Non-IID Recommender Systems in Practice with Modern AI Techniques

Liang Hu, Longbing Cao and Songlei Jian
University of Technology Sydney

www.datasciences.org

Tutorial Website

- https://sites.google.com/view/lianghu/home/tutorials/pakdd2018

Goal

In this tutorial, we aim to provide a comprehensive understanding of how to apply the state-of-the-art ***AI (especially machine learning)*** techniques to build ***non-IID recommender systems*** by modeling ***heterogeneities*** and ***couplings*** of ***users, items,*** and ***between users and items***.

Agenda

- Overview of Non-IIDness in RS
  - 20 mins
- Data representation in RSs with ML approach
  - 30 mins
- Non-IID RS on modeling heterogeneities and couplings over users
  - 40 mins
- Break
  - 30 mins
- No-IID RS on modeling heterogeneities and couplings over items
  - 45 mins
- Non-IID RS on modeling heterogeneities and couplings over implicit user-item interactions
  - 50 mins
- Non-IID RS in practice
  - 25 mins

Outline

Non-IIDness in RS

Overview
Challenges
Non-IIDness in RS

Data Representation

Attributes

- Text
- Rating table
- Image
- Sequence

Non-IIDness on users

- Social RS
- Group RS

Non-IIDness on items

- Cross-domain RS
- Session-based RS

Non-IIDness on implicit user-item interactions

- Context-aware RS
- Multi-objective RS
- Attraction RS

Non-IID RS in practice

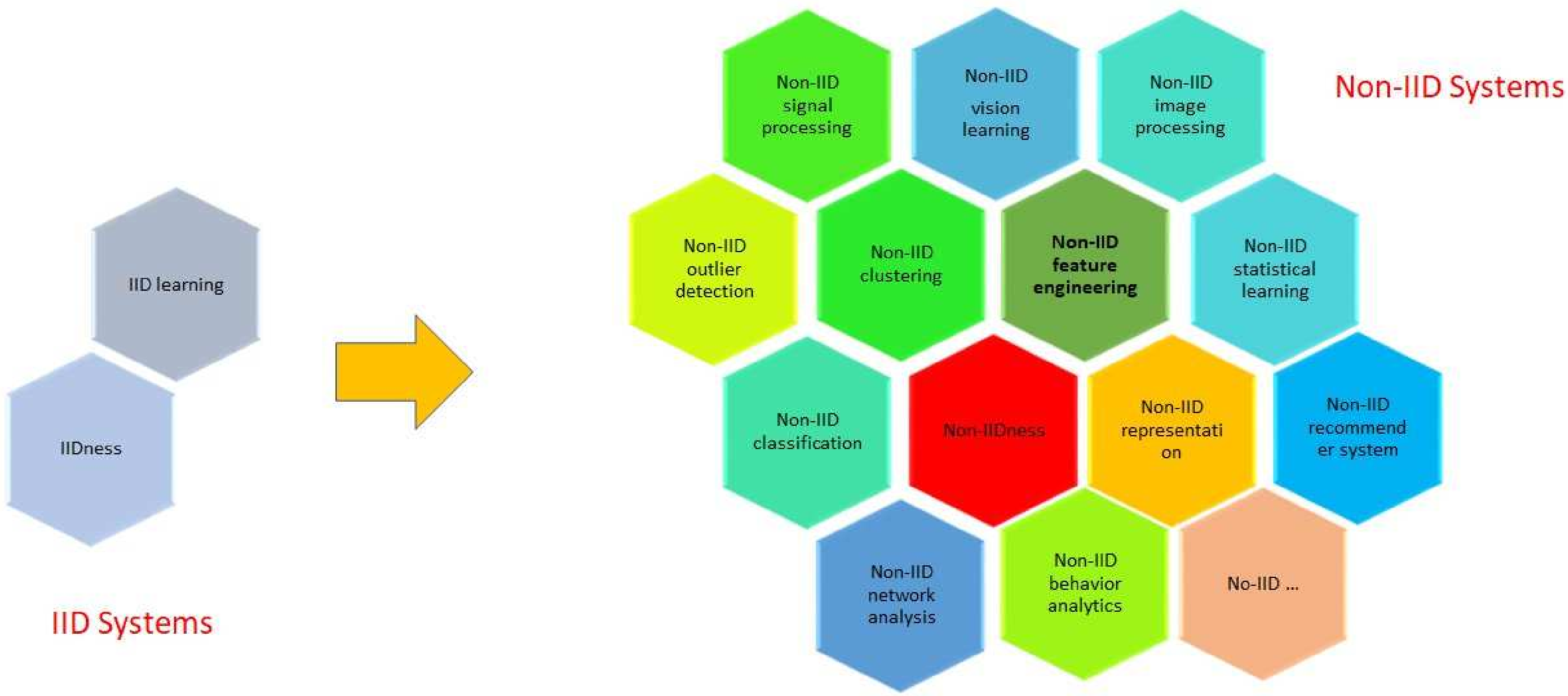
Non-IIDness in RS

- Overview of recommender systems
- Challenges of recommender systems
- Non-IIDness in recommender systems

Non-IID learning

Non-IID learning, KDD2017 tutorial

www.datasciences.org



What are Recommender Systems

- Recommender systems (push information) are the evolution of information retrieval systems (pull information).

Information Age → Recommendation Age

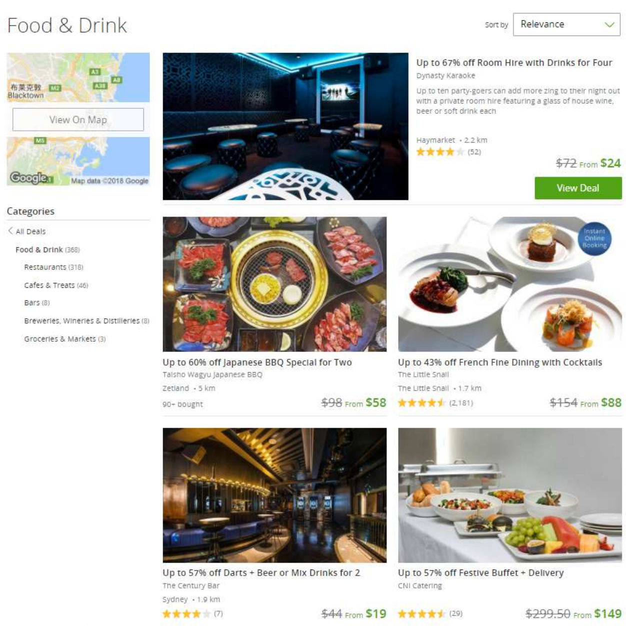
Pull mode (IR):
Query → Matched Results → Manual Filtering

Push mode (RS):
Potential Requirement → Machine Filtering → Recommendation

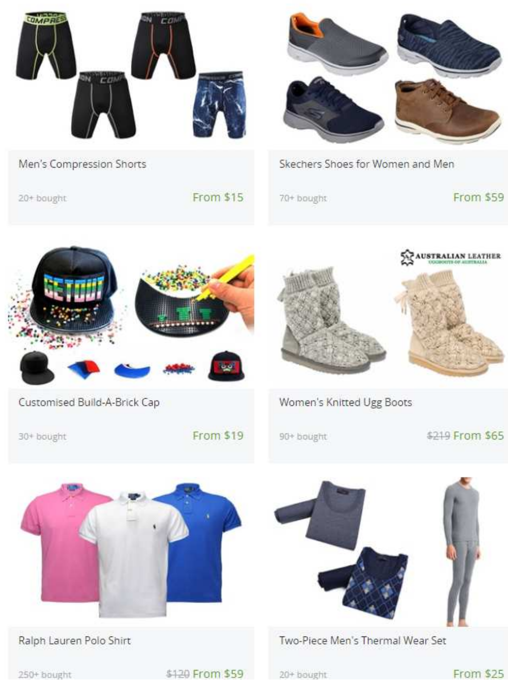
Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*

Recommender Systems have occupied our life

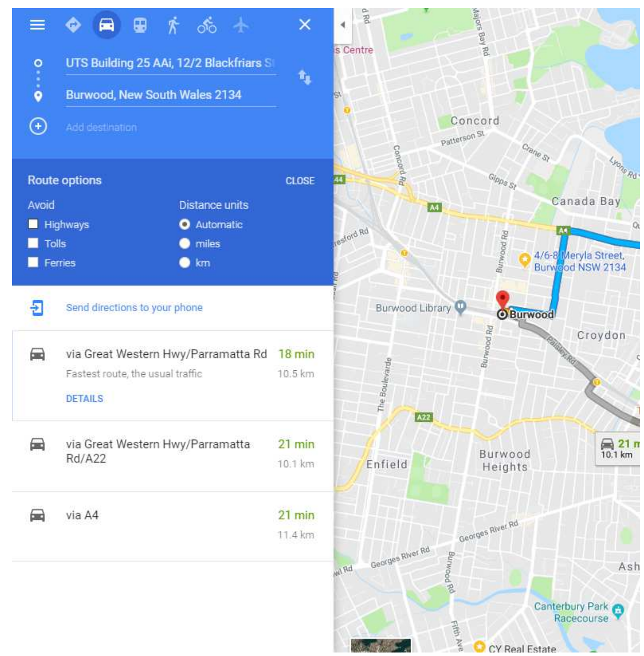
What to eat



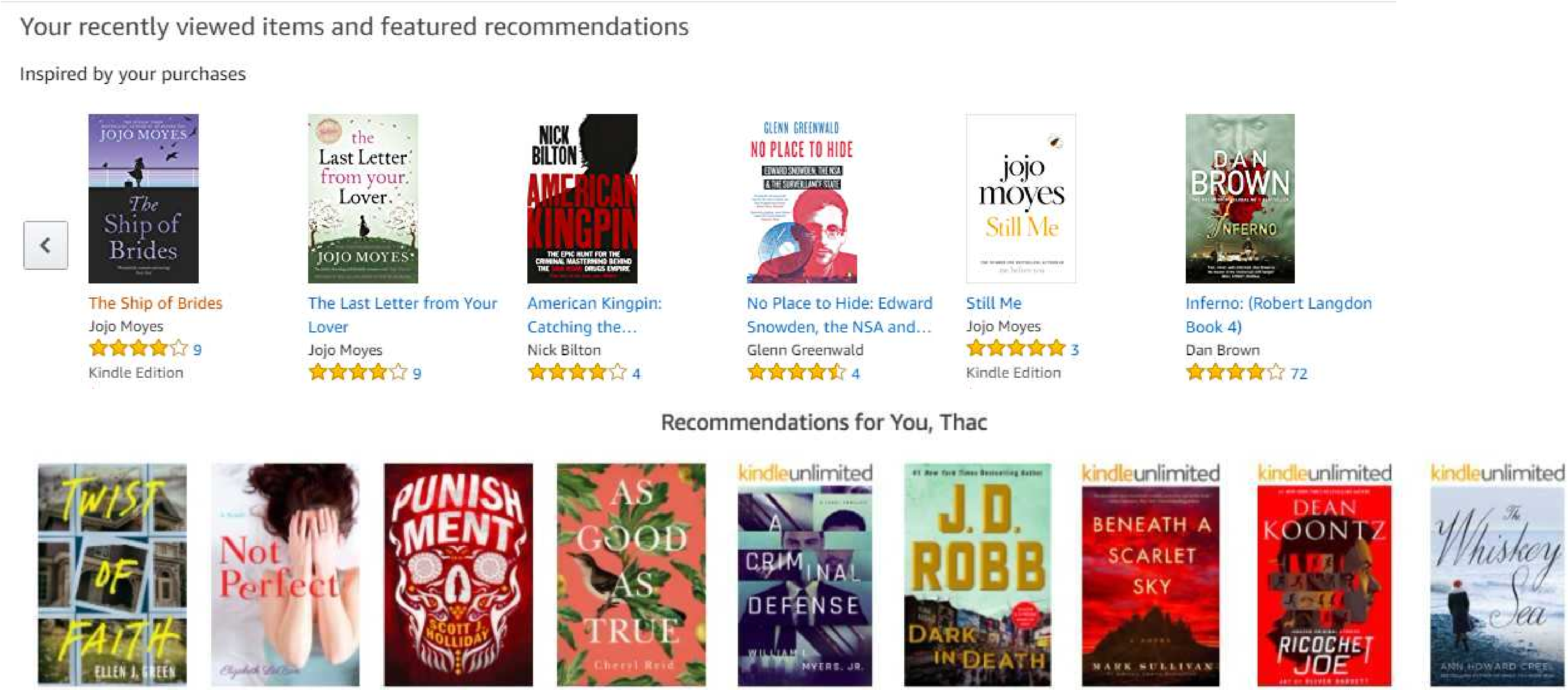
What to buy



Where to go



Personalized Recommendation



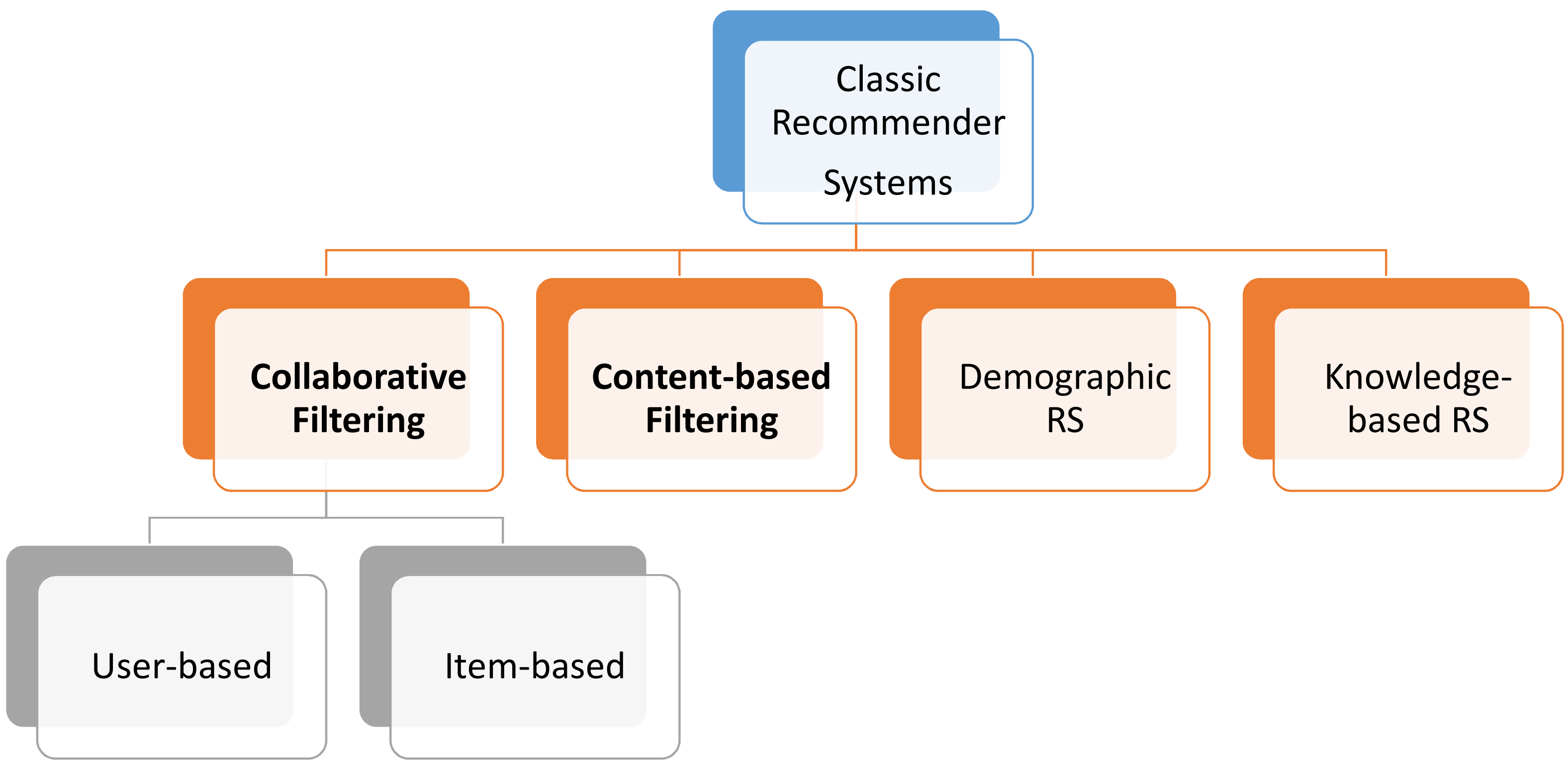
Problems

Amazon

Recommendation problems:

- Duplicated
- Irrelevant
- Missing
- Falsified
- …

Classic Recommender Systems

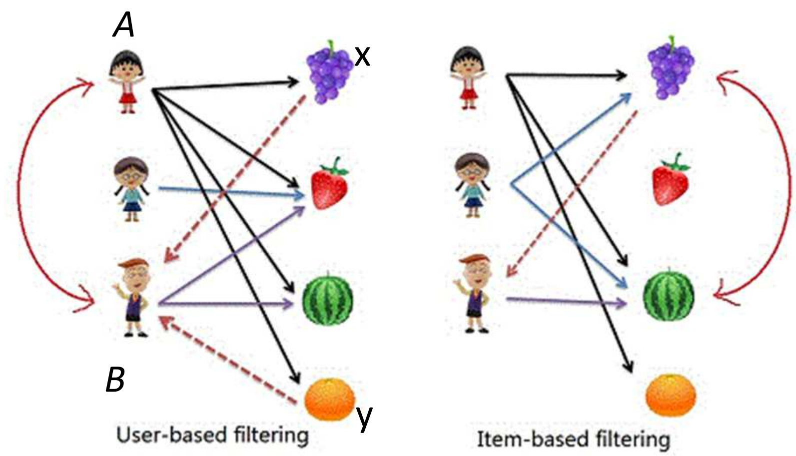


Collaborative Filtering (CF)

- Intuition (user-based filtering): If user ***A*** related to user ***B*** and ***A*** bought x and y, then ***B*** bought x tends to buy y.

- Famous examples (item-based filtering): Amazon.com's recommender system

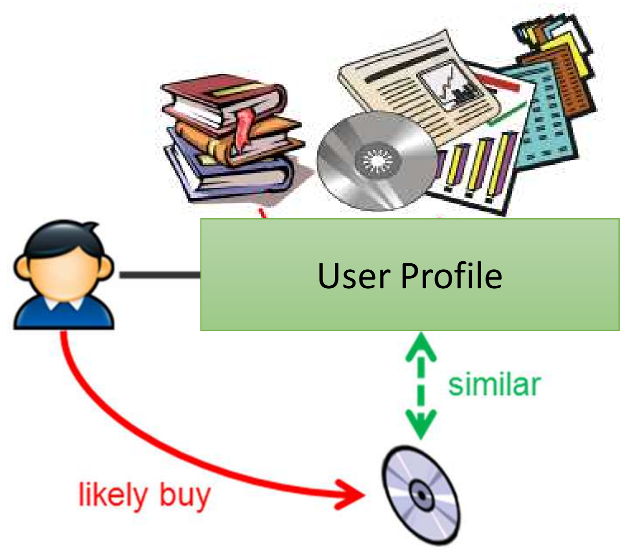
- Facebook, MySpace, LinkedIn use collaborative filtering to recommend new friends, groups, and other social connections.



Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). *Item-based collaborative filtering recommendation algorithms*. Paper presented at the Proceedings of the 10th international conference on World Wide Web, Hong Kong.

Content-based Filtering (CBF)

- CBF is based on the features of items
  - Attributes of items
  - Description of items
  - Text of an article

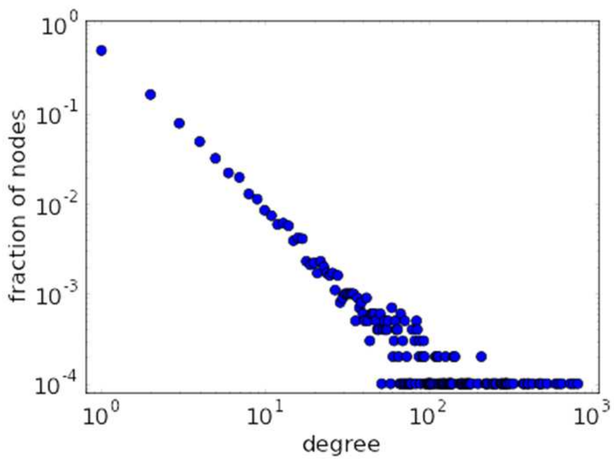


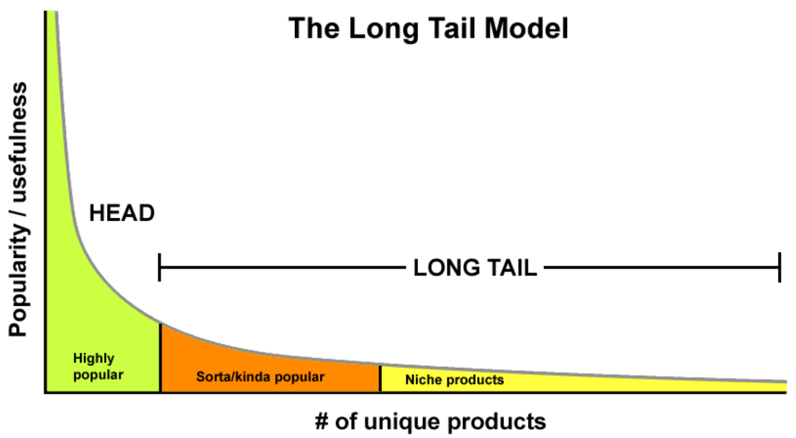
- User profile is built with the features of historical items

- Recommend items according to user profile

Data Characteristics in Recommender Systems

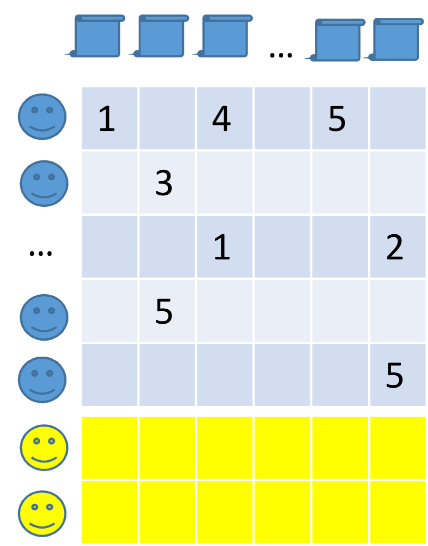
- Power law or Long tail distribution
  - Data associated with the majority of users are insufficient and even absent in real world.
  - In most recommender systems, the majority of users/items only associated with very few data while only the minority of users/items have sufficient data





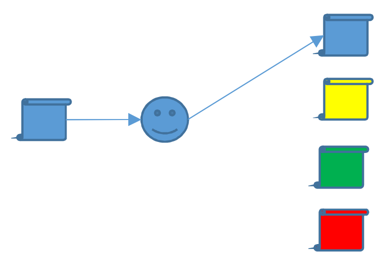
Challenges in Collaborative Filtering

- Data Sparsity
  - In real-world recommender systems, the user-item matrix is very sparse.
- Cold Start
  - When new users or new items are added, the system cannot recommend to these users and these items.
- Scalability
  - There are millions of users and products in real systems.
    - Large amount of computation
    - Large storage



Challenges in Content-based Filtering

- Limited Content Analysis
  - System has a limited amount of information on its users or the content of its items.
- Over-specialization
  - The system can only recommend items that highly similar with user's profile, the user is limited to be recommended items similar to those already rated.



Question: what's the main cause of these challenges?

- Data Sparsity
- Cold Start
- Limited Content Analysis
- Over-specialization

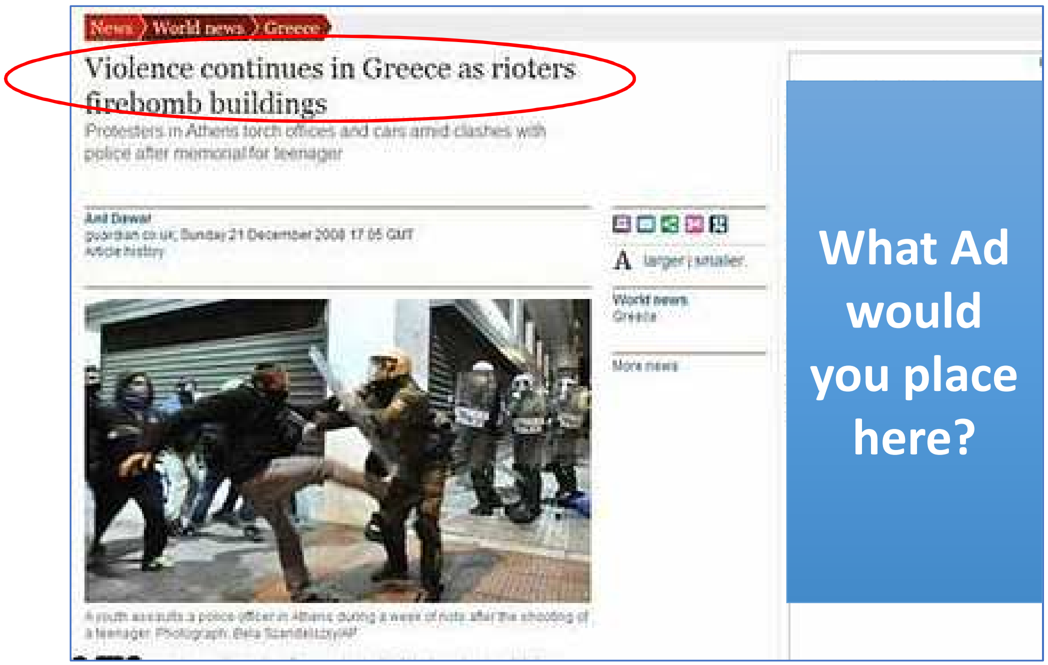
Insufficient and simple data

Prospects: modeling RS with more complex data

- Built on More Complex Data
  - Multiple data types
    - Ratings
    - Images
    - Text
  - Multisource
    - Multiple domains
    - Multiple systems
  - Social data
    - Acquire data from user social media
  - Multiple criteria
    - Multi-objectives: accuracy, novelty…

Data complexity

Data complexity challenges existing theories and systems



Irrelevant and Damaging to Brand

Non-IIDness in Complex Data

- Heterogeneity:
  - Data types, attributes, sources, aspects, …
  - Formats, structures, distributions, relations, …
  - Learning outcomes

Not identically distributed.

- Coupling relationships:
  - Within and between values, attributes, objects, sources, aspects, …
  - Structures, distributions, relations, …
  - Methods, models, …
  - Outcomes, impact, …

Not independent distributed.

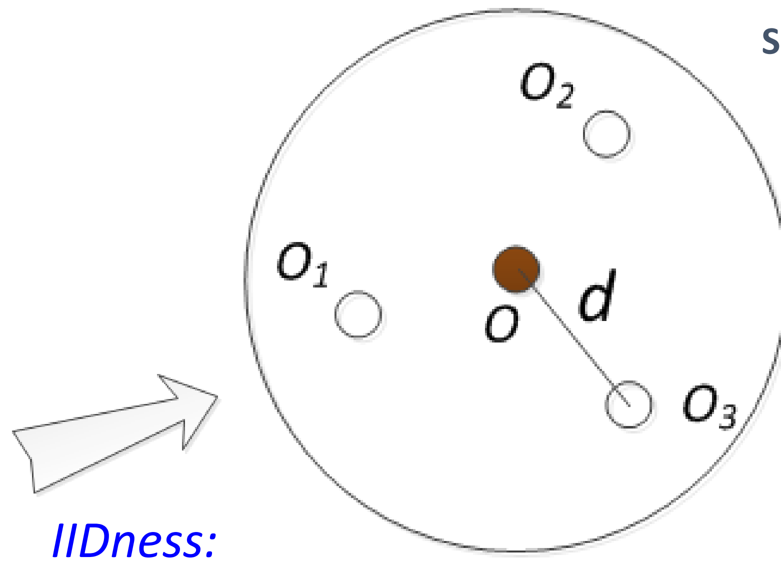
Non-IIDness

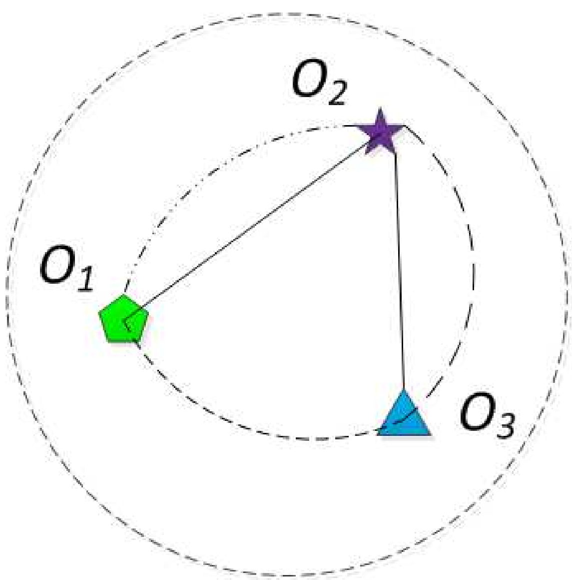
Longbing Cao. *Non-IIDness Learning in Behavioral and Social Data*, The Computer Journal, 57(9): 1358-1370 (2014).

Classic Assumption – IIDness & IID Learning

IID learning:
Dominates classic analytics, AI/KDD/ML/CVPR/Statistics research & development

Solution





Problem

IIDness:

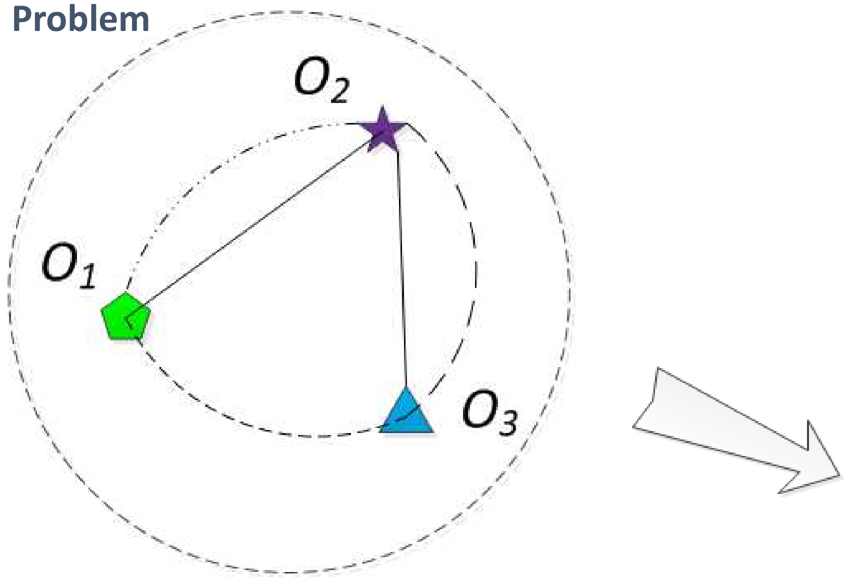
Independence +
Identical Distribution

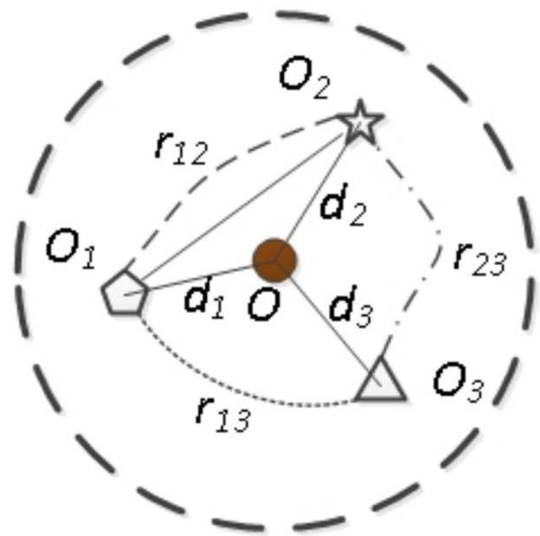
O_1, O_2, O_3 are iid

$d_3 = ||O_3 - O||$

A Foundational Issue: Non-IID Learning

Problem





O_1, O_2, O_3 share different distributions

$d_3 = ||O_3 - O||$

$= || O_3(r_{13},r_{23}) - O(d_1,d_2) ||$

A Systematic View of Recommendation

NS	SS	AS	CS	**Subcategory**
NC	SC	AC	CC	**Category**
NP	SP	AP	CP	**Price**
Name	**Sex**	**Age**	**City**	

(D). Implicit user-item interactions

Subcategory	C1.6	C2.2	C2.3
Category	C1	C2	C2
Price	100	800	1200
	i1	**i2**	**i3**

(C). Item properties

Name	**Sex**	**Age**	**City**	
John	M	45	Sydney	**u1**
Cindy	F	42	Sydney	**u2**
Julie	F	20	Sydney	**u3**

(B). User demographics

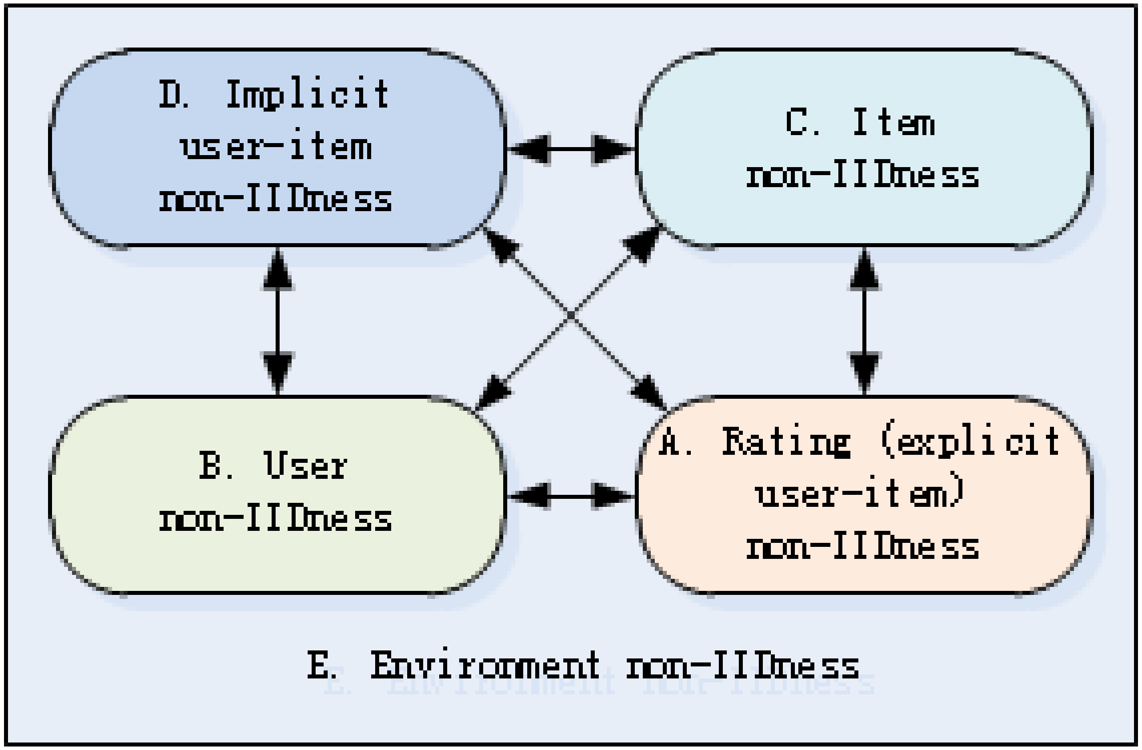
	i1	**i2**	**i3**
u1	5	3	4
u2	4	5	4
u3	4	5	5

(A). Ratings

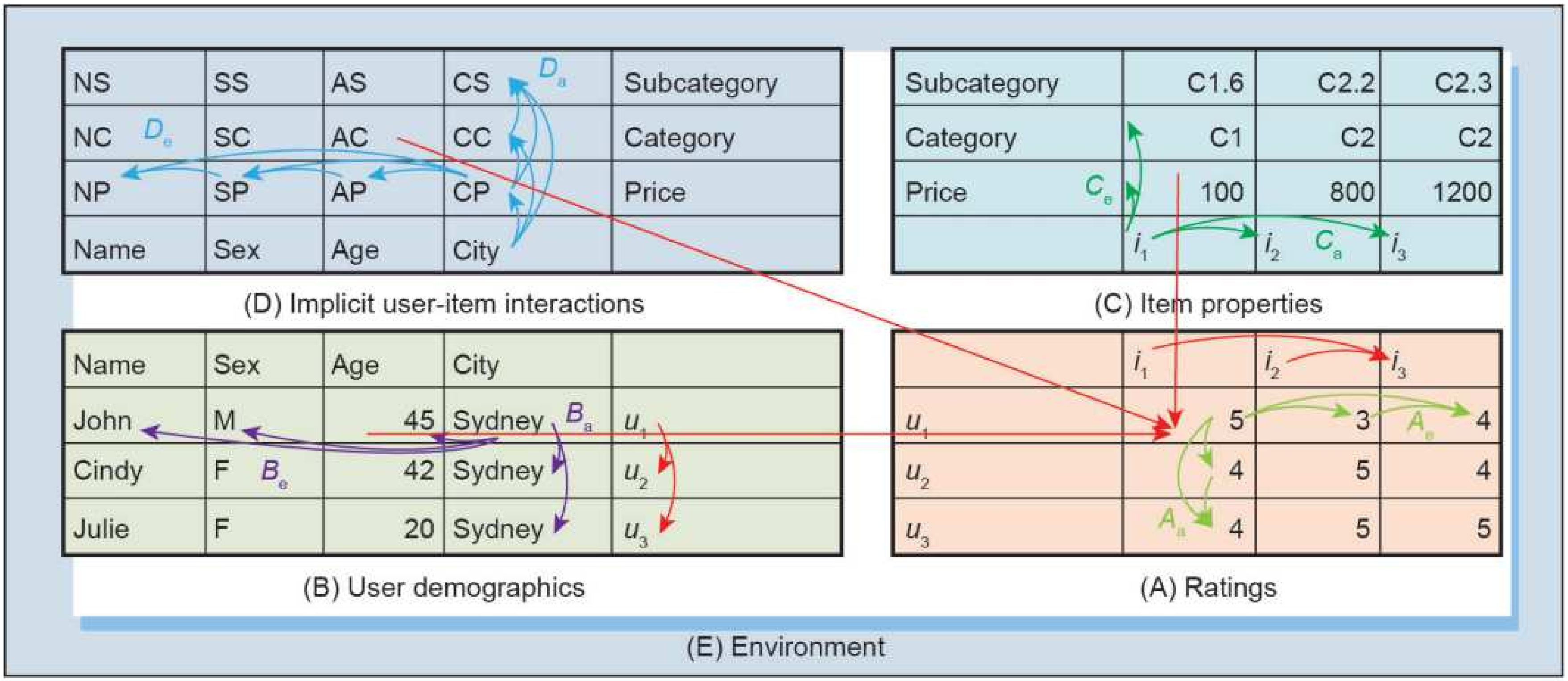
(E). Environment

Longbing Cao. *Non-IID Recommender Systems: A Review and Framework of Recommendation Paradigm Shifting*. Engineering, 2: 212-224, 2016.

