_{ij} is the average sentiment of these mentions

Multi-Matrix Factorization

- Integrating the explicit and implicit features



$$\underset{U_1,U_2,V,H_1,H_2}{\text{minimize}} \left\{ \|PQ^T - A\|_F^2 + \lambda_x \|U_1 V^T - X\|_F^2 + \lambda_y \|U_2 V^T - Y\|_F^2 + \lambda_u (\|U_1\|_F^2 + \|U_2\|_F^2) + \lambda_h (\|H_1\|_F^2 + \|H_2\|_F^2) + \lambda_v \|V\|_F^2 \right\}$$

$$P = [U_1 \ H_1], \ Q = [U_2 \ H_2]$$

Explicit Factors

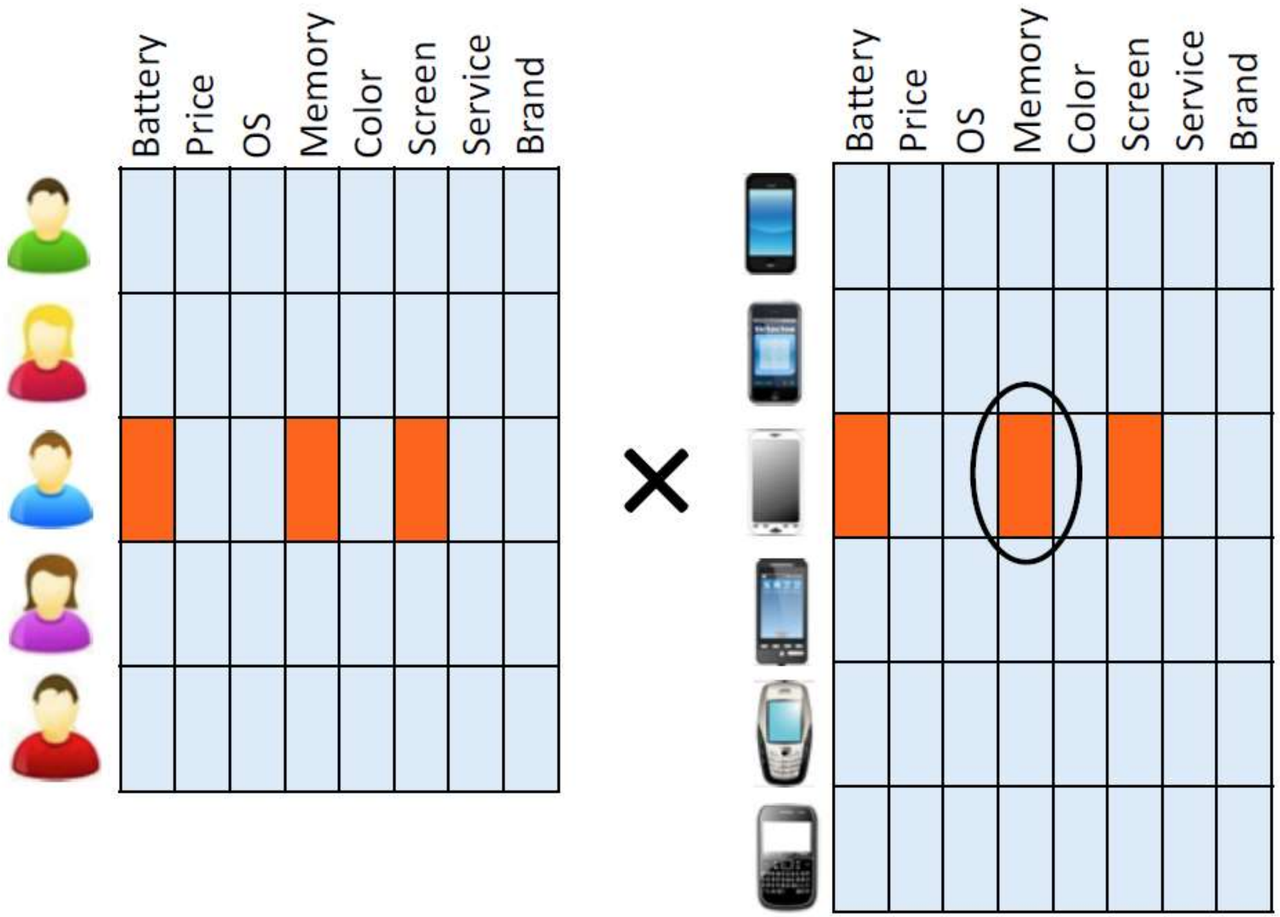
Hidden Factors

Recommendation model | Explanation method

Personalized Explanations

Feature-level explanation for a recommended item

```
You might be interested in [feature],
on which this product performs well.
```

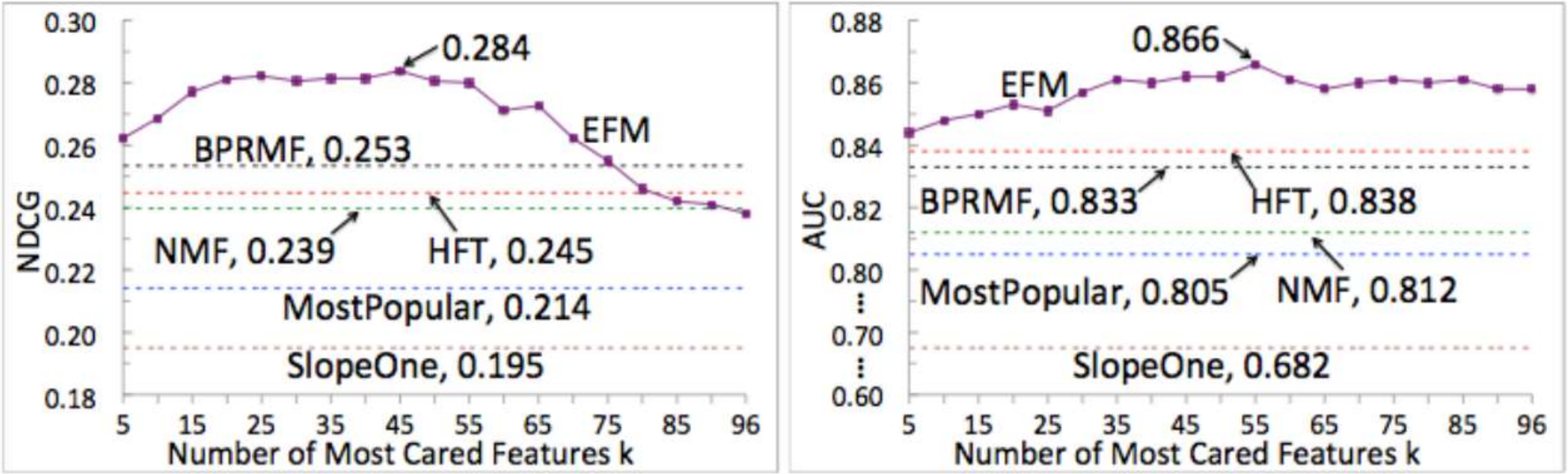


Recommendation model | Explanation method

Offline Experiment

- Top-N recommendation is improved
- k: number of most cared features

$$R_{ij} = \alpha \cdot \frac{\sum_{c \in \mathcal{C}_i} \tilde{X}_{ic} \tilde{Y}_{jc}}{kN} + (1 - \alpha) \tilde{A}_{ij}$$

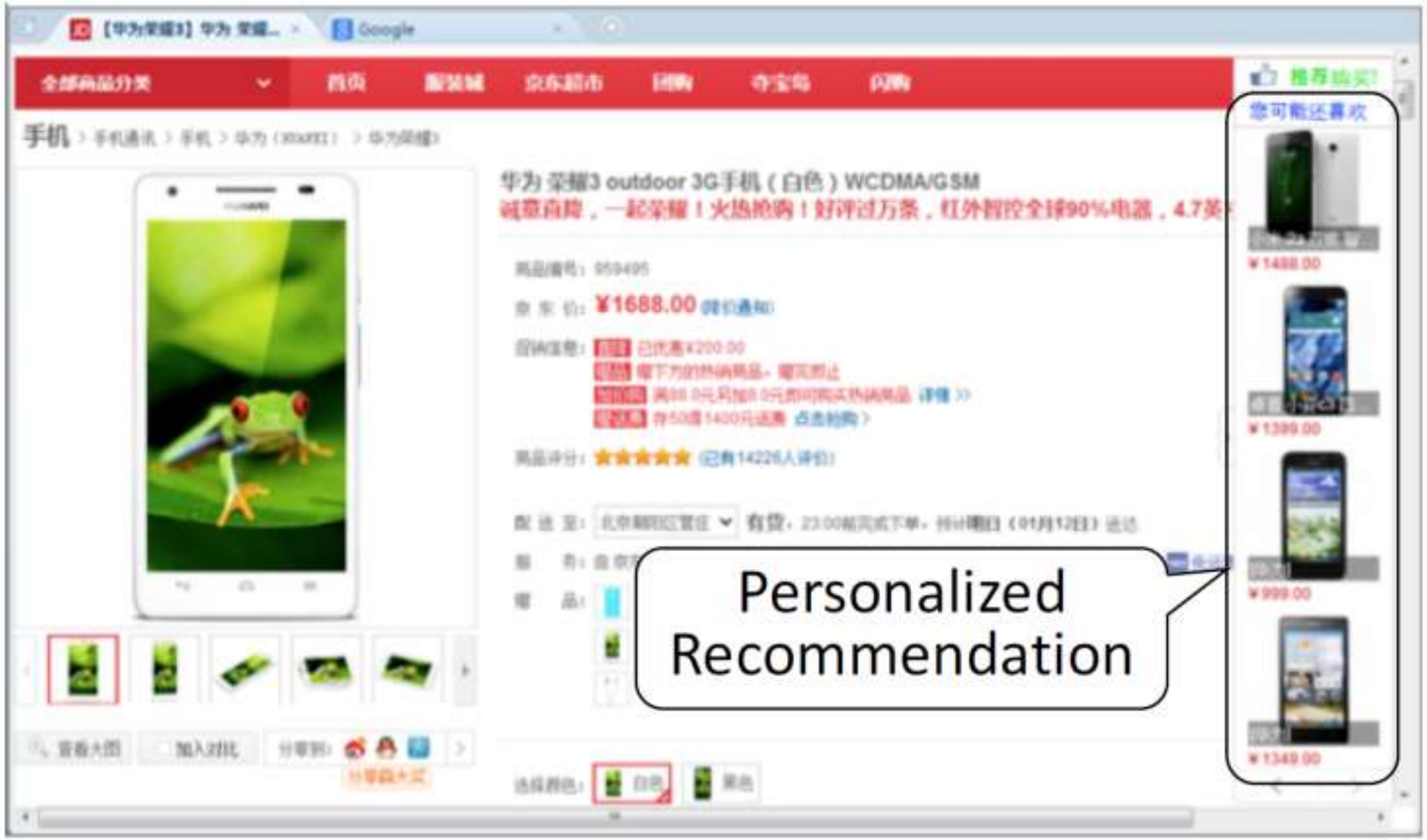


(a) NDCG vs k (b) AUC vs k

Recommendation model | Explanation method

Applied in Commercial Systems

- Provide personalized recommendations by a popular commercial web browser in an e-commerce website



Recommendation model | Explanation method

Click Through Rate Improvement

3 user groups

- A (experimental group): Receive our personalized explanations
- B (comparison group): Receive the 'people also viewed' explanation
- C (control group): Receive no explanation

User Set	A		B		C	
Records	#Record	#Click	#Record	#Click	#Record	#Click
	15,933	691	11,483	370	17,265	552
CTR	4.34%		3.22%		3.20%	

Recommendation model | Explanation method

Embedded Methods

- Most embedded methods are feature-based
  - Features are usually parts from the auxiliary information (review, images)
  - It fits well with existing recommendation models (can even improve accuracy)
- Types of features
  - Phrases
  - Sentences
  - Images

EFM: Phrase-level explanation

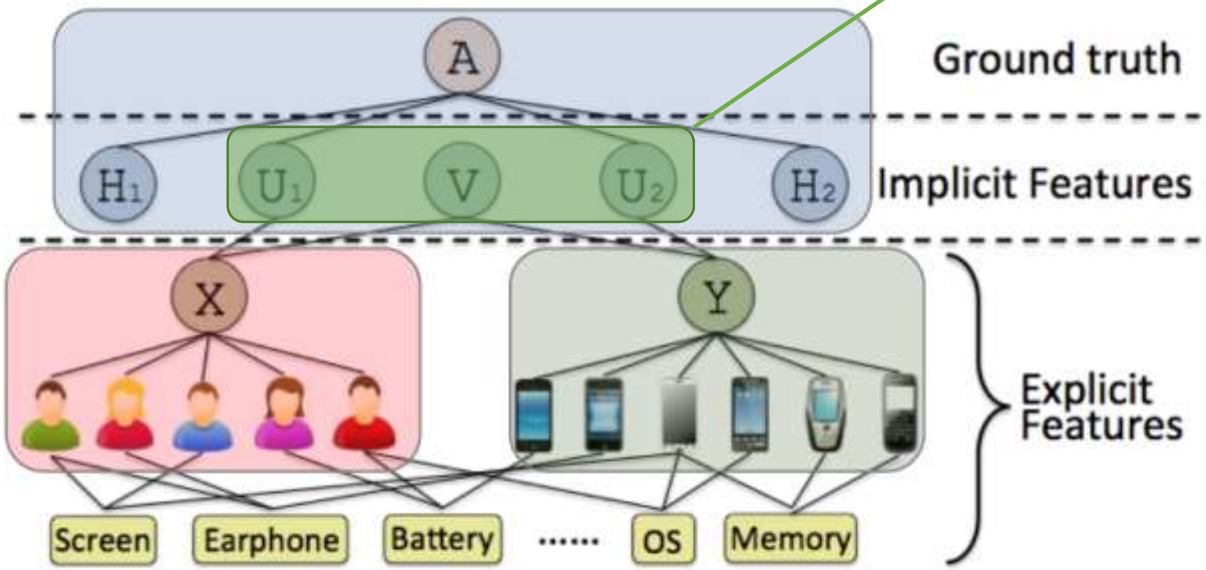
Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis

Yongfeng Zhang[†], Guokun Lai[†], Min Zhang[†], Yi Zhang[‡], Yiqun Liu[†], Shaoping Ma[†]
[†]State Key Laboratory of Intelligent Technology and Systems
[†]Department of Computer Science & Technology, Tsinghua University, Beijing, 100084, China
[‡]School of Engineering, University of California, Santa Cruz, CA 95060, USA
{zhangyf07,laiguokun}@gmail.com,{z-m,yiqunliu,msp}@tsinghua.edu.cn,yiz@soe.ucsc.edu

[SIGIR2014]

Recommendation Model

Explanation



Recommendation model | Explanation method

Embedded Methods

- Most embedded methods are feature-based
  - Features are usually parts from the auxiliary information (review, images)
  - It fits well with existing recommendation models (can even improve accuracy)
- Types of features
  - Phrases
  - Sentences
  - Images

NARRE: Review-level explanation

Neural Attentional Rating Regression with Review-level Explanations

Chong Chen
Yiqun Liu
Min Zhang*
Shaoping Ma

[WWW2018]

Recommendation model

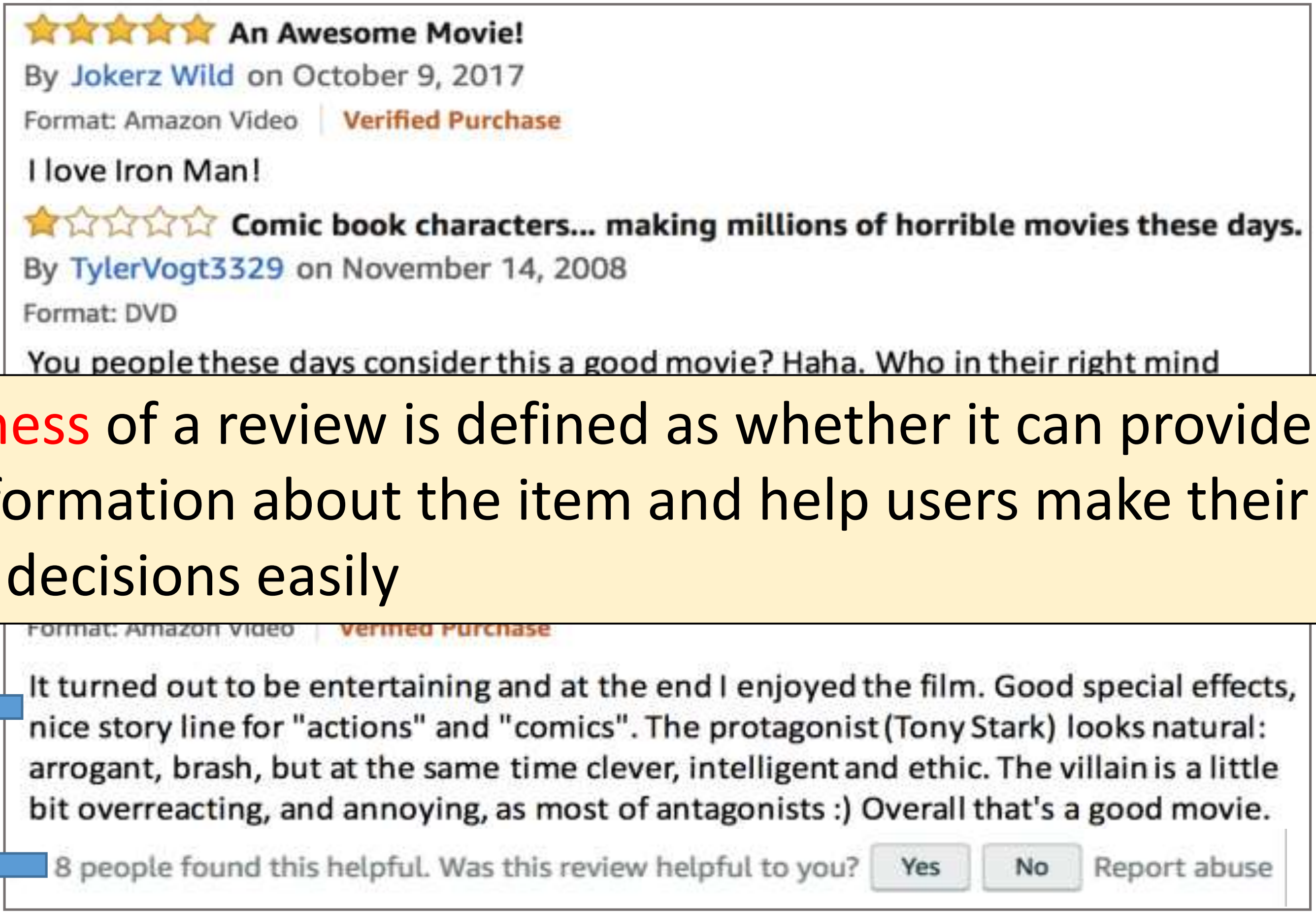
Explanation method

Usefulness of Review

A

B

C



The usefulness of a review is defined as whether it can provide detailed information about the item and help users make their purchasing decisions easily

Review
(to the item)

Rated usefulness
(to the review)

Recommendation model | Explanation method

Limitations of Previous Work

Incorporating Textual Review

- ✓✓ Most efforts are focused on how to combin with LDA model to improve the recommendation performance and generate word/feature-level explanation
- ✓✓ Manual pre-processing is usually required
- ✓✓ Damage the integrity of the sentence and lead to loss of information

---

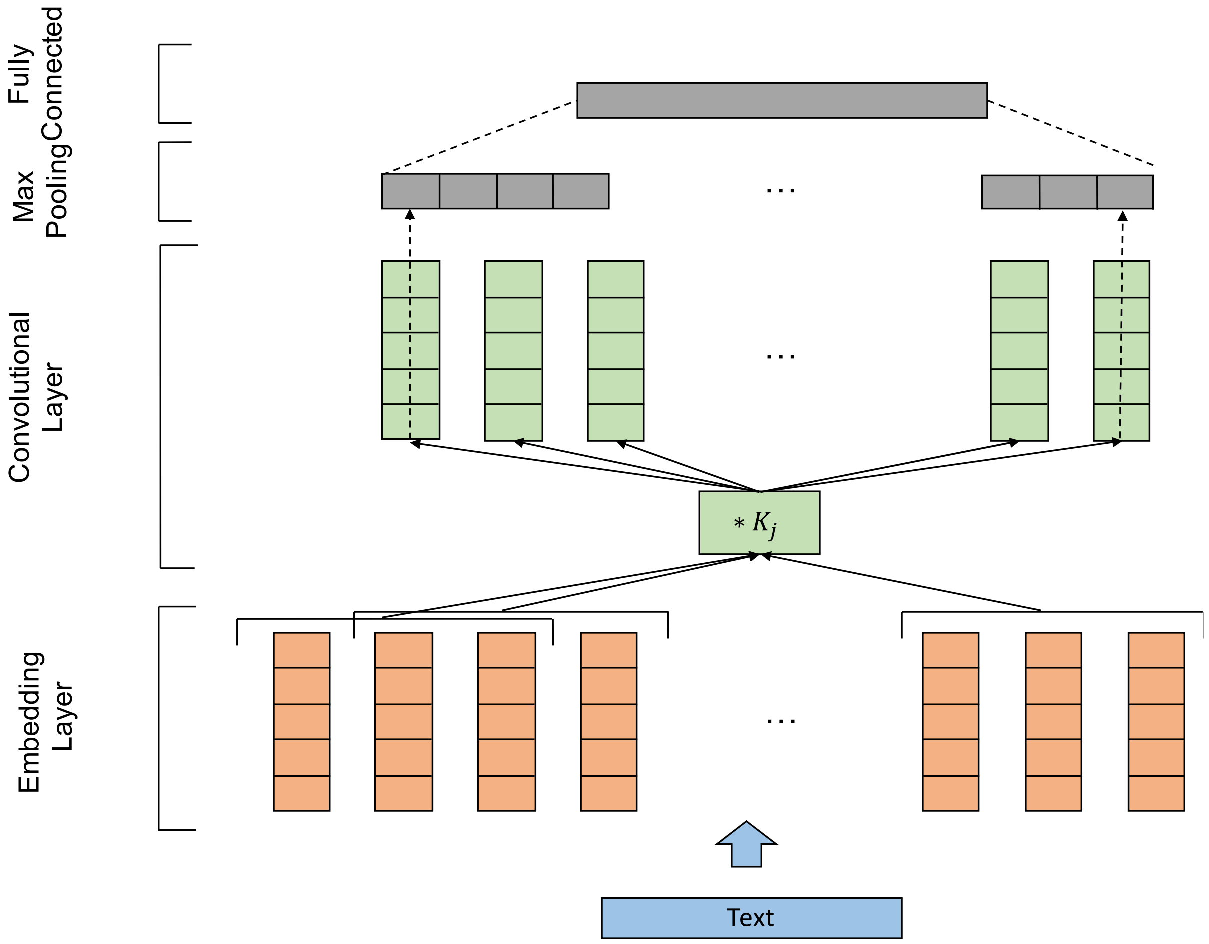
Usefulness of review

- ✓✓ Previous work only focuses on filtering spam in reviews as pre-processing

Recommendation model

Explanation method

CNN Text Processor



$$X = WO + g$$

$$O = [o_1, o_2, \ldots o_m]$$

$$o_j = max(z_1, z_2, \ldots z_j^{(T-t+1)})$$

$$z_j = ReLU(V_{1:T} * K_j + b_j)$$

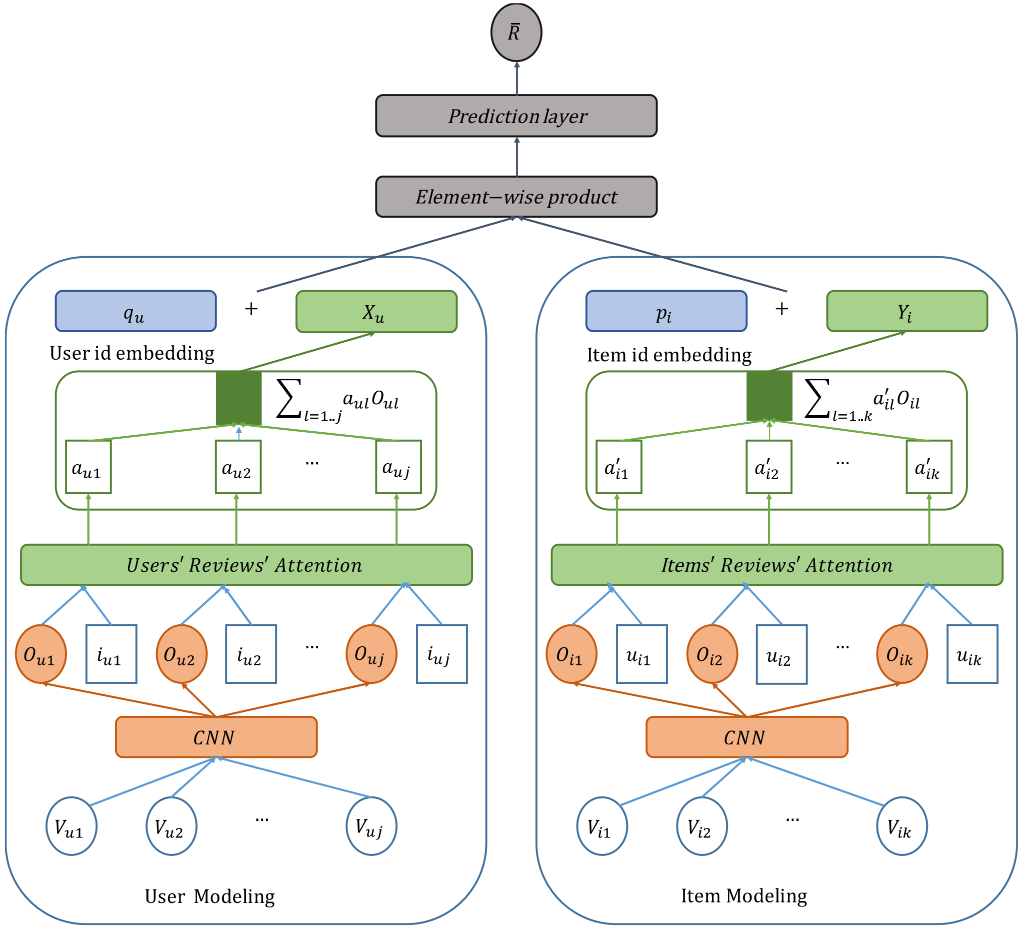
Recommendation model | Explanation method