Non-IIDness in Recommendation

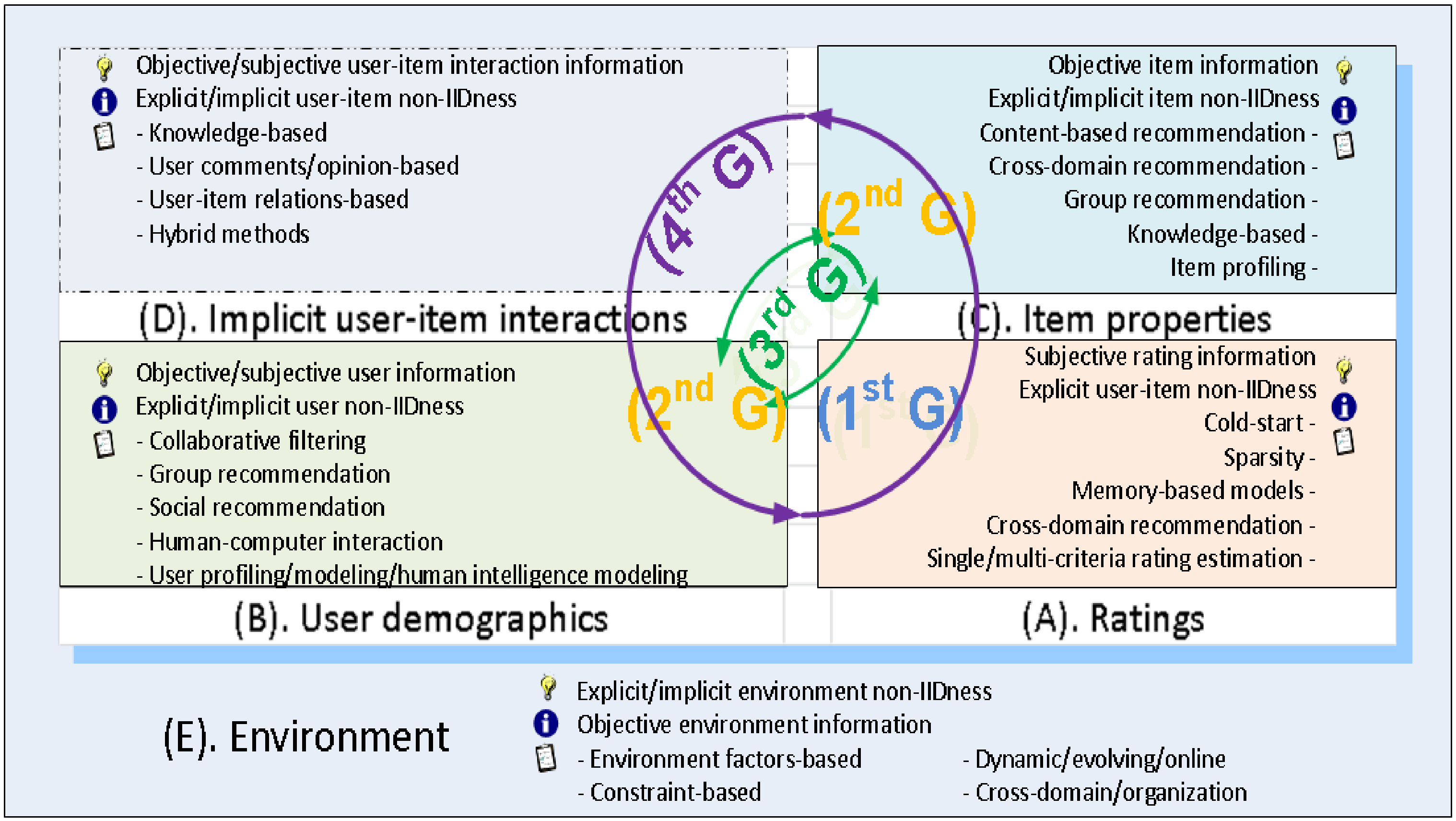


Non-IIDness in Recommendation



Cao, L. (2016). Non-IID Recommender Systems: A Review and Framework of Recommendation Paradigm Shifting. *Engineering*, *2*(2), 212-224.

Four-generation Recommendation Research



Modeling Non-IIDness for Advanced RS

- Heterogeneity modeling:
  - The heterogeneity over users
    social RS, group-based RS
  - The heterogeneity over items
    cross-domain RS, multi-modal RS
  - The heterogeneity of data types
    multi-modal RS
  - The heterogeneity of domains
    cross-domain RS
  - The heterogeneity of objectives
    multi-objective RS

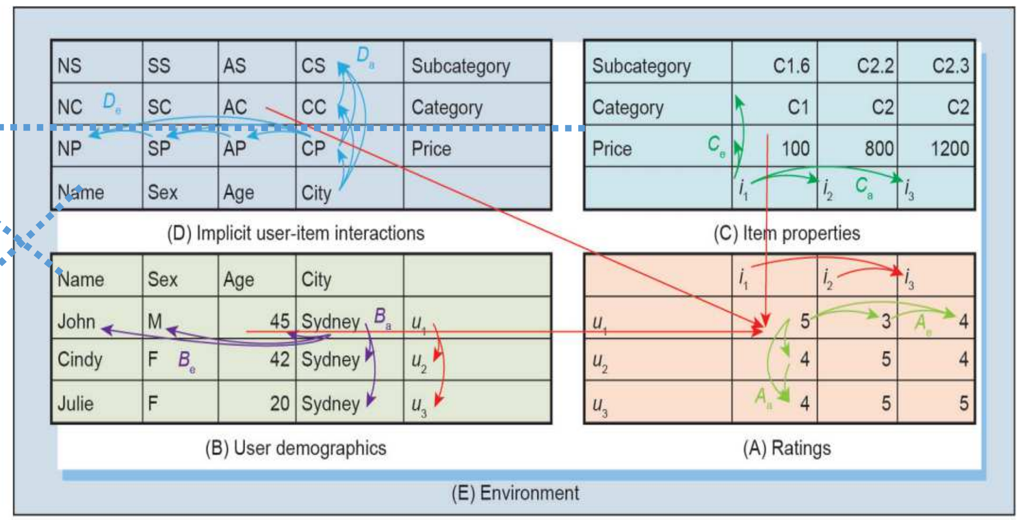
- Coupling modeling :
  - The coupling between users
    social RS, group-based RS
  - The coupling between items
    session-based RS, cross-domain RS
  - The coupling between data types
    multi-modal RS
  - The coupling between domains
    cross-domain RS
  - The coupling between objectives
    multi-objective RS

Modeling the Non-IIDness in RS

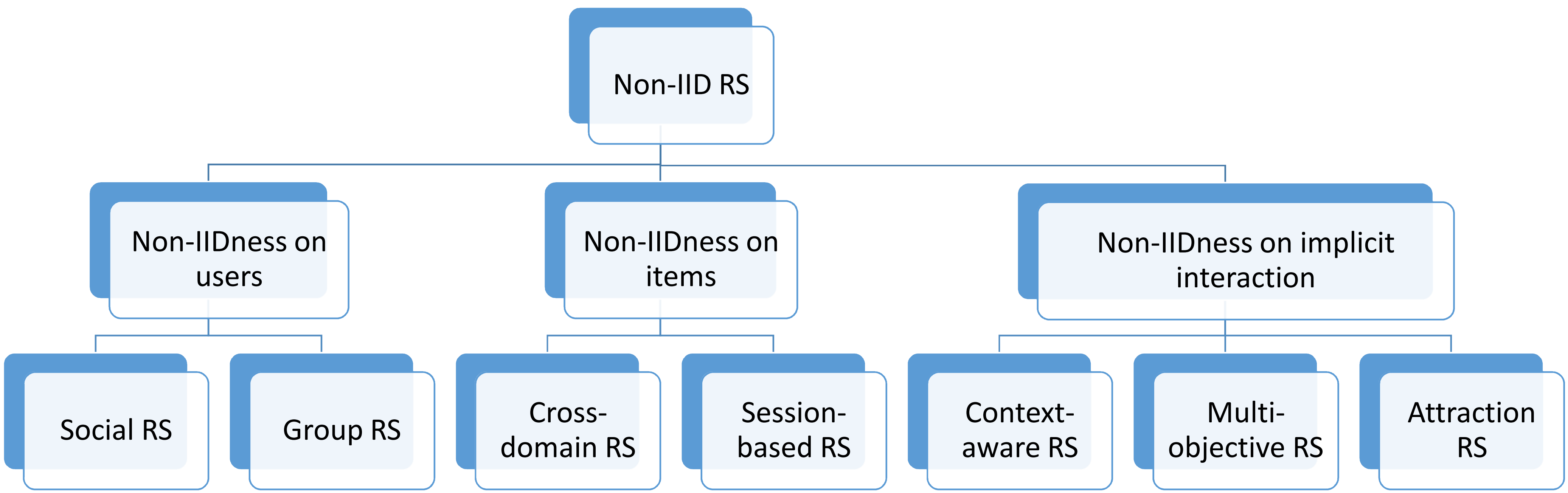
A. Non-IIDness on users:
   - Social RS: user mutual influence
   - Group RS: group joint decision

B. Non-IIDness on items:
   - Cross-domain RS: domain coupling
   - Session-based RS: sequential coupling

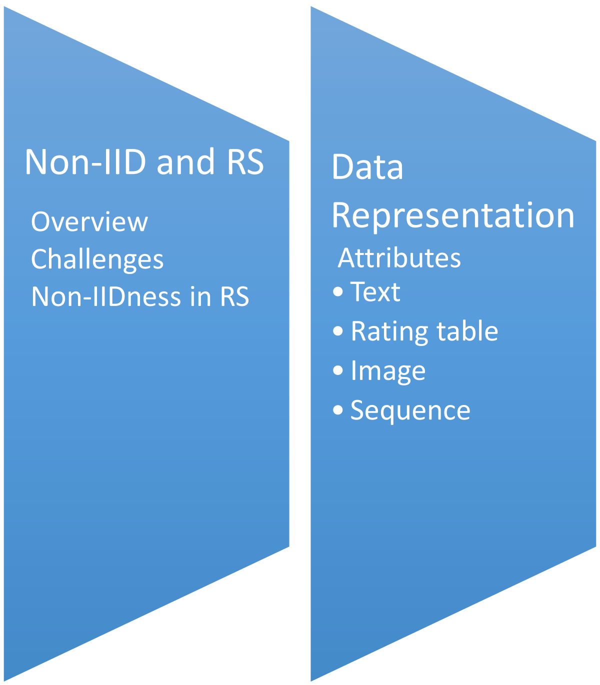
C. Non-IIDness on implicit interaction:
   - Context-aware RS: contextual dependency
   - Multi-objective RS: multi-aspect ratings
   - Attraction RS: subjective attention



Non-IID RS covered in this tutorial

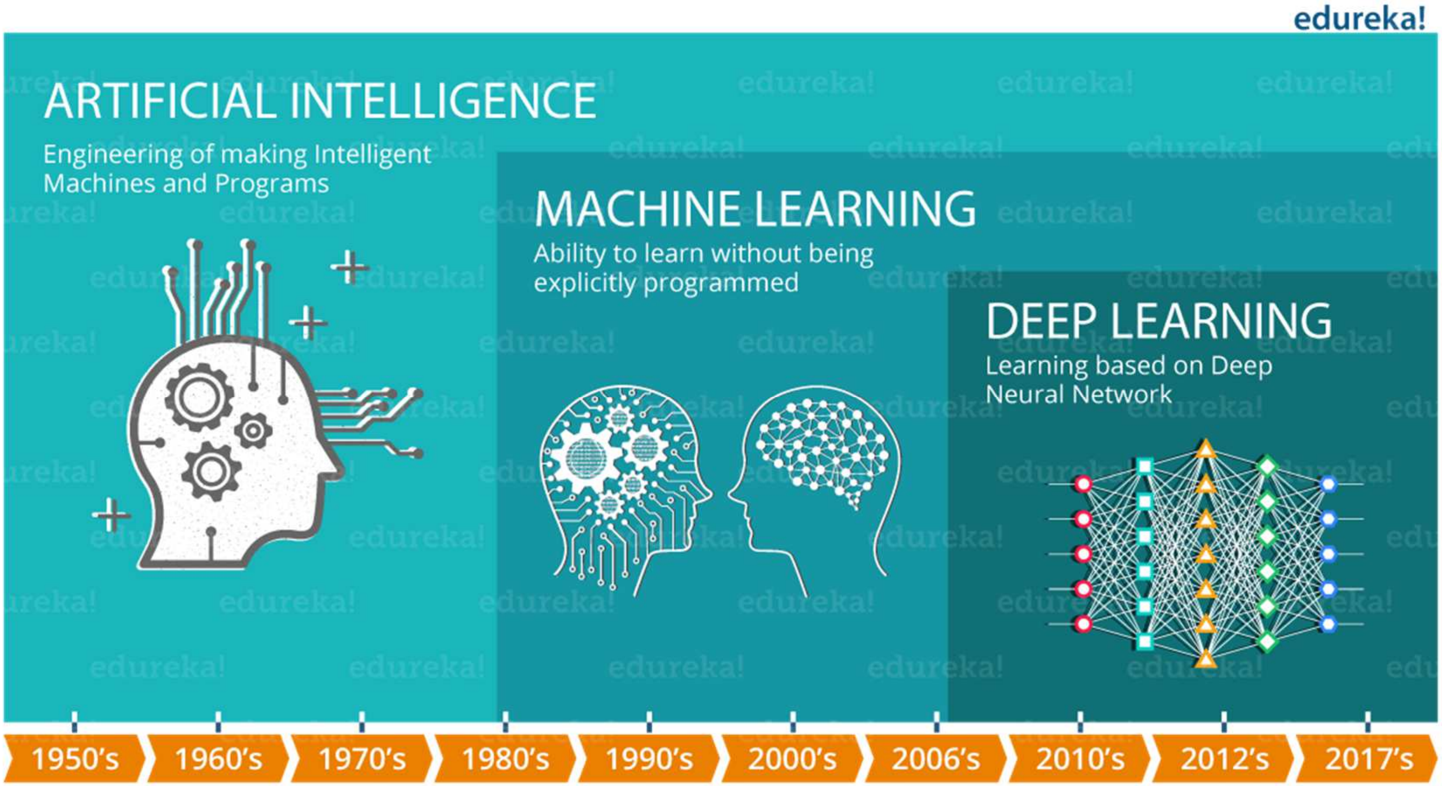


Data Representation



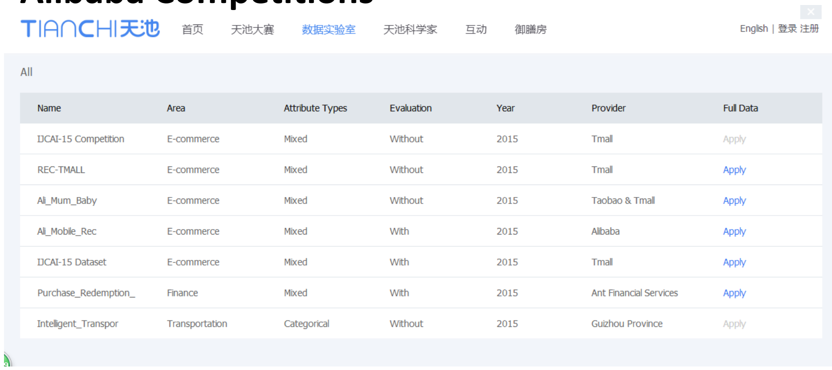
- AI-related preliminaries: from data representation perspective
  - Representing attributes
  - Representing text
  - Representing rating table
  - Representing image
  - Representing sequence

AI, Machine Learning and Deep Learning



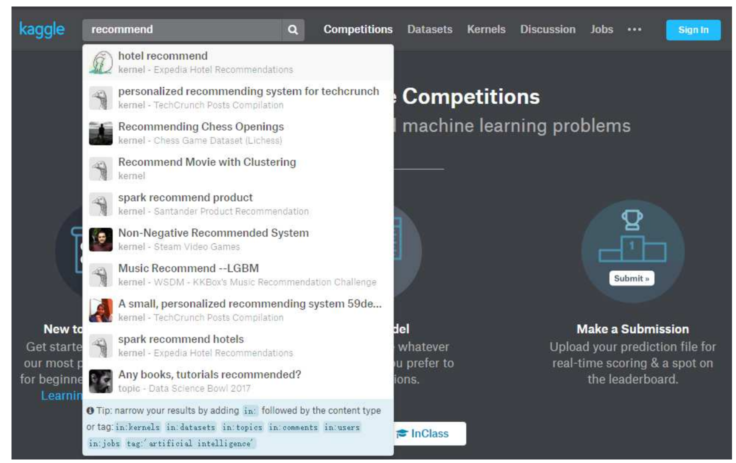
Machine Learning Methods Dominate RS Competitions

Alibaba Competitions


TIANCHI天池 首页 天池大赛 数据实验室 天池科学家 互动 御膳房 English | 登录 注册

All

Name	Area	Attribute Types	Evaluation	Year	Provider	Full Data
IJCAI-15 Competition	E-commerce	Mixed	Without	2015	Tmall	Apply
REC-TMALL	E-commerce	Mixed	Without	2015	Tmall	Apply
Ali_Mum_Baby	E-commerce	Mixed	Without	2015	Taobao & Tmall	Apply
Ali_Mobile_Rec	E-commerce	Mixed	With	2015	Alibaba	Apply
IJCAI-15 Dataset	E-commerce	Mixed	With	2015	Tmall	Apply
Purchase_Redemption_	Finance	Mixed	With	2015	Ant Financial Services	Apply
Intelligent_Transpor	Transportation	Categorical	Without	2015	Guizhou Province	Apply



RecSys 2018 – Challenge – RecSys
https://recsys.acm.org/recsys18/challenge/ ▾ 翻译此页
RecSys Challenge 2018. The RecSys Challenge 2018 will be organized by Spotify, The University of Massachusetts, Amherst, and Johannes Kepler University, Linz. Spotify is an online music streaming service with over 140 million active users and over 30 million tracks. One of its popular features is the ability to create ...

RecSys 2017 – Challenge – RecSys
https://recsys.acm.org/recsys17/challenge/ ▾ 翻译此页
RecSys Challenge 2017. The RecSys Challenge 2017 is organized by XING, Politecnico Milano and Free University of Bozen-Bolzano. XING is a social network for business. People use XING, for example, to find a job and recruiters use XING to find the right candidate for a job. At the moment, XING has more than 18 ...

RecSys 2016 – Challenge – RecSys
https://recsys.acm.org/recsys16/challenge/ ▾ 翻译此页
In this year's edition of the RecSys Challenge, the task is: given a XING user, predict those job postings that a user will click on. Submitted solutions will be evaluated offline and online. A detailed description of the challenge can be found on the website of the RecSys Challenge 2016. Accepted contributions will be ...

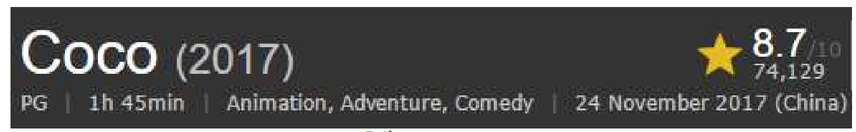
Machine Learning: Tell Truth from Data
Recommender Systems: Recommend Truth from Data

Data with Machine Learning Methods

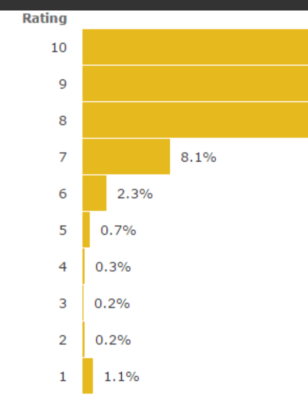
- Attributes
  - Regression
  - Clustering
  - Factor Analysis
- Labels
  - Classification
  - Learning to Rank
- Images, Videos
  - Computer Vision Approach

Data with Recommender Systems

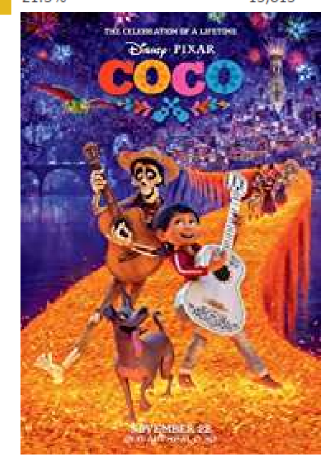
- User/Item features



- Ratings



- Item pictures



Machine Learning: Tell Truth from Data (Cont.)
Recommender Systems: Recommend Truth from Data

Data with Machine Learning Methods

- Text
  - Natural Language Processing (NLP)
  - Sentiment Analysis
- Sequence
  - Time Series Analysis
- Network
  - Link Prediction, Network Embedding

Data with Recommender Systems

- Reviews

- Transaction

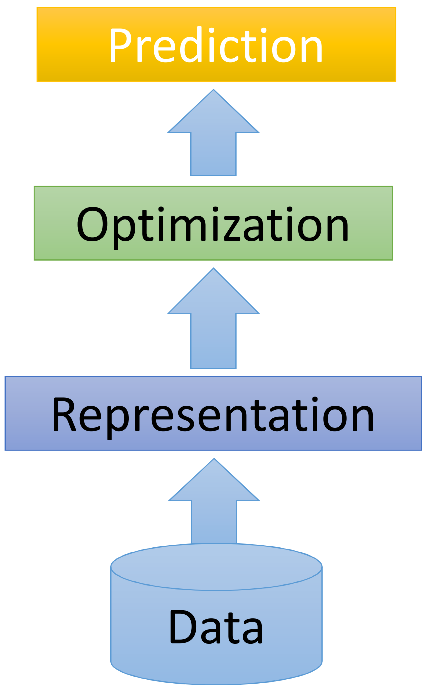
- User/Item Network

In one word

- ***Data*** is the matchmaker to bring advanced machine learning methods to recommender systems

Representation is the foundation of machine learning

- Machine learning concerns the construction and study of systems that can learn from data.
- Machine learning focuses on prediction, based on *known* properties learned from training data
- The *core* of machine learning deals with *representation* and generalization.
- Good representation are essential for successful ML: 90% of effort



Data Representation

Non-IID and RS

Overview
Challenges
Non-IIDness in RS

Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

- AI-related preliminaries: from data representation perspective
  - Representing attributes(User/Item features)
  - Representing text (reviews, comments)
  - Representing rating table
  - Representing image (User/Item pictures)
  - Representing sequence (Transactions)

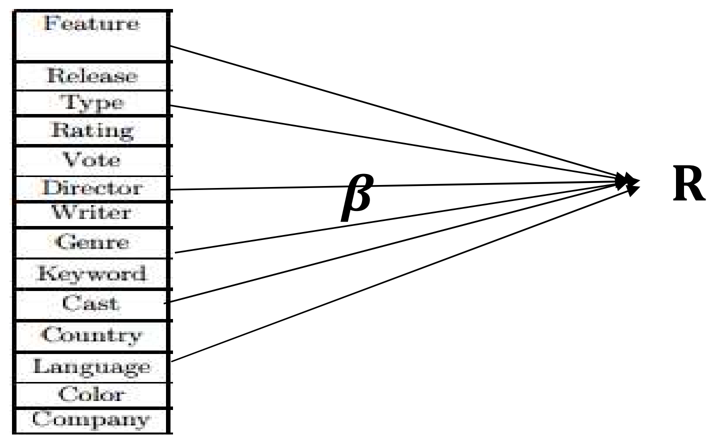
Representing attributes

- Attributes are most commonly used in RS
  - User feature or Item feature
  - Categorical feature or Numerical feature
- Modeling the relationship between a target, e.g. rating, and given item attributes.

Shallow model: rating regression

- $\boldsymbol{\beta}$ is the parameters used to model the importance of each feature.
- $\mathbf{R}$ is the ratings given by a user
- Disadvantage: fails to capture the coupling between features

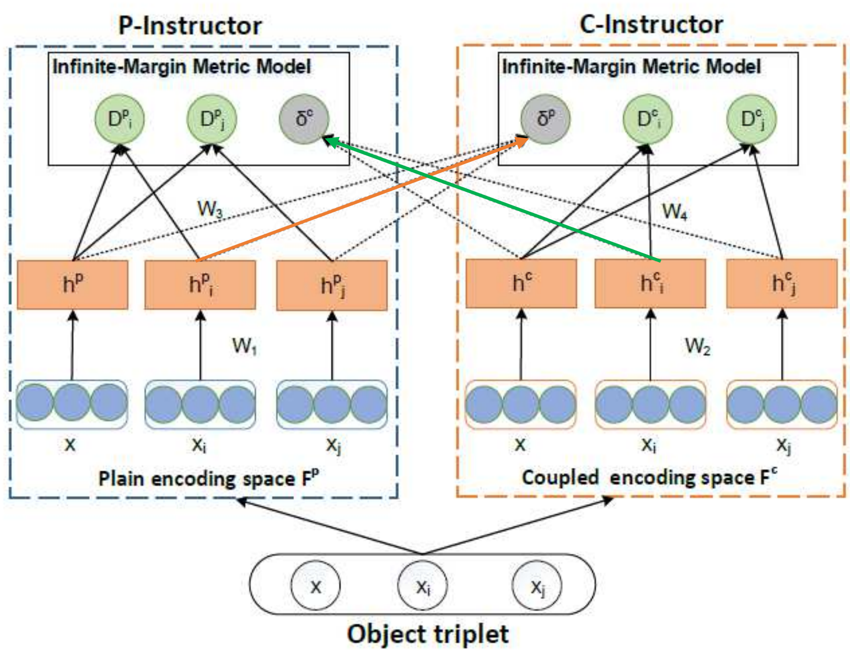
$$\mathbf{R} = \mathbf{f}(\boldsymbol{\beta}|\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}$$



Statistical estimation and inference focuses on $\boldsymbol{\beta}$

Metric-based Auto-Instructor for Learning Mixed Data Representation

- Representing ***categorical feature*** and ***numerical feature*** in one unified feature space
  - At the feature level: capture the heterogeneous coupling between features
  - At the object level: express the discrimination and margins between objects
- P-Instructor provides supervision for C-Instructor learning, and vice versa, to reach consensus mixed representation.



Songlei Jian, Liang Hu, Longbing Cao & Kai Lu. AAAI 2018. Metric-based Auto-Instructor for Learning Mixed Data Representation

Data Representation

Non-IID and RS

- Overview
- Challenges
- Non-IIDness in RS

Data Representation

Attributes

- Text
- Rating table
- Image
- Sequence

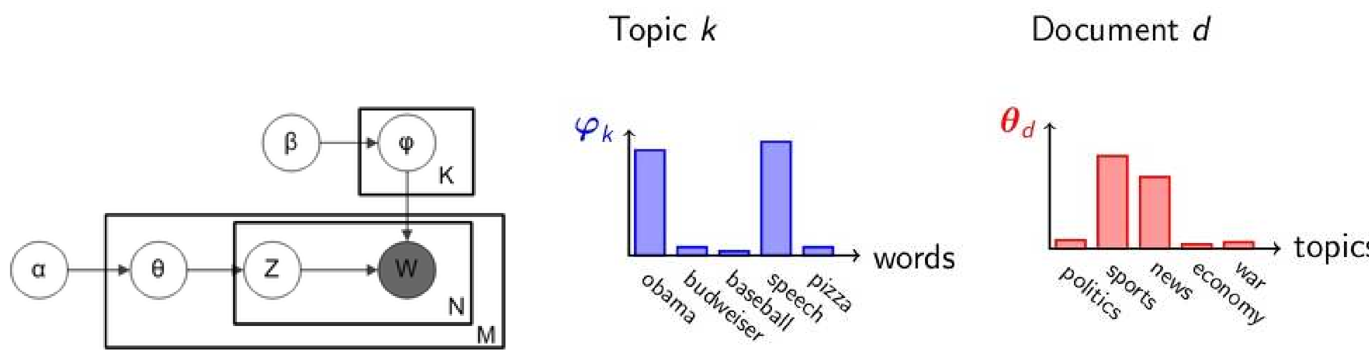
- AI-related preliminaries: from data representation perspective
  - Representing attributes(User/Item features)
  - Representing text (reviews, comments)
  - Representing rating table
  - Representing image (User/Item pictures)
  - Representing sequence (Transactions)

Representing Text Data

- TF-IDF
- Topic model
  - LSA
  - LDA
  - HDP
- Word embedding
  - Skip-gram
  - CBOW

Recommended TF-IDF weighting schemes

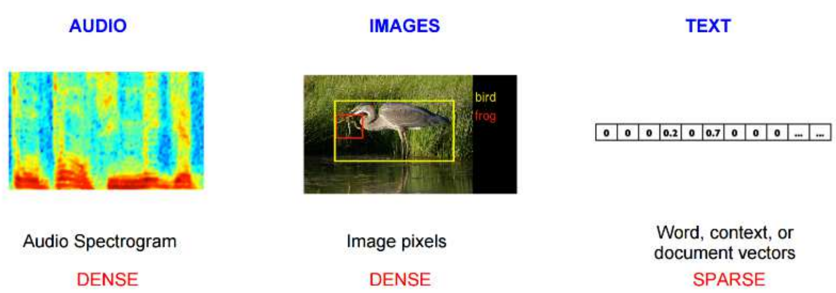
weighting scheme	document term weight	query term weight
1	$f_{t,d} \cdot \log \frac{N}{n_t}$	$\left(0.5 + 0.5 \frac{f_{t,q}}{\max_t f_{t,q}}\right) \cdot \log \frac{N}{n_t}$
2	$1 + \log f_{t,d}$	$\log(1 + \frac{N}{n_t})$
3	$(1 + \log f_{t,d}) \cdot \log \frac{N}{n_t}$	$(1 + \log f_{t,q}) \cdot \log \frac{N}{n_t}$



Latent Dirichlet Allocation

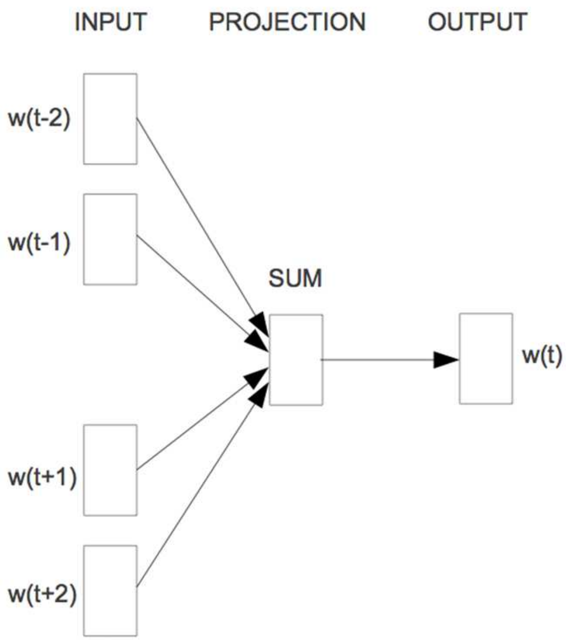
Word Embedding

- *Why Learn Word Embedding?*
  - NLP systems traditionally treat words as discrete atomic symbols
    - E.g. 'cats': id 22, 'dogs': id 23, while they are both animals, four-legged, etc.
  - Using vector representations can overcome some of these obstacles.
- A word embedding $W: \text{words} \to \mathbb{R}^n$ is a parameterized function mapping words in to low-dimensional vectors.