Framework

Prediction layer

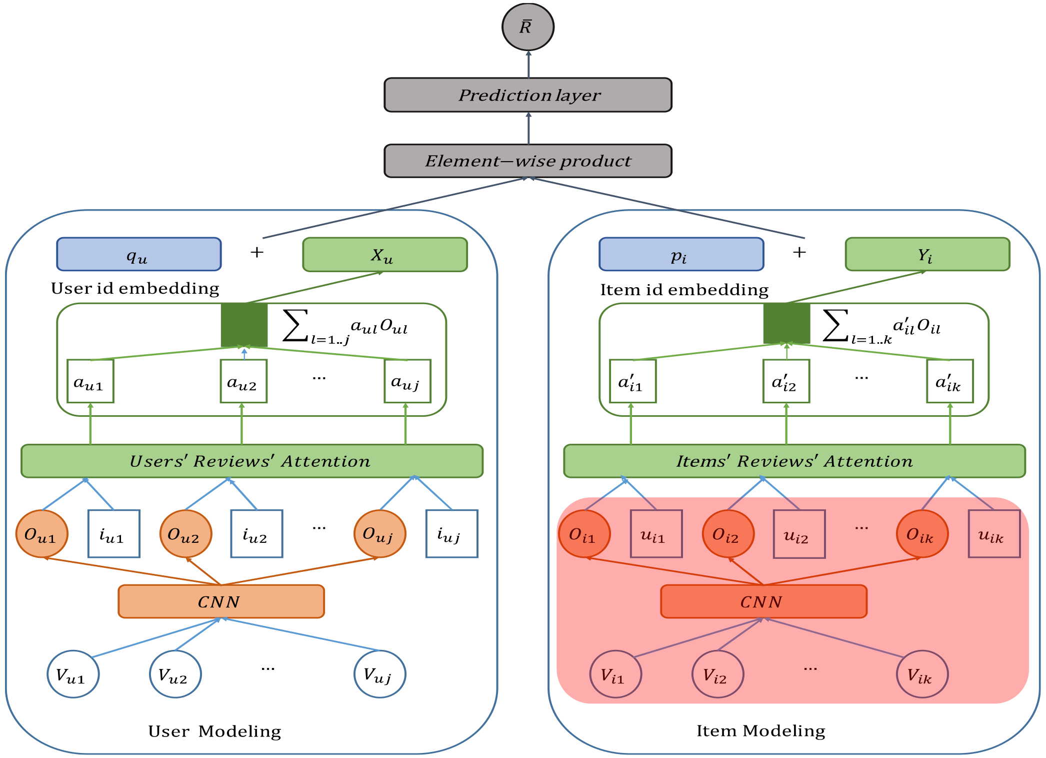
Attention-based review pooling

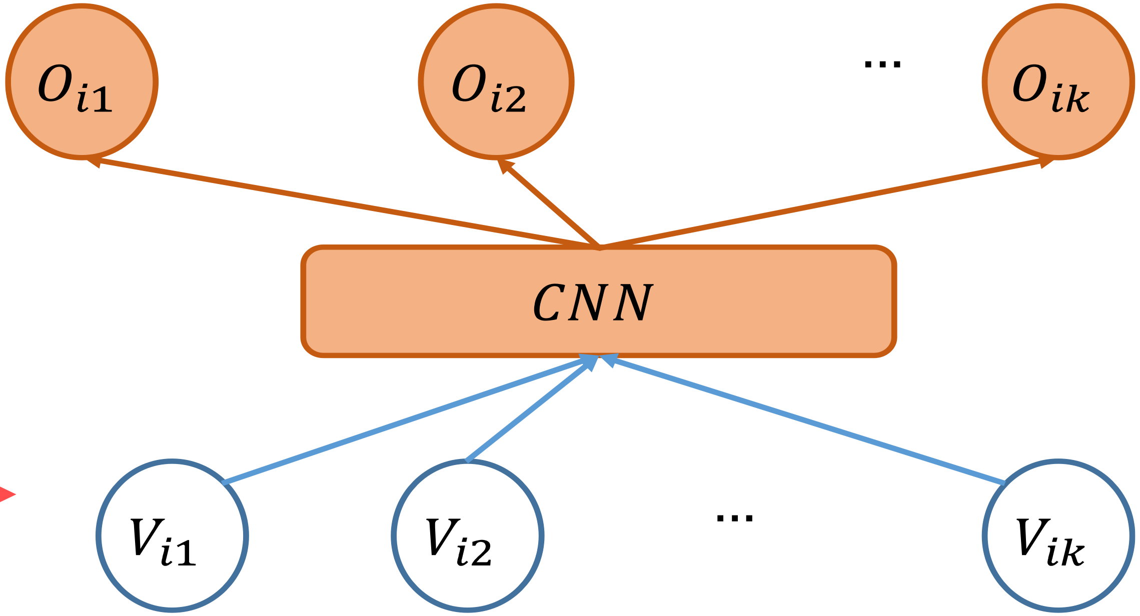
Text processing



Recommendation model | Explanation method

1. Text Processing





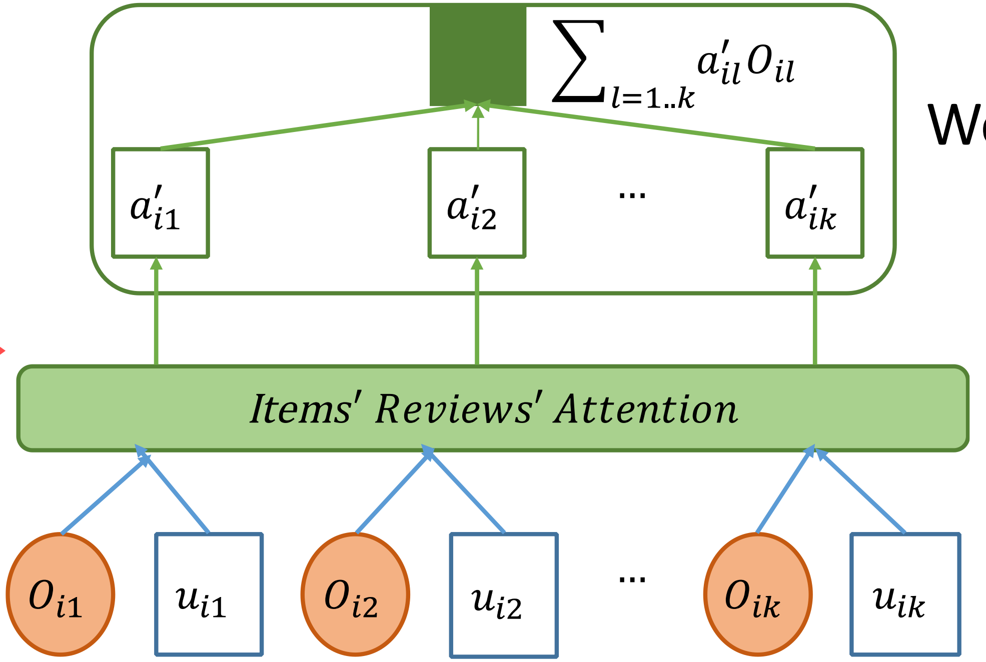
Input: Review list of item i $[V_{i_1}, V_{i_2}, \ldots V_{i_k}]$

Output: $[O_{i_1}, O_{i_2}, \ldots O_{i_k}]$

Recommendation model | Explanation method

2. Attention-based Review pooling

Fully connected: $Y_i = W_0 O_i + b_0$



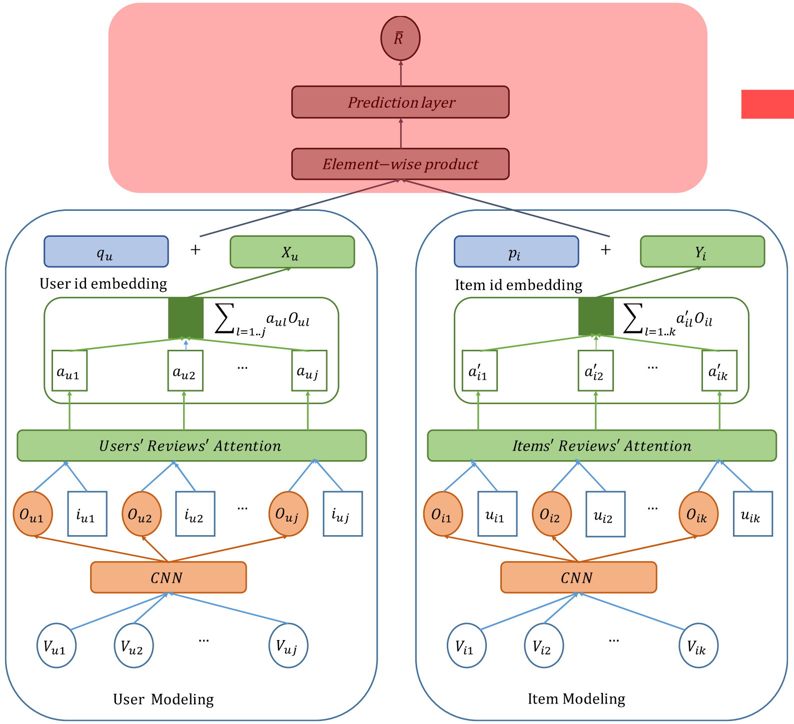
Weighted sum *Oi*

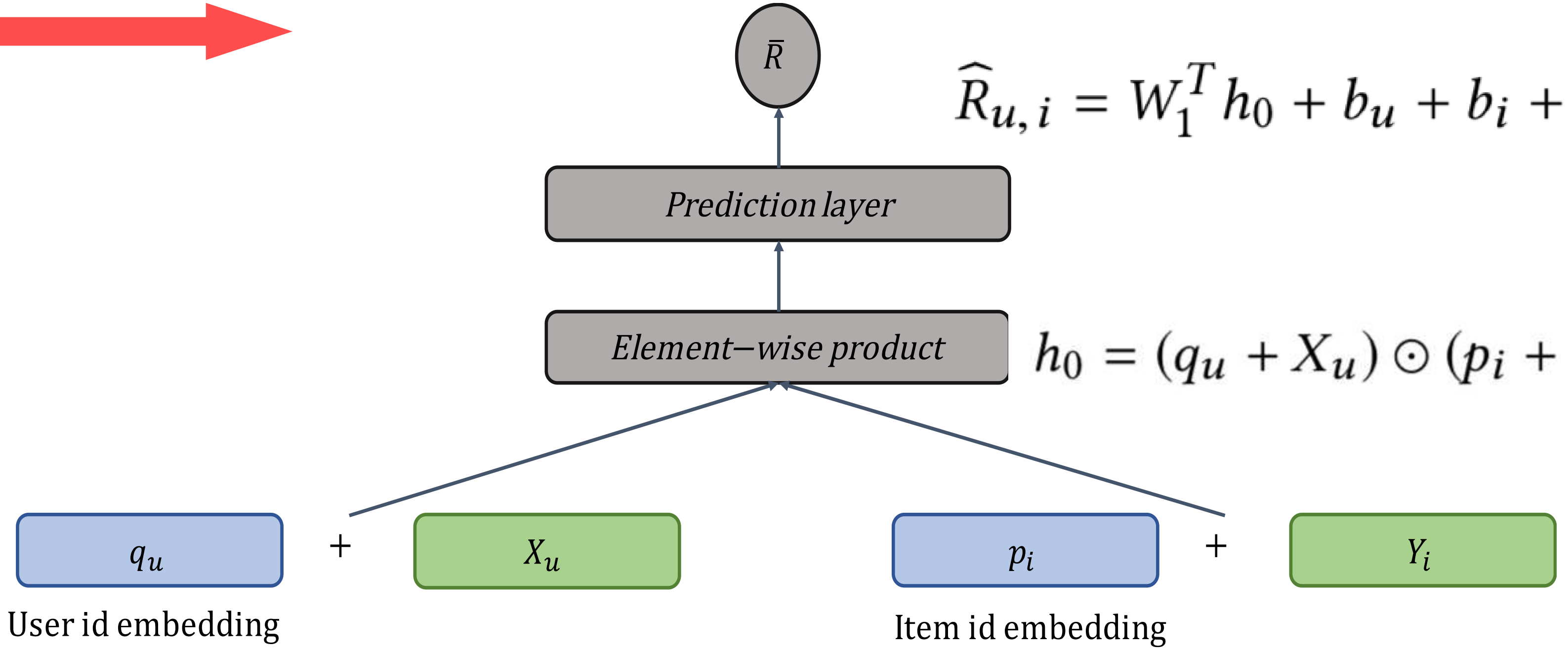
Attention score: $a^*_{il} = h^T ReLU(W_O O_{il} + W_u u_{il} + b_1) + b_2$

Normalization: $a_{il} = \frac{exp(a^*_{il})}{\sum_{l=0}^{k} exp(a^*_{il})}$

Recommendation model | Explanation method

3. Prediction Layer





$$\hat{R}_{u,i} = W_1^T h_0 + b_u + b_i + \mu$$

$$h_0 = (q_u + X_u) \odot (p_i + Y_i)$$

Recommendation model | Explanation method

Experiments: Data & Metric

Datasets:

	Toys_and_Games	Kindle_Store	Movies_and_TV	Yelp_2017
users	19,412	68,223	123,960	199,445
items	11,924	61,935	50,052	119,441
ratings & reviews	167,597	982,619	1,679,533	3.072.129

Evaluation Metric :

– RMSE

$$RMSE = \sqrt{\frac{1}{N}\sum_{u,i}(\widehat{R}_{u,i} - R_{u,i})^2}$$

Recommendation model | Explanation method

Baselines

- CF-based Methods
  - PMF, NMF, SVD++
- LDA-based Method
  - HFT
- Deep learning Method
  - DeepCoNN

Characteristics	PMF	NMF	SVD++	HFT	DeepCoNN	NARRE
Ratings	√	√	√	√	√	√
Textual Reviews	\	\	\	√	√	√
Deep Learning	\	\	\	\	√	√
Review Usefulness	\	\	\	\	\	√

NARRE: Neural Attentional Regression model with Review-level Explanations

Recommendation model | Explanation method

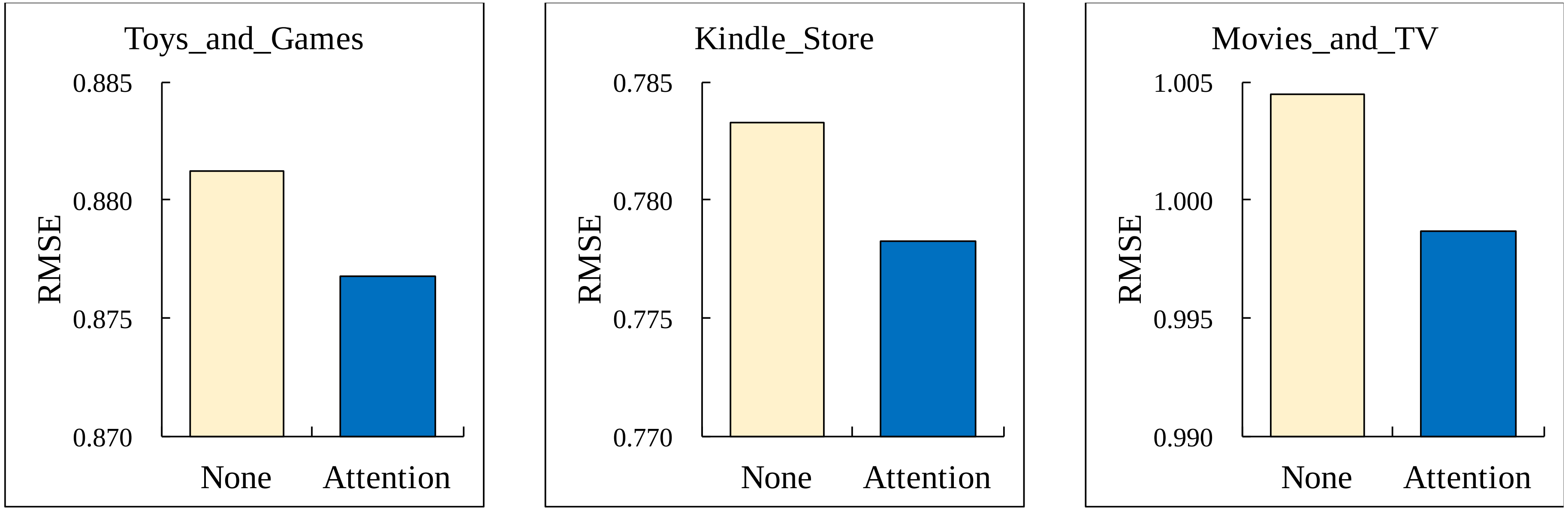
Model Comparisons

- Performance comparison on four datasets for all methods (RMSE)
- 80% training, 10% validation, 10% test

	Toys_and_Games	Kindle_Store	Movies_and_TV	Yelp-2017
PMF	1.3076	0.9914	1.2920	1.3340
NMF	1.0399	0.9023	1.1125	1.2916
SVD++	0.8860	0.7928	1.0447	1.1735
HFT	0.8925	0.7917	1.0291	1.1699
DeepCoNN	0.8890	0.7875	1.0128	1.1642
NARRE	**0.8769****	**0.7783****	**0.9965****	**1.1559***

Recommendation model | Explanation method

Effect of Attention



- None: $$O_i = \sum_{l=1,\ldots k} \frac{1}{k} O_{il}$$
- Attention: $$O_i = \sum_{l=1,\ldots k} a_{il} O_{il}$$

Recommendation model | Explanation method

Case Study

Item 1	a (a_{ij}=0.1932)	These brushes are great quality for children's art work. They seem to last well and the bristles stay in place very well even with tough use.
	b (a_{ij}=0.0161)	I bought it for my daughter as a gift.
Item 2	a (a_{ij}=0.2143)	From beginning to end this book is a joy to read. Full of mystery, mayhem, and a bit of magic for good measure. Perfect flow with excellent writing and editing.
	b (a_{ij}=0.0319)	I like reading in my spare time, and I think this book is very suitable for me.