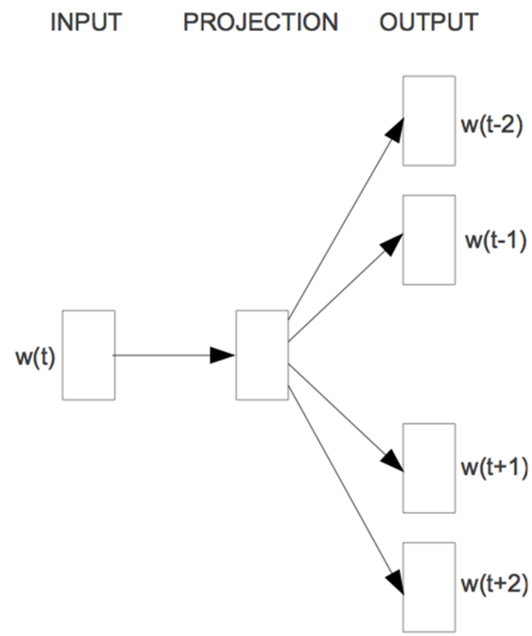


https://www.tensorflow.org/tutorials/word2vec

Word2Vec



Continuous bag-of-words (Mikolov et al., 2013)



Skip-gram (Mikolov et al., 2013)

Mikolov, T., Corrado, G., Chen, K., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality

Data Representation

Non-IID and RS

Overview
Challenges
Non-IIDness in RS

Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

- AI-related preliminaries: from data representation perspective
  - Representing attributes(User/Item features)
  - Representing text (reviews, comments)
  - Representing rating table
  - Representing image (User/Item pictures)
  - Representing sequence (Transactions)

User-item rating

- A full matrix $\boldsymbol{Y} \in \mathbb{R}^{N \times M}$

		5			4		2		
	5					5		1	
		4							

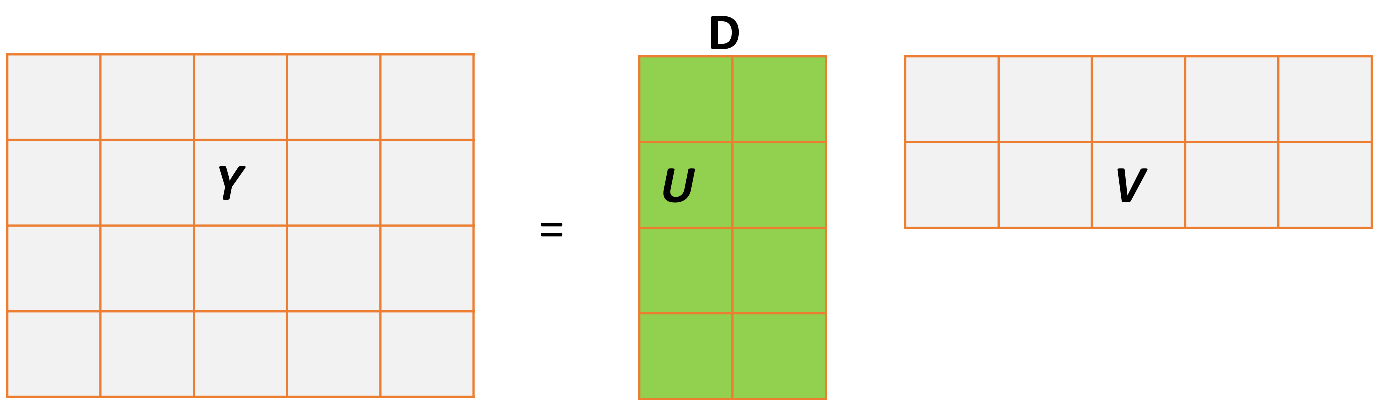
Is there any efficient way to represent rating table?

	1			5		4			1
		4							
	2					4			
		3							5

$\boldsymbol{O(NM)}$, if N=100,000 users, M=50,000 items, 4GB memory is needed.

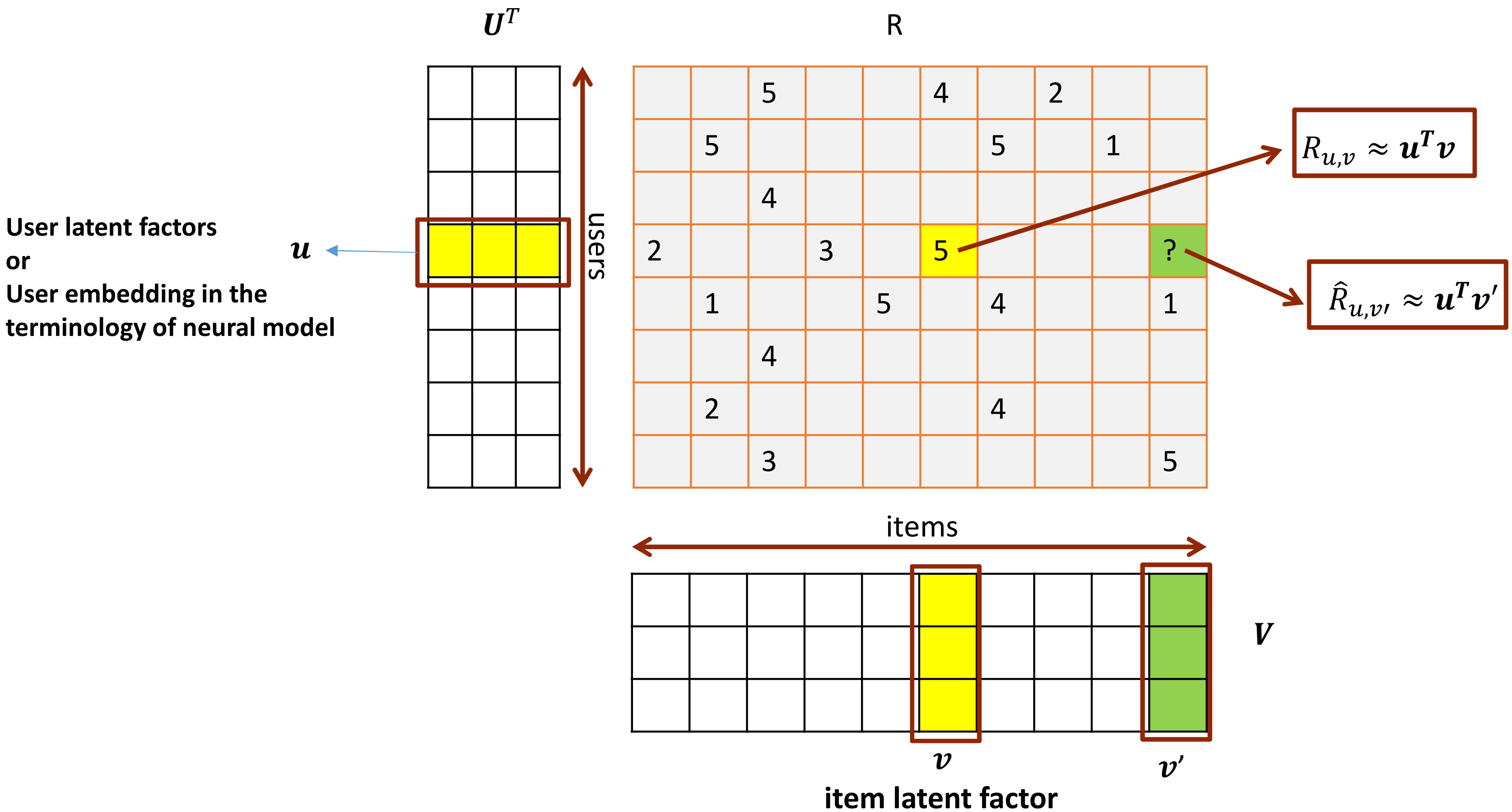
Matrix Factorization

- Approximated by low-rank matrices
  - Given a matrix $\boldsymbol{Y} \in \mathbb{R}^{N \times M}$, we have
    - $\boldsymbol{Y} = \boldsymbol{U}^T\boldsymbol{V}$ where $\boldsymbol{U} = [\boldsymbol{u_1}, \dots, \boldsymbol{u_N}], \boldsymbol{V} = [\boldsymbol{v_1}, \dots, \boldsymbol{v_M}], \boldsymbol{u_i}, \boldsymbol{v_j} \in \mathbb{R}^D$



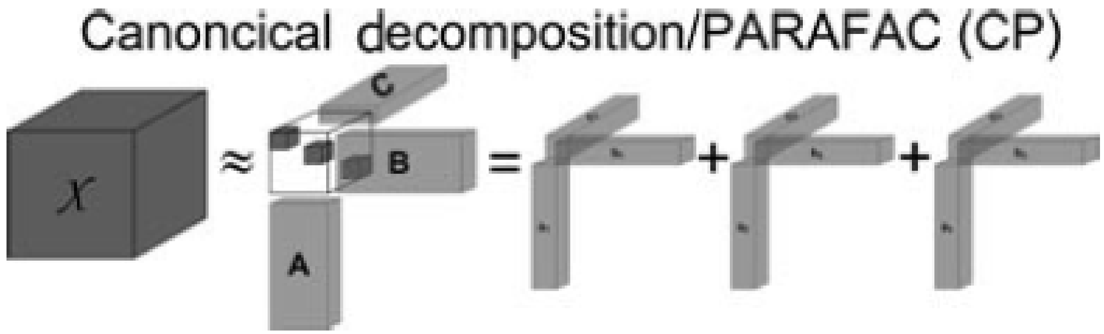
$\boldsymbol{O(ND + MD)}$, if N=100,000 users, M=50,000 items, D=10, only 8MB memory is needed.

Applying MF for Recommender Systems



Tensor Factorization

- **CP Model:** $\boldsymbol{Y} = \boldsymbol{A} \circ \boldsymbol{B} \circ \boldsymbol{C}$



Full Storage: $O(\prod_i N_i)$, if N_i=100,000, 8PB (10^{15}) memory is needed

Low-rank Storage: $O(D \sum_i N_i)$, if D=10, only 24MB memory is needed

Data Representation

Non-IID and RS

Overview
Challenges
Non-IIDness in RS

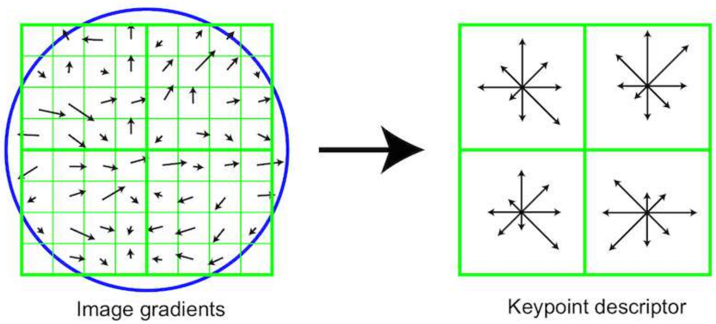
Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

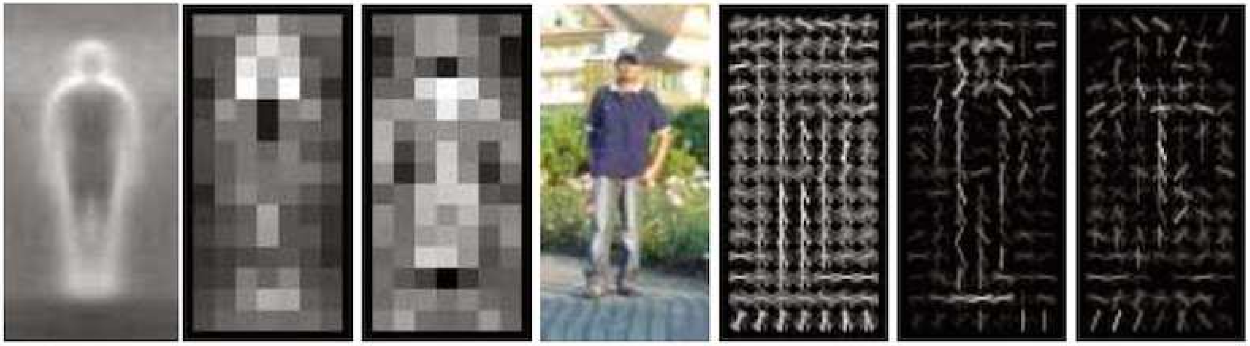
- AI-related preliminaries: from data representation perspective
  - Representing attributes(User/Item features)
  - Representing text (reviews, comments)
  - Representing rating table
  - Representing image (User/Item pictures)
  - Representing sequence (Transactions)

Representing Image

- SIFT (Scale Invariant Feature Transform)

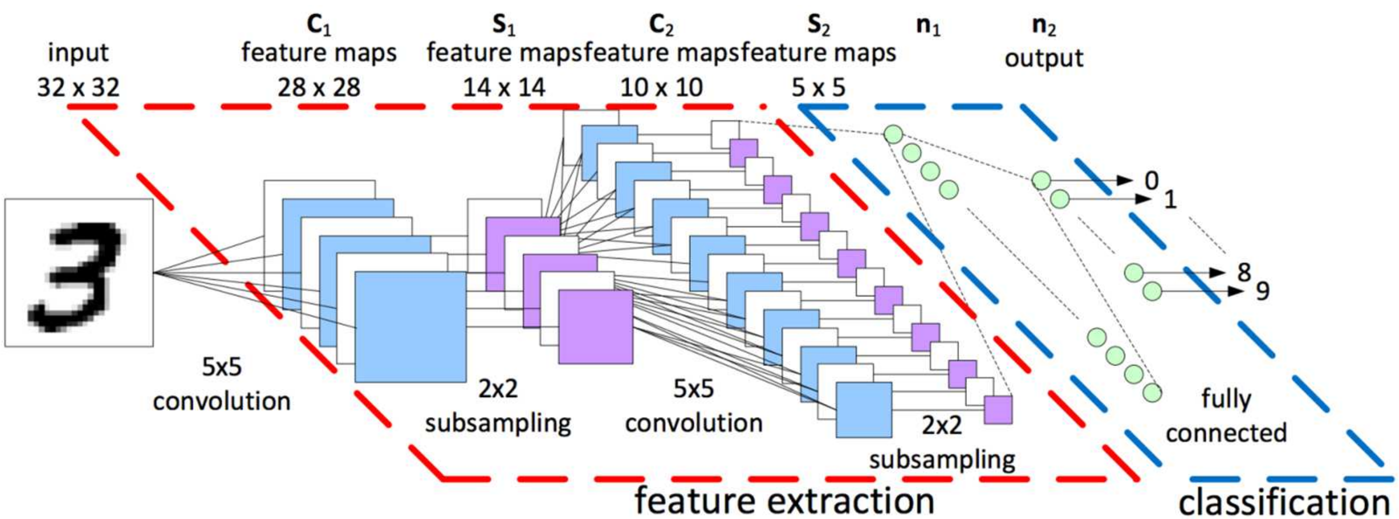


- HOG (Histogram of Oriented Gradients)

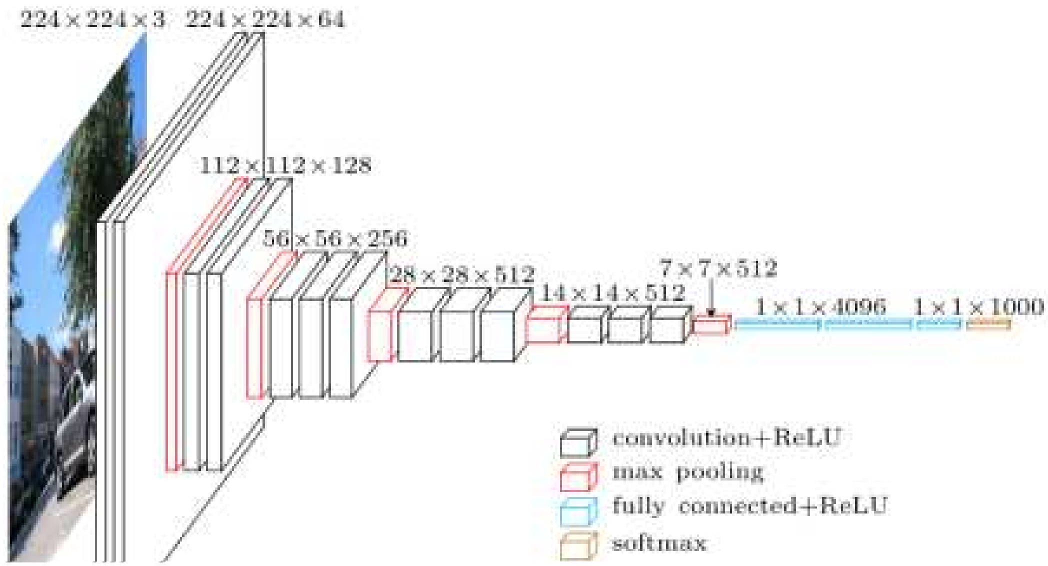


Feature learning for image

- Deep Learning
  - CNN (Convolutional Neural Networks)

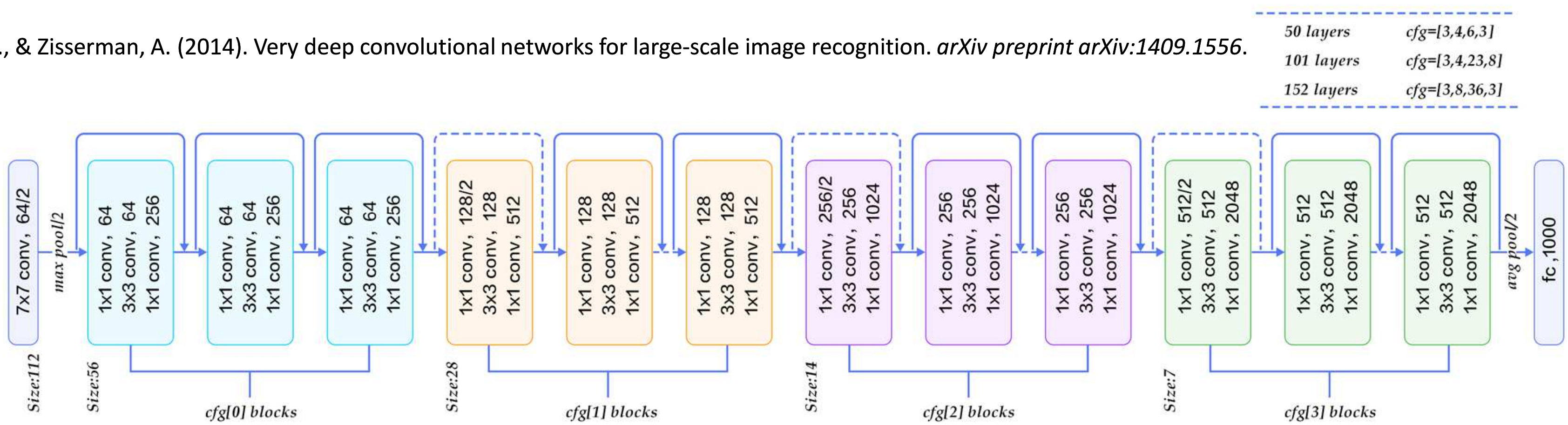


VGG



Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

ResNet



He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*

Data Representation

Non-IID and RS

Overview
Challenges
- Non-IIDness in RS

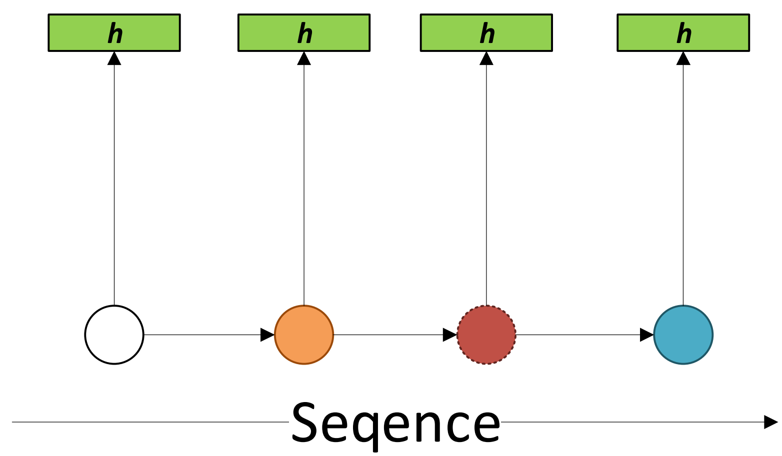
Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

- AI-related preliminaries: from data representation perspective
  - Representing attributes(User/Item features)
  - Representing text (reviews, comments)
  - Representing rating table
  - Representing image (User/Item pictures)
  - Representing sequence (Transactions)

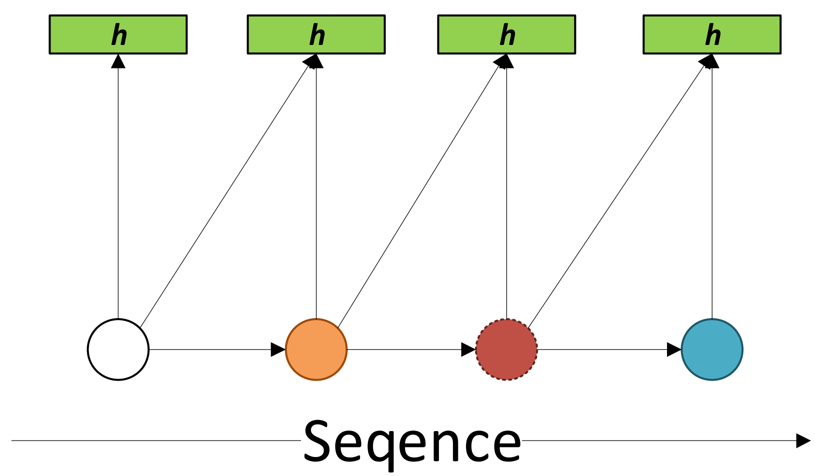
Represent zero-order information

- Map each item into a real-valued vector without considering the sequential dependency
- $h_i = f(o_i)$



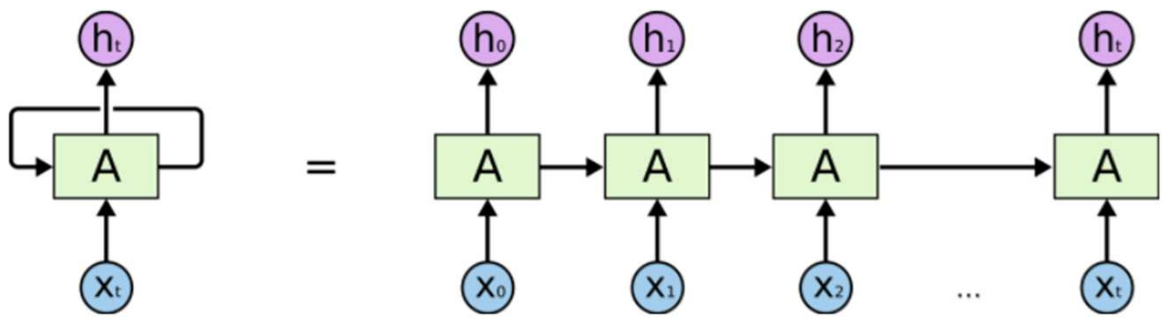
Representing 1st-order information

- Map an object to embedding conditional on the last object
- $h_t = f(o_t | o_{t-1})$



Representing higher order information

- RNN: the representation accumulate the information of recent states.
- $h_t = f(o_t|h_{t-1}) = f\big(o_t|f(o_{t-1}|h_{t-2})\big) = \cdots$

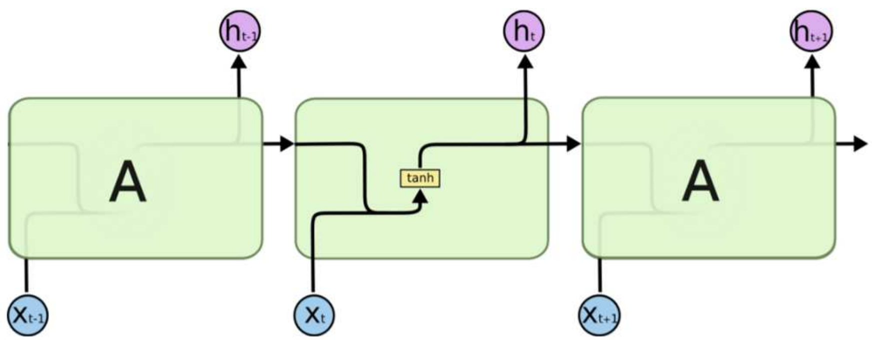


An unrolled recurrent neural network.

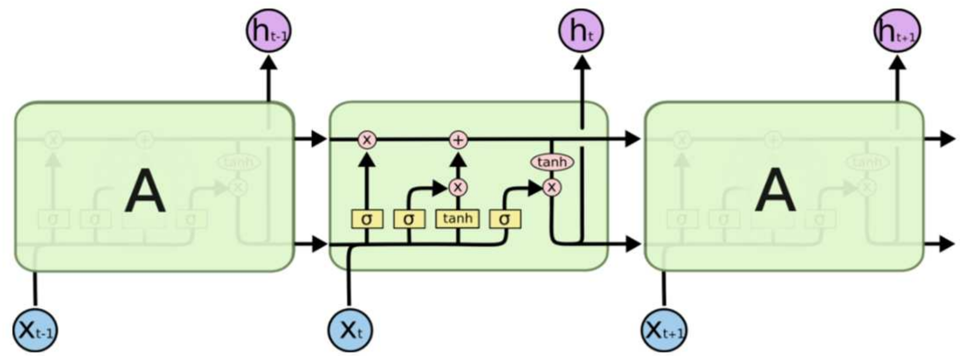
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Representing long-short term information

- LSTM models the long and short-term dependencies, where the accumulation of information is controlled by gate modules



The repeating module in a standard RNN contains a single layer.



The repeating module in an LSTM contains four interacting layers.

Forget gate: $f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$

Input gate: $i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$

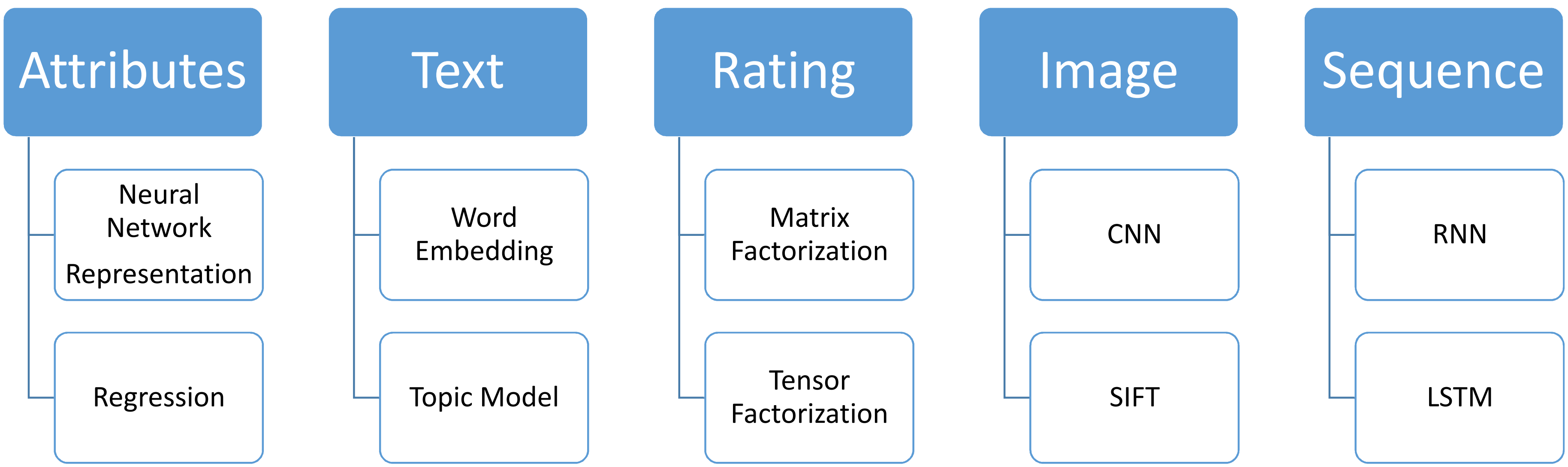
Output gate: $o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = o_t \circ \sigma_h(c_t)$$

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Conclusion of data representation for RS



Non-IIDness on users

Non-IID and RS

Overview
Challenges
- Non-IIDness in RS

Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

Non-IIDness on users
- Social RS
- Group RS

- Social RS: user mutual influence
  - Latent factor model
    - Sorec
    - SocialMF
    - Soreg
  - Deep learning model
    - Item Silk Road
  - Open issues and directions
- Group RS: group joint decision

Non-IIDness on Users

- Heterogeneity
  - Different users often have different tastes in nature,
    - E.g. Some users like sci-fi movies, and others like action movies
- Coupling
  - User choices are often influenced by other users,
    - E.g. Friends' choices often have impact on our choices
  - The choice made by a group is dependent on all group members,
    - E.g. The selection of a movie to a household

Social Recommendation

- The growth of social media usage

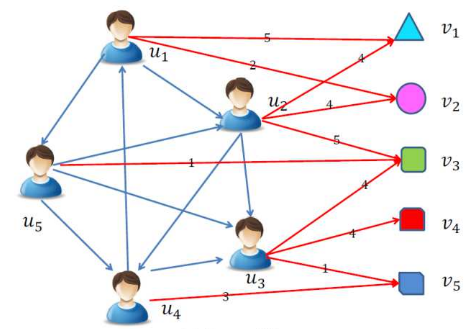
https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

Social Recommendation

- Recommendation + social relations
  - Latent factor model:
    - Co-factorization
    - Regularization methods
  - Deep learning model

	v_1	v_2	v_3	v_4	v_5
u_1	5	?	2	?	?
u_2	4	4	5	?	?
u_3	?	?	4	4	1
u_4	?	?	?	?	3
u_5	?	?	1	?	?

(b) Traditional Recommender Systems



(a) Connected Users

	v_1	v_2	v_3	v_4	v_5
u_1	5	?	2	?	?
u_2	4	4	5	?	?
u_3	?	?	4	4	1
u_4	?	?	?	?	3
u_5	?	?	1	?	?

	u_1	u_2	u_3	u_4	u_5
u_1	0	1	0	0	1
u_2	0	0	1	1	0
u_3	0	0	0	0	0
u_4	1	0	1	0	0
u_5	0	1	1	1	0

(c) Social Recommender Systems

Tang, J., Hu, X., & Liu, H. (2013). Social recommendation: a review. *Social Network Analysis and Mining*, *3*(4), 1113-1133.

Non-IIDness on users

Non-IID and RS

Overview
Challenges
Non-IIDness in RS

Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

Non-IIDness on users

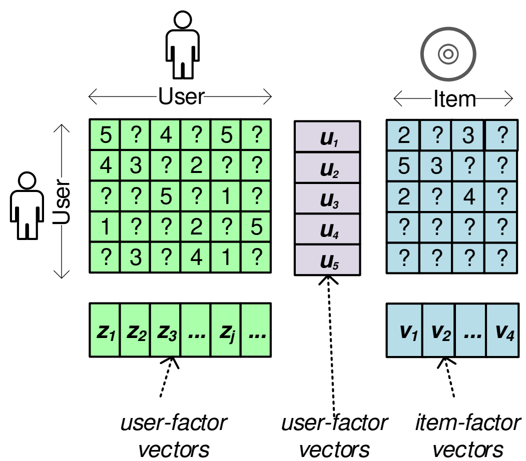
- Social RS
- Group RS

- Social RS: user mutual influence
  - Latent factor model
    - Sorec
    - SocialMF
    - Soreg
  - Deep learning model
    - Item Silk Road
  - Open issues and directions
- Group RS: group joint decision

Sorec: social recommendation using probabilistic matrix factorization

- Integrating social network structure and the user-item rating matrix
  - Connecting through the shared user latent feature space

$$\mathbf{U} = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_N]$$



$$\min \sum_{i=1}^{n} \sum_{u_k \in \mathcal{N}_i} (\mathbf{S}_{ik} - \mathbf{u}_i^\top \mathbf{z}_k)^2,$$

$$\min_{\mathbf{U},\mathbf{V},\mathbf{Z}} \|\mathbf{W} \odot (\mathbf{R} - \mathbf{U}^\top \mathbf{V})\|_F^2 + \alpha \sum_{i=1}^{n} \sum_{u_k \in \mathcal{N}_i} (\mathbf{S}_{ik} - \mathbf{u}_i^\top \mathbf{z}_k)^2$$

$$+ \lambda(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2 + \|\mathbf{Z}\|_F^2)$$

Ma, H., Yang, H., Lyu, M. R., & King, I. (2008, October). Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 931-940). ACM.

SocialMF: MF with social trust propagation

- Based on the assumption of trust-aware recommender
  - Users have similar tastes with other users they trust
  - The transitivity of trust, i.e., trust propagation, is taken into account.

$$\min \sum_{i=1}^{n} (\mathbf{u}_i - \sum_{u_k \in \mathcal{N}_i} \mathbf{S}_{ik} \mathbf{u}_k)^2$$

$$\min_{\mathbf{U},\mathbf{V}} \|\mathbf{W} \odot (\mathbf{R} - \mathbf{U}^\top \mathbf{V})\|_F^2 + \alpha \sum_{i=1}^{n} (\mathbf{u}_i - \sum_{u_k \in \mathcal{N}_i} \mathbf{S}_{ik} \mathbf{u}_k)^2 + \lambda(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2)$$

Jamali, M., & Ester, M. (2010, September). A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 135-142). ACM.

SoReg: recommender systems with social regularization

- Two important assumptions
  - “trust relationships” are different from “social friendships”.
  - The tastes of one user’s friends may vary significantly.

$$\min \sum_{i=1}^{n} \sum_{u_k \in \mathcal{N}_i} \mathbf{S}_{ik} (\mathbf{u}_i - \mathbf{u}_k)^2$$

$$\min_{\mathbf{U},\mathbf{V}} \|\mathbf{W} \odot (\mathbf{R} - \mathbf{U}^\top \mathbf{V})\|_F^2 + \alpha \sum_{i=1}^{n} \sum_{u_k \in \mathcal{N}_i} \mathbf{S}_{ik} (\mathbf{u}_i - \mathbf{u}_k)^2$$
$$+ \lambda(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2)$$

Ma, H., Zhou, D., Liu, C., Lyu, M. R., & King, I. (2011, February). Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 287-296). ACM.

Non-IIDness on users

Non-IID and RS

Overview
Challenges
Non-IIDness in RS

Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

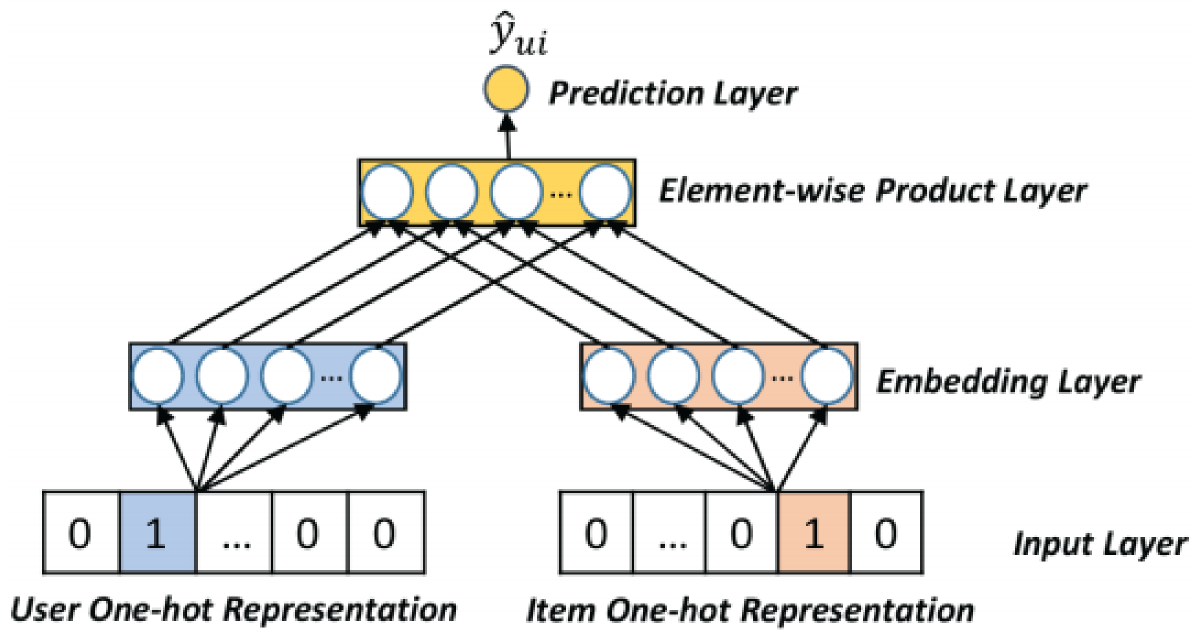
Non-IIDness on users

- Social RS
- Group RS

- Social RS: user mutual influence
  - Latent factor model
    - Sorec
    - SocialMF
    - Soreg
  - Deep learning model
    - Item Silk Road
  - Open issues and directions
- Group RS: group joint decision

Neural network with social recommendation

- Matrix factorization can be regarded as a shallow neural network

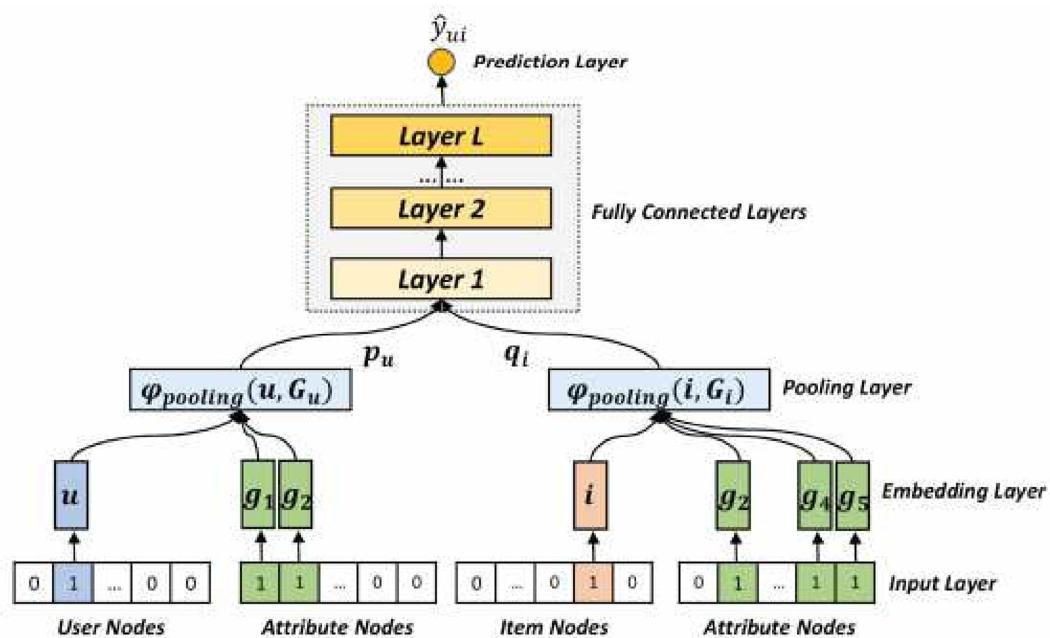


Wang, X., He, X., Nie, L., & Chua, T. S. (2017). Item Silk Road: Recommending Items from Information Domains to Social Users. *arXiv preprint arXiv:1706.03205*.

Deep neural network with social recommendation

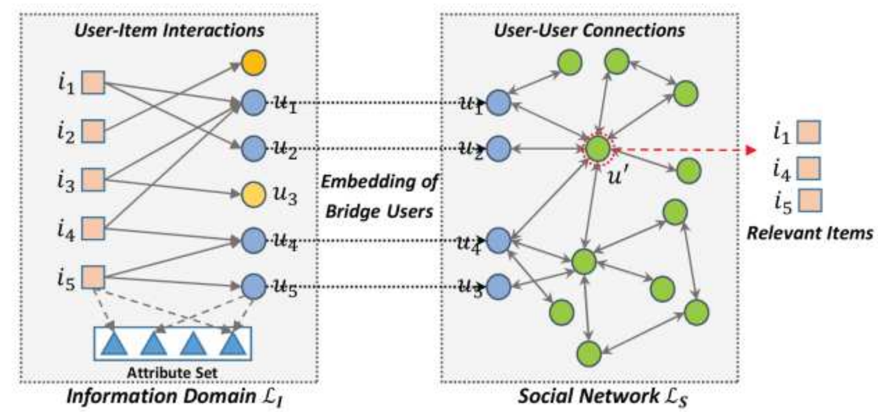
- The key of integrating rating and social information is representation
  - How to project items and users in rating domain and users in social domain into the same embedding space
- Deep neural network is a better option than MF
  - More complex relation
  - Non-linear relation
  - Higher-order interactions

Recommending Items from Information Domains to Social Users



Learning the attribute-aware representation

Representation propagation with graph regularization through bridge nodes



Neighbor regularization $\theta(\mathcal{U}_2) = \frac{1}{2} \sum_{u', u'' \in \mathcal{U}_2} s_{u'u''} \left\| \frac{\mathbf{p}_{u'}}{\sqrt{d_{u'}}} - \frac{\mathbf{p}_{u''}}{\sqrt{d_{u''}}} \right\|^2$

Self regularization $\theta(\mathcal{U}) = \frac{1}{2} \sum_{u' \in \mathcal{U}} \left\| \mathbf{p}_{u'} - \mathbf{p}_{u'}^{(0)} \right\|^2$

Wang, X., He, X., Nie, L., & Chua, T. S. (2017). Item Silk Road: Recommending Items from Information Domains to Social Users. *arXiv preprint arXiv:1706.03205*.

Non-IIDness on users

Non-IID and RS

Overview
Challenges

Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

Non-IIDness on users

- Social RS
- Group RS

- Social RS: user mutual influence
  - Latent factor model
    - Sorec
    - SocialMF
    - Soreg
  - Deep learning model
    - Item Silk Road
  - Open issues and directions
- Group RS: group joint decision

Open issues

- How to deal with the large amount of social data?
  - Billions of nodes
  - Links change every day, every hour, every second
- How to recognize the influential nodes for recommendation?
  - Not all nodes contribute to the recommendation
  - The same nodes may have different influence on different targets
- How to incorporate more social information without harm to users' privacy?
  - More personal information may benefit the recommendation quality
  - Keeping the users' privacy is the top priority

Directions

- Network embedding and learning
  - Mapping user raw information and social relationships to high-level representation
- Memory mechanism
  - Representation learning on social activity sequence
- Dynamic model
  - Using temporal model to capture the shift of social relationships
  - Using neural networks to capture dynamic group coupling

Non-IIDness on users

Non-IID and RS

Overview
Challenges
Non-IIDness in RS

Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

Non-IIDness on users

- Social RS
- Group RS

- Social RS: user mutual influence
- Group RS: group joint decision
  - Profile Aggregation
  - Latent factor model
    - Household Recommendation
  - Deep learning model
    - DLGR model
  - Open issues and directions

Group choices are joint decision

- Group activities are observed throughout life
  - e.g., watching a family movie, planning family travel
- Each member of a group may have **different opinions** on the same items, so the main challenge in GRSs is to satisfy most group members with **diverse preferences**.
- This cannot be achieved through an individual-based recommendation method.

Non-IIDness on users

Non-IID and RS

Overview
Challenges
Non-IIDness in RS

Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

Non-IIDness on users

- Social RS
- Group RS

- Social RS: user mutual influence
- Group RS: group joint decision
  - Profile Aggregation
  - Latent factor model
    - Household Recommendation
  - Deep learning model
    - DLGR model
  - Open issues and directions

Profile Aggregation

- Group Preference Aggregation (GPA) (Pre-aggregation)
  - Aggregating all members' ratings into a group profile
  - Groups are regarded as virtual individual users.
  - Disadvantage: the preference is biased to active users with more data

- Individual Preference Aggregation (IPA) (Post-aggregation)
  - Predicting the individual ratings over candidate items
  - Aggregating the predicted ratings of members via predefined strategies.
  - Disadvantage: IPA fails to consider the group behavior

Overview of Aggregation Strategies for Group Recommendation

- Many strategies exist for aggregating individual ratings into a group rating (e.g. used in elections and when selecting a party leader)

Strategy	How it works	Example
Plurality voting	Uses 'first past the post': repetitively, the item with the most votes is chosen.	A is chosen first, as it has the highest rating for the majority of the group, followed by E (which has the highest rating for the majority when excluding A).
Average	Averages individual ratings	B's group rating is 6, namely (4 + 9 + 5)/3.
Multiplicative	Multiplies individual ratings	B's group rating is 180, namely 4*9*5.
Borda count	Counts points from items' rankings in the individuals' preference lists, with bottom item getting 0 points, next one up getting one point, etc.	A's group rating is 17, namely 0 (last for Jane) + 9 (first for Mary) + 8 (shared top 3 for Peter)
Copeland rule	Counts how often an item beats other items (using majority vote[a]) minus how often it looses	F's group rating is 5, as F beats 7 items (B,C,D,G,H,I,J) and looses from 2 (A,E).
Approval voting	Counts the individuals with ratings for the item above a approval threshold (e.g. 6)	B's group rating is 1 and F's is 3.
Least misery	Takes the minimum of individual ratings	B's group rating is 4, namely the smallest of 4,9,5.
Most pleasure	Takes the maximum of individual ratings	B's group rating is 9, namely the largest of 4,9,5.
Average without misery	Averages individual ratings, after excluding items with individual ratings below a certain threshold (say 4).	J's group rating is 7.3 (the average of 8,8,6), while A is excluded because Jane hates it.
Fairness	Items are ranked as if individuals are choosing them in turn.	Item E may be chosen first (highest for Peter), followed by F (highest for Jane) and A (highest for Mary).
Most respected person (or Dictatorship)	Uses the rating of the most respected individual.	If Jane is the most respected person, then A's group rating is 1. If Mary is most respected, then it is 10.

Masthoff, J. (2015). Group recommender systems: aggregation, satisfaction and group attributes. In *Recommender Systems Handbook* (pp. 743-776).

Most Frequently Used Aggregation Strategies

- *Average* and *Least misery* are the two most prevalent strategies.
- ***Least misery* strategy** assumes a group tends to be as happy as its **least happy member**.
- ***Average* strategy** recommends items with the highest average ratings over all members.

Masthoff, J. (2015). Group recommender systems: aggregation, satisfaction and group attributes. In *Recommender Systems Handbook* (pp. 743-776).

Group Recommender Systems

System	Usage scenario	Classification				Strategy used
		Preferences known	Direct experience	Group active	Recommends sequence	
MusicFX [33]	Chooses radio station in fitness center based on people working out	Yes	Yes	No	No	Average Without Misery
PolyLens [36]	Proposes movies for a group to view	Yes	No	No	No	Least Misery
Intrigue [2]	Proposes tourist attractions to visit for a group based on characteristics of subgroups (such as children and the disabled)	Yes	No	No	Yes	Average
Travel Decision Forum [22]	Proposes a group model of desired attributes of a planned joint vacation and helps a group of users to agree on these	Yes	No	Yes	No	Median
Yu's TV Rec. [49]	Proposes a TV program for a group to watch based on individuals' ratings for multiple features	Yes	No	No	No	Average
CATS [34]	Helps users choose a joint holiday, based on individuals' critiques	No	No	Yes	No	Counts requirements met Uses Without Misery
Masthoff's [28, 30]	Chooses a sequence of music video clips for a group to watch	Yes	Yes	No	Yes	Multiplicative etc
GAIN [11]	Displays information and advertisements adapted to the group present	Yes	Yes	No	Yes	Average
REMPAD [7]	Proposes multimedia material for a group reminiscence therapy session	Yes	No	No	No	Least Misery
HappyMovie [39]	Recommends movies to groups	Yes	No	No	No	Average
IntelliReq [14]	Supports groups in deciding which requirements to implement	No	No	Yes	Yes	Plurality Voting

Masthoff, J. (2015). Group recommender systems: aggregation, satisfaction and group attributes. In *Recommender Systems Handbook* (pp. 743-776).

Non-IIDness on users

Non-IID and RS
- Overview
- Challenges

Data Representation
- Attributes
  - Text
  - Rating table
  - Image
  - Sequence

Non-IIDness on users
- Social RS
- Group RS

- Social RS: user mutual influence
- Group RS: group joint decision
  - Profile Aggregation
  - Latent factor model
    - Movie recommendation for household
  - Deep learning model
    - DLGR model
  - Open issues and directions

Fixed group and Flexible group

- Fixed group-based recommendation
  - A family
  - A working group

- Flexible group-based recommendation
  - Friends meetup
  - Conference attenders

Ranking oriented feature-based matrix factorization

- Feature-based matrix factorization

$$y = f\Big(\mu + (\sum_j b_j^{(g)}\gamma_j + \sum_j b_j^{(u)}\alpha_j + \sum_j b_j^{(i)}\beta_j) + (\sum_j p_j\alpha_j)^T(\sum_j q_j\beta_j)\Big)$$

where *α, β, γ* denote user, item, and global features respectively

Lu, Q., Yang, D., Cheng, T., Zhang, W., Yu, Y. Informative Household Recommendation with Feature-based Matrix Factorization. In *CAMRa2011*, 2011.

Ranking matrix factorization

- Pairwise preference generation rule:

$$\delta_{u,i,j} = \begin{cases} +1 & i,j \in I_u \text{ and } r_{u,i} > r_{u,j} \\ +1 & i \in I_u \text{ and } j \notin I_u \\ -1 & i,j \in I_u \text{ and } r_{u,i} < r_{u,j} \\ -1 & i \notin I_u \text{ and } j \in I_u \end{cases}$$

- MF parameterization:

$$\hat{r}_{u,i} = p_u^T q_i + b_u + b_i$$

- Using Bayesian Personalization Ranking (BPR) for optimization

$$P(\delta_{u,i,j} = +1) = \frac{1}{1 + e^{-(\hat{r}_{u,i} - \hat{r}_{u,j})}}$$

Movie recommendation for household

- **Individual Preference Aggregation (post-aggregation):**
  - First predict users' rating for items by MF
  - Then combining the household members' ratings to get household rating

$$\hat{r}_{h,i} = \sum_{u \in H(h)} w_u \cdot \hat{r}_{u,i}$$

where w is set to 1 for each member (Average strategy)

- **Group Preference Aggregation (pre-aggregation):**
  - First build household profile
  - Then adopt MF on household-item ratings

$$r_{h,i} = \frac{\sum_{u \in H(h)} r_{u,i}}{|H(h)|}$$

The rating of a household equals to the average rating of its members (Average strategy)

Dataset for group recommendation

- CAMRa2011 dataset containing the movie watching records of households and the ratings on each watched movie given by some group members.

- The dataset for track 1 of CAMRa2011 has 290 households with a total of 602 users who gave ratings (on a scale 1~100) over 7,740 movies.

Experimental results

- Comparisons of recommendation performance:

- BMF: This model represents the basic matrix factorization approach which is equivalent to Equation 3. We set the parameter λ_1 to 0.004 and λ_2 to 0, which is optimal on the evaluation set. The dimensionality of user and item factors is 64 here for efficiency. The same parameters are used in the ranking matrix factorization model.
- RMF: This model represents the ranking matrix factorization approach, stated in Section 3.1. We try the two sampling schemes in our experiments, denoting RMF-S1(as in Equation 6) and RMF-S2(as in Equation 7, $r_1 = 100, r_2 = 70$) respectively.
- BMF-3N: This approach represents the BMF model which add three fold negative examples to the training set as stated in Section 3.2.
- BMF-3N-I100NN(HIR, IMFB, ALL): These approaches represent the informative models which integrate the item neighborhood information, user household hierarchy and user implicit feedback into the BMF-3N model respectively, as stated in Section 4. BMF-3N-ALL denotes the model that integrates all the useful information into the BMF-3N model.
- RMF-S2-I100NN(HIR, IMFB, ALL): These approaches represent the informative models which integrate the item neighborhood information, user household hierarchy and user implicit feedback into the RMF-S2 model respectively, as stated in Section 4. RMF-S2-ALL denotes the model that integrates all the useful information into the basic RMF-S2 model.

Models	MAP	AUC	P@5	P@10
BMF	0.1390	0.8374	0.1344	0.1051
BMF-3N	0.2268	0.9926	0.2039	0.1680
BMF-3N-HIR	0.2315	0.9910	0.2124	0.1718
BMF-3N-IMFB	0.2383	0.9940	0.2150	0.1727
BMF-3N-I100NN	0.2614	0.9922	0.2402	0.1968
BMF-3N-ALL	0.2639	0.9924	0.2435	0.1970
RMF-S1	0.2053	0.9931	0.1931	0.1608
RMF-S2	0.2275	0.9939	0.2065	0.1741
RMF-S2-HIR	0.2387	0.9943	0.2167	0.1814
RMF-S2-IMFB	0.2477	0.9943	0.2322	0.1893
RMF-S2-I100NN	0.2847	0.9936	0.2550	0.2021
RMF-S2-ALL	0.3096	0.9956	0.2872	0.2190

Non-IIDness on users

Non-IID and RS

Overview
Challenges
Non-IIDness in RS

Data Representation

Attributes
- Text
- Rating table
- Image
- Sequence

Non-IIDness on users

- Social RS
- Group RS

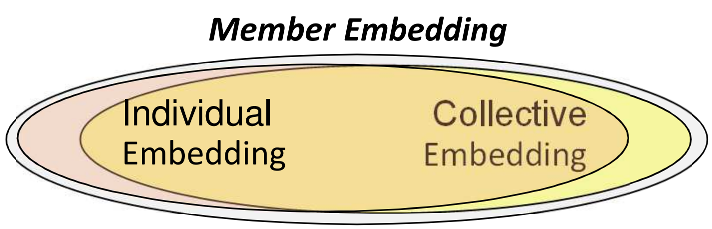
- Social RS: user mutual influence
- Group RS: group joint decision
  - Profile Aggregation
  - Latent factor model
    - Household Recommendation
  - Deep learning model
    - DLGR model
  - Open issues and directions

DLGR : Modeling Features in Group-based Decision

- ***Member Embedding***: which model the individual preference of a user when she/he makes choices as a group member, which can be regarded as **a mixture of *Collective Embedding* and *Individual Embedding*** .

- ***Collective Embedding***: which represent compromised preferences of a group, which are **shared among all members** and can be disentangled from the *Member Embedding*.

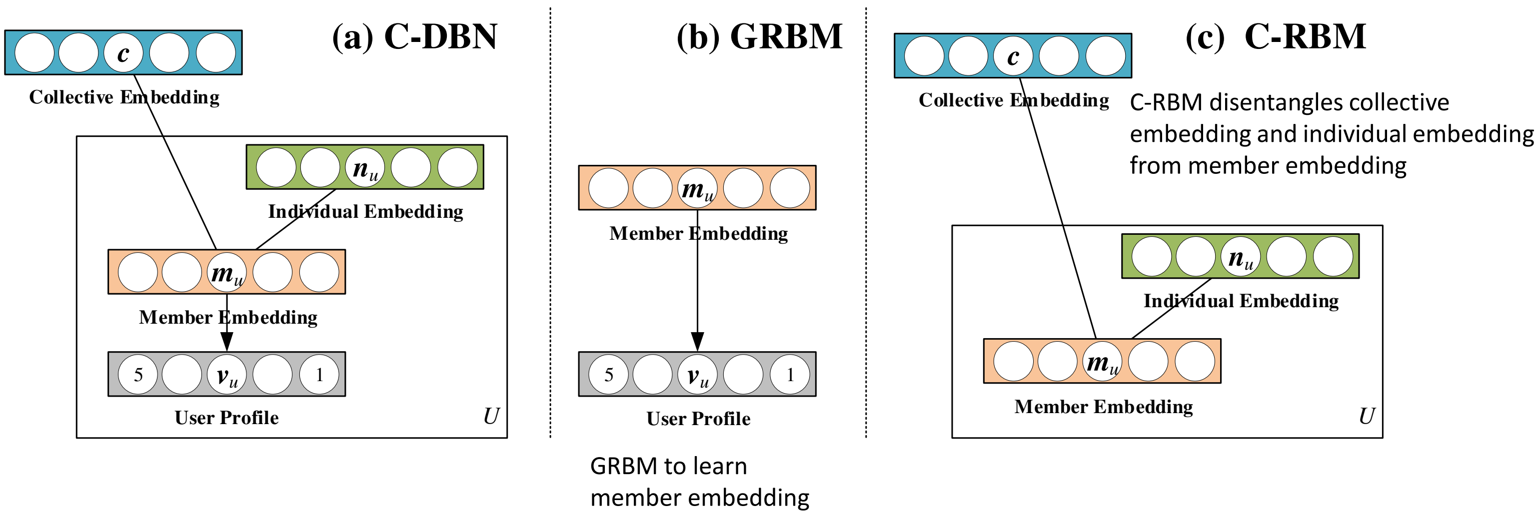
- ***Individual Embedding*** : these represent **independent individual-specific preference**, which can be disentangled from the *Member Embedding* w.r.t. this user.



Hu, L., Cao, J., Xu, G., Cao, L., Gu, Z., and Cao, W. Deep modeling of group preferences for group-based recommendation. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

Disentangling Collective and Individual Embedding

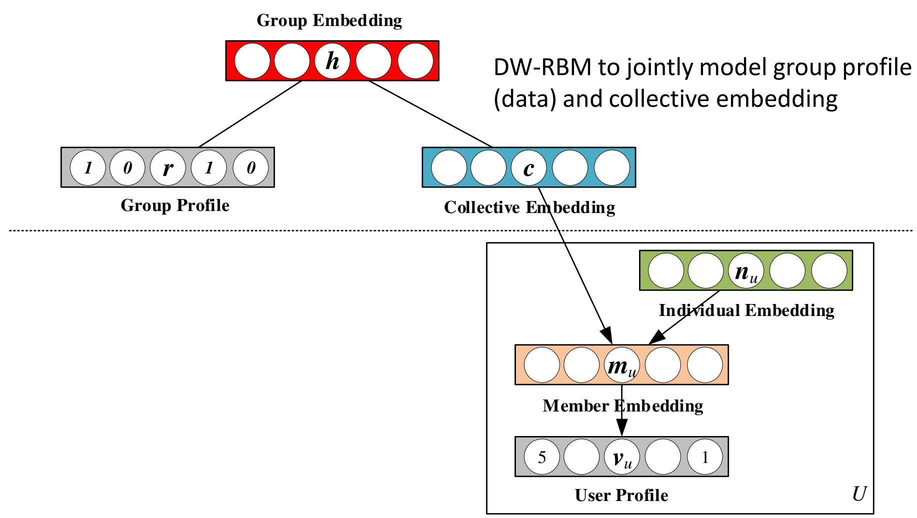
- Each group choice can be regarded as a **joint decision** by all members



Hu, L., Cao, J., Xu, G., Cao, L., Gu, Z., and Cao, W. Deep modeling of group preferences for group-based recommendation. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

Comprehensive Representation of Group Preferences

- A dual-wing RBM is placed on the top of DBN, which jointly models the group choices and collective features to learn the comprehensive features of group preference



DW-RBM to jointly model group profile (data) and collective embedding

Results

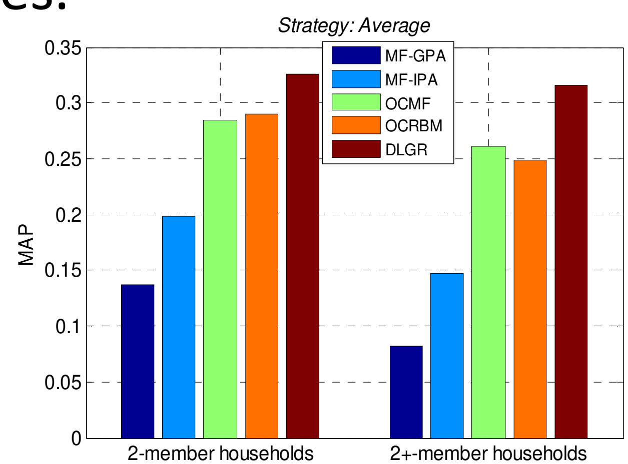
Dataset: CAMRa2011 dataset

MAP and mean AUC of all comparative models with different strategies

	MAP			AUC		
Model/Strategy	No Strategy	Average	Least Misery	No Strategy	Average	Least Misery
kNN (k=5)	0.1595	N/A	N/A	0.9367	N/A	N/A
MF-GPA	N/A	0.1341	0.0628	N/A	0.9535	0.9297
MF-IPA	N/A	0.1952	0.1617	N/A	0.9635	0.9503
OCMF	0.2811	0.2858	0.2801	0.9811	0.9813	0.9810
OCRBM	0.2823	0.2922	0.2951	0.9761	0.9778	0.9782
DLGR	0.3236	0.3252	0.3258	0.9880	0.9892	0.9897

Group with different number of members

- A group with **more members implies more different preferences**, so it is harder to find recommendations satisfying all members.
- Each household may contain 2~4 members in this dataset. We additionally evaluated the MAP w.r.t. 2-member households and the 2^+-member (>2) households under *Average* and *Least Misery* strategies.



Non-IIDness on users

Non-IID and RS

Overview
Challenges
Non-IIDness in RS

Data Representation

Attributes

- Text
- Rating table
- Image
- Sequence

Non-IIDness on users

- Social RS
- Group RS

- Social RS: user mutual influence
- Group RS: group joint decision
  - Profile Aggregation
  - Latent factor model
    - Household Recommendation
  - Deep learning model
    - DLGR model
  - Open issues and directions

Open issues and directions

- Lack of group feedback data
  - There are very few real-world public datasets
  - Most datasets are synthetic from personal feedback, which does not contain the features of group decision

- Learning group context representation given a group of any users

- Dynamic group recommendation with contextual information
  - Flexible group-based recommendation, e.g., friends meetup, conference attenders

Attention Mechanism

- Visual attention: many animals only focus on specific parts of their visual inputs to compute the adequate responses.

Figure 2. Attention over time. As the model generates each word, its attention changes to reflect the relevant parts of the image. "soft" (top row) vs "hard" (bottom row) attention. (Note that both models generated the same captions in this example.)

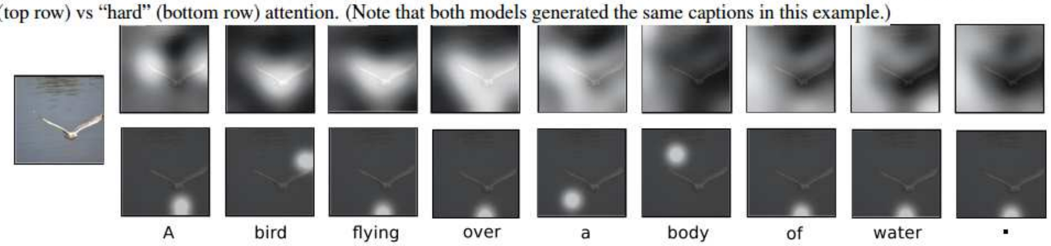
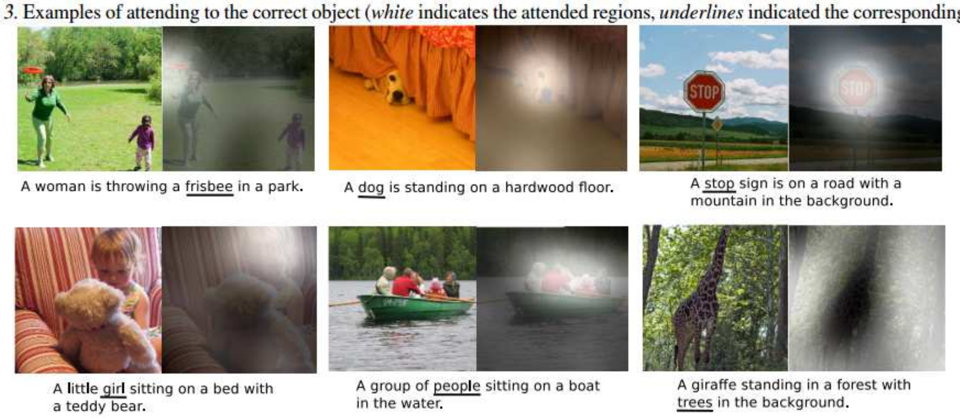
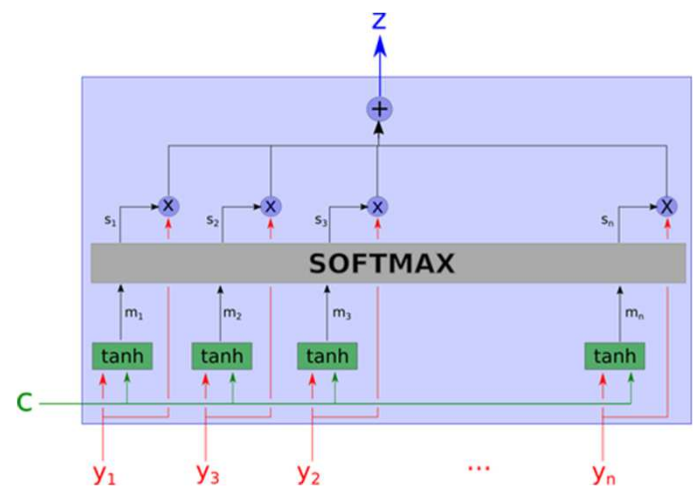


Figure 3. Examples of attending to the correct object (*white* indicates the attended regions, *underlines* indicated the corresponding word)

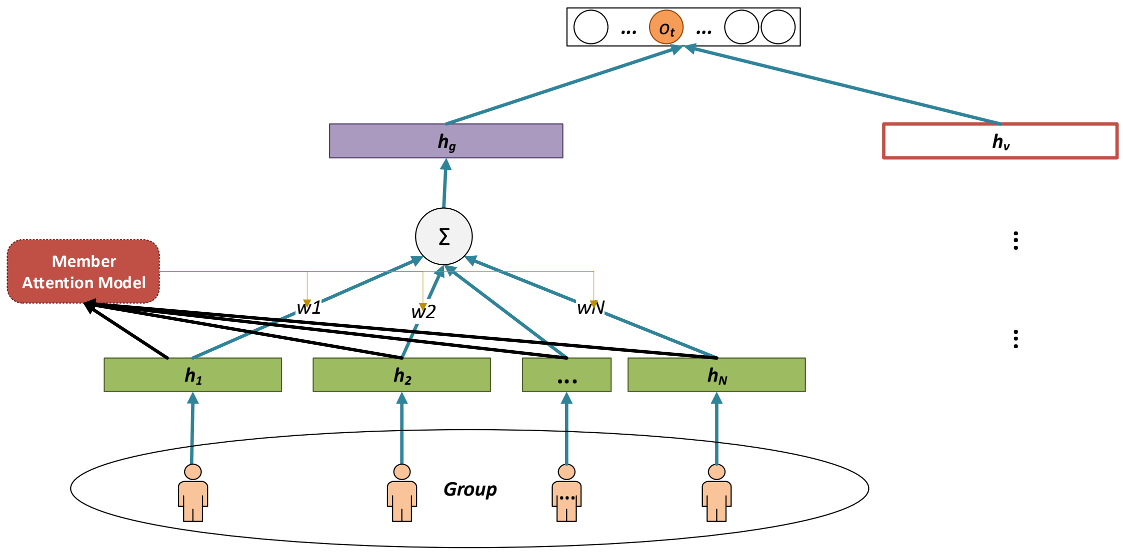




Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... & Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In *ICML 2015*.