Recommendation model | Explanation method

Review Usefulness Evaluation 1

- Baselines:
  - Latest
  - Random Selected
  - Length
- Ground truth:
  - Top_rated_useful

	Toys_and_Games				Kindle_Store				Movies_and_TV			
	Latest	Random	Length	NARRE	Latest	Random	Length	NARRE	Latest	Random	Length	NARRE
Precision@1	0.1487	0.3255	0.2476	**0.3860****	0.2447	0.4574	0.4041	**0.5235****	0.3040	0.4908	0.3903	**0.6576****
Recall@1	0.0362	0.0952	0.0771	**0.1398****	0.0400	0.0992	0.0852	**0.1131****	0.0436	0.0976	0.0677	**0.1445****
Precision@10	0.1550	0.2000	0.2316	**0.2697****	0.2228	0.2707	0.2933	**0.3530****	0.2325	0.2925	0.3369	**0.3459****
Recall@10	0.4367	0.5763	0.6763	**0.8601****	0.4510	0.5551	0.6168	**0.8317****	0.3716	0.4673	0.5403	**0.7674****

**:p<0.01 in statistical significance test, compared to the best baseline

Recommendation model | Explanation method

Review Usefulness Evaluation 2

- Crowd-sourcing based usefulness labeling

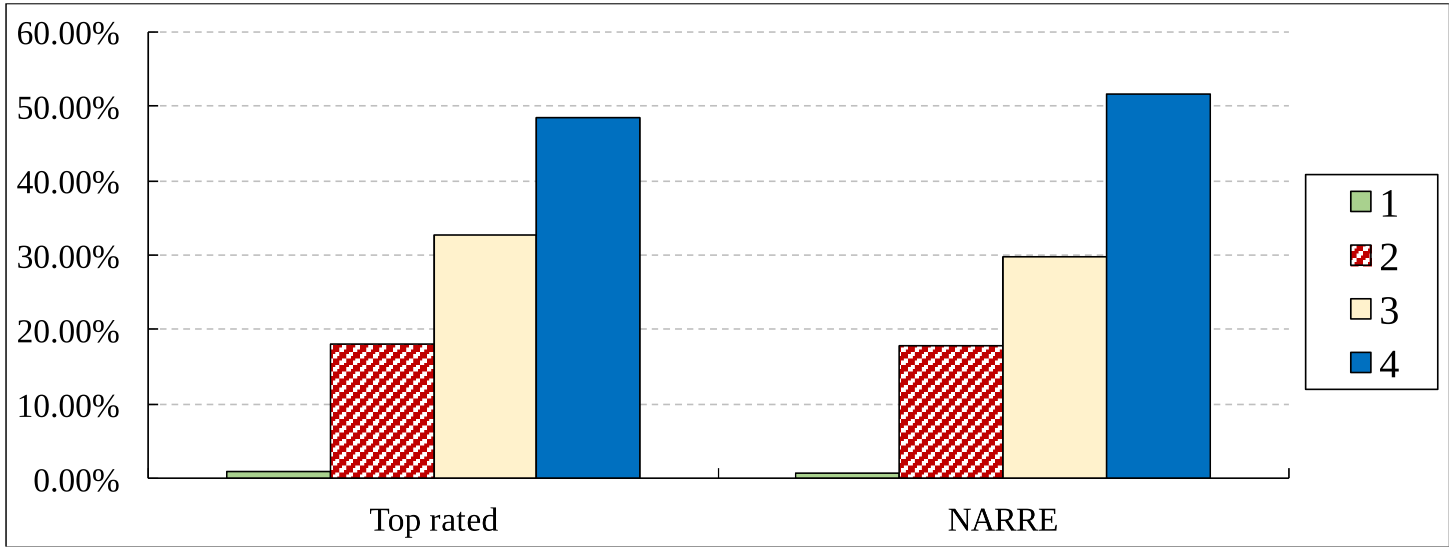
Annotation Instructions 1:
Background: You are going to buy an item, so you want to refer to the reviews written by previous consumers to know more about this item.
Task1: You need to browse each of the reviews below and then determine whether it is useful for your purchasing.
The review can be classified as follows:
- **1 star**: Not useful at all.
- **2 stars**: Somewhat useful.
- **3 stars**: Fairly useful.
- **4 stars**: Very useful.

	Items	Reviews	Reviews of each method	Annotations	Weighted κ
U_a	100	1264	745	3792	0.4112

Recommendation model | Explanation method

Review Usefulness Evaluation 2

- Crowd-sourcing based usefulness labeling



Ua = 1: not useful at all;
2: somewhat useful;
3: fairly useful;
4: very useful.

	Precision@1	Precision@5	Precision@10	Recall@1	Recall@5	Recall@10	NDCG@1	NDCG@5	NDCG@10
Top_Rated_Useful	0.4800	0.4440	0.3610	0.0821	0.3453	0.4953	0.6640	0.6906	0.7076
NARRE	**0.5900****	**0.4760****	**0.3850****	**0.1067****	**0.3532****	**0.5046****	**0.7413****	**0.7231****	**0.7358****

Recommendation model | Explanation method

Review Usefulness Evaluation 3

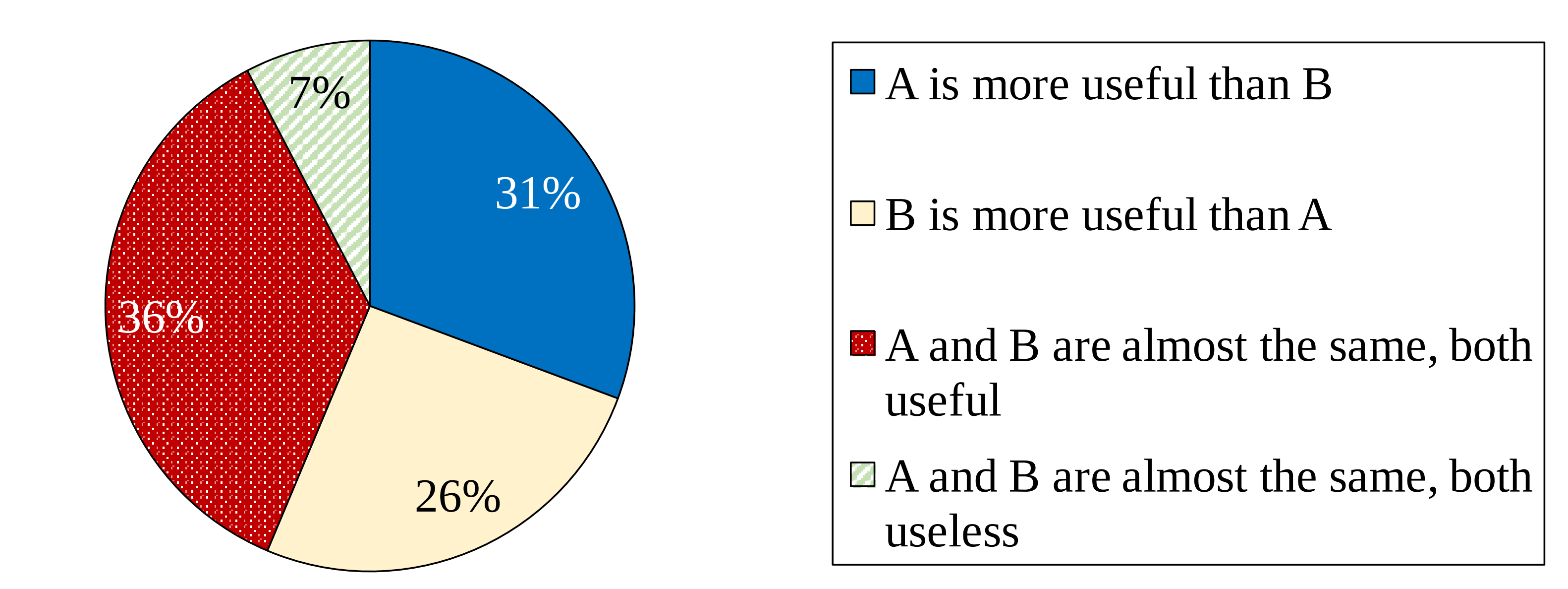
- Crowd-sourcing based pairwise evaluation

Annotation Instructions 2:
Task2: You will see two groups of reviews, and each group contains 5 reviews. You need to browse each group and annotate pairwise usefulness between Group A and Group B.

- A is more useful than B.
- B is more useful than A.
- A and B are almost the same, both useful.
- A and B are almost the same, both useless.

In the evaluation: randomly shown as A or B

In the Figure →
A: Ours;
B: Top_rated_useful;



Recommendation model | Explanation method

Embedded Methods

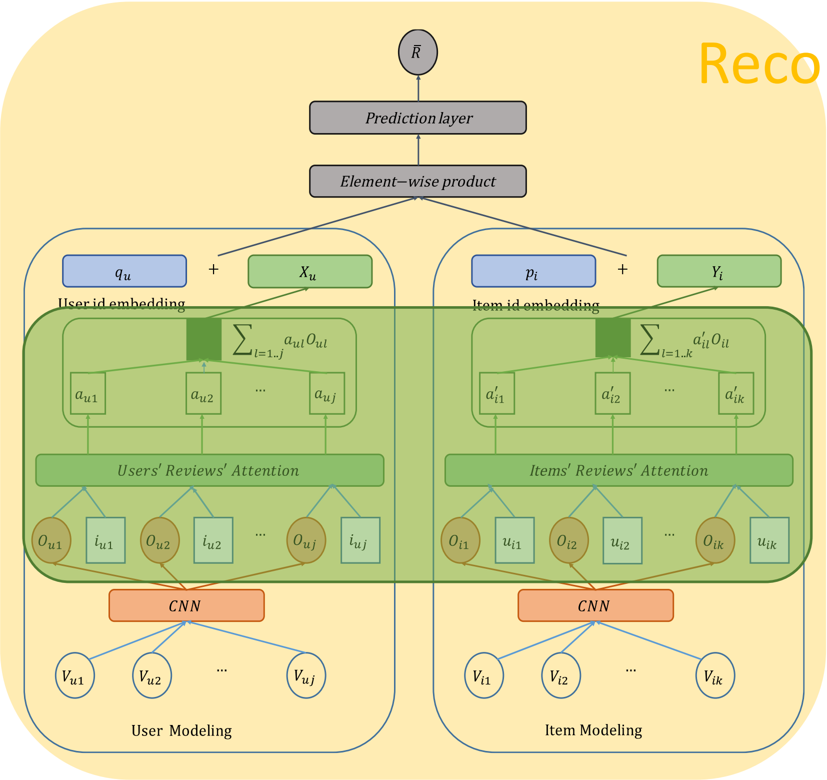
- Most embedded methods are feature-based
  - Features are usually parts from the auxiliary information (review, images)
  - It fits well with existing recommendation models (can even improve accuracy)
- Types of features
  - Phrases
  - Sentences
  - Images

NARRE: Review-level explanation

Neural Attentional Rating Regression with Review-level Explanations

Chong Chen, Min Zhang*
Yiqun Liu, Shaoping Ma

[WWW2018]



Recommendation model

Explanation method

Recommendation model | Explanation method

Embedded Methods

- Most embedded methods are feature-based
  - Features are usually parts from the auxiliary information (review, images)
  - It fits well with existing recommendation models (can even improve accuracy)
- Types of features
  - Phrases
  - Sentences
  - Images

Re-VECF: Review enhanced visual explanation

#	Target Item	Historical Records	Textual Review	Visual Explanation VECF	Visual Explanation Re-VECF
1			this is a large watch... nearly as large as my suunto but due to ***its articulated strap it fits on the wrist very well.***		
2			***this is a really comfortable v-neck. i found that the size and location of the v are just right for me.*** i'm 5'8 & #34, but 200 lbs (and dropping :))		
3			***Great leggings. perfect for fly fishing or hunting or running.*** just perfect anytime you are cold!		
4			The socks on the shoes are a perfect fit for me. ***first time with a shoe with the speed laces and i like them a lot***		
5			Really like these socks! they are really thick woolen socks and are good for cold days. ***they cover a good portion of your feet as they go a little (halfway) above the calf muscle area.***		
6			***I like the front pocket~!*** Very cool!		

Visually Explainable Recommendation

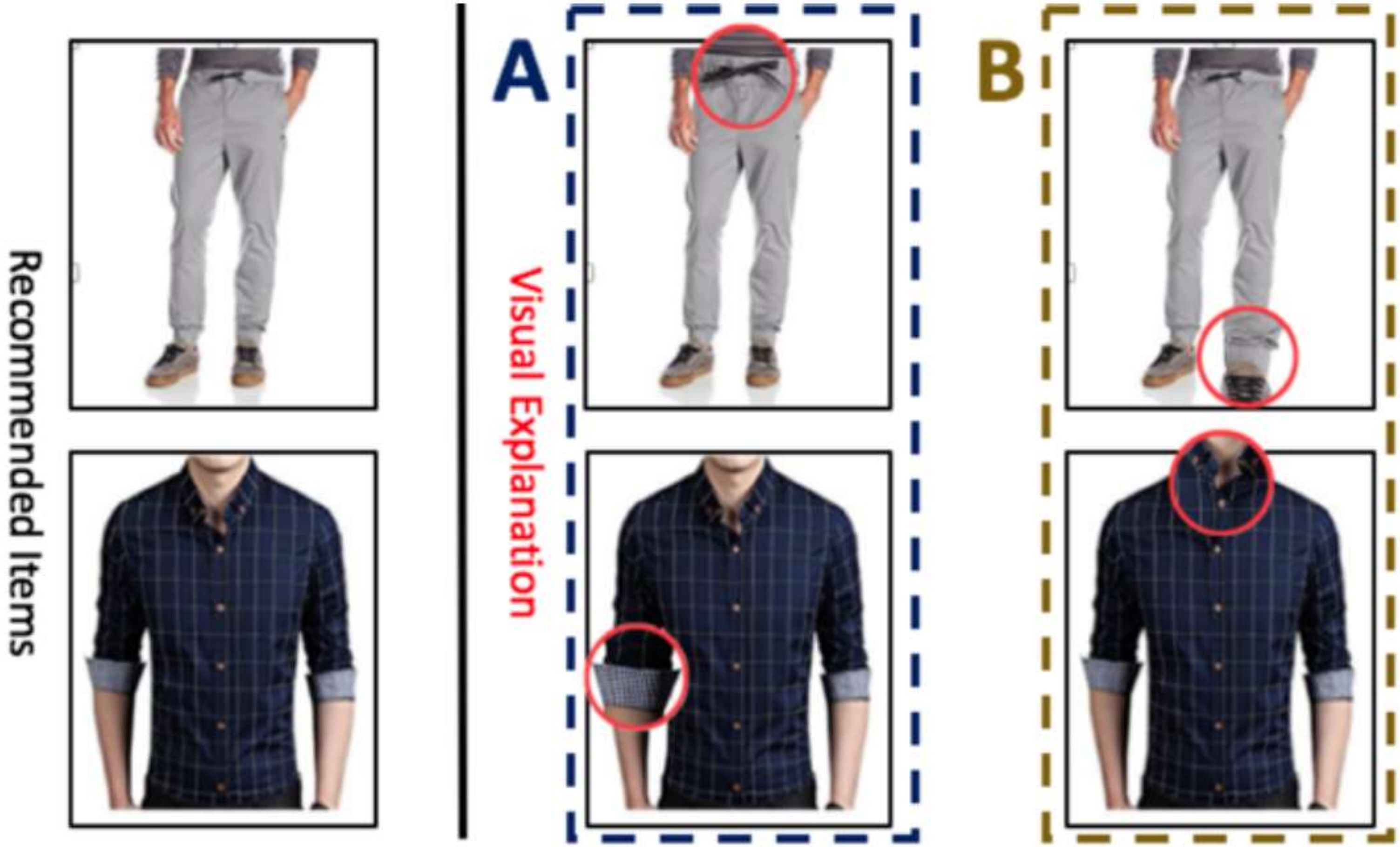
Xu Chen, Yixin Cao, Yongfeng Zhang, Zheng Qin, Hongteng Xu, Hongyuan Zha

[Arxiv2018]