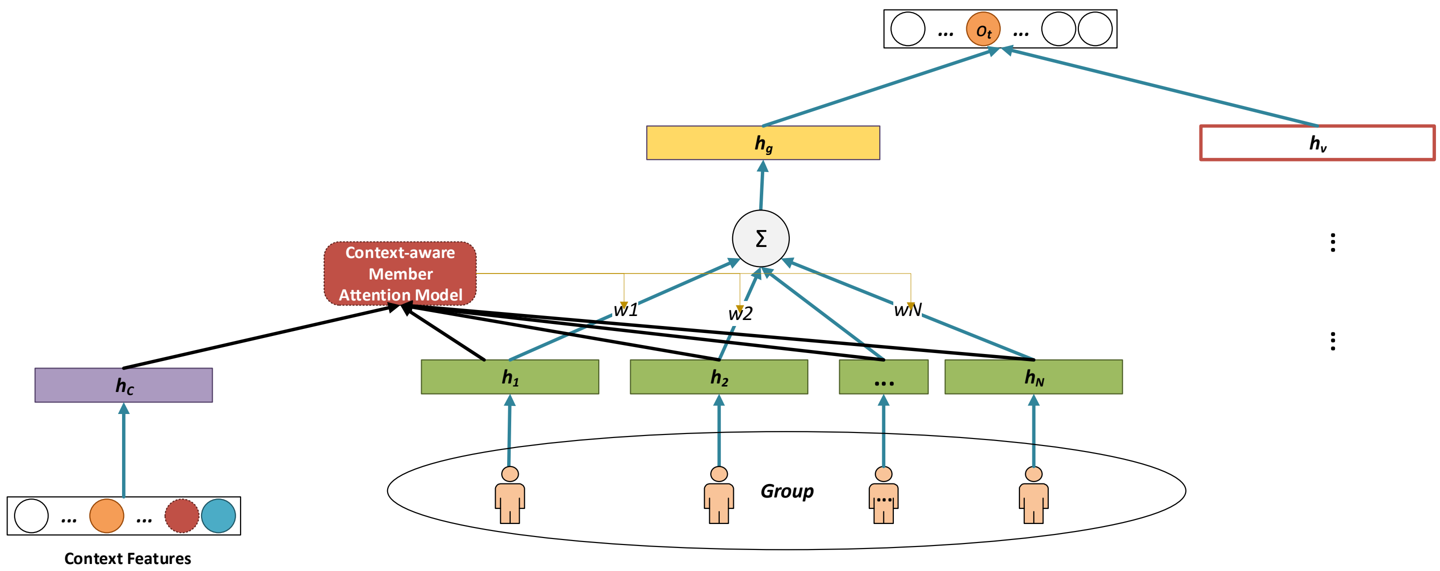
Using attention mechanism

- Most aggregation strategies are about how to weight group members
- Member attention model to learn how to assign weights on members



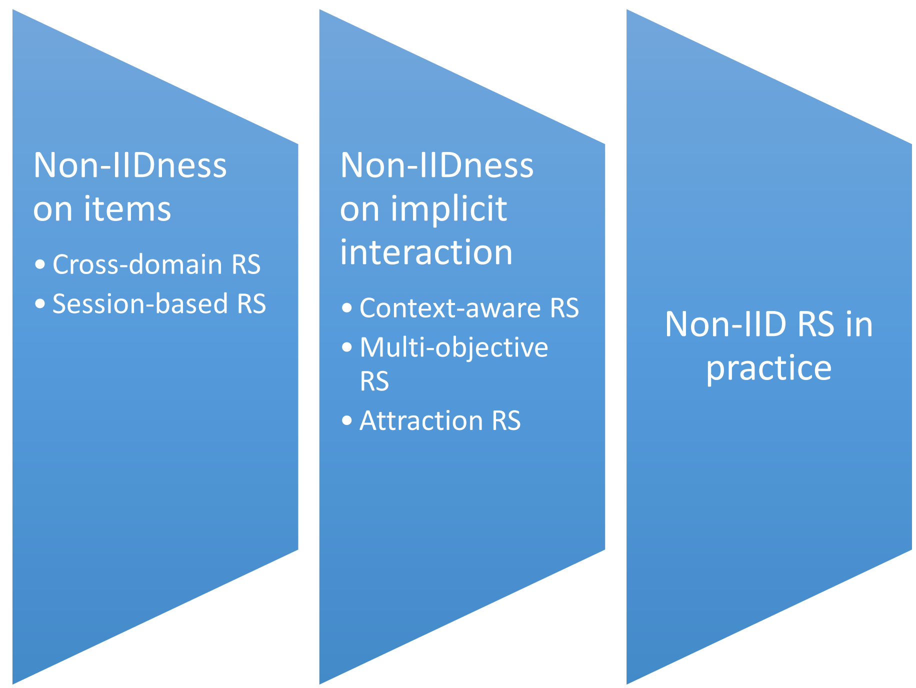
Context-aware group recommendation

- Each group member plays different role in different context
  - Assign different weights in different context



Non-IIDness on items

- Cross-domain RS: domain coupling
  - Item domains
    - Latent factor model
    - Deep learning model
  - Modality domains
    - Multimodal RS
  - Open issues and directions
- Session-based RS: sequential coupling

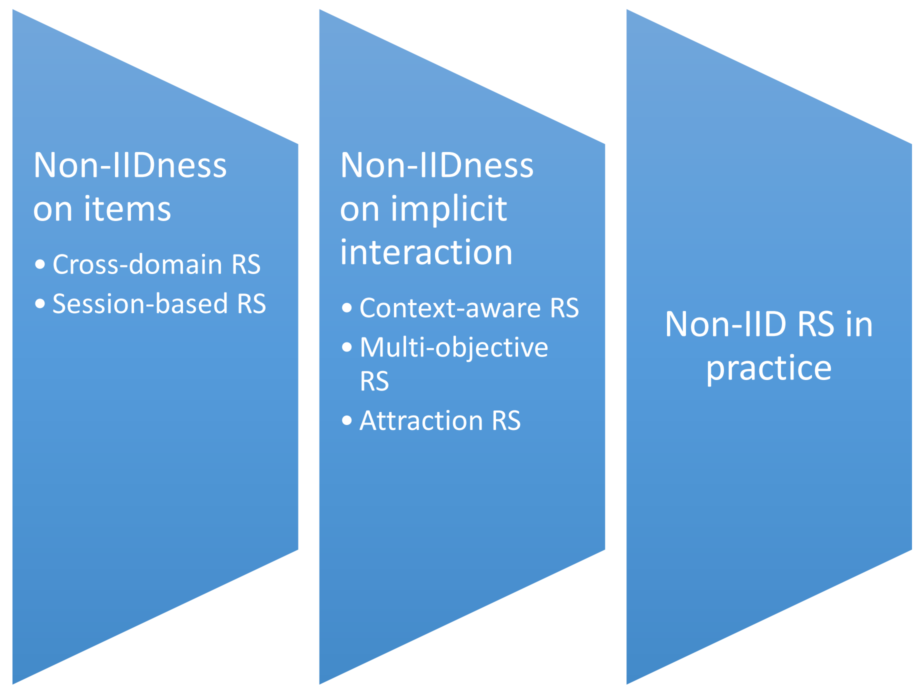


Non-IIDness on items

- Heterogeneity
  - Items in different domains have different meaningful attributes,
    - E.g. color is a critical attribute for clothes but not for books
  - One item is often associated with multi-modal data,
    - E.g. For a movie, there are rating, poster (image), comments (text), prevue (videos)
- Coupling
  - Items in different domain often share some common patterns
    - E.g. users who like sci-fi novels (book domain) may also like relevant sci-fi movies (movie domain)
  - Different types of data can provide complementary information
    - E.g. the description and the pictures of an item can provide more comprehensive information
  - The choice of items in a transaction are dependent
    - E.g. a user has selected milk, s/he may select bread in a transaction

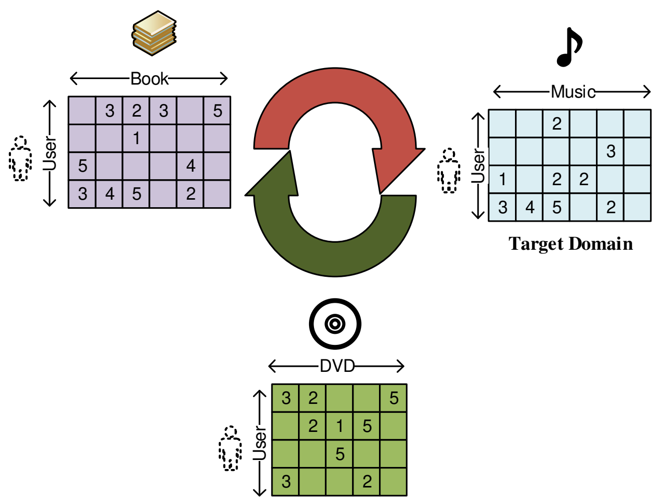
Non-IIDness on items

- Cross-domain RS: domain coupling
  - Item domains
    - Latent factor model
    - Deep learning model
  - Modality domains
    - Multimodal RS
  - Open issues and directions
- Session-based RS: sequential coupling



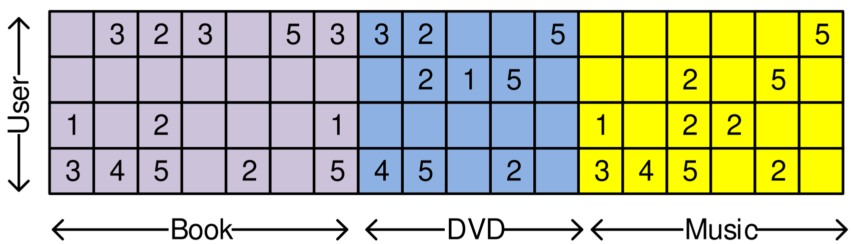
Cross item-domain assumption

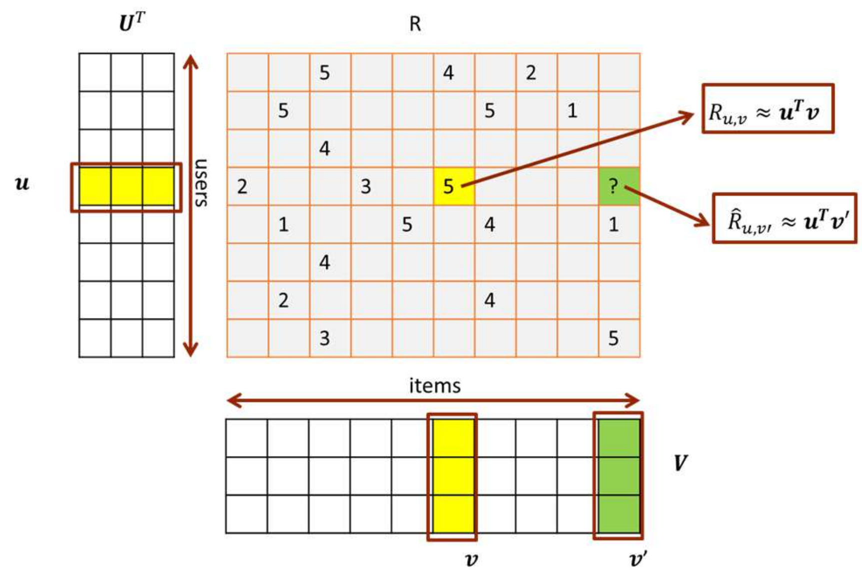
- The assumption of leveraging cross-domain information in RS
  - The existence of multiple *related* domains
  - The user preference from each domain is not independent
- Two main methods for item domains
  - Latent factor model
    - MF-based transfer learning
    - Weighted irregular tensor factorization
  - Deep learning model
    - A multi-view deep learning approach
    - DiscoGAN



Naïve MF for cross domains

- Concatenating the rating matrices for all domains





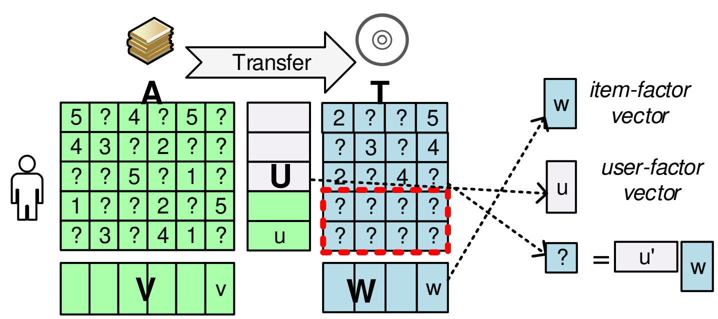
Deficiency

- Each domain has different characteristics
  - The factor of *color* has huge impact on the user preference in *clothes* domain
  - But factor of *color* has little impact on the user preference in *book* domain

- Above method using the single domain model implicitly assume the *homogeneity* of items.
  - Obviously, such assumption may decrease the prediction accuracy due to the ***heterogeneities*** of different domains.

MF based Transfer Learning

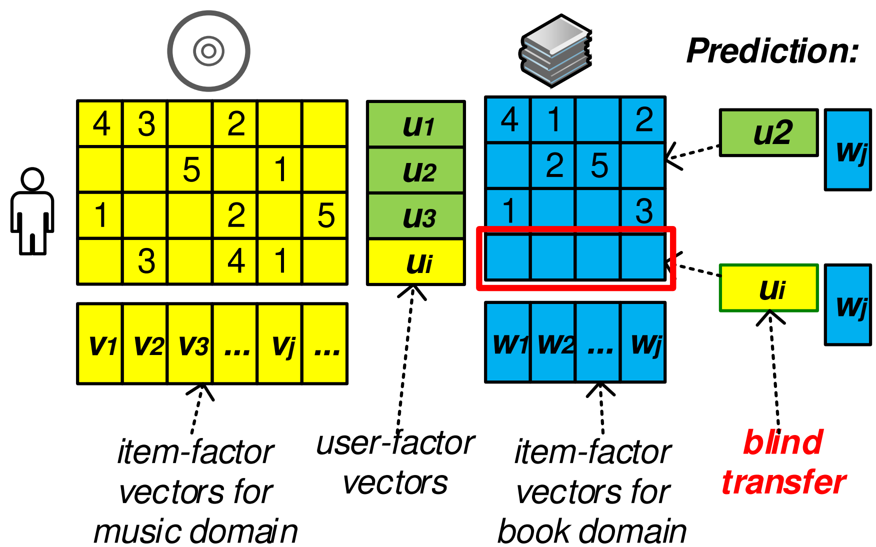
- Transfer the knowledge learned from the auxiliary domain to the target domain
  - The user-factor vectors are *co-determined* by the feedback in auxiliary and target domains



Pan, W., Xiang, E. W., Liu, N. N., & Yang, Q. (2010, July). Transfer Learning in Collaborative Filtering for Sparsity Reduction. In *AAAI* (Vol. 10, pp. 230-235).

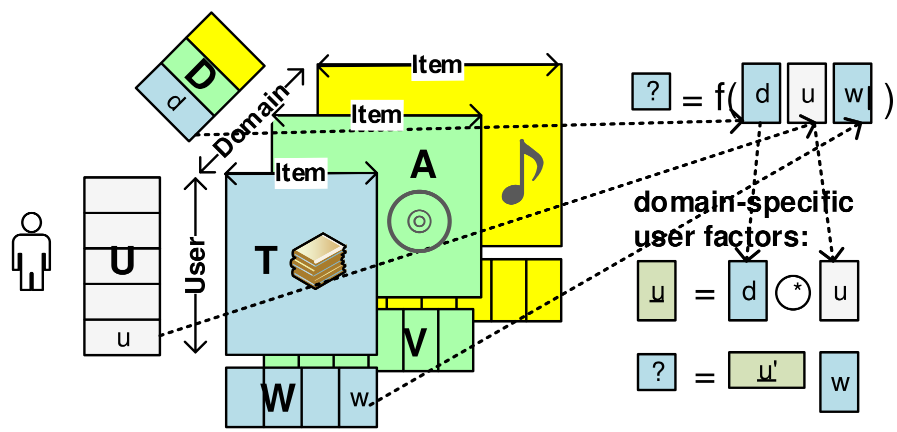
Deficiency

- Blind Transfer
  - If $\boldsymbol{u}_i$ is transferred to the target domain and interacts with heterogeneous item factors, it may yield a poor prediction.



Modeling Domain Heterogeneity

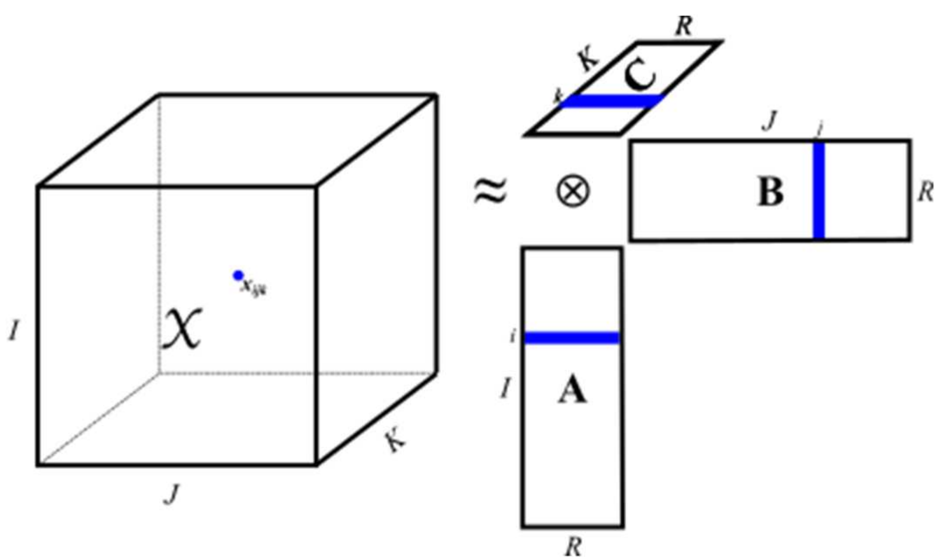
- **Domain factors** is an essential element in cross "domain" problem to model domain heterogeneity
- Triadic relation ***user-item-domain*** to reveal the domain-specific user preference

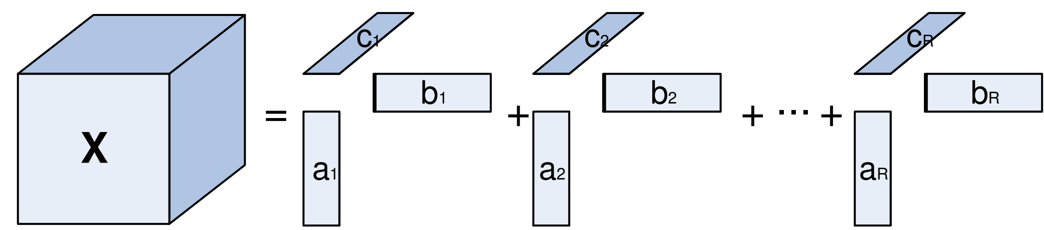


Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., & Yang, D. (2016). Learning Informative Priors from Heterogeneous Domains to Improve Recommendation in Cold-Start User Domains. *ACM Transactions on Information Systems (TOIS)*, *35*(2), 13.

Tensor Factorization over Triadic Relation

- Decompose a tensor into a sum of rank-one components
  - $\boldsymbol{\mathcal{X}} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \sum_{r=1}^{R} \mathbf{A}_{\cdot,r} \circ \mathbf{B}_{\cdot,r} \circ \mathbf{C}_{\cdot,r}$

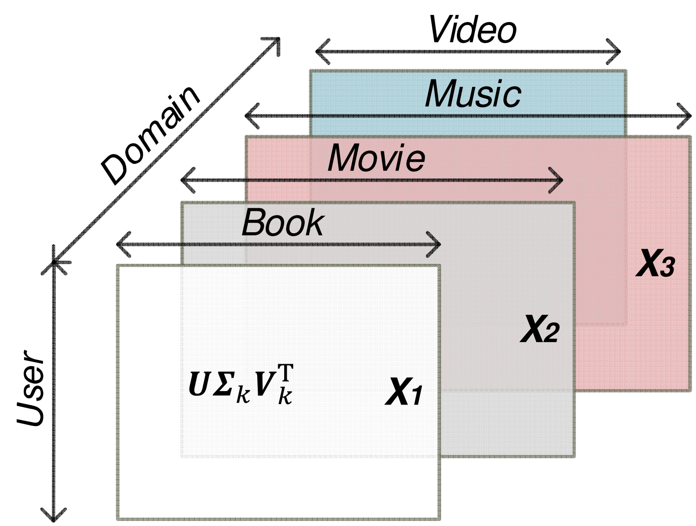




Collective Matrix Factorization (CMF)

Sum loss over all domains:

$$\underset{\boldsymbol{U},\boldsymbol{V},\boldsymbol{C}}{argmin}\frac{1}{2}\sum_{k=1}^{K}\left\|\boldsymbol{W}_k \circledast \left(\boldsymbol{X}_k - \boldsymbol{U}\boldsymbol{V}_k^{\mathrm{T}}\right)\right\|_F^2 + \frac{\lambda_U}{2}\|\boldsymbol{U}\|^2 + \frac{\lambda_V}{2}\sum_{k=1}^{K}\|\boldsymbol{V}_k\|_F^2$$



Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., & Yang, D. (2016). Learning Informative Priors from Heterogeneous Domains to Improve Recommendation in Cold-Start User Domains. *ACM Transactions on Information Systems (TOIS)*, *35*(2), 13.

Weighted Irregular Tensor Factorization

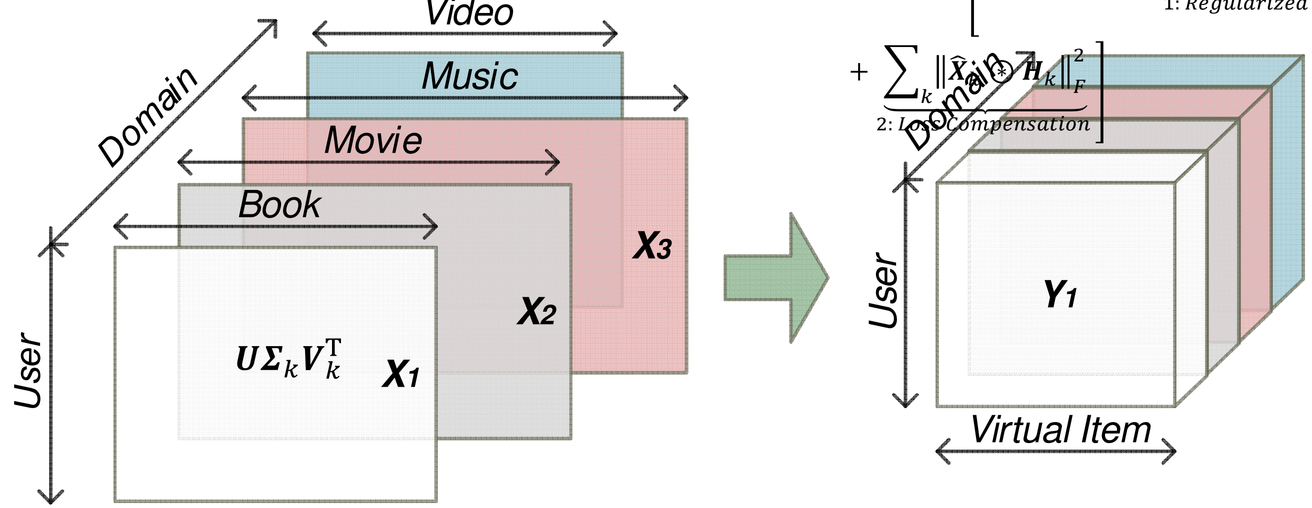
Sum loss over all domains:

$$\underset{\boldsymbol{U},\boldsymbol{V},\boldsymbol{C}}{argmin}\frac{1}{2}\sum_{k=1}^{K}\left\|\boldsymbol{W}_k \circledast \left(\boldsymbol{X}_k - \boldsymbol{U}\boldsymbol{\Sigma}_k\boldsymbol{V}_k^{\mathrm{T}}\right)\right\|_F^2 + \frac{\lambda_U}{2}\|\boldsymbol{U}\|^2 + \frac{\lambda_V}{2}\|\boldsymbol{V}\|^2 + \frac{\lambda_C}{2}\|\boldsymbol{C}\|^2$$

where $\boldsymbol{\Sigma}_k = diag(\boldsymbol{C}_k)$

With orthonormal constraints, we can obtain equivalent loss:

$$\underset{\boldsymbol{U},\boldsymbol{V},\boldsymbol{C}}{argmin}\frac{1}{2}\left[\underbrace{\left(\|\boldsymbol{\mathcal{Y}} - [\![\boldsymbol{U},\boldsymbol{V},\boldsymbol{C}]\!]\|^2 + \lambda_U\|\boldsymbol{U}\|_F^2 + \lambda_V\|\boldsymbol{V}\|_F^2 + \lambda_C\|\boldsymbol{C}\|_F^2\right)}_{1:\, Regularized\ TF\ Model} + \underbrace{\sum_k\left\|\widehat{\boldsymbol{X}}_k \circledast \boldsymbol{H}_k\right\|_F^2}_{2:\, Loss\ Compensation}\right]$$



Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., & Yang, D. (2016). Learning Informative Priors from Heterogeneous Domains to Improve Recommendation in Cold-Start User Domains. *ACM Transactions on Information Systems (TOIS)*, *35*(2), 13.

Handling miss values

- For rating data
  - Add weight matrix
    - $w_{k,i,j} = \begin{cases} 1 & (k,i,j)\ is\ an\ observation \\ a & (k,i,j)\ is\ a\ noisy\ example \\ 0 & else \end{cases}$
    - Noisy data act as regularization
- For one-class data
  - Users may deliberately choose to access which items [Marlin *et al*, 2007]
  - Confidence Modeling[Hu *et al*, 2008]
    - $w_{k,i,j} = \begin{cases} c_{k,i,j} + 1 & (k,i,j)\ is\ observed \\ 1 & else \end{cases}$

Marlin, B.M., Zemel, R.S., Roweis, S., and Slaney, M. Collaborative filtering and the missing at random assumption. In *Proceeding 23rd Conference on Uncertainty in Artificial Intelligence*, 2007.
Hu, Y., Koren, Y., and Volinsky, C. Collaborative Filtering for Implicit Feedback Datasets. In *Eighth IEEE International Conference on Data Mining*, 263-272, 2008.

Epinions dataset (ratings)

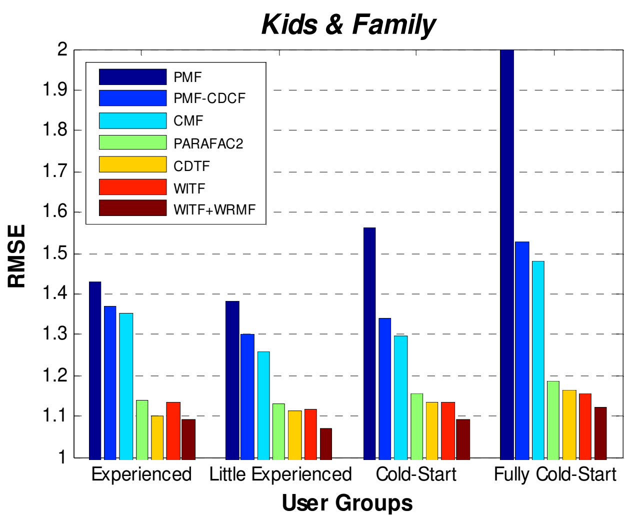
- Covering 5 domains

Domain	# Items	# Ratings / # Users	# Ratings / # Items	Sparsity
Kids & Family*	3,769	4.9309	9.9077	0.0013
Hotels & Travel*	2,545	3.9210	11.6676	0.0015
Restaurants & Gourmet	2,543	3.3394	9.9446	0.0013
Wellness & Beauty	3,852	3.5481	6.9756	0.0009
Home and Garden	2,785	2.6003	7.0707	0.0009

http://liris.cnrs.fr/red/

Performance over users grouped by # ratings

- RMSE of comparative methods (the smaller the better)



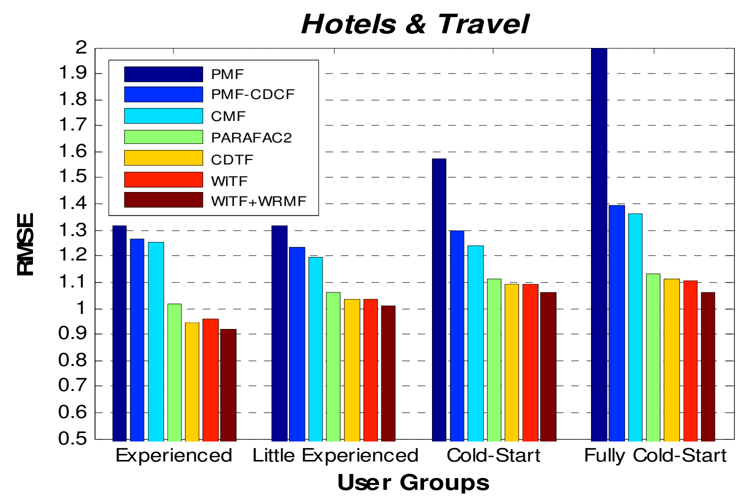


Table IV. Statistics of Testing Users Grouped by the Number of Ratings

Target Domain / Group	# Ratings	*Kids & Family*	*Hotels & Travel*
		# testing users in TS-50%	**# testing users in TS-50%**
Experienced	> 20	120	55
Little Experienced	6 ~ 20	816	517
Cold-Start	1 ~ 5	2,260	2,807
Fully Cold-Start	0	695	1,072

Tmall.com dataset (clicks)

- One-class preference problem

Domain	# Items	# Clicks / # Users	# Clicks / # Items	Sparsity
*D1**	8,179	23.2003	19.7170	0.0028
*D2**	6,940	18.5455	18.5749	0.0027
D3	5,561	22.5005	28.1246	0.0040
D4	6,145	16.0606	18.1671	0.0026

https://tianchi.aliyun.com/datalab/dataSet.htm?id=5

The Mean AP@5,10 and nDCG@5,10

Target Domain / Method	D1							
	TR-80%				TR-50%			
	AP@5	AP@20	nDCG@5	nDCG@20	AP@5	AP@20	nDCG@5	nDCG@20
Most-Pop	0.0161^	0.0175^	0.0269^	0.0382^	0.0322^	0.0223^	0.0567^	0.0577^
N-CDCF	0.0252*	0.0240*	0.0441*	0.0465*	0.0352*	0.0210	0.0604*	0.0534
MF-IF	0.0263*	0.0293*	0.0432*	0.0631*	0.0455*	0.0324	0.0813*	0.0854*
MF-IF-CDCF	0.0242*	0.0258*	0.0399*	0.0552*	0.0431*	0.0296	0.0763*	0.0775*
PARAFAC2	0.0213*	0.0226*	0.0350*	0.0476*	0.0395*	0.0267	0.0691*	0.0687*
CDTF-IF	0.0258*	0.0276*	0.0425*	0.0587*	0.0423*	0.0294	0.0758*	0.0767*
WITF	**0.0267***	**0.0285***	**0.0451***	**0.0623***	**0.0484***	**0.0340**	**0.0849***	**0.0872***
WITF+WRMF	**0.0271****	**0.0290****	**0.0462****	**0.0643****	**0.0486****	**0.0343****	**0.0851****	**0.0879****

Target Domain / Method	D2							
	TR-80%				TR-50%			
	AP@5	AP@20	nDCG@5	nDCG@20	AP@5	AP@20	nDCG@5	nDCG@20
Most-Pop	0.0175^	0.0194^	0.0288^	0.0424^	0.0297^	0.0231^	0.0530^	0.0591^
N-CDCF	0.0281*	0.0261*	0.0435*	0.0520*	0.0228	0.0243*	0.0380	0.0357
MF-IF	0.0320*	0.0354*	0.0528*	0.0747*	0.0501*	0.0370*	0.0872**	0.0924**
MF-IF-CDCF	0.0240*	0.0262*	0.0397*	0.0563*	0.0380*	0.0285*	0.0675	0.0724*
PARAFAC2	0.0215*	0.0234*	0.0356*	0.0506*	0.0327*	0.0251*	0.0589*	0.0638*
CDTF-IF	0.0326*	0.0337*	0.0526*	0.0662*	0.0454*	0.0316*	0.0761*	0.0750*
WITF	**0.0338***	**0.0363***	**0.0552***	**0.0753***	**0.0538***	**0.0383***	**0.0905***	**0.0909***
WITF+WRMF	**0.0343****	**0.0369****	**0.0556****	**0.0758****	**0.0542****	**0.0386****	**0.0907****	**0.0915***

Non-IIDness on items

- Cross-domain RS: domain coupling
  - Item domains
    - Latent factor model
    - Deep learning model
  - Modality domains
    - Multimodal RS
  - Open issues and directions
- Session-based RS: sequential coupling

Non-IIDness on items

- Cross-domain RS
- Session-based RS

Non-IIDness on implicit interaction

- Context-aware RS
- Multi-objective RS
- Attraction RS

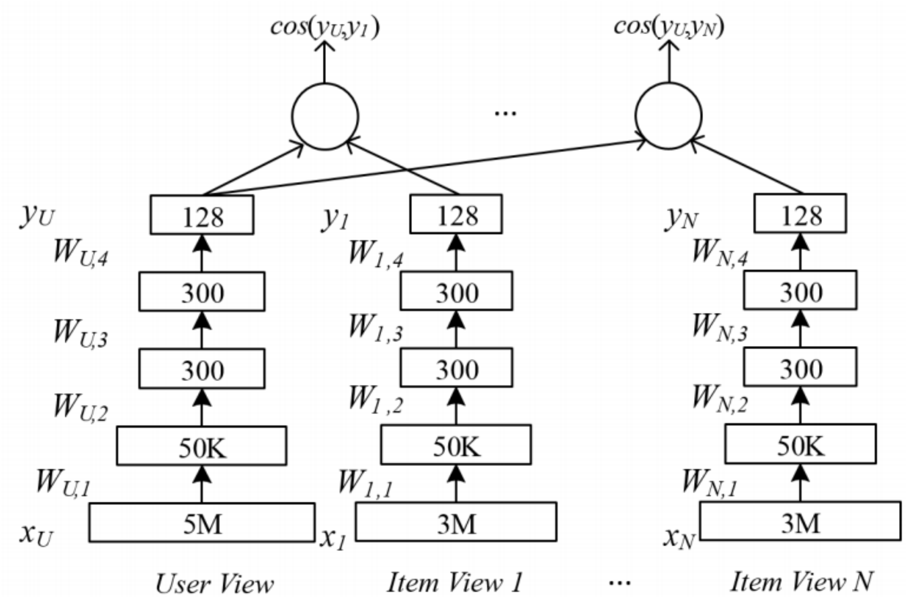
Non-IID RS in practice

A Multi-View Deep Learning Approach

- Multi-learning Framework
  - One user view VS. multiple item views
  - DNN to map high-dimensional sparse features (e.g., raw features of users and items) into low dimensional dense features in a joint semantic space

Type	DataSet	UserCnt	Feature Size	Joint Users
User View	Search	20M	3.5M	/
Item View	News	5M	100K	1.5M
	Apps	1M	50K	210K
	Movie/TV	60K	50K	16K

Table 1: Statistics of the four data sets used in this paper. The *Joint Users* column indicates the number of common users between each item view and the user view.



Elkahky, A.M., Song, Y., and He, X. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th International Conference on World Wide Web*, 278-288, 2015.

User log for Microsoft products

- The data sets:
  - Search engine logs from Bing Web vertical
  - News article browsing history from Bing News vertical
  - App download logs from Windows AppStore
  - Movie/TV view logs from Xbox.

Data Set	Training			Testing		
	Number Of unique users	Number of unique items	Number of training pairs	Number of new users	Number of test pairs for old users	Number of test pairs for new users
Apps Data	200K	55k	2.5M	1K	11K	2K
News Data	1.5M	5M	> 1B	5K	50K	10K
Xbox Data	16K	10K	45K	1K	10K	3K

Mapping between URL domains, News articles and Apps

User View with Single Domain ID Feature	Top Matched News	Top Matched Apps
barackobama.com	Obama to Delay Obamacare Again to Help Democrats Froma Harrop: Democrats should not run away from Obamacare Democratic Senator: I am willing to defy Obama Governor Jindal proposes Republican alternative to Obamacare	7 Minutes Fitter Relax Meditate Escape Sleep Sleep Tracker U.S. Constitution
spiegel.de	Nazi-Era Jerseys on View in World Cup Exhibit 2014 World Cup Day 3 Lessons: Colombia Fun In The Sun... Belgium Vs. Algeria World Cup 2014: Live Stream... Colombia vs. Ivory Coast: Tactical Preview ...	ESPN Cricinfo Golf News RSS Pulse News Dinamalar - Tamil News Paper
linkedin.com	RectorSeal, ... Acquires Assets of Resource Conservation... Berkshire Partners Teams With Glen T. Senk To Co-Invest ... TF Financial: National Penn Bancshares, Inc. to Acquire ... H.I.G. Capital Portfolio Company Surgery Partners to Acquire ...	LinkedIn App LinkedIn Touch The Economist on Windows The Wall Street Journal
babycenter.com	Jenelle Evans' Baby Name: What We Know Catelynn Lowell ... Are Reportedly Pregnant With Baby #2! Jenelle Evans Can Take Drugs During Pregnancy If She Wants Pregnant Jenelle Evans: What Should She Name Her Baby?	Parents Pregnancy & Baby Guide ANIMALS FOR KIDS GAME Minecraft Fan Hub GS Preschool Games

MRR and Precision@1

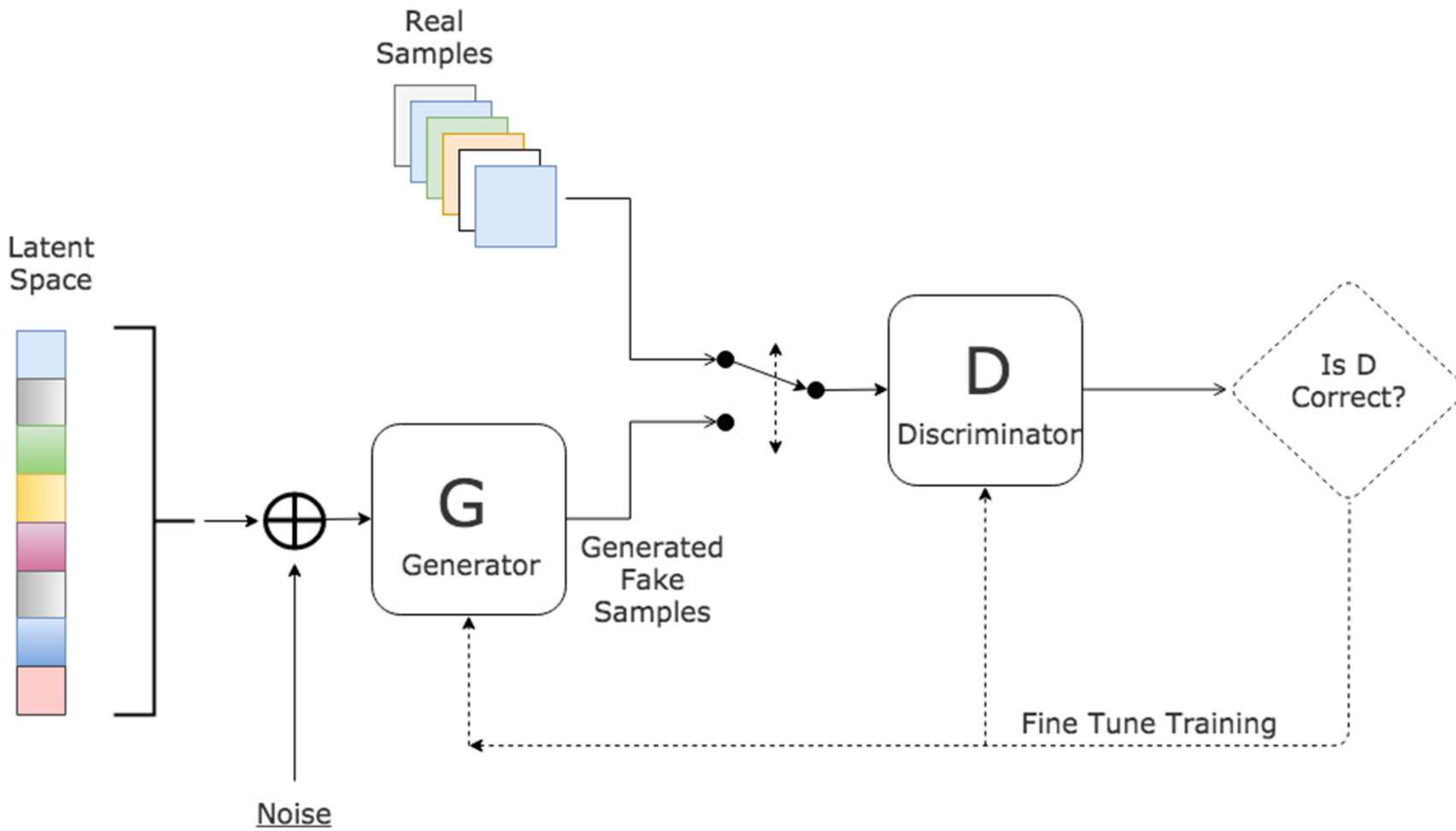
	Algorithm	All Users		New Users	
		MRR	P@1	MRR	P@1
I	Most Frequent	0.298	0.103	0.303	0.119
	CF	0.337	0.142	/	/
	CCA (TopK) [29]	0.295	0.105	0.295	0.104
	CTR [32]	0.448	0.277	0.319	0.142
II	SV- Kmeans	0.359	0.159	0.336	0.154
	SV-LSH	0.372	0.169	0.339	0.158
	SV-TopK	**0.497**	**0.315**	**0.436**	**0.268**
III	MV-Kmeans	0.362	0.16	0.339	0.156
	MV-TopK	0.517	0.335	0.466	0.297
	MV-TopK w/ Xbox	**0.527**	**0.347**	**0.473**	**0.306**

Table 3: Results for different algorithms on Windows Apps Data Set. Type *I* algorithms are baseline methods we compare with. Type *II* are single user-item view methods trained using the original DSSM framework. Type *III* are multi-view DNN models we proposed. The best performance is achieved by training a MV-DNN on all three user-item views with TopK as feature selection method.

	Algorithm	All Users		New Users	
		MRR	P@1	MRR	P@1
I	Most Frequent	0.301	0.111	0.305	0.111
	CTR [32]	0.427	0.215	0.276	0.123
II	SV-Kmeans	0.386	0.192	0.294	0.143
	SV-LSH	0.45	0.247	0.34	0.186
	SV-TopK	**0.486**	**0.286**	**0.358**	**0.208**
III	MV-Kmeans	0.391	0.194	0.296	0.145
	MV-TopK	0.494	0.303	0.368	0.222
	MV-TopK w/ Xbox	**0.503**	**0.321**	**0.398**	**0.245**

Table 4: Results for the News Data Set. Similarly, the best performance is achieved by our multi-view models. Note that due to the extreme big size of this data set ($>$ 1B entries), traditional algorithms like CF (SVD) and CCA failed to handle it due to memory constraint.

Generative Adversarial Network

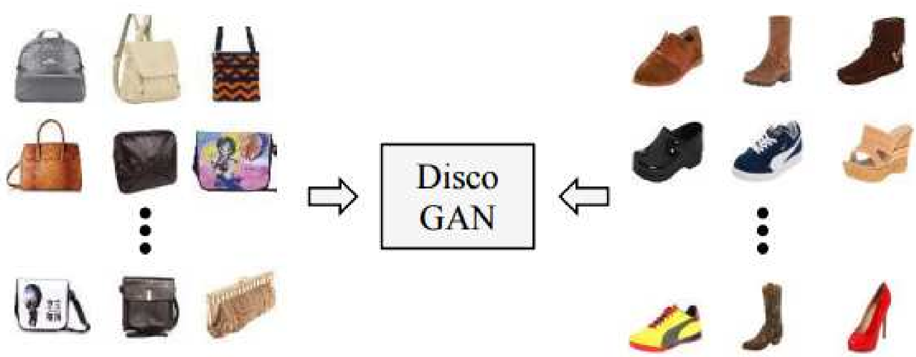


https://www.kdnuggets.com/2017/01/generative-adversarial-networks-hot-topic-machine-learning.html

Ian Goodfellow, NIPS 2016 Tutorial: Generative Adversarial Networks

Content-based Cross-domain Recommendation with Generative Adversarial Networks

- Discovering cross-domain relations given unpaired data.



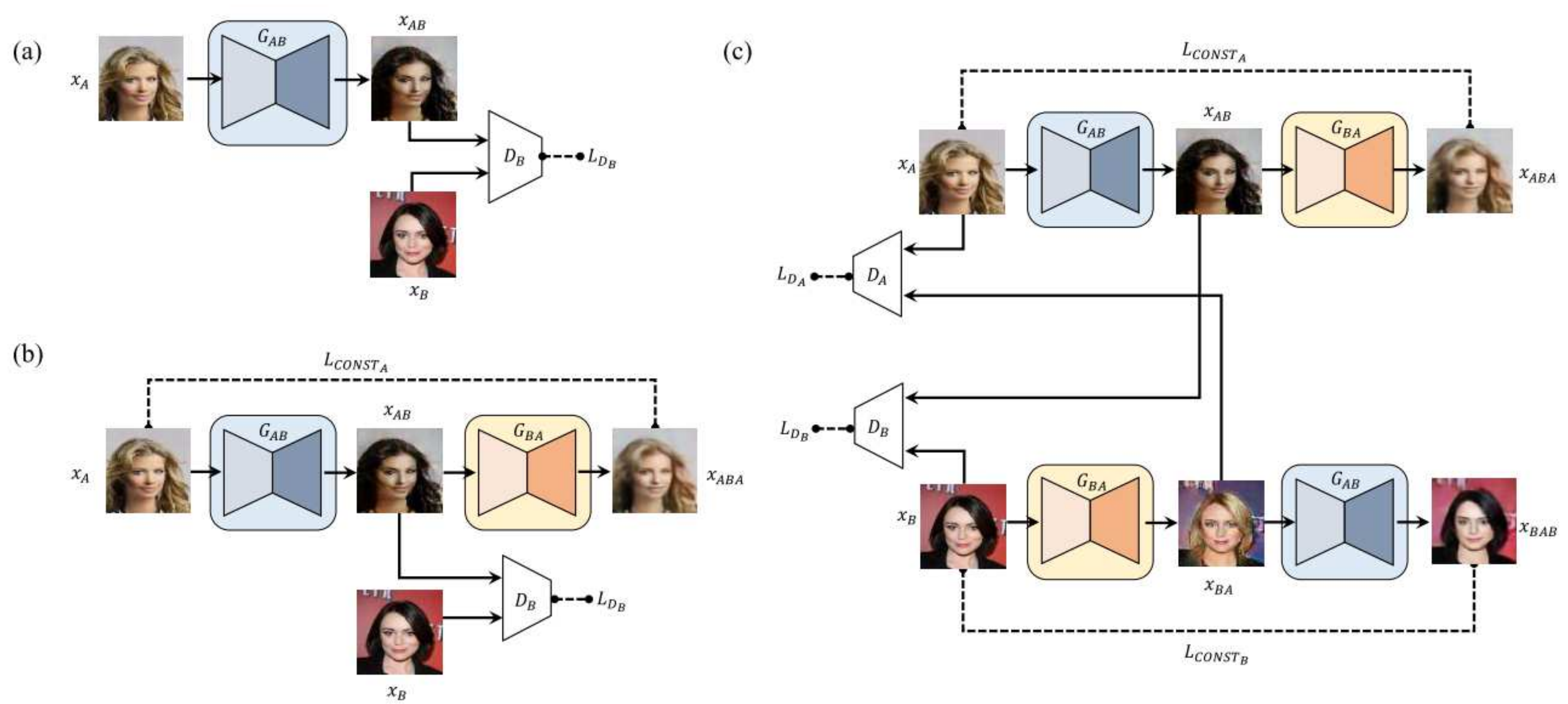
(a) Learning cross-domain relations **without any extra label**

(b) Handbag images (input) & **Generated** shoe images (output)

(c) Shoe images (input) & **Generated** handbag images (output)

Kim, T., Cha, M., Kim, H., Lee, J., & Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. *arXiv preprint arXiv:1703.05192*.

DiscoGAN for unpaired, unlabeled datasets



(a) Standard GAN (b) GAN with a reconstruction loss, (c) DiscoGAN designs two coupled GAN between two unpaired, unlabeled datasets.

Datasets

- Car dataset (Fidler et al., 2012)
  - Fidler, S., Dickinson, S., and Urtasun, R. 3d object detection and viewpoint estimation with a deformable 3d cuboid model. In NIPS, 2012.
- Chair dataset (Paysan et al., 2009)
  - Aubry, M., Maturana, D., Efros, A. A., Russell, B., and Sivic, J. Seeing 3d chairs: Exemplar part-based 2d-3d alignment using a large dataset of cad models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
- Handbags dataset
  - Zhu, Jun-Yan, Kr¨ahenb¨uhl, Philipp, Shechtman, Eli, and Efros, Alexei A. Generative visual anipulation on the natural image manifold. In roceedings of EuropeanConference on Computer Vision (ECCV), 2016.
- Shoe dataset
  - Yu, A. and Grauman, K. Fine-grained visual comparisons with local learning. In Computer Vision and Pattern Recognition (CVPR), June 2014.

Chair to Car Translation

(a) Chair to Car

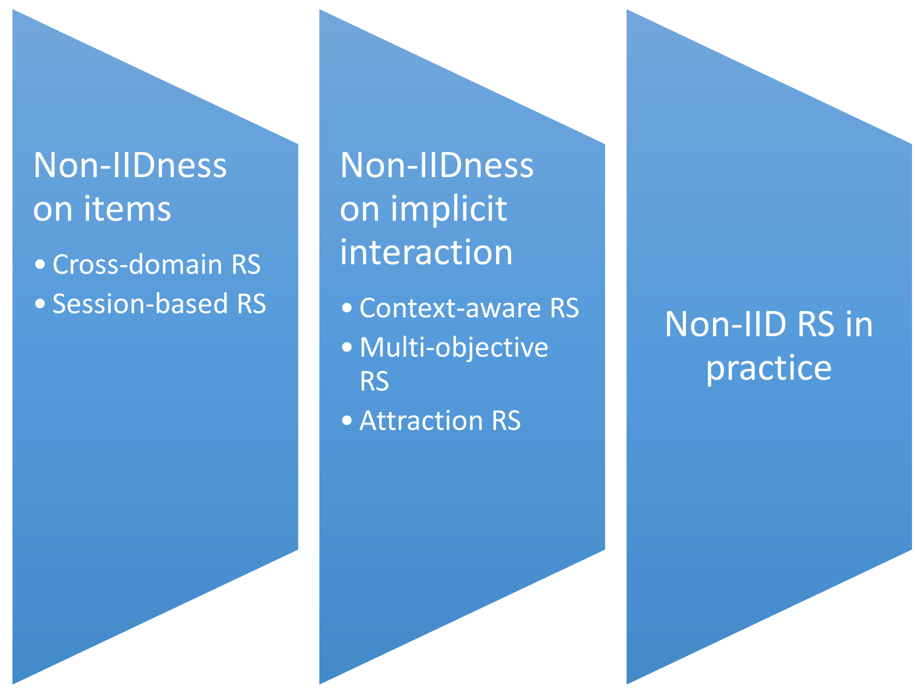
Discovering relations of images from visually very different object classes.
DiscoGAN is trained on chair and car images

Recommend Items from Sketches

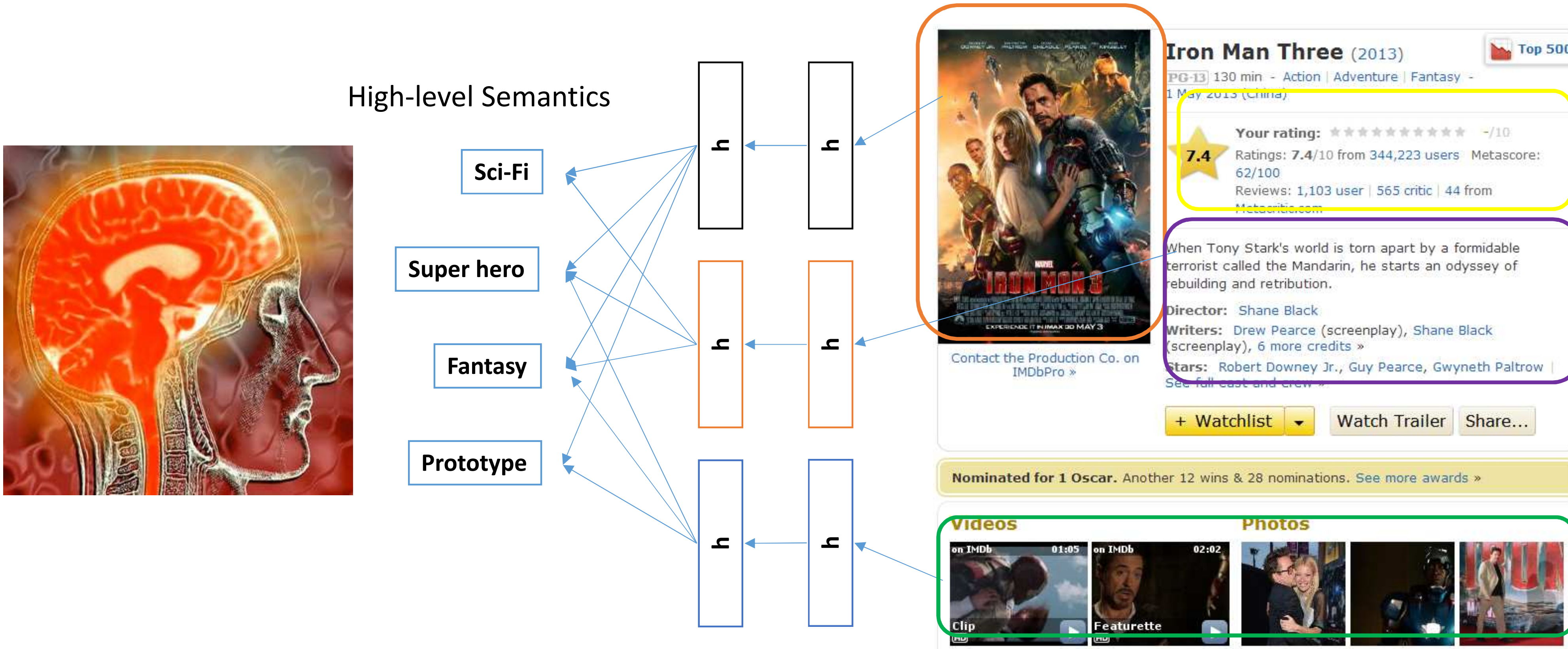
- (a) colored images of handbags are generated from sketches of handbags
- (b) colored images of shoes are generated from sketches of shoes
- (c) sketches of handbags are generated from colored images of handbags

Non-IIDness on items

- Cross-domain RS: domain coupling
  - Item domains
    - Latent factor model
    - Deep learning model
  - Modality domains
    - Multimodal RS
  - Open issues and directions
- Session-based RS: sequential coupling



Human are Joint Thinking with Related Data



Multimodal Learning

- The information in real world usually comes as different modalities.
  - Images are usually associated with tags and text;
  - Texts contain images to more clearly express the main idea of the article.

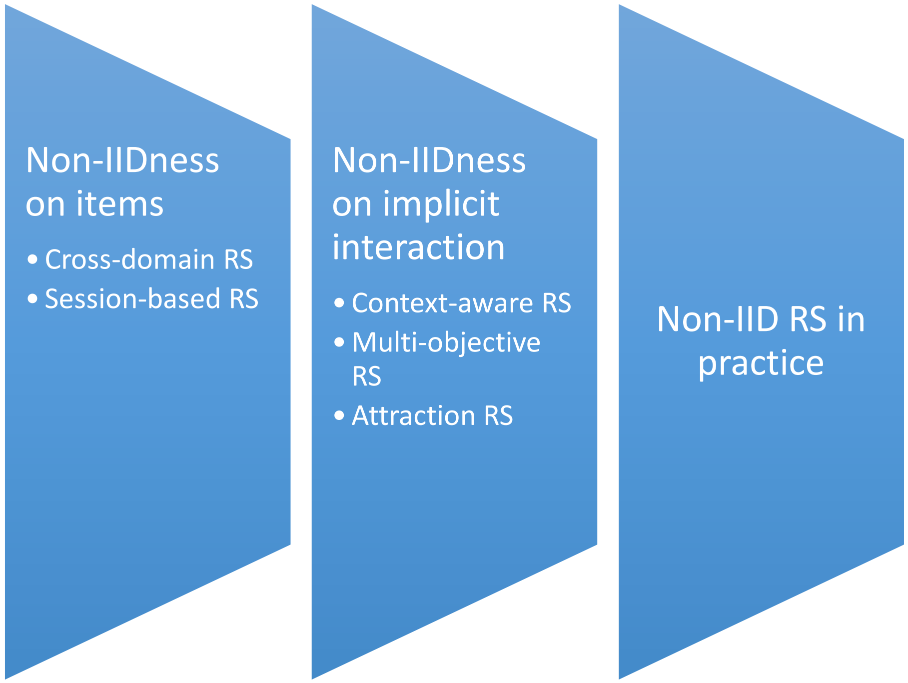
- Different modalities are characterized by very different statistical properties.

- **Multimodal learning** aims to learn a joint representation of different modalities.

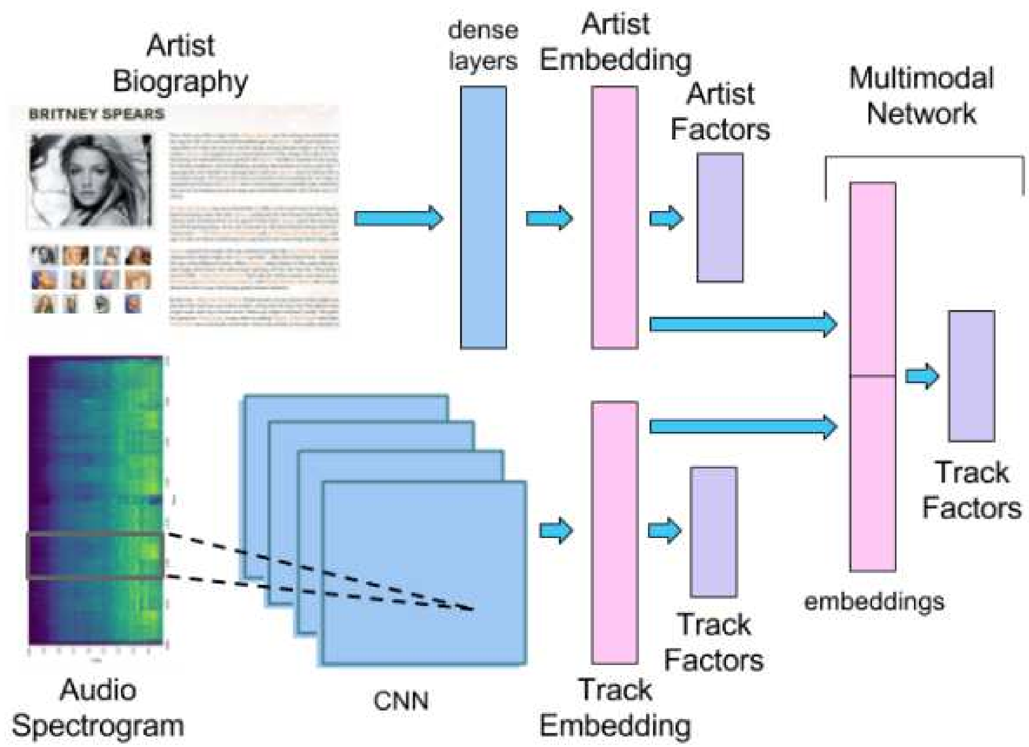
https://en.wikipedia.org/wiki/Multimodal_learning

Non-IIDness on items

- Cross-domain RS: domain coupling
  - Item domains
    - Latent factor model
    - Deep learning model
  - Modality domains
    - Multimodal RS
      - Multimodal Music Recommendation
      - Multimodal learning for images and texts
  - Open issues and directions
- Session-based RS: sequential coupling



Multimodal Music Recommendation



Oramas, S., Nieto, O., Sordo, M., & Serra, X. (2017). A deep multimodal approach for cold-start music recommendation. *arXiv preprint arXiv:1706.09739*.

Datasets

- Million Song Dataset (MSD)
  - https://labrosa.ee.columbia.edu/millionsong/
  - Echo Nest Taste Profile Subset provides play counts of 1 million users on more than 380,000 songs from the MSD
  - Biographies and social tags are collected from Last.fm for all the artists that have at least one song in the dataset.
- Final Dataset (MSD-A)
  - https://zenodo.org/record/831348
  - The dataset consists of 328,821 tracks from 24,043 artists. Each track has at least 15 seconds of audio, each biography is at least 50 characters long, and each artist has at least 1 tag associated with it.

Results of Artist and Song Recommendation

Table 1: Artist Recommendation Results

Aproach	Input	Data model	Arch	MAP
A-TEXT	Bio	VSM	FF	0.0161
A-SEM	**Sem Bio**	**VSM**	**FF**	**0.0201**
A-W2V-GOO	Bio	w2v-pretrain	CNN	0.0119
A-W2V	Bio	w2v-trained	CNN	0.0145
A-TAGS	Tags	VSM	FF	0.0314
TAGS-ITEMKNN	Tags	-	itemKnn	0.0161
TEXT-RF	Bio	VSM	RF	0.0089
RANDOM	-	-	-	0.0014
UPPER-BOUND	-	-	-	0.5528

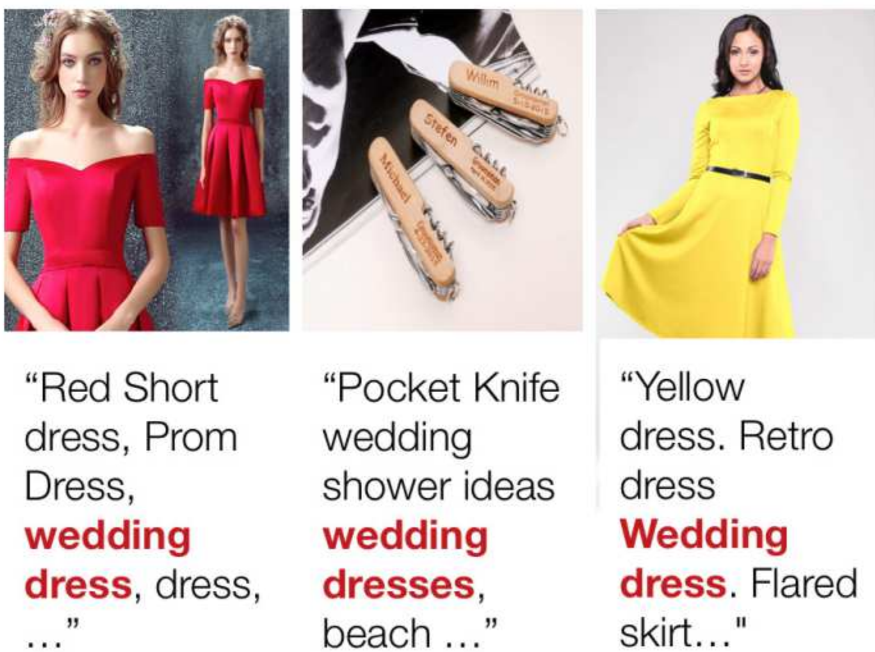
Mean average precision (MAP) at 500 for the predictions of artist recommendations in 1M users. VSM refers to Vector Space Model, FF to Feedforward, RF to Random Forest, CNN to Convolutional Neural Network, and itemKnn to itemAttributeKnn approach. Bio refers to biography texts and Sem Bio to semantically enriched texts.

Table 2: Song Recommendation Results

Approach	Artist Input	Track Input	Arch	MAP
AUDIO	-	audio spec	CNN	0.0015
SEM-VSM	Sem Bio	-	FF	0.0032
SEM-EMB	A-SEM	-	FF	0.0034
MM-LF-LIN	**A-SEM**	**AUDIO emb**	**MLP**	**0.0036**
MM-LF-H1	**A-SEM**	AUDIO emb	MLP	0.0035
MM	Sem Bio	audio spec	CNN	0.0014
TAGS-VSM	Tags	-	FF	0.0043
TAGS-EMB	A-TAGS	-	FF	0.0049
RANDOM	rnd emb	-	FF	0.0002
UPPER-BOUND	-	-	-	0.1649

Mean average precision (MAP) at 500 for the predictions of song recommendations in 1M users. AUDIO emb refers to the track embedding of AUDIO approach, SEM to artist embedding of SEM approach, TAGS to artist embedding of TAGS approach, spec to spectrogram, mm to multimodal, lf to late fusion, lin to linear, and h1 to one hidden layer.

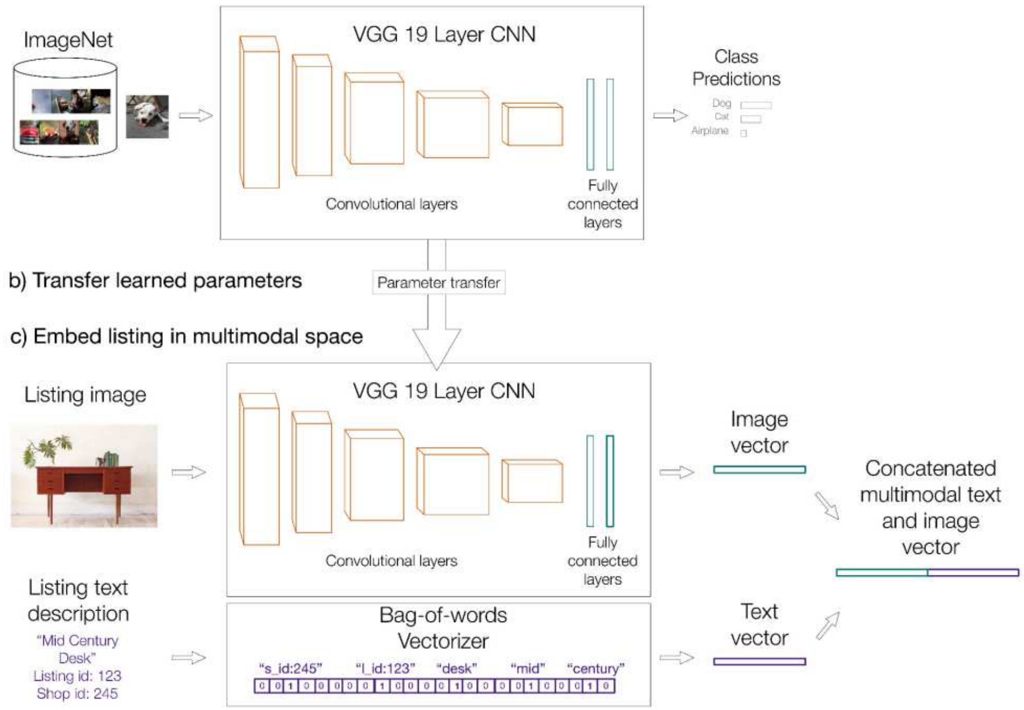
Multimodal learning for images and texts



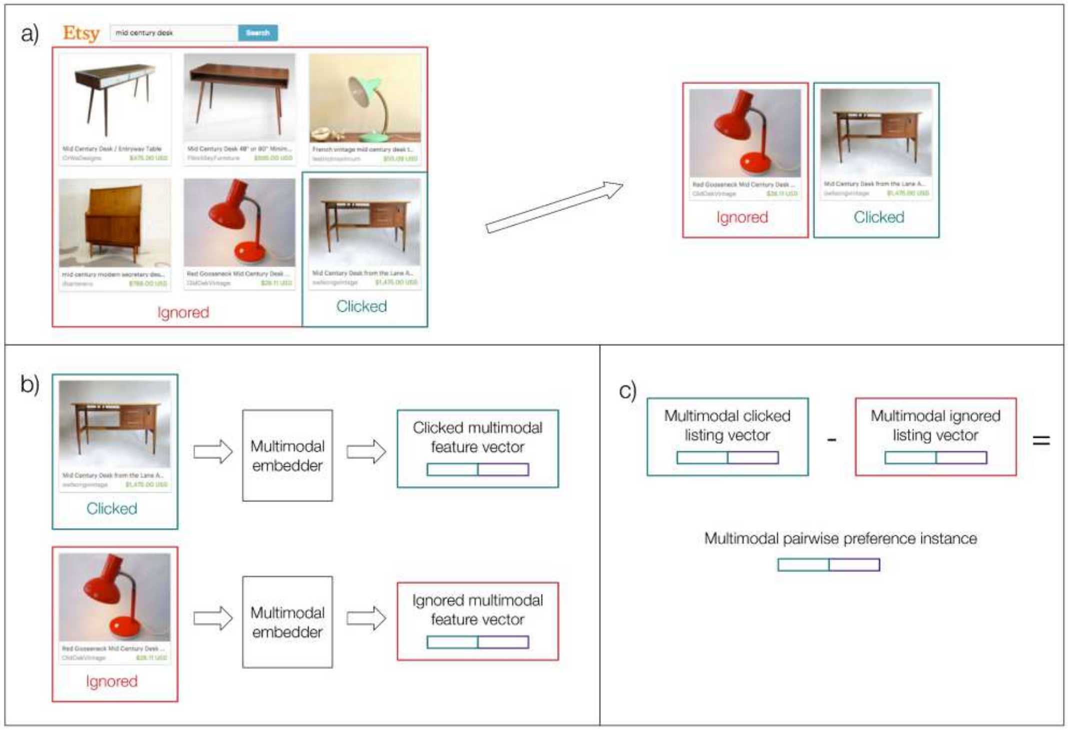
- Irrelevant search results for the query "wedding dress"
- Even though it's apparent in the images that these are not wedding dresses

Lynch, C., Aryafar, K., and Attenberg, J. Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 541-548, 2016.