Recommendation model | Explanation method

Visually Explainable Recommendation

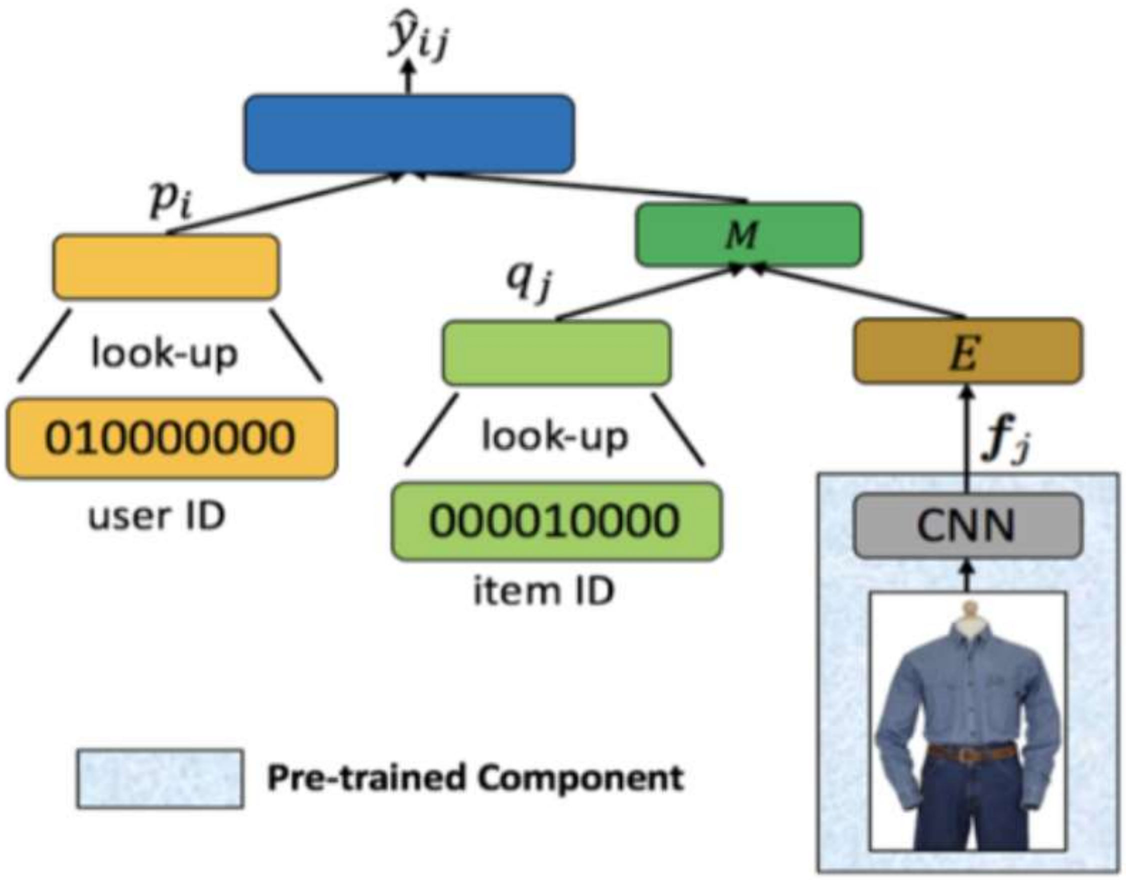
- Users may care about different visual features even on the same item



Recommendation model

Explanation method

Visual Collaborative Filtering (VCF)



If we do not consider image feature :

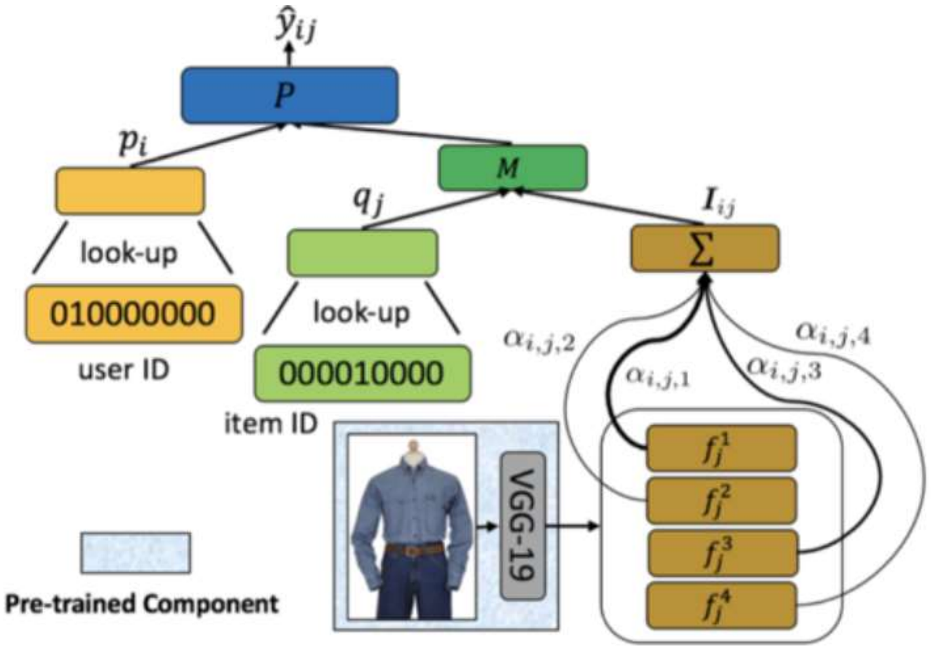
$$\hat{y}_{ij} = \boldsymbol{p}_i^T \cdot \boldsymbol{q}_j$$

Integrating pre-trained image feature:

$$\hat{y}_{ij} = \boldsymbol{p}_i^T \cdot M(\boldsymbol{q}_j, E(\boldsymbol{f}_j))$$

Visually Explainable Collaborative Filtering (VECF)

Recommendation model | Explanation method



- 1. Image feature extraction: divide image by 14*14, each region is fed into pre-trained VGG network [1] to generate a 512-dim vector.
- 2. Using attention mechanism to learn a unified image vector

$$IMAGE_j = \sum_{k=1}^{h} \alpha_{i,j,k} \cdot f_j^k$$

$$a_{i,j,k} = g\left((\boldsymbol{w}_{att}^u)^T \cdot \boldsymbol{p}_i + (\boldsymbol{w}_{att}^r)^T \cdot f_j^k + b_{att}\right)$$

$$\alpha_{i,j,k} = \frac{a_{i,j,k}}{\sum_{\kappa=1}^{h} a_{i,j,\kappa}}$$

- 3. Merge image feature with randomly initialized item vector (we use element-wise multiplication) $\boldsymbol{q}_j^* = MERGE(\boldsymbol{q}_j, \boldsymbol{IMAGE}_j)$
- 4. Predict user-item ratings by maximizing the cross-entropy

$$\hat{y}_{ij} = PREDICT(\boldsymbol{p}_i, \boldsymbol{q}_j^*) \quad l_1 = \log \prod_{(i,j)} (\hat{y}_{ij})^{y_{ij}} (1-\hat{y}_{ij})^{1-y_{ij}} - \lambda \|\Theta\|_F^2$$

Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. ICLR (2015)

Recommendation model | Explanation method

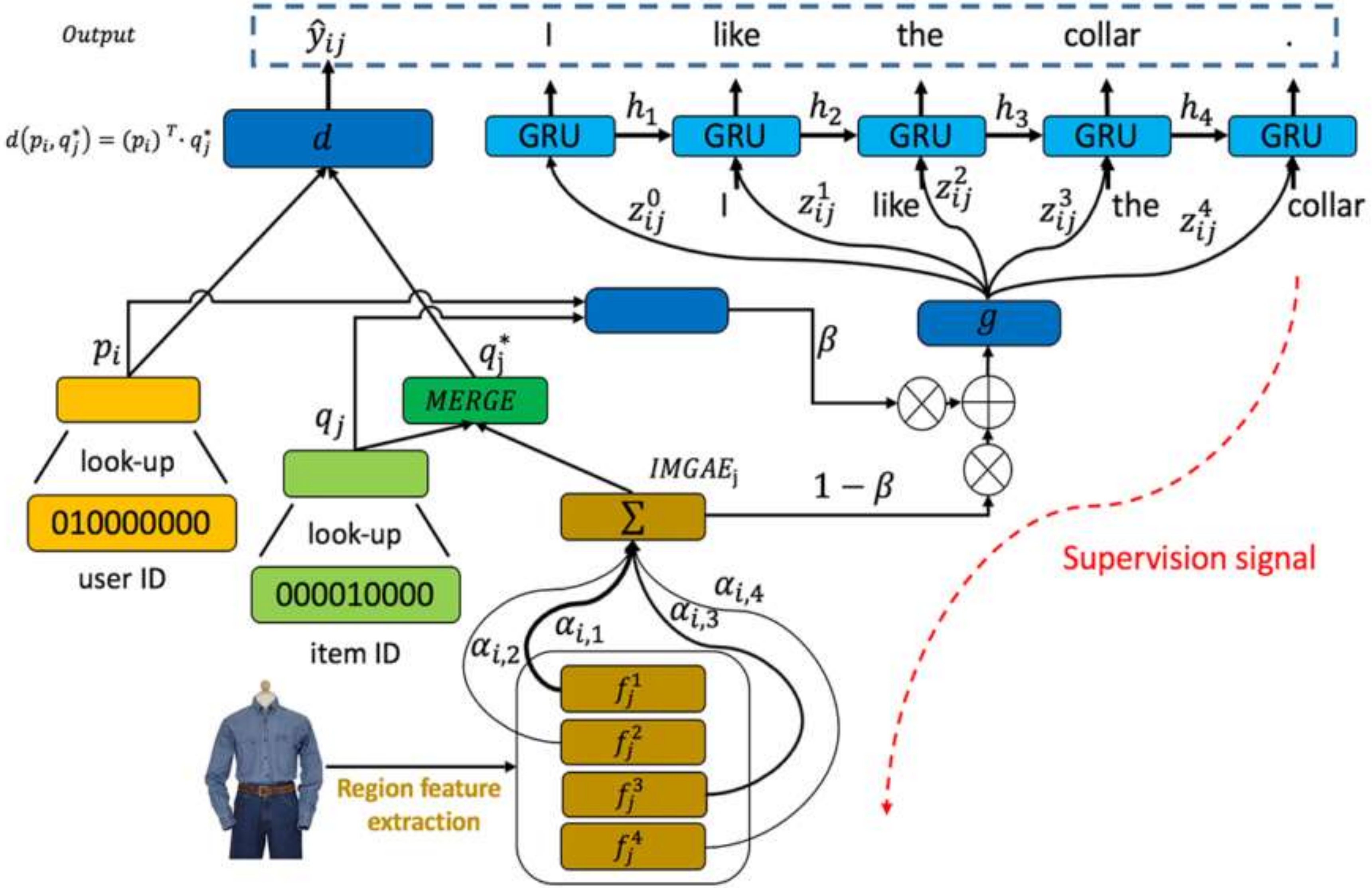
Incorporating the Text Signal

- People comment on image features that they care about in their textual reviews

Recommendation model

Explanation method

Review-Enhanced VECF



Recommendation model | Explanation method

Experiments

- Recommendation accuracy

Table 2: Statistics of the datasets.

Datasets	#Users	#Items	#Interactions	Density	#Words
Men	643	2454	6359	0.403%	21600
Women	570	3346	7640	0.401%	17614

Dataset	Men			Women		
Measure@5(%)	F_1	HR	NDCG	F_1	HR	NDCG
BPR	1.209	3.901	0.740	0.897	3.342	0.611
HFT	1.242	4.243	0.757	0.915	3.371	0.631
VBPR	1.361	4.261	0.773	0.929	3.402	0.648
NRT	1.399	4.469	0.802	0.952	3.527	0.674
JRL	1.424*	4.703*	0.813*	0.967*	3.542*	0.686*
Re-CF	1.370	4.364	0.781	0.937	3.451	0.651
VECF	1.378	4.373	0.791	0.948	3.523	0.669
Re-VECF	**1.442**	**4.803**	**0.846**	**0.985**	**3.587**	**0.712**

Recommendation model

Explanation method

Experiments

- Evaluation of visual explanations
  - Use crowd sourcing (Amazon MTurk) to label the images
    - Worker is asked to identify the top 5 relevant regions given an image and the corresponding review.
    - Each image is labeled by at least two worker
    - Keep the common regions
    - Evaluate the identified regions by our algorithm

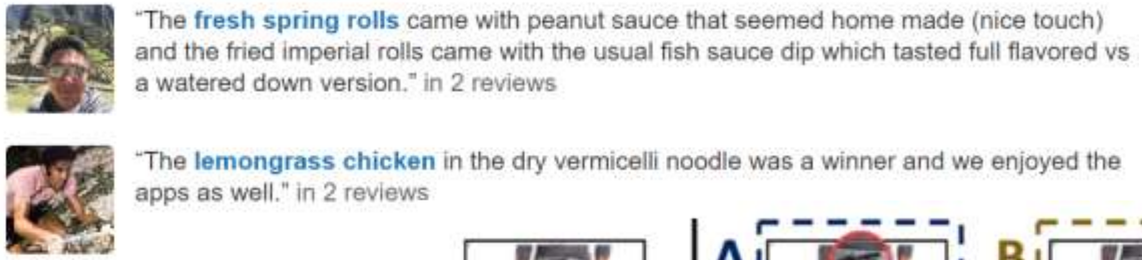
Method	Top-5		Top-10	
	F_1(%)	NDCG(%)	F_1(%)	NDCG(%)
Random	3.22	8.24	7.41	11.46
VECF	6.70	17.37	10.38	16.40
Re-VECF	8.35	20.53	12.99	19.95

- Though identified region may not be the exact true region, they are usually very close

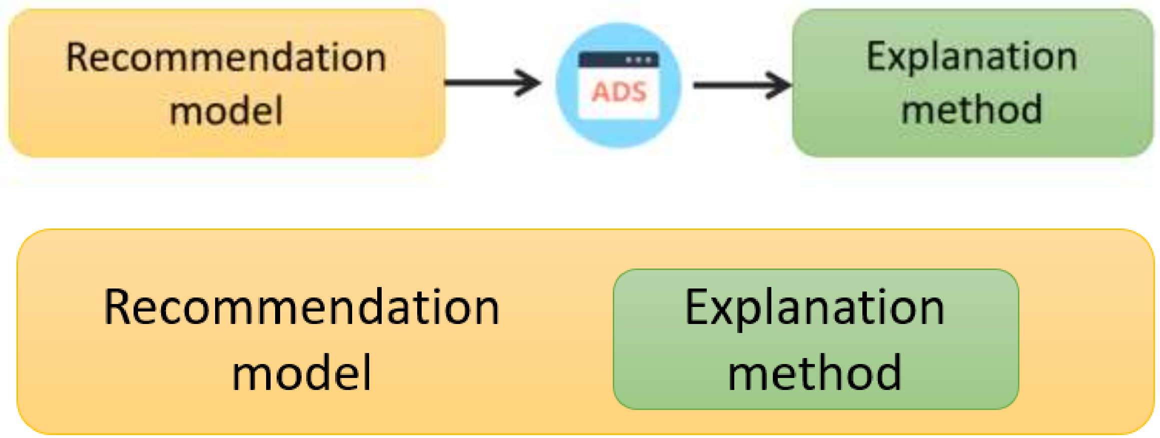
Recommendation model | Explanation method