Transferring Parameters of A CNN to The Task of Multimodal Embedding



From search logs to multimodal pairwise classification instances



Datasets

- https://www.etsy.com/
- 2 week period in search logs, 1.4 million Etsy listings with images.
- Related dataset:
  - http://vision.is.tohoku.ac.jp/~kyamagu/research/etsy-dataset/

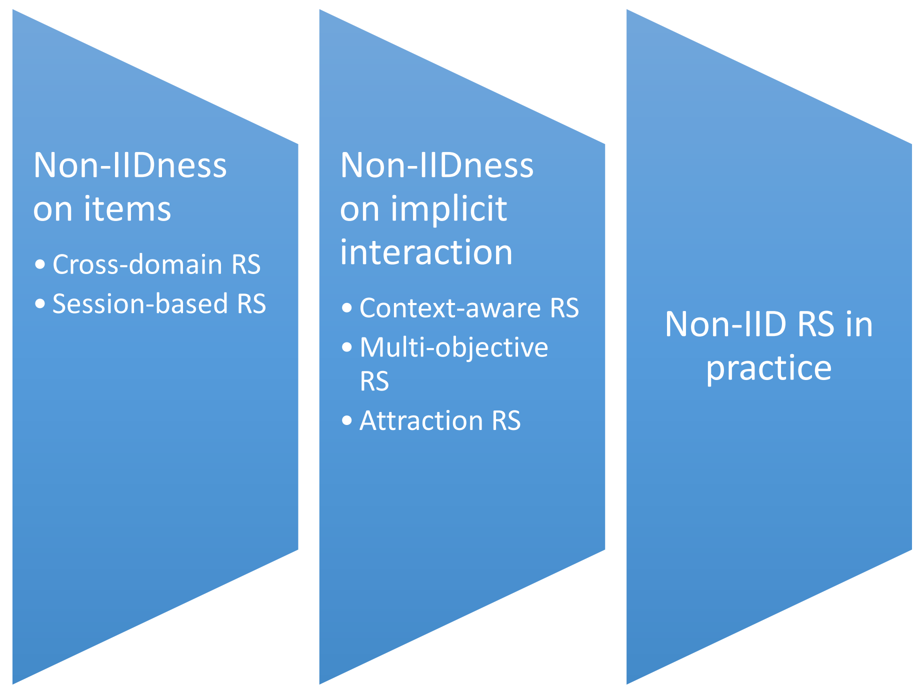
Image information can help disentangle different listings considered similar by a text model

(a) (b) (c)

Visualizing ranking changing by incorporating image information

Non-IIDness on items

- Cross-domain RS: domain coupling
  - Item domains
    - Latent factor model
    - Deep learning model
  - Modality domains
    - Multimodal RS
  - Open issues and directions
- Session-based RS: sequential coupling



Open issues and directions

- Information and Influence adaptation
  - What information should be imposed from which domains?
  - How much information should be imposed for each domain?
  - How to integrate the heterogeneous information from multiple domains?
- Non-overlap cross-domain learning
  - Joint learning complementary information without overlapped users and items
- How to utilize multi-modal data in RS?
  - Appling GAN-based models to generate multiple types of samples

Application of GAN in RS

- Preview virtual images of item from NLP description

- Generate virtual data to relieve data sparsity

GAN for Generating Images by Text

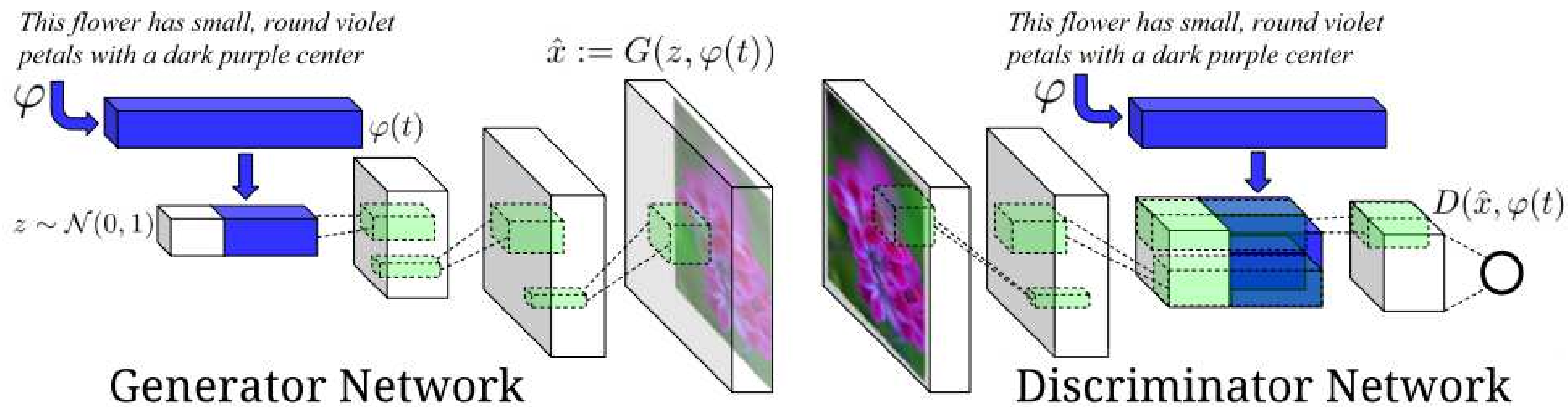


Figure 2. Our text-conditional convolutional GAN architecture. Text encoding $\varphi(t)$ is used by both generator and discriminator. It is projected to a lower-dimensions and depth concatenated with image feature maps for further stages of convolutional processing.

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis. *arXiv preprint arXiv:1605.05396*.

Text-to-image Synthesis

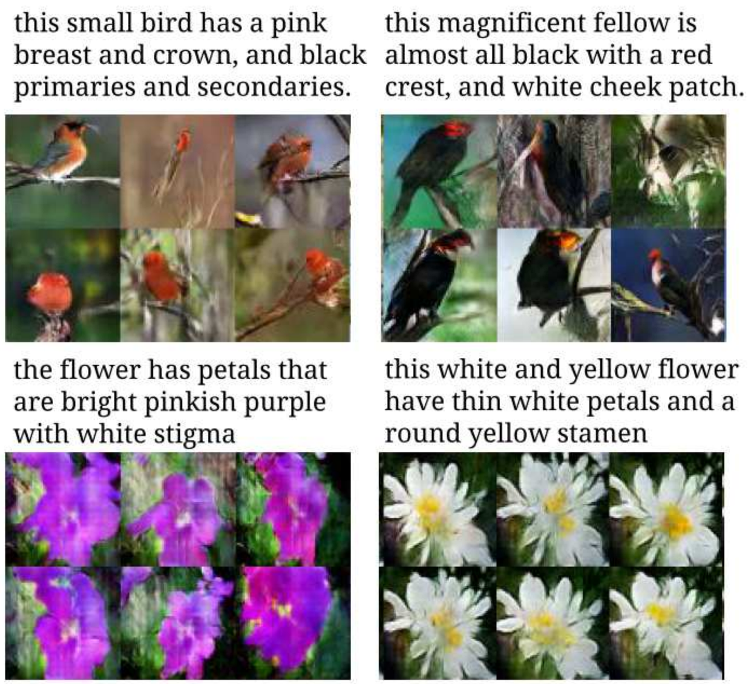


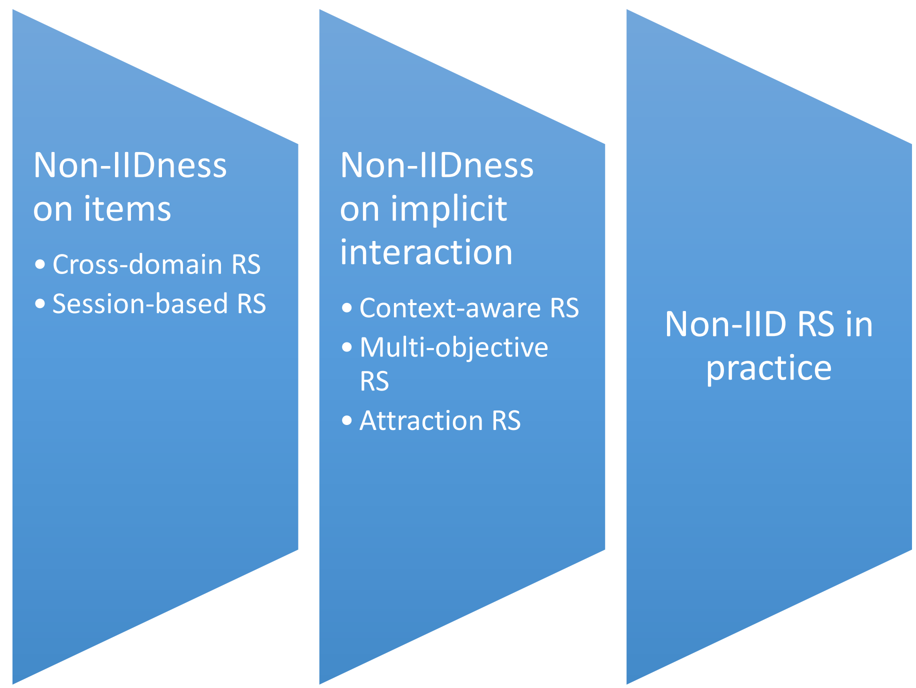
Figure 23: Text-to-image synthesis with GANs. Image reproduced from Reed *et al.* (2016b).

Figure 25: StackGANs are able to achieve higher output diversity than other GAN-based text-to-image models. Image reproduced from Zhang *et al.* (2016).

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis.
Zhang, H., Xu, T., Li, H., Zhang, S., Huang, X., Wang, X., and Metaxas, D. (2016). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks

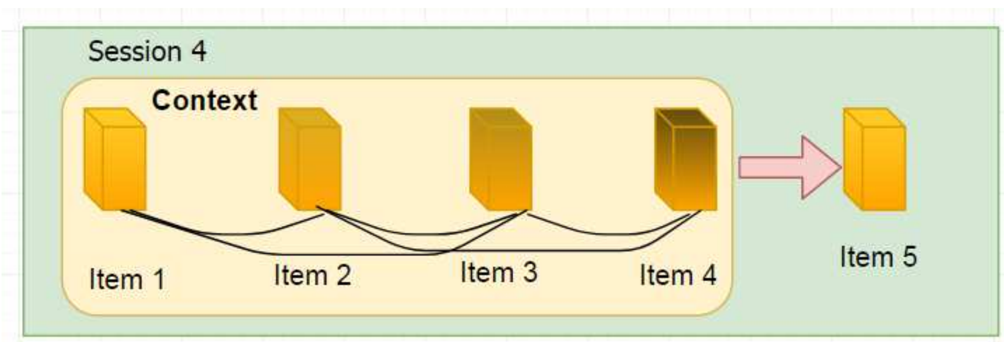
Non-IIDness on items

- Cross-domain RS: domain coupling
- Session-based RS: sequential coupling
  - What is a session?
  - First-order dependency modeling
    - Markov chain-based matrix factorization
  - Higher-order dependency modeling
    - RNN based session modeling
    - Encoder-decoder based session modeling
  - Loosely ordered dependency modeling
    - SWIWO model and its extensions
  - Open issues and directions



What is a session?

- A session consists of observed sequence that leads to the consequent actions.
  - There is couplings between the items within a session.
    - e.g., clicked pages in browsing history, or chosen items in a transaction.

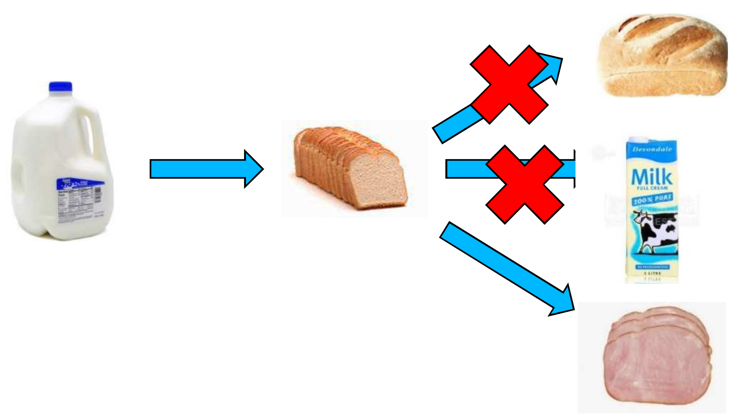


Why modeling session?

- Recommender systems built on historical profile are often **repeatedly recommended similar items**.
  - E.g. neighborhood-based methods, matrix factorization methods
- In most real-world scenarios, we prefer to find items that are **relevant** to our recent activities instead of only **similar** items.
- A system makes more sensible and relevant recommendations if the **session** was taken into consideration.

Diversifying recommendations

- Users prefer **more diversified options** than those they have had.
  - It is unlikely that a custom will purchase another loaf of bread if they have purchased one, whereas butter or ham may be a more appealing recommendation.
- A system makes more sensible and relevant recommendations if the **session** was taken into consideration.



Non-IIDness on items

- Cross-domain RS: domain coupling
- Session-based RS: sequential coupling
  - What is a session?
  - First-order dependency modeling
    - Markov chain-based matrix factorization
  - Higher-order dependency modeling
    - RNN based session modeling
    - Encoder-decoder based session modeling
  - Loosely ordered dependency modeling
    - SWIWO model and its extensions
  - Open issues and directions

Non-IIDness on items

- Cross-domain RS
- Session-based RS

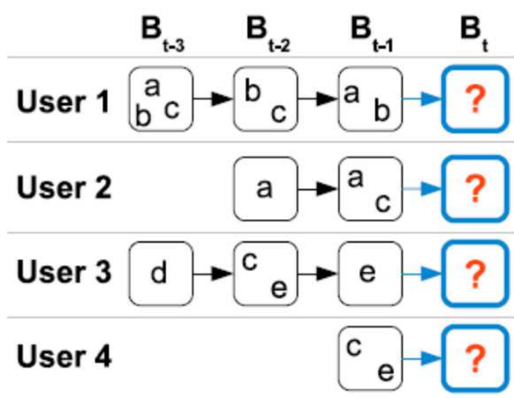
Non-IIDness on implicit interaction

- Context-aware RS
- Multi-objective RS
- Attraction RS

Non-IID RS in practice

Next-basket recommendation

- Sequential shopping basket data is given per user
- To recommend the items which the user may buy in his next visit



Rendle, S., Freudenthaler, C., and Schmidt-Thieme , L. (2010, August). Factorizing Personalized Markov Chains for Next-Basket Recommendation. WWW2010.

Markov chain based matrix factorization

- It models the *pairwise* interaction in <***user u, item i, item l***>:

- For each interaction mode, the pair of factorization matrices are :

$$\hat{a}_{u,l,i} := \langle v_u^{U,I}, v_i^{I,U} \rangle + \langle v_i^{I,L}, v_l^{L,I} \rangle + \langle v_u^{U,L}, v_l^{L,U} \rangle$$

$$U - I : V^{U,I} \in \mathbb{R}^{|U| * k_{U,I}}, V^{I,U} \in \mathbb{R}^{|I| * k_{U,I}}$$
$$I - L : V^{I,L} \in \mathbb{R}^{|I| * k_{I,L}}, V^{L,I} \in \mathbb{R}^{|I| * k_{I,L}}$$
$$U - L : V^{U,L} \in \mathbb{R}^{|U| * k_{U,L}}, V^{L,U} \in \mathbb{R}^{|I| * k_{U,L}}$$

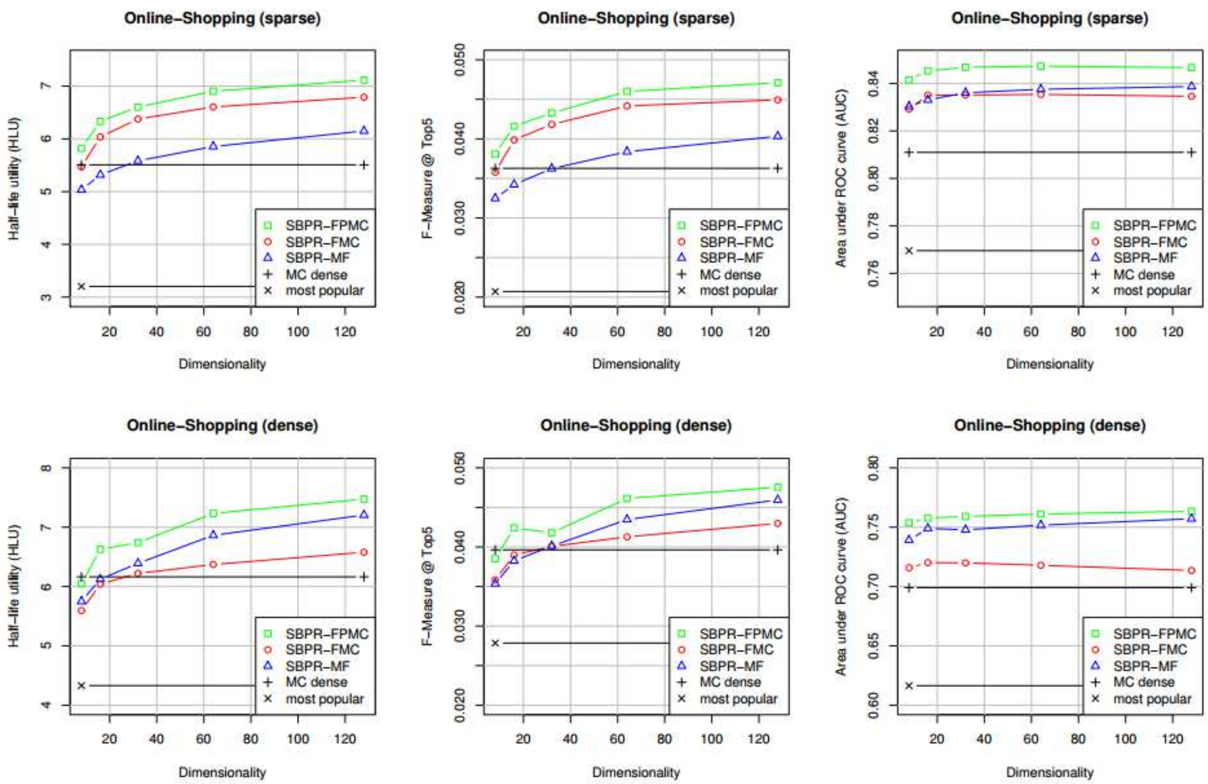
Rendle, S., Freudenthaler, C., and Schmidt-Thieme , L. (2010, August). Factorizing Personalized Markov Chains for Next-Basket Recommendation. WWW2010.

Experiment datasets

- The evaluation is performed on an anonymized purchase data of online drug store. http://www.rossmannversand.de
- The dataset is 10-core subset, i.e. every user bought at least 10 items and vice versa each item was bought by 10 users.

dataset	users $\|U\|$	items $\|I\|$	baskets	avg. basket size	avg. baskets per user	triples
Drug store 10-core (sparse)	71,602	7,180	233,476	11.3	3.2	2,635,125
Drug store (dense)	10,000	1,002	90,655	9.2	9.0	831,442

Experimental results



- Factorized Personalized Markov Chains (FPMC)
- Factorized Markov Chain (FMC)
- Matrix Factorization (MF)
- A standard dense Markov Chain (MC dense)
- baseline 'most-popular'

Non-IIDness on items

- Cross-domain RS: domain coupling
- Session-based RS: sequential coupling
  - What is a session?
  - First-order dependency modeling
    - Markov chain-based matrix factorization
  - Higher-order dependency modeling
    - RNN based session modeling
    - Encoder-decoder based session modeling
  - Loosely ordered dependency modeling
    - SWIWO model and its extensions
  - Open issues and directions

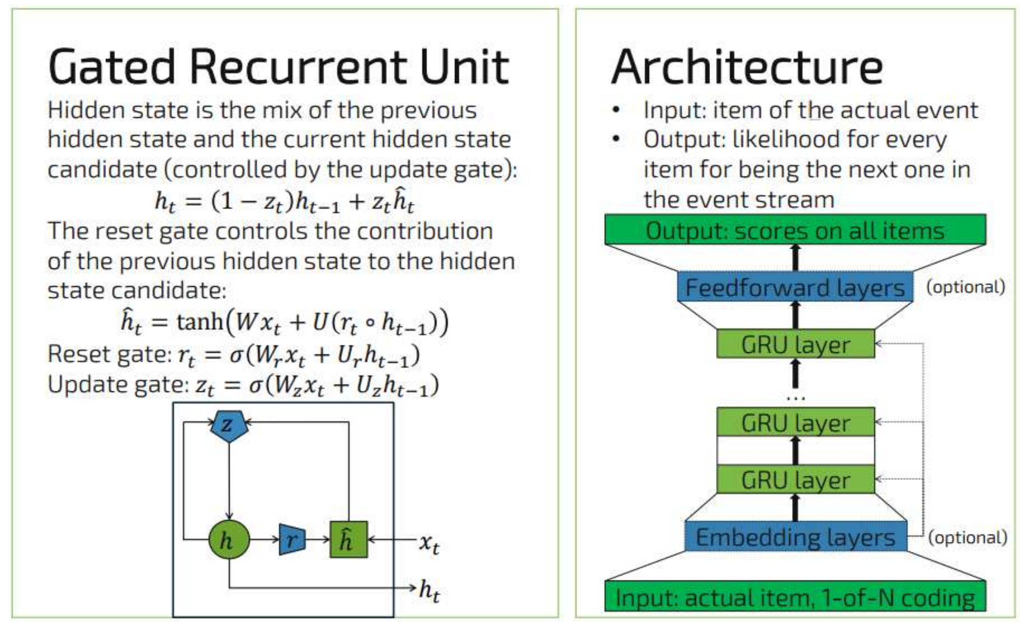
Non-IIDness on items
- Cross-domain RS
- Session-based RS

Non-IIDness on implicit interaction
- Context-aware RS
- Multi-objective RS
- Attraction RS

Non-IID RS in practice

GRU4Rec network architecture

- By modeling the whole session, more accurate recommendations can be provided.
- Applying GRU-RNN to model session.
- Treating the clicks on items as a sequence.
- Modeling the transition between items with GRU.



Hidasi, B., Karatzoglou,A., Baltrunas, L., and Tikk, D. (2016, May). Session-based Recommendations with Recurrent Neural Networks. ICLR2016.

Dataset

- RecSys Challenge 2015:
  - http://2015.recsyschallenge.com/.
  - This dataset contains click-streams of an ecommerce site.

Experiments

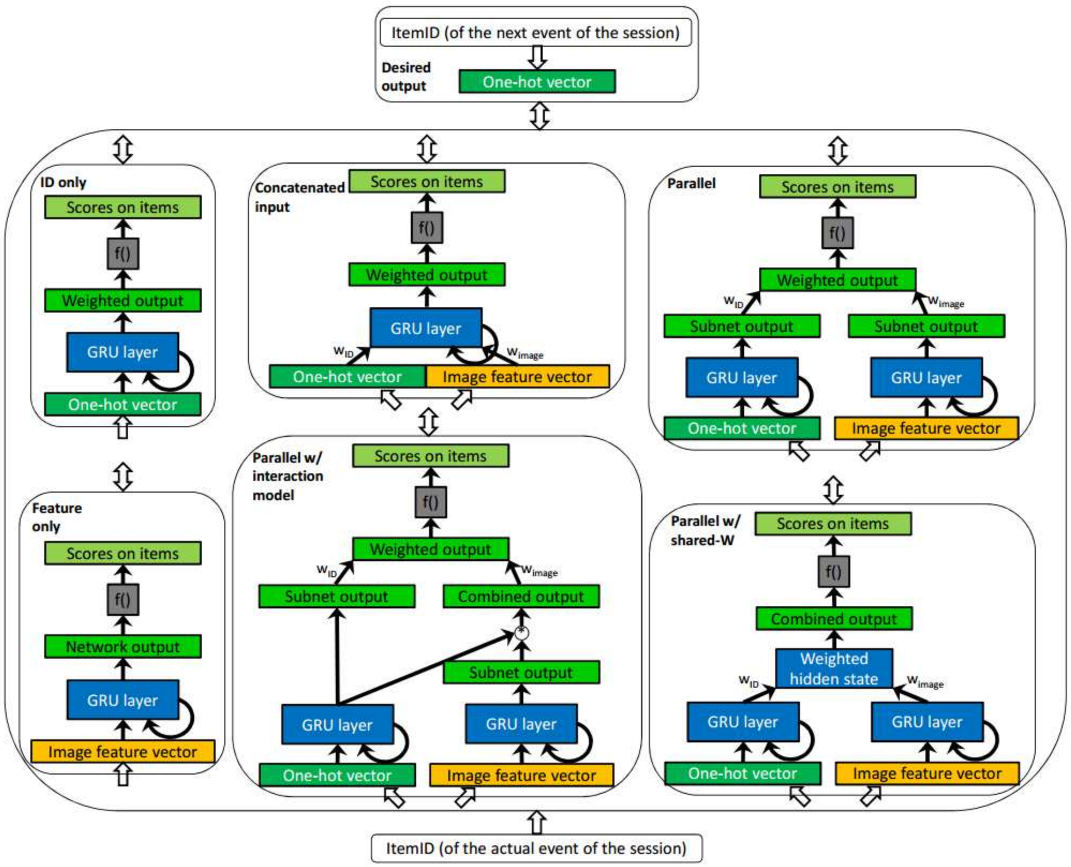
Table 1: Recall@20 and MRR@20 using the baseline methods

Baseline	RSC15		VIDEO	
	Recall@20	MRR@20	Recall@20	MRR@20
POP	0.0050	0.0012	0.0499	0.0117
S-POP	0.2672	0.1775	0.1301	0.0863
Item-KNN	0.5065	0.2048	0.5508	0.3381
BPR-MF	0.2574	0.0618	0.0692	0.0374

Table 3: Recall@20 and MRR@20 for different types of a single layer of GRU, compared to the best baseline (item-KNN). Best results per dataset are highlighted.

Loss / #Units	RSC15		VIDEO	
	Recall@20	MRR@20	Recall@20	MRR@20
TOP1 100	0.5853 (+15.55%)	0.2305 (+12.58%)	0.6141 (+11.50%)	0.3511 (+3.84%)
BPR 100	0.6069 (+19.82%)	0.2407 (+17.54%)	0.5999 (+8.92%)	0.3260 (-3.56%)
Cross-entropy 100	0.6074 (+19.91%)	0.2430 (+18.65%)	0.6372 (+15.69%)	0.3720 (+10.04%)
TOP1 1000	0.6206 (+22.53%)	**0.2693 (+31.49%)**	**0.6624 (+20.27%)**	**0.3891 (+15.08%)**
BPR 1000	**0.6322 (+24.82%)**	0.2467 (+20.47%)	0.6311 (+14.58%)	0.3136 (-7.23%)
Cross-entropy 1000	0.5777 (+14.06%)	0.2153 (+5.16%)	–	–

Parallel RNN for Feature-rich Session Recommendations



- Incorporate **item features** (e.g., text, image) into RNN-based session models.
- Introduce a number of **parallel RNN** (p-RNN) architectures to model sessions and item features at the same time.
- Propose alternative **training strategies**.

Gravity R, B., Quadrana, M., Karatzoglou, A., and Tikk, D. (2016 August). Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations. RecSys’2016.

Datasets

- Two datasets
  - Coined VIDXL : it was collected over a 2-month period from a Youtube-like video site, and contains video watching events having at least a predefined length
  - Class: it consists of product view events of an online website

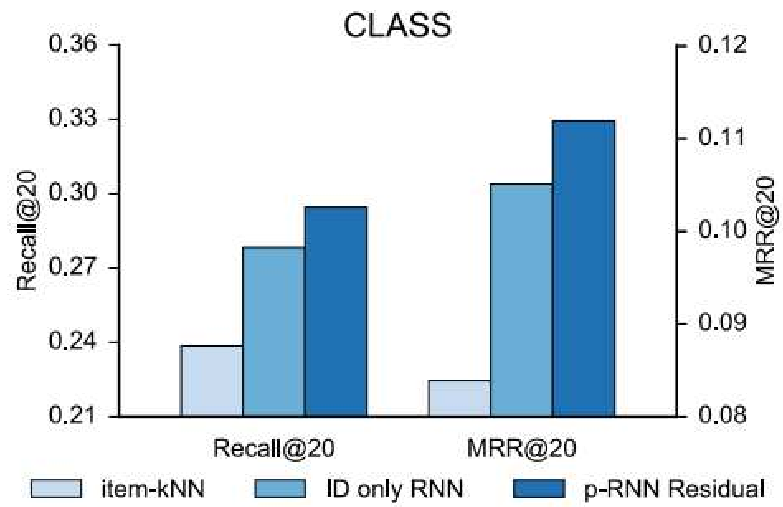
Table 1: Properties of the datasets.

Data	Train set		Test set		Items
	Sessions	Events	Sessions	Events	
VIDXL	17,419,964	69,312,698	216,725	921,202	712,824
CLASS	1,173,094	9,011,321	35,741	254,857	339,055

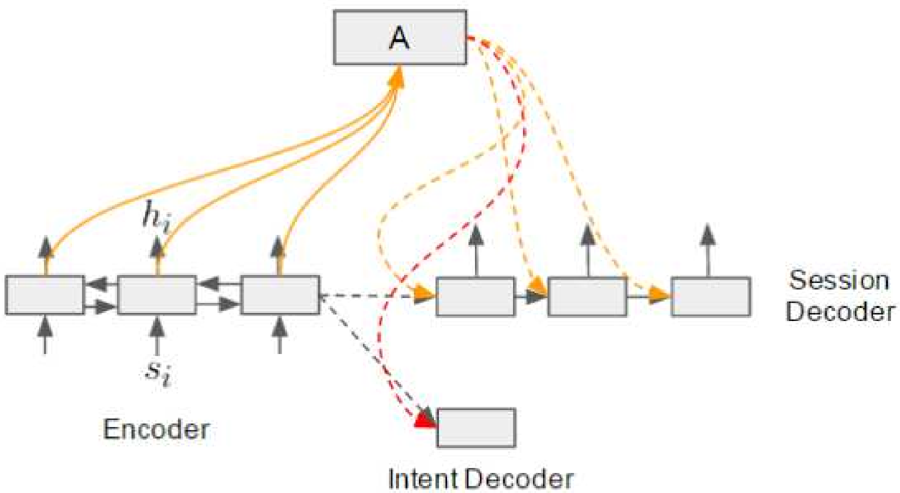
Experiments

- Performance of p-RNN, ID only RNN, and item-KNN.
- p-RNN with features incorporated clearly outperforms the other two approaches.

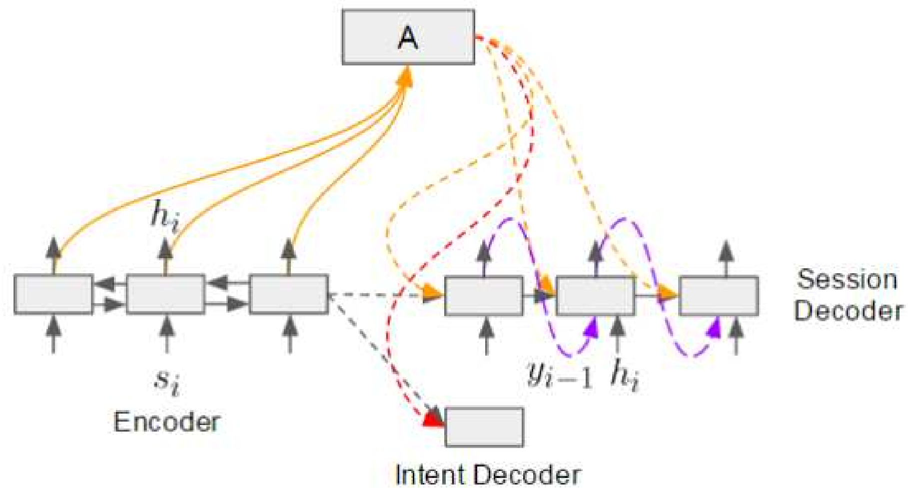
Method	Recall@20	MRR@20
Item-kNN	0.6263	0.3740
ID only	0.6831 (+9.07%)	0.3847 (+2.85%)
ID only (200)	0.6963 (+11.17%)	0.3881 (+3.77%)
Feature only	0.5367 (−14.30%)	0.3065 (−18.05%)
Concatenated	0.6766 (+8.03%)	0.3850 (+2.94%)
Parallel (sim)	0.6765 (+8.01%)	0.4014 (+7.34%)
Parallel (alt)	0.6874 (+9.76%)	0.4331 (+15.81%)
Parallel (res)	0.7028 (+12.21%)	**0.4440** (+18.72%)
Parallel (int)	**0.7040** (+12.41%)	0.4361 (+16.60%)
Shared-W (sim)	0.6681 (+6.66%)	0.4007 (+7.13%)
Shared-W (alt)	0.6804 (+8.63%)	0.4035 (+7.88%)
Shared-W (res)	0.6425 (+2.58%)	0.3541 (−5.31%)
Shared-W (int)	0.6658 (+6.31%)	0.3715 (−0.66%)
Int. model (sim)	0.6751 (+7.78%)	0.3998 (+6.90%)
Int. model (alt)	0.6847 (+9.32%)	0.4104 (+9.74%)
Int. model (res)	0.6749 (+7.76%)	0.4098 (+9.56%)
Int. model (int)	0.6843 (+9.25%)	0.4040 (+8.02%)



Encoder-Decoder for Session Modeling



model for session and intent modeling with attention



model for session and intent modeling with attention + alignments

A bidirectional RNN is used for the encoder to load the item sequence. Decoder is a unidirectional RNN.

Explicit information transfer with alignment, passing both the emitted label y_{i-1} and the internal hidden state h_i at time t to the decoder.

Loyola, P., Liu, C., and Hirate, Y. (2017 August). Modeling User Session and Intent with an Attention-based Encoder-Decoder Architecture. RecSys'2017.

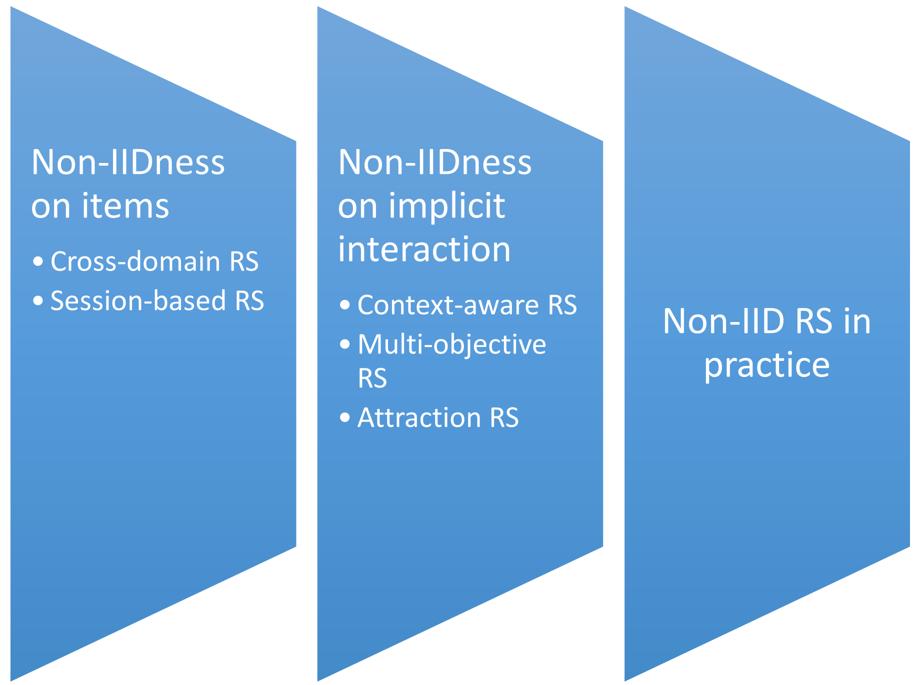
Experiments

- Results of different approaches and their variations, encoder-decoder with attention achieves the best performance.