Embedded Methods

- Most embedded methods are feature-based
  - Features are usually parts from the auxiliary information (review, images)
  - It fits well with existing recommendation models (can even improve accuracy)
- Types of features
  - Phrases
  - Sentences
  - Images

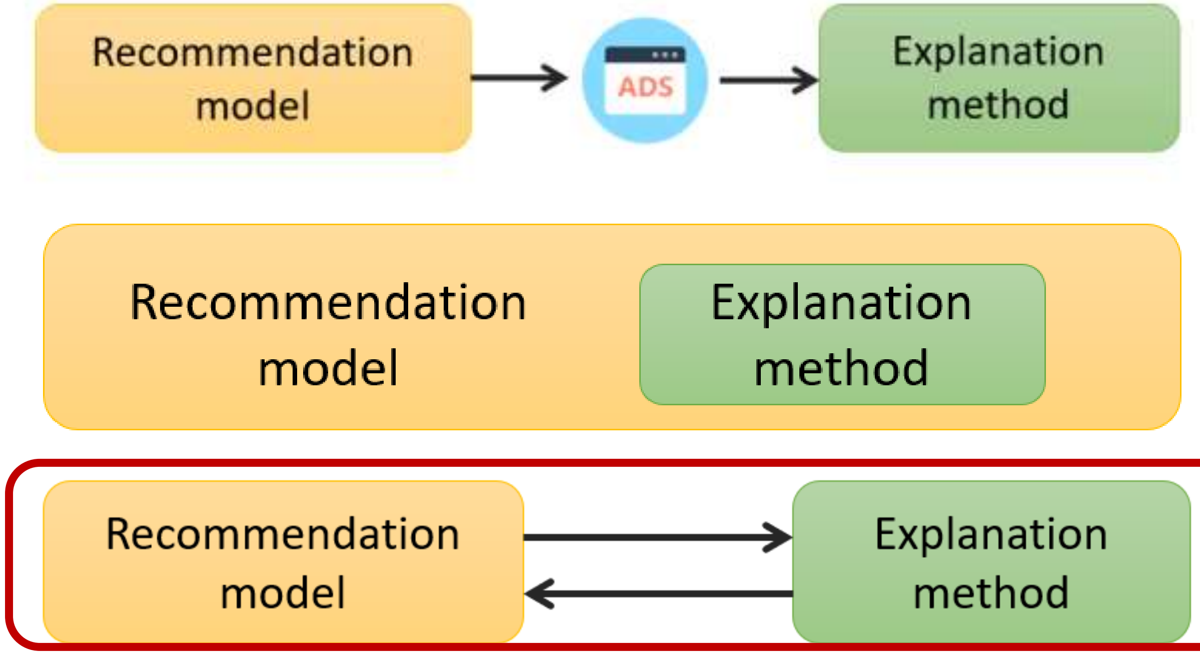


Comparison of Pipelines



	Model explainability	Presentation quality	Model agnostic
Post-hoc	x	√	√
Embedded	√	x	x

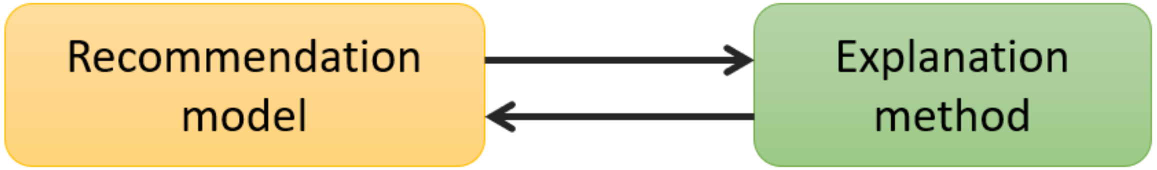
Comparison of Pipelines



	Model explainability	Presentation quality	Model agnostic
Post-hoc	x	√	√
Embedded	√	x	x
Wrapper	√	√	√

Our Pipeline

Our Wrapper Method



A Reinforcement Learning Framework for Explainable Recommendation

Xiting Wang
Microsoft Research Asia
xitwan@microsoft.com

Yiru Chen
Peking University
chen1ru@pku.edu.cn

Jie Yang
Tsinghua University
yangj16@mails.tsinghua.edu.cn

Le Wu
Hefei University of Technology
lewu@hfut.edu.cn

Zhengtao Wu
University of Science and Technology of China
wzt@mail.ustc.edu.cn

Xing Xie
Microsoft Research Asia
xing.xie@microsoft.com

Recommendation model ⇄ Explanation method

Problem Definition

- Input
  - User set U, $u \in U$ is a user ---------------------- u: user ID and/or some side information
  - Item set V, $v \in V$ is an item ---------------------- $\boldsymbol{v} = (i, \boldsymbol{l}_1, \boldsymbol{l}_2, ..., \boldsymbol{l}_m)$
  - A recommendation model to be explained $f(u, v)$

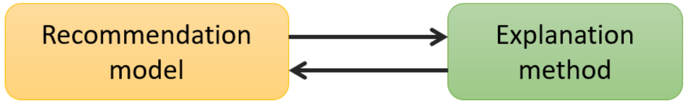
i: item ID l_j: interpretable component

EFM: phrases like "图像-清晰"
NARRE: a user review
Re-VECF: part of an image
Can also be: key features of an item, like five-star rating

- Output
  - Explanation $z = (z_1, z_2, ...z_m)$

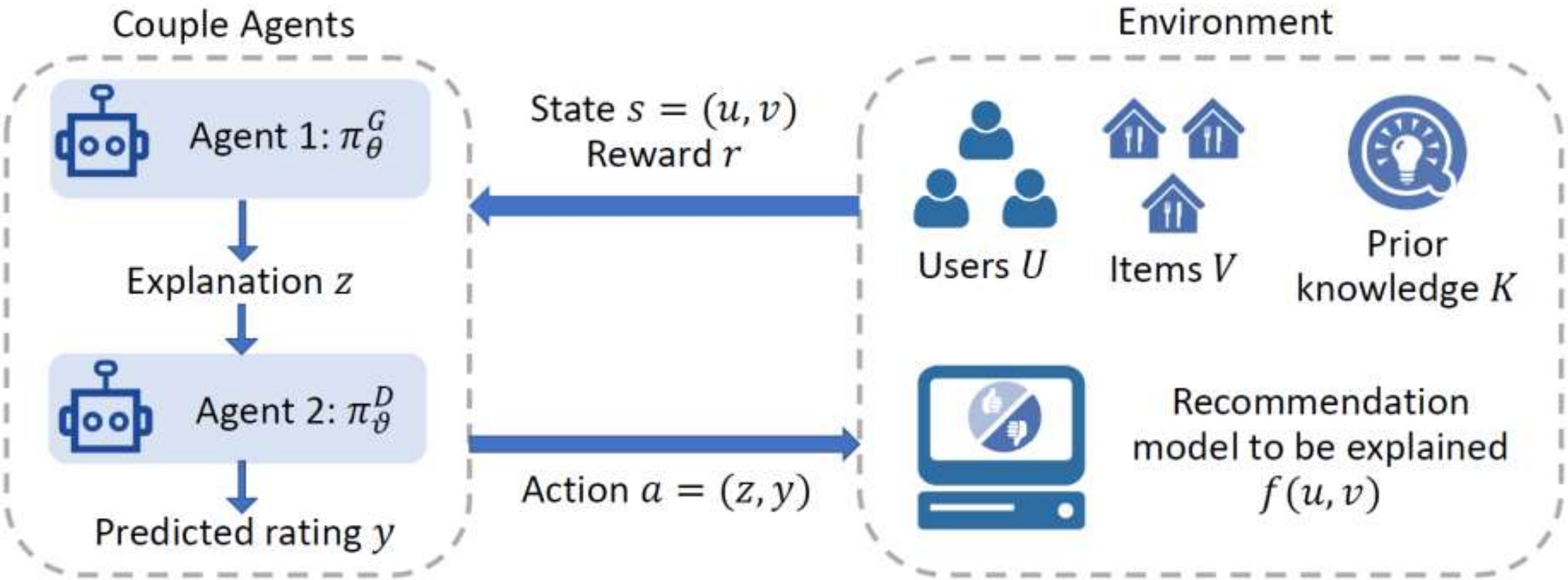
$z_j = 1$ The jth interpretable component is selected

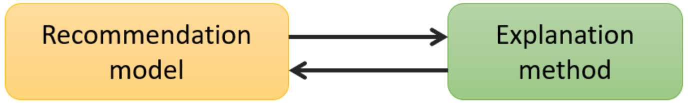
$z_j = 0$ The jth interpretable component is not selected



Reinforcement Learning Framework

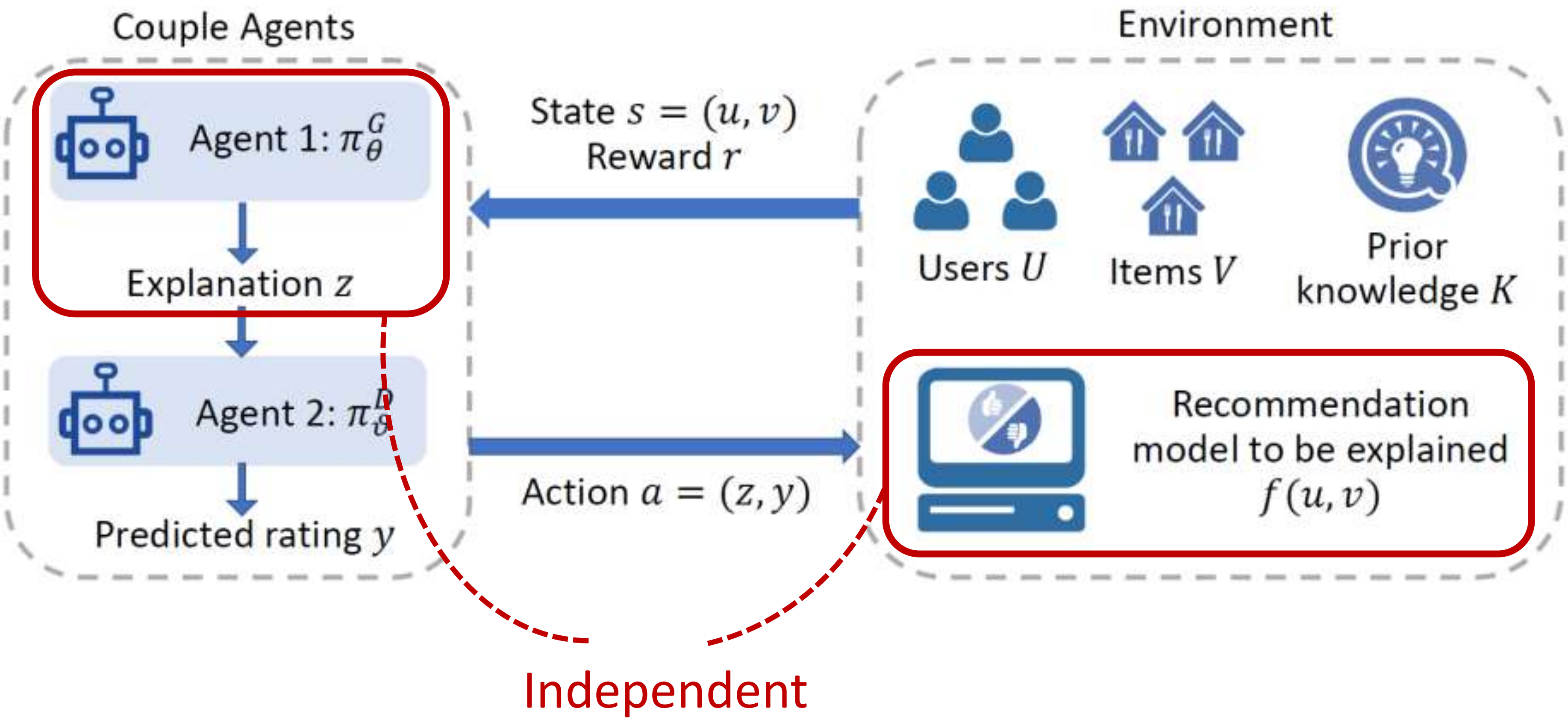
- Advantages: model-agnostic, model-explainability, presentation quality

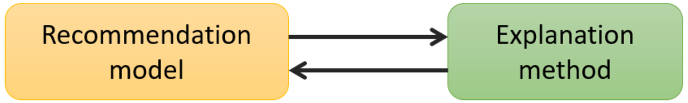




Reinforcement Learning Framework

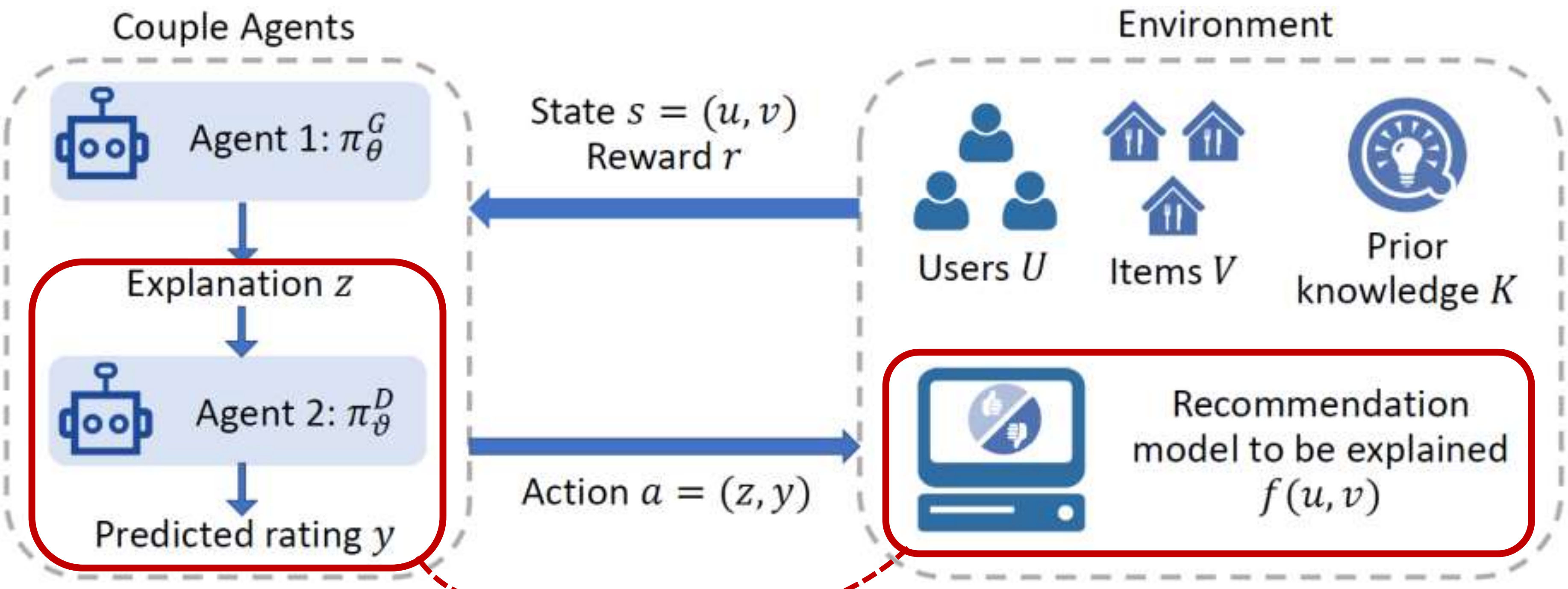
- Advantages: **model-agnostic**, model-explainability, presentation quality



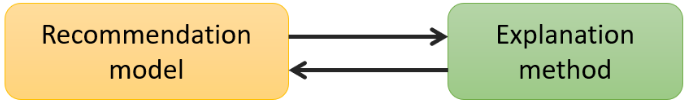


Reinforcement Learning Framework

- Advantages: model-agnostic, **<u>model-explainability</u>**, presentation quality

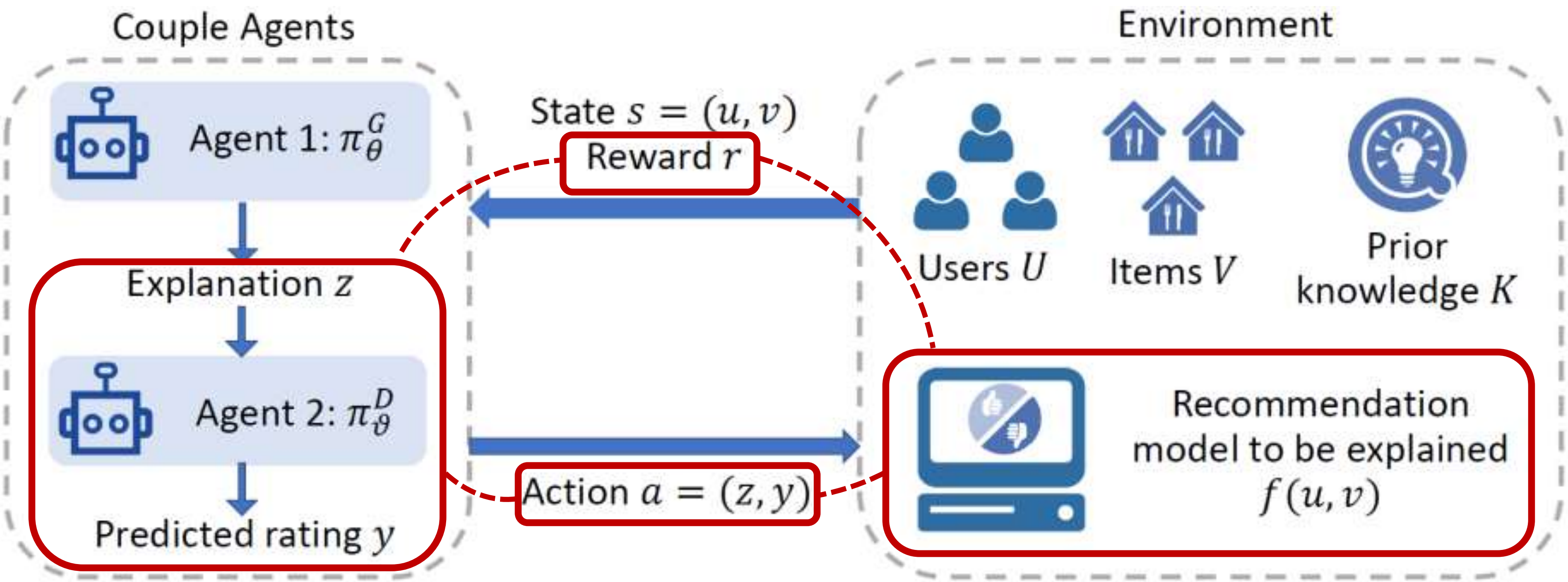


Agent 2 can predict $f(u, v)$



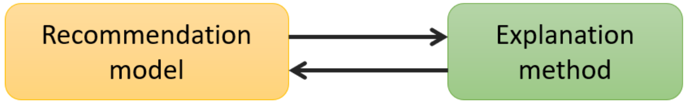
Reinforcement Learning Framework

- Advantages: model-agnostic, **model-explainability**, presentation quality



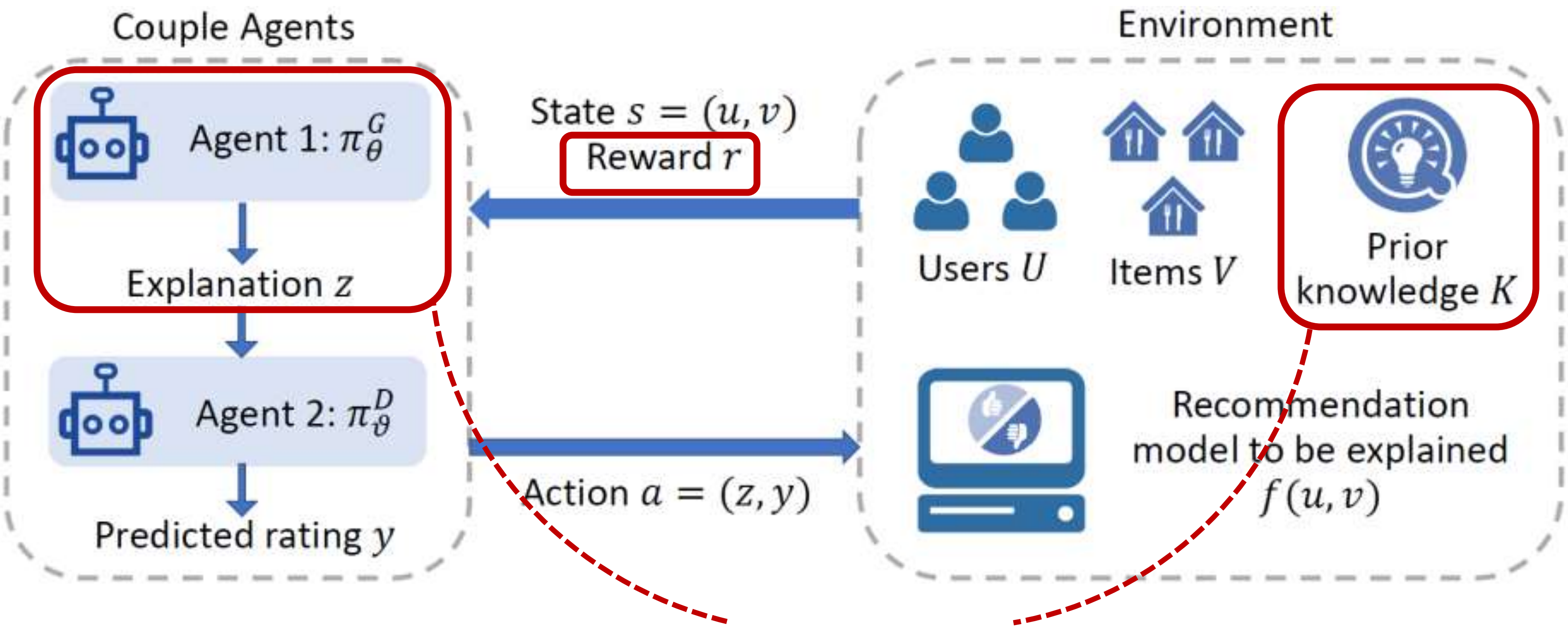
Agent 2 can predict $f(u, v)$

$$r = \mathcal{L}(f(\boldsymbol{u}, \boldsymbol{v}), y) + \Omega(\boldsymbol{z}) \qquad \mathcal{L}(f(\boldsymbol{u}, \boldsymbol{v}), y) = -(y - f(\boldsymbol{u}, \boldsymbol{v}))^2$$



Reinforcement Learning Framework

- Advantages: model-agnostic, model-explainability, **presentation quality**



r will increase if z is good given K

Reinforcement Learning Framework

- Advantages: model-agnostic, model-explainability, **presentation quality**

$$r = \mathcal{L}(f(\boldsymbol{u}, \boldsymbol{v}), y) + \Omega(\boldsymbol{z})$$

$$\Omega(\boldsymbol{z}) = \lambda_r \Omega_r(\boldsymbol{z}) + \lambda_c \Omega_c(\boldsymbol{z})$$

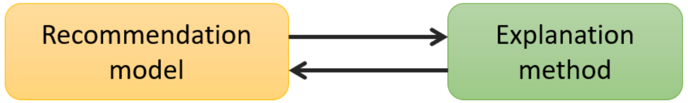
Readability

$$\Omega_r(\boldsymbol{z}) = -|z^* - \sum_{j=1}^{m} z_j| - \lambda_b \sum_{j=2}^{m} |z_j - z_{j-1}|$$

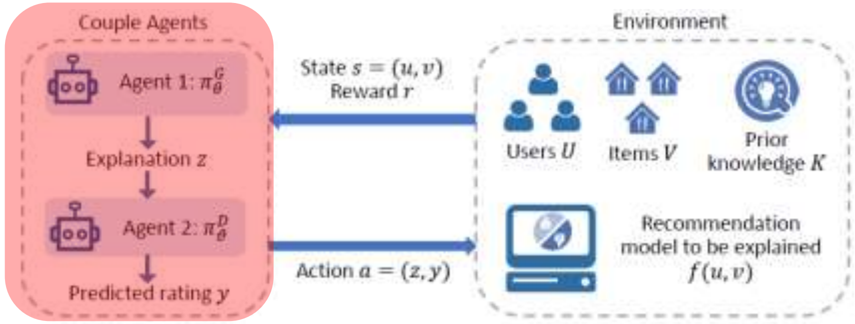
Consistency

$$\Omega_c(\boldsymbol{z}) = (\frac{\sum_{j=1}^{m} z_j \varphi(\boldsymbol{l}_j)}{\sum_{j=1}^{m} z_j} - \bar{\varphi})(f(\boldsymbol{u}, \boldsymbol{v}) - \bar{f})$$

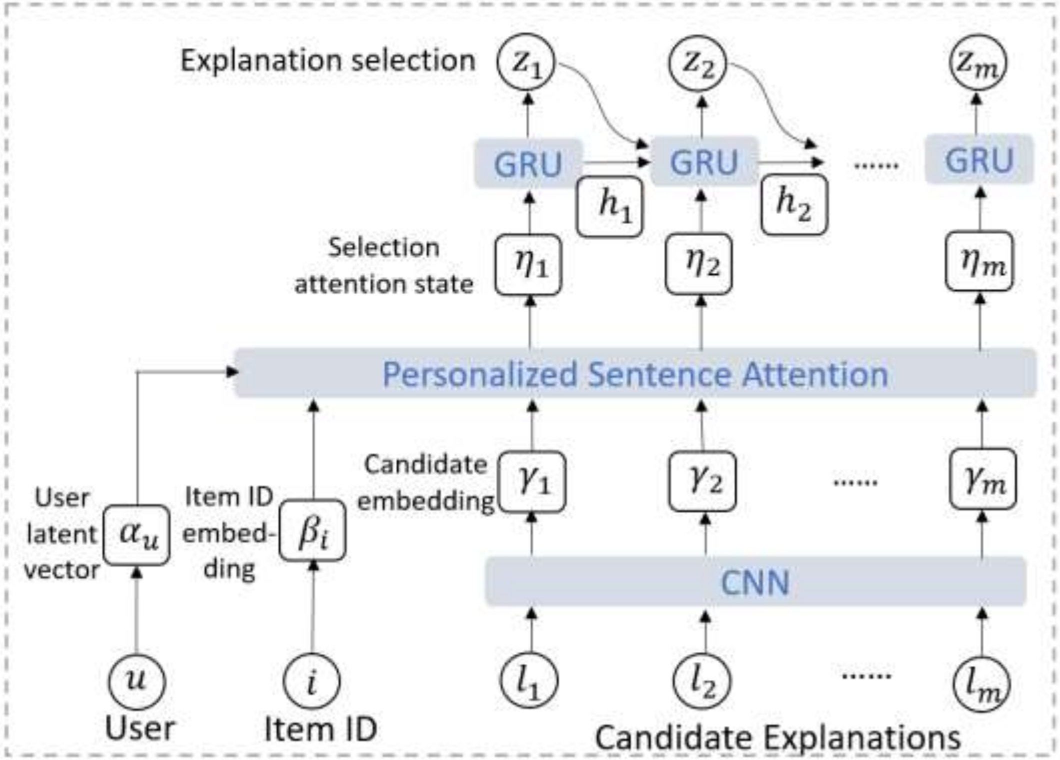
Reinforcement learning framework [ICDM2018]



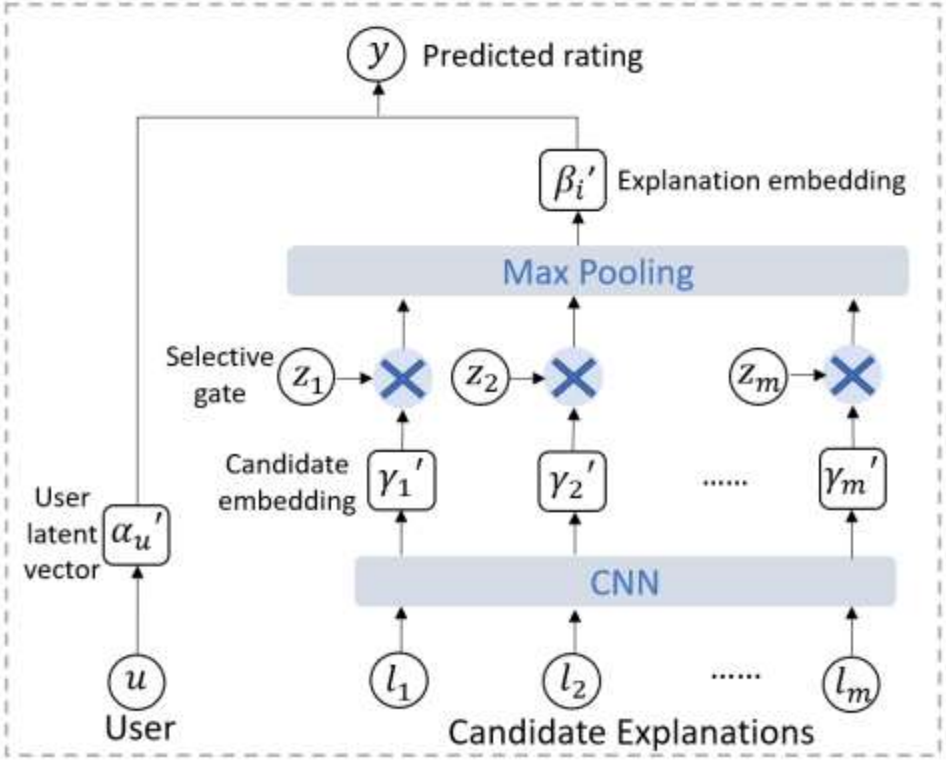
Couple Agents



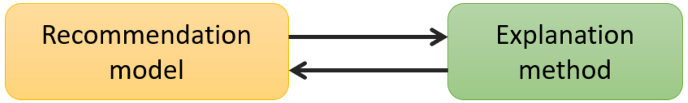
Agent 1: $\pi_\theta^G(z, u, v) = p(z|u, v, \theta)$



Agent 2: $y = \pi_\vartheta^D(u, v, z)$



Sentence-level Explanation



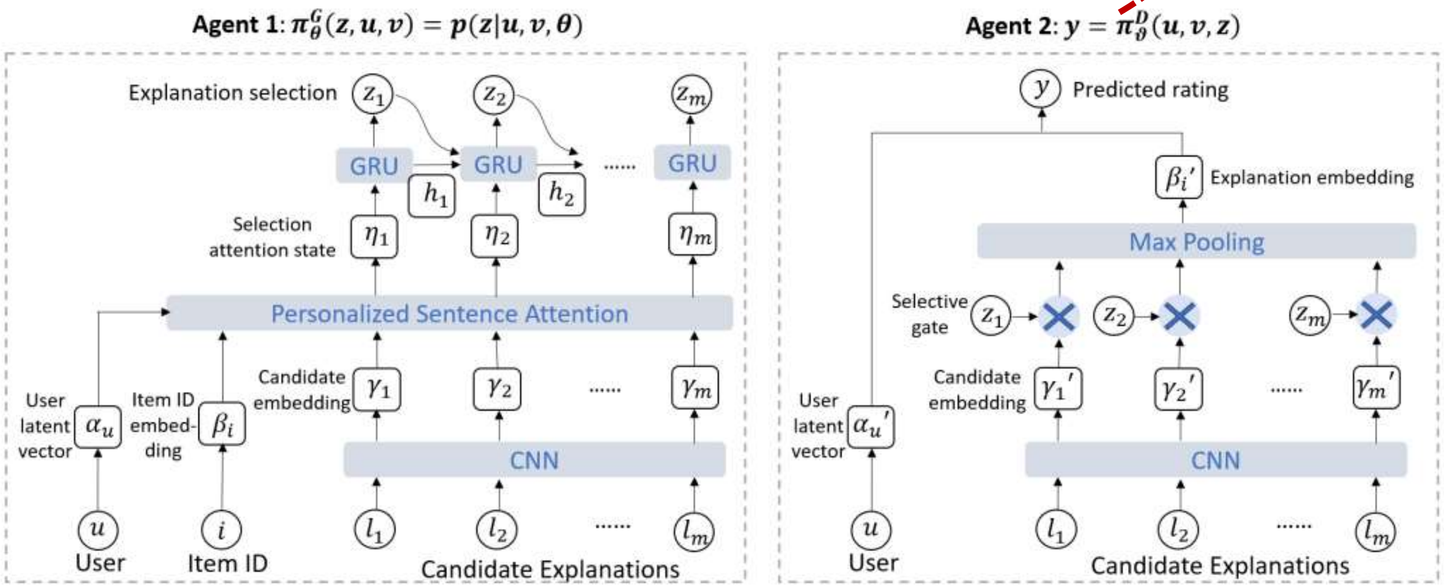
Optimization Goal

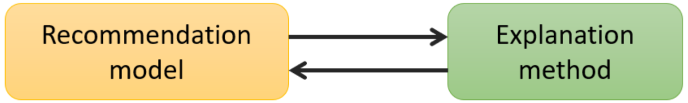
- Maximizing expected reward

$$\arg\max_{\boldsymbol{\theta},\boldsymbol{\vartheta}} \sum_{\boldsymbol{u},\boldsymbol{v}} E_{\boldsymbol{z}\sim p(\cdot|\boldsymbol{u},\boldsymbol{v},\boldsymbol{\theta})} \overbrace{[\mathcal{L}(f(\boldsymbol{u},\boldsymbol{v}), \pi_{\boldsymbol{\vartheta}}^{D}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{z})) + \Omega(\boldsymbol{z})]}^{\text{Reward } r}.$$

Model-explainability

Presentation quality





Optimization Method

- Doubly Stochastic Policy Gradient

Agent 1

$$\begin{aligned}
& \nabla_{\boldsymbol{\theta}} E_{\boldsymbol{z}\sim p(\cdot|\boldsymbol{u},\boldsymbol{v},\boldsymbol{\theta})} \psi_{\boldsymbol{\vartheta}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}) \\
\approx\ & \nabla_{\boldsymbol{\theta}} \sum_{\boldsymbol{z}'} p(\boldsymbol{z}'|\boldsymbol{u},\boldsymbol{v},\boldsymbol{\theta}) \psi_{\boldsymbol{\vartheta}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}') \\
=\ & \sum_{\boldsymbol{z}'} \nabla_{\boldsymbol{\theta}} p(\boldsymbol{z}'|\boldsymbol{u},\boldsymbol{v},\boldsymbol{\theta}) \psi_{\boldsymbol{\vartheta}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}') \\
=\ & \sum_{\boldsymbol{z}'} p(\boldsymbol{z}'|\boldsymbol{u},\boldsymbol{v},\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{z}'|\boldsymbol{u},\boldsymbol{v},\boldsymbol{\theta}) \psi_{\boldsymbol{\vartheta}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}') \\
\approx\ & E_{\boldsymbol{z}\sim p(\cdot|\boldsymbol{u},\boldsymbol{v},\boldsymbol{\theta})} \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{z}|\boldsymbol{u},\boldsymbol{v},\boldsymbol{\theta}) \psi_{\boldsymbol{\vartheta}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}).
\end{aligned}$$

Agent 2

$$\begin{aligned}
& \nabla_{\boldsymbol{\vartheta}} E_{\boldsymbol{z}\sim p(\cdot|\boldsymbol{u},\boldsymbol{v},\boldsymbol{\theta})} \psi_{\boldsymbol{\vartheta}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}) \\
=\ & E_{\boldsymbol{z}\sim p(\cdot|\boldsymbol{u},\boldsymbol{v},\boldsymbol{\theta})} \nabla_{\boldsymbol{\vartheta}} \mathcal{L}(f(\boldsymbol{u},\boldsymbol{v}), \pi_{\boldsymbol{\vartheta}}^{D}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{z})).
\end{aligned}$$

Offline Evaluation

Statistics of the datasets.

	Amazon_Toys_and_Games	Yelp_2018_LasVegas
#users	19,412	23,196
#items	11,924	13,433
#reviews and ratings	167,597	568,454

Explaining different recommendation models trained on the **Amazon_Toys_and_Games** dataset. Here NMF, PMF, SVD++, and CDL are recommendation models to be explained. Larger M_c and M_e indicate better consistency and explainability, respectively.

	M_c					M_e				
	NMF	PMF	SVD++	CDL	GT	NMF	PMF	SVD++	CDL	GT
Random	0.006	0.007	0.035	0.010	0.030	-1.329	-1.046	-0.150	-1.080	-0.981
NARRE	0.012	0.022	0.038	0.043	0.048	-1.271	-1.032	-0.142	-0.967	-0.927
Ours	**0.025**	**0.028**	**0.048**	**0.079**	**0.155**	**-1.234**	**-0.956**	**-0.130**	**-0.956**	**-0.903**

Explaining different recommendation models trained on the **Yelp_2018_LasVegas** dataset. Here NMF, PMF, SVD++, CDL, and GT are recommendation models to be explained. Larger M_c and M_e indicate better consistency and explainability, respectively.

	M_c					M_e				
	NMF	PMF	SVD++	CDL	GT	NMF	PMF	SVD++	CDL	GT
Random	-0.030	-0.030	-0.031	0.012	0.007	-0.478	-0.287	-0.266	-0.517	-1.488
NARRE	-0.015	-0.000	0.018	0.031	0.038	-0.448	-0.266	-0.239	-0.482	-1.424
Ours	**0.018**	**0.037**	**0.041**	**0.227**	**0.168**	**-0.421**	**-0.258**	**-0.232**	**-0.460**	**-1.380**

M_c: presentation quality M_e: explainability

Offline Evaluation

Statistics of the datasets.

	Amazon_Toys_and_Games	Yelp_2018_LasVegas
#users	19,412	23,196
#items	11,924	13,433
#reviews and ratings	167,597	568,454

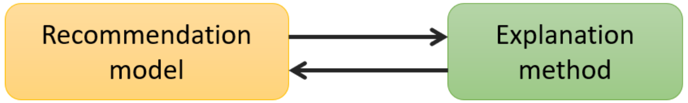
Comparison of M_c and M_e at different explanation lengths (the **Amazon_Toys_and_Games** dataset).

	M_c					M_e				
	$z^* = 1$	$z^* = 2$	$z^* = 3$	$z^* = 4$	$z^* = 5$	$z^* = 1$	$z^* = 2$	$z^* = 3$	$z^* = 4$	$z^* = 5$
Random	0.030	0.013	0.029	0.037	0.037	-0.981	-0.991	-0.973	-0.962	-0.995
NARRE	0.048	0.064	0.089	0.110	0.133	-0.927	-0.919	-0.910	-0.911	-0.906
Ours	**0.155**	**0.142**	**0.140**	**0.160**	**0.161**	**-0.903**	**-0.901**	**-0.898**	**-0.898**	**-0.894**

Comparison of M_c and M_e at different explanation lengths (the **Yelp_2018_LasVegas** dataset).

	M_c					M_e				
	$z^* = 1$	$z^* = 2$	$z^* = 3$	$z^* = 4$	$z^* = 5$	$z^* = 1$	$z^* = 2$	$z^* = 3$	$z^* = 4$	$z^* = 5$
Random	0.007	0.011	0.012	0.032	0.030	-1.488	-1.405	-1.403	-1.400	-1.406
NARRE	0.038	0.035	0.044	0.057	0.054	-1.424	-1.390	-1.377	-1.378	-1.372
Ours	**0.168**	**0.172**	**0.183**	**0.188**	**0.160**	**-1.380**	**-1.377**	**-1.370**	**-1.366**	**-1.353**

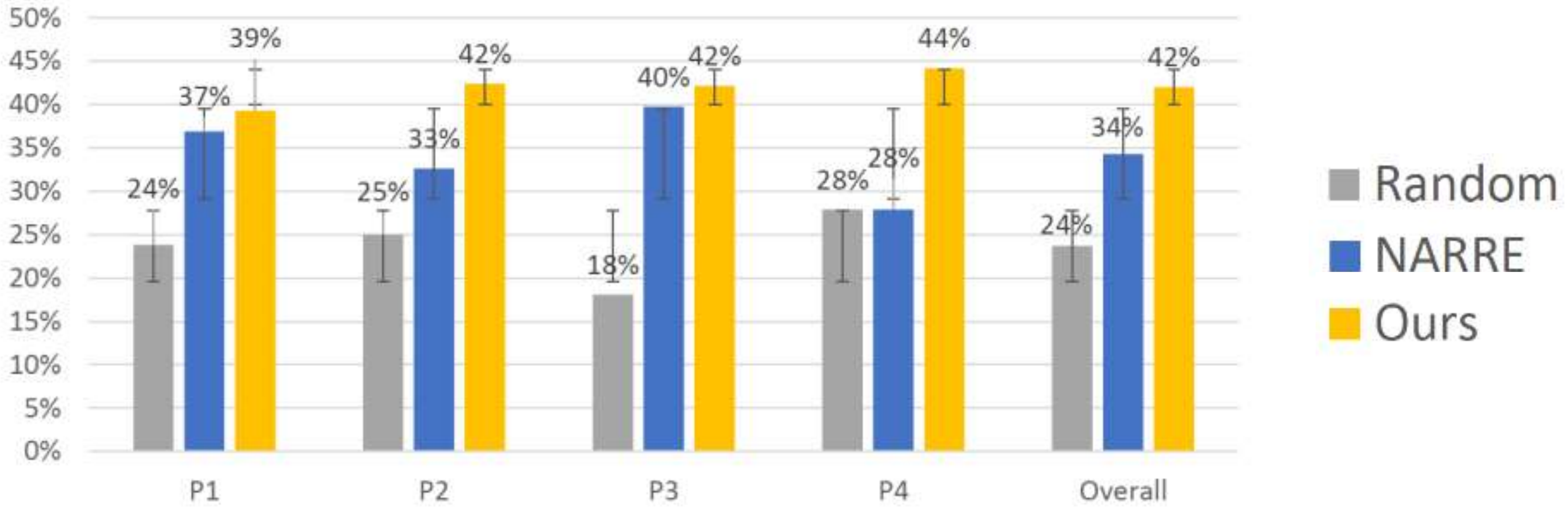
M_c: presentation quality M_e: explainability z^*:explanation length



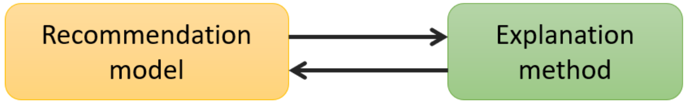
Evaluation with Human Subjects

- Ask the participants to choose the explanations that are **most useful** in helping them decide whether they will go to the restaurants

Frequency of a method (Random, NARRE, or ours) being considered the most useful. We show the results of individual participants (P1 to P4) as well as the overall summarization.



Reinforcement learning framework [ICDM2018]



Evaluation with Human Subjects

Frequent words in reviews:

P3 chicken, buffet, portions, sushi, beef

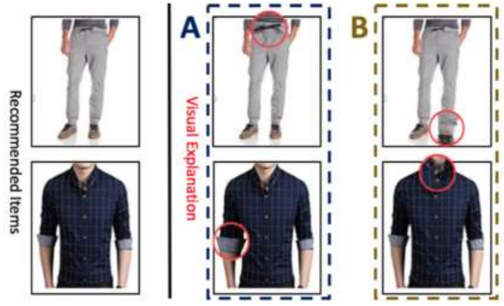
P4 service, pizza, server, table, clean

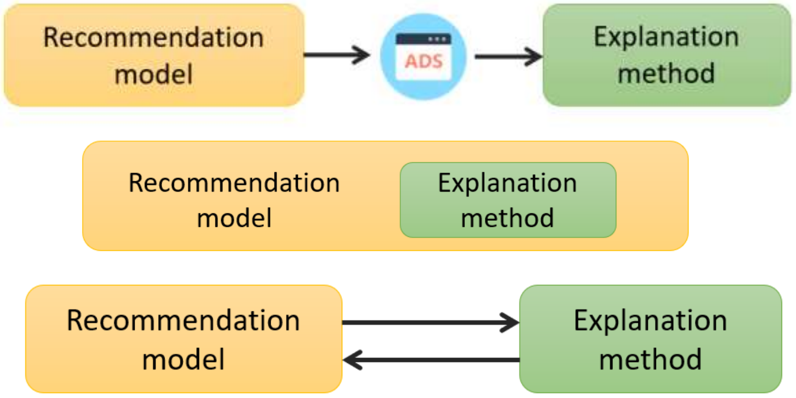
	NARRE	Ours - P3	Ours - P4
Item 1	By the way, try to park at the side of gold coast farthest from the rio if you want to have a shorter walk, which is healthier than it sounds due to less secondhand smoke exposure.	The chicken's feet was tasty, so were the har gow.	In the past we had trouble communicating with the staff because they usually speak in their own language , this last time though it seems they have hired more English speaking staff and it was considerably easier to order.
Item 2	If you needa fajita, your search should end here.	They came with red & green peppers and onions. First, I thought the salsa was delicious, and i appreciated it was actually spicy versus the mild you typically receive.	Overall, the service throughout our meal was swift & friendly.
Item 3	Unfortunately, after living in the city for a few years and trying a lot of wonderful food that this city has to offer, we returned for a visit and I was less than impressed.	It was the perfect burger, cheesy with just the right amount of dressing and chips!	At least put the stuff in a fancy container?

■ Words related to food ■ Words related to services

Conclusion

- Definition and goals
- Forms of explanations
- Explainable recommendation pipelines





	Model explainability	Presentation quality	Model agnostic
Post-hoc	x	√x	√
Embedded	√	x	x
Wrapper	√	√	√

Thanks!