Model	Recall@20	MRR@20
Item-KNN	0.327	0.139
BPR-MF	0.310	0.135
GRU4Rec	0.3481	0189
GRU4Rec (cross-entropy)	0.3506	0.207
EDRec	0.3775	0.214
EDRec w/ alignment	0.3905	0.249
EDRec w/ alignment and attention	0.3914	0.231

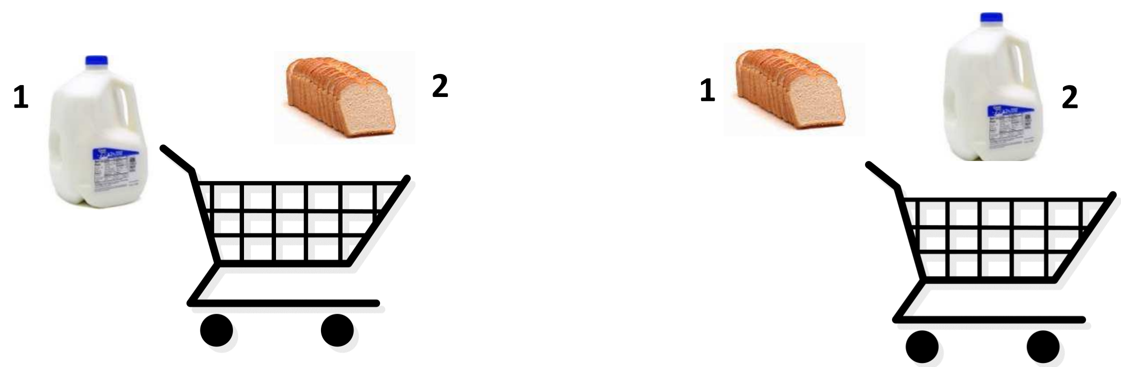
Non-IIDness on items

- Cross-domain RS: domain
- Session-based RS: temporal
  - What is a session?
  - First-order dependency modeling
    - Markov chain-based matrix factorization
  - Higher-order dependency modeling
    - RNN based session modeling
    - Encoder-decoder based session modeling
  - Loosely ordered dependency modeling
    - SWIWO model and its extensions
  - Open issues and directions



Loosely ordered sequence in session

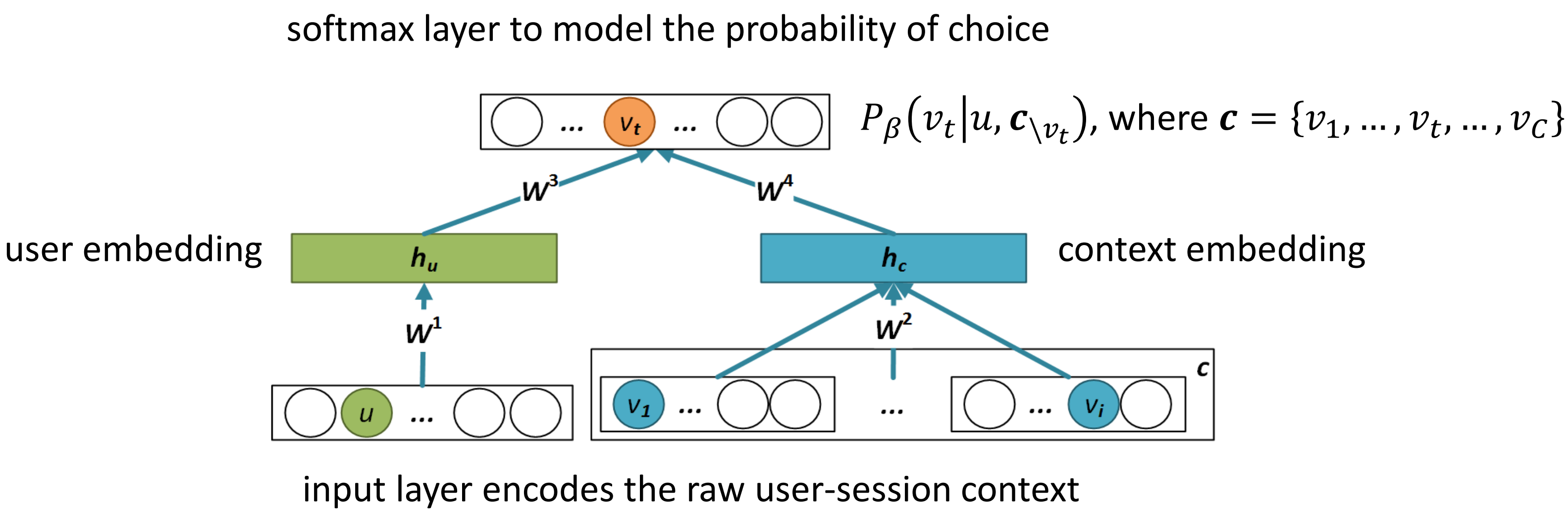
- The choices of items in a session may **not follow a rigidly ordered sequence**
  - For example, toast and milk, which is first put into a shopping cart is not sensitive to the next choice.



Hu, L., Cao, L., Wang, S., Cao, J., Gu, Z., Xu, G., and Wang, J. Diversifying Personalized Recommendation with User-session Context. IJCAI2017

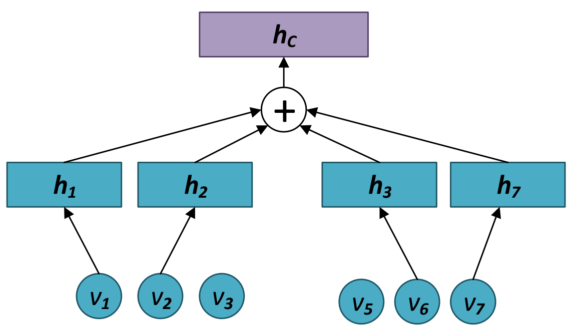
Wide-in-wide-out Shallow Networks

- SWIWO Architecture (Inspired by CBOW)
  - Three-layer shallow wide-in-wide-out networks

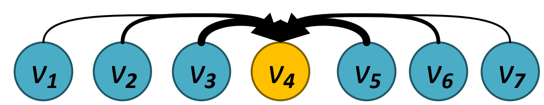


Hu, L., Cao, L., Wang, S., Cao, J., Gu, Z., Xu, G., and Wang, J. Diversifying Personalized Recommendation with User-session Context. IJCAI2017

Weight assignment on context items



$$\mathbf{h}_c = \sum_{v \in \mathbf{c}} w_v \mathbf{h}_v$$



The context items previous and next to the target item v_t, i.e. v_{t-1} and v_{t+1}, have the largest weight, and those context items farther from v_t are assigned smaller weights.

$$w_v \propto \exp[-\lambda(|v-t|-1)]$$

Dataset

- IJCAI-15 Dataset: https://tianchi.aliyun.com/datalab/dataSet.htm?id=5
  - This real-world dataset was collected from Tmall.com which is the largest online B2C platform in China, and it contains anonymized users' shopping logs for the six months before and on the "Double 11" day (November 11th).

Statistic of IJCAI-15 dataset for evaluation
#users: 50K
#items: 52K
avg. session length: 2.99
#training sessions: & 0.20M
#training examples: & 0.59M
#testing cases (***LAST***): 4.5K
#testing cases (***LOO***): 11.9K

Accuracy Evaluation

- The result of *REC@10*, *REC@20* and *MRR* over the testing sets
  - ***Last:*** predict the last item in a testing session
  - ***LOO:*** predict the leave-one-out item

LAST			
Model	**REC@10**	**REC@20**	**MRR**
POP	0.0185	0.0317	0.0104
FPMC	0.0023	0.0068	0.0021
PRME	0.0670	0.0821	0.0363
GRU4Rec	0.2283	0.2464	0.1586
SWIWO-I	**0.3223**	**0.3797**	**0.1918**
SWIWO	**0.3131**	**0.3689**	**0.1896**
LOO			
Model	**REC@10**	**REC@20**	**MRR**
POP	0.0234	0.0420	0.0123
FPMC	0.0064	0.0117	0.0044
PRME	0.0757	0.0976	0.0431
GRU4Rec	0.2242	0.2425	0.1574
SWIWO-I	**0.3177**	**0.3810**	**0.1903**
SWIWO	**0.3082**	**0.3703**	**0.1885**

Diversity Evaluation

- SWIWO considers the whole session context, so it recommends more diverse items.

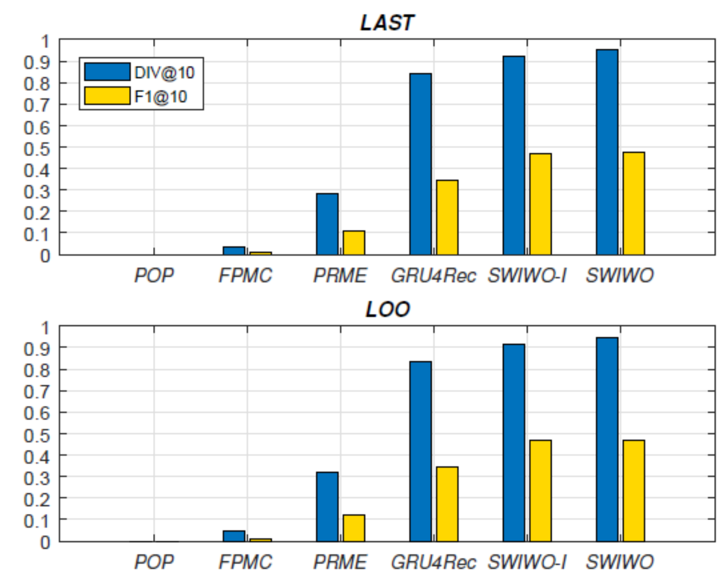
- **DIV@K**: This diversity measures the mean non-overlap ratio between each pair of recommendations $\langle \mathbf{R}_i, \mathbf{R}_j \rangle$ over all N top-K recommendations (note that the number of all possible pairs is $N(N-1)/2$).

$$DIV@K = \frac{2}{N(N-1)} \sum_{i \neq j} (1 - \frac{|\mathbf{R}_i \bigcap \mathbf{R}_j|}{K})$$

- **F1@K**: The traditional F1 score is the harmonic mean of recall and precision. Here, we replace precision with diversity to jointly consider accuracy and diversity metrics.

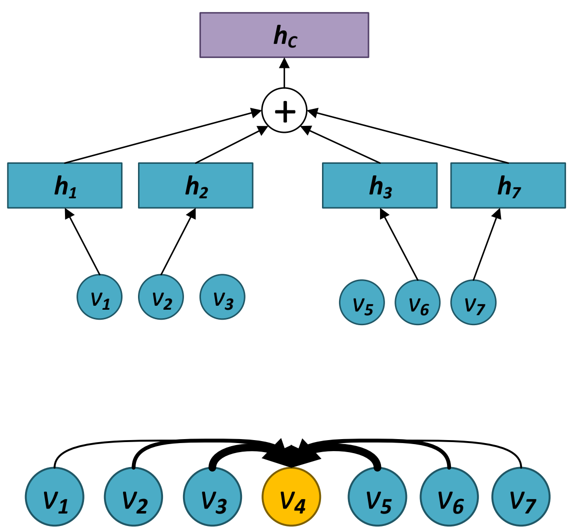
$$F1_{MRR-DIV}@K = \frac{2(MRR@K \times DIV@K)}{MRR@K + DIV@K}$$

$$F1_{REC-DIV}@K = \frac{2(REC@K \times DIV@K)}{REC@K + DIV@K}$$

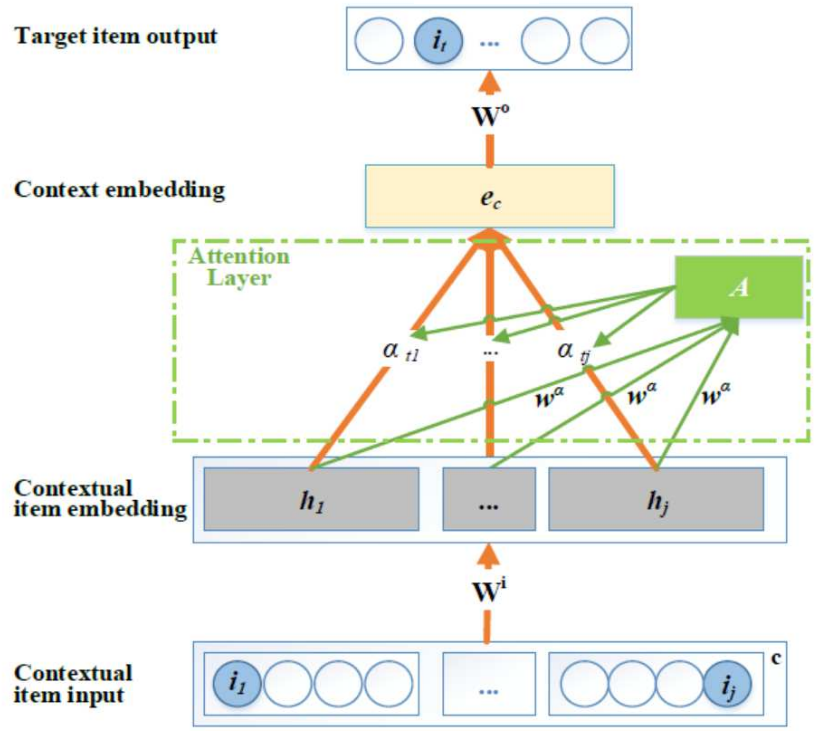


Extension 1:
Weight transaction embedding with attention mechanism

Context items contribute differently to the next choice



ATEM



Wang, S., Hu, L., & Cao, L. *Attention-based Transactional Context Embeddings for Next-Item Recommendation*. AAAI2018

Datasets

- IJCAI-15 Dataset
- Tafang Dataset
  - This real-world dataset is a grocery shopping-supermarket dataset collected from a supermarket from November 2001 to February 2002.

Table 1: Statistics of experimental datasets

Statistics	**IJCAI-15**	**Tafang**
#Transactions	144,936	19,538
#Items	27,863	5,263
Avg. Transaction Length	2.91	7.41
#Training Transactions	141,840	18,840
#Training Instances	412,679	141,768
#Testing Transactions	3,096	698
#Testing Instances	9,030	3,150

Experiments

- ATEM achieves best performance compared to baselines.
- Attention mechanism contributes greatly by comparing ATEM and TEM, a simplified model without attention mechanism.

Table 2: Accuracy comparisons on IJCAI-15

Model	**REC@10**	**REC@50**	**MRR**
PBRS	0.0780	0.0998	0.0245
FPMC	0.0211	0.0602	0.0232
PRME	0.0555	0.0612	0.0405
GRU4Rec	0.2283	0.3021	0.1586
ATEM	**0.3542**	**0.5134**	**0.2041**
TEM	0.3177	0.3796	0.1918

Table 3: Accuracy comparisons on Tafang

Model	**REC@10**	**REC@50**	**MRR**
PBRS	0.0307	0.0307	0.0133
FPMC	0.0191	0.0263	0.0190
PRME	0.0212	0.0305	0.0102
GRU4Rec	0.0628	0.0907	0.0271
ATEM	**0.1089**	**0.2016**	**0.0347**
TEM	0.0789	0.1716	0.0231

Experiments

- Test the robustness to the item order within session
- ATEM is almost not affected when randomly disordering items.

Table 2: Accuracy comparisons on IJCAI-15

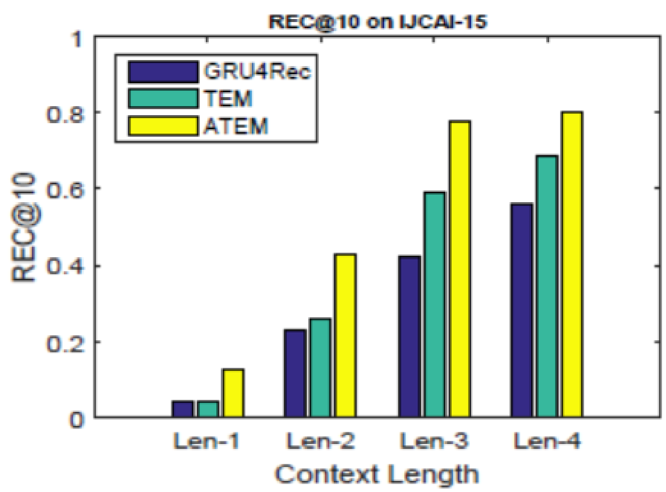
Model	REC@10	REC@50	MRR
PBRS	0.0780	0.0998	0.0245
FPMC	0.0211	0.0602	0.0232
PRME	0.0555	0.0612	0.0405
GRU4Rec	0.2283	0.3021	0.1586
ATEM	**0.3542**	**0.5134**	**0.2041**
TEM	0.3177	0.3796	0.1918

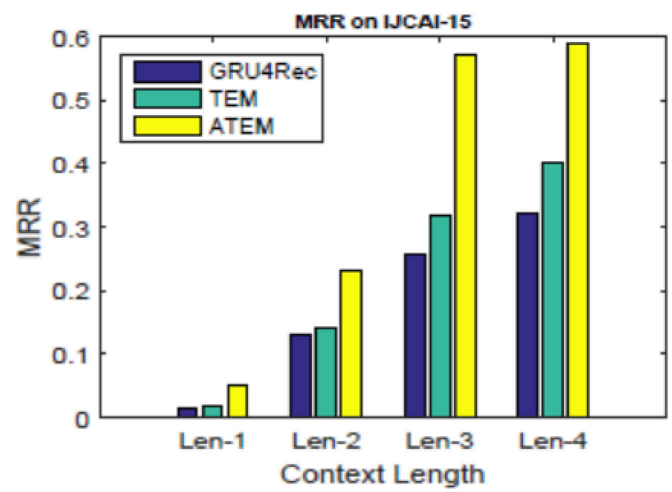
Table 4: Accuracy on disordered IJCAI-15

Model	REC@10	REC@50	MRR
PBRS	0.0500	0.0559	0.0185
FPMC	0.0151	0.0412	0.0183
PRME	0.0346	0.0389	0.0351
GRU4Rec	0.1636	0.2121	0.1022
ATEM	**0.3423**	**0.4981**	**0.1960**
TEM	0.2660	0.3012	0.1431

Experiments

- Test the effect of context length
  - ATEM outperforms other methods on longer context, which proves attention mechanism effectively choose the most related items in a session.





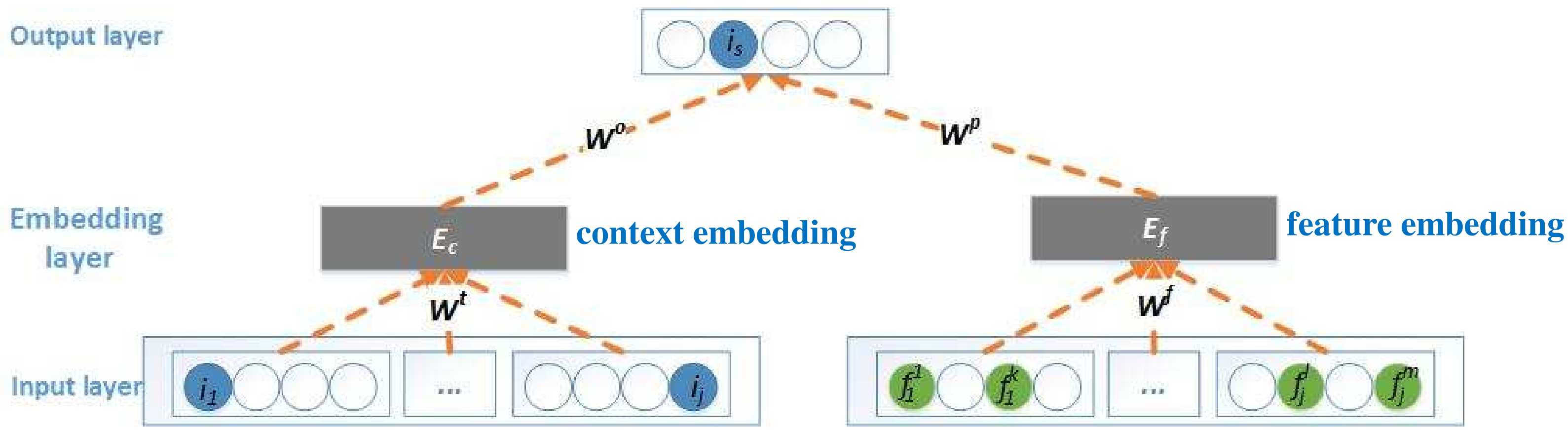
Extension 2:
Embedding attributes for cold start recommendations

- Previous models cannot recommend items which rarely occurred or totally new items.

- We incorporate the item features into the embedding model to handle such cold-start item recommendation issue.

NTEM Architecture

- Three-layer shallow wide-in-wide-out networks



input layer encodes the raw contextual item set and the corresponding features

Wang, S., Hu, L., & Cao, L. (2017, September). *Perceiving the Next Choice with Comprehensive Transaction Embeddings for Online Recommendation*. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*(pp. 285-302). Springer, Cham.

Accuracy Evaluation

- The result of *REC@10*, *REC@50* and *MRR* over the testing sets of two real-world datasets.

We build datasets with different cold start levels to test our model's capability on cold start recommendations.

		IJCAI-15			Tafang		
Scenario	**Model**	**REC@10**	**REC@50**	**MRR**	**REC@10**	**REC@50**	**MRR**
drop 0	FPMC	0.0016	0.0025	0.0031	0.0189	0.0216	0.0089
	PRME	0.0555	0.0612	0.0405	0.0212	0.0305	0.0102
	GRU4Rec	0.1182	0.1566	0.0965	0.0428	0.0887	0.0221
	NTEM	0.2026	0.3224	0.1125	0.0689	0.1716	0.0231
drop 40%	FPMC	0.0012	0.0021	0.0026	0.0008	0.0010	0.0058
	PRME	0.0327	0.0411	0.0312	0.0102	0.0205	0.0095
	GRU4Rec	0.1108	0.1356	0.0868	0.0330	0.0659	0.0196
	NTEM	0.1928	0.2794	0.1117	0.0575	0.1049	0.0377
drop 80%	FPMC	0.0009	0.0017	0.0021	0.0005	0.0008	0.0020
	PRME	0.0212	0.0287	0.0215	0.0084	0.0125	0.0056
	GRU4Rec	0.0493	0.0611	0.0398	0.0110	0.0244	0.0054
	NTEM	0.1098	0.1450	0.0686	0.0254	0.0494	0.0072
drop 95%	FPMC	0.0003	0.0008	0.0012	0.0002	0.0004	0.0008
	PRME	0.0089	0.0113	0.0105	0.0071	0.0096	0.0043
	GRU4Rec	0.0233	0.0337	0.0173	0.0101	0.0172	0.0042
	NTEM	0.0318	0.0639	0.0173	0.0215	0.0305	0.0068

Novelty Evaluation

- We aim to recommend some novel items with the consideration of transactional context and the incorporation of item features.
- Now, let's consider the following metrics.

Global novelty M^2ITF: the opposite of item popularity w.r.t the whole population.

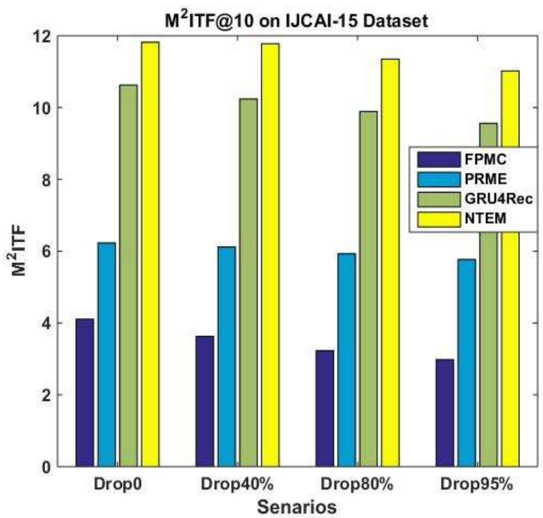
$$MITF = -\frac{1}{|R|}\sum_{i\in R} log_2 \frac{|T_i|}{|T|} \qquad M^2ITF = \frac{1}{N}\sum MITF$$

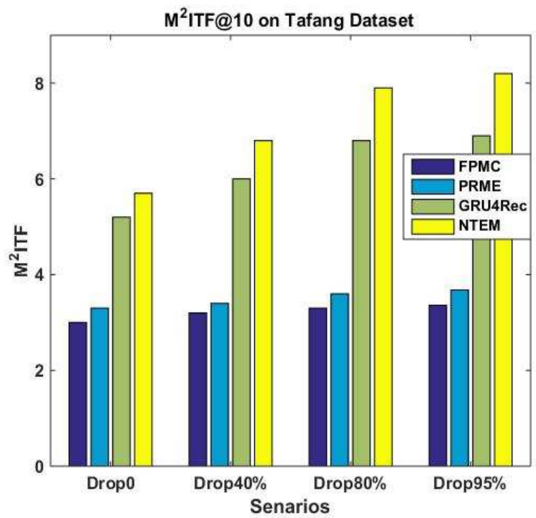
Local novelty MCAN: the difference of recommended list R w.r.t the corresponding context **c**.

$$CAN = 1 - \frac{|R \cap \mathbf{c}|}{|R|} \qquad MCAN = \frac{1}{N}\sum CAN$$

Novelty Evaluation

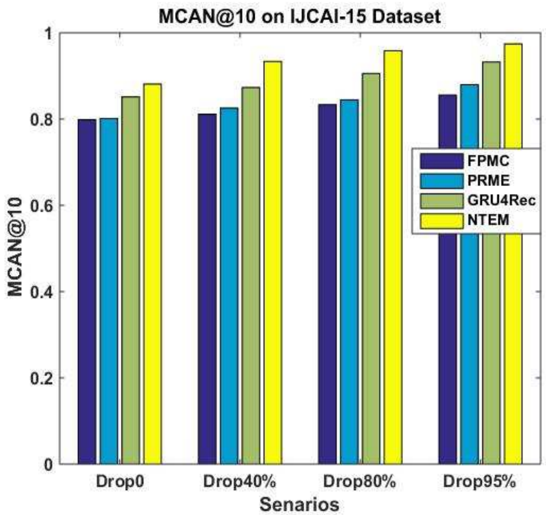
- NTEM incorporates the item features, so it is easier to discover and recommend those unpopular but relevant items.

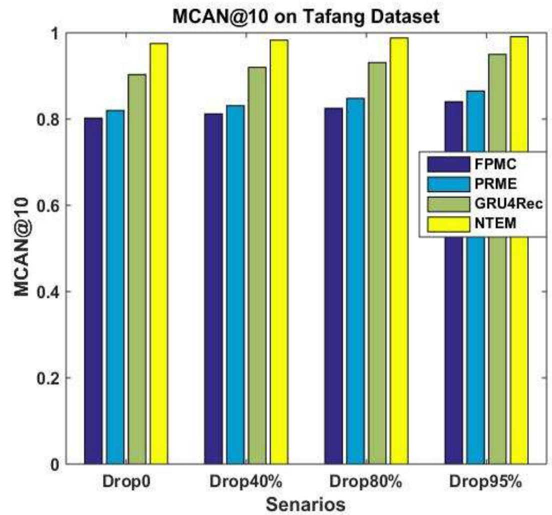




Novelty Evaluation: local novelty

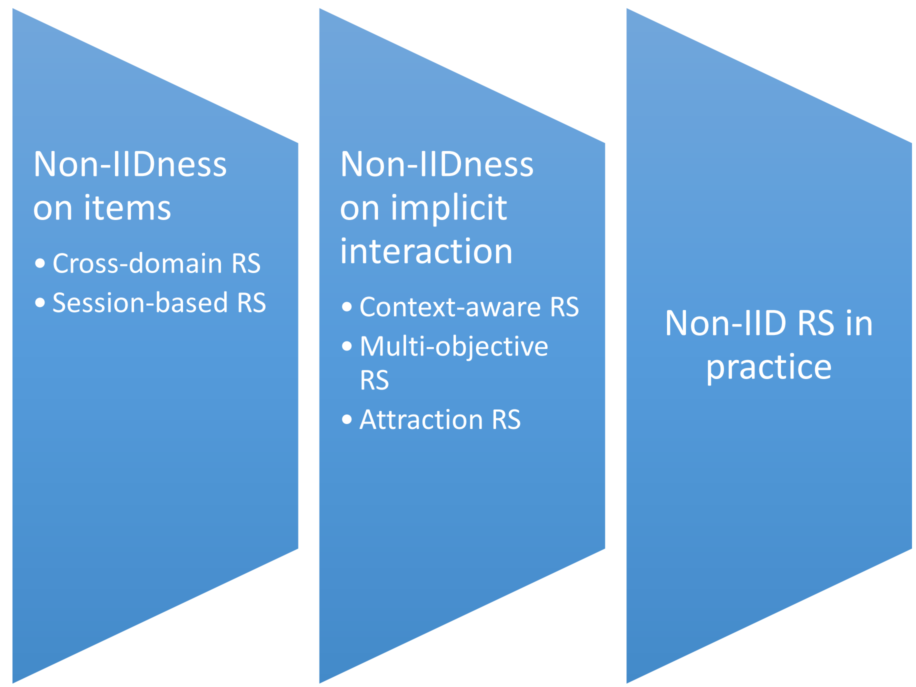
- NTEM considers the whole transaction context, so it more easily to avoid duplicate recommendations and thus recommend something different from the context.





Non-IIDness on items

- Cross-domain RS: domain
- Session-based RS: temporal
  - What is a session?
  - First-order dependency modeling
    - Markov chain-based matrix factorization
  - Higher-order dependency modeling
    - RNN based session modeling
    - Encoder-decoder based session modeling
  - Loosely ordered dependency modeling
    - SWIWO model and its extensions
  - Open issues and directions

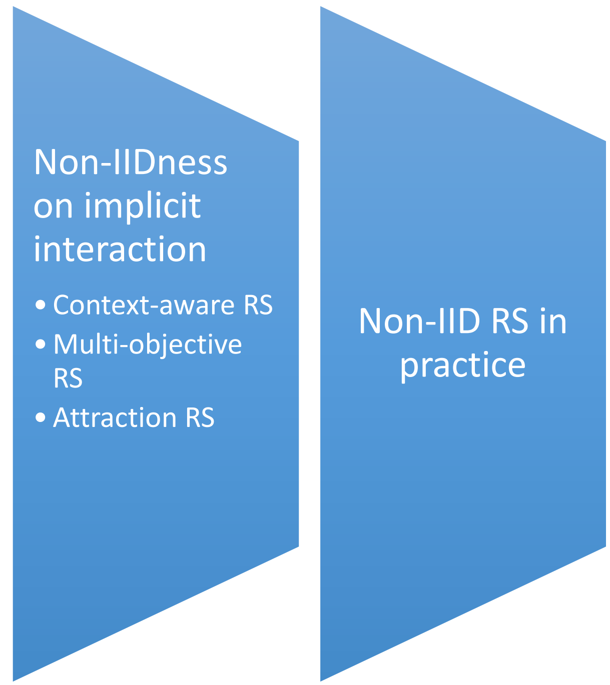


Open issues and future directions

- Open issues
  - How to deal with long-term dependency, e.g., long sessions?
  - How to reduce the influence from irrelevant items in a session?
- Future directions
  - Incorporating cross-session dependency
  - Involving more side information, e.g., attributes, text, images
  - Involving additional contextual information, e.g., weather, locations

Non-IIDness on implicit interaction

- Context-aware RS: contextual information
  - Context-aware Recommender Systems
  - factorization machines
  - Open issues and direction
- Multi-objective RS
  - Recurrent Mutual Regularization Model (RMRM)
  - Open issues and directions
- Attraction RS: subjective attention

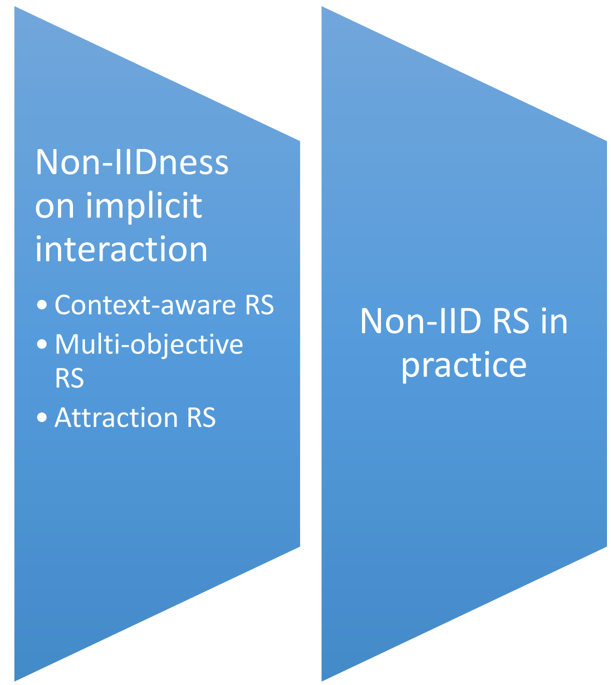


Non-IIDness on Implicit Interaction

- Heterogeneity
  - User choices are often quite different in different context,
    - E.g. the time, the place, the companion
  - An item can be rated by different criteria,
    - E.g. rating on price, rating on usability
  - The attraction points to select an item are often different,
    - E.g. For a paper, one may be attracted by its applications, and others may be attracted by its model
- Coupling
  - Recommendation should consider contextual information
    - E.g. a user often prefer different food for breakfast and dinner
  - The final user choices are often made according to multiple criteria
    - E.g. Novelty, accuracy, diversity are jointly considered when making recommendation
  - User selection is quite dependent on the attraction points
    - E.g. a touching sentence of a song, a favorite actor of a movie

Non-IIDness on implicit interaction

- Context-aware RS: contextual information
  - Context-aware Recommender Systems
  - Factorization machines
  - Open issues and direction
- Multi-objective RS
  - Recurrent Mutual Regularization Model (RMRM)
  - Open issues and directions
- Attraction RS: subjective attention



What is context?

- There are many definitions of context across various disciplines and even within specific subfields of these disciplines.
- The **representational view** assumes that the contextual attributes are identifiable and known a priori and, hence, can be captured and used within the context-aware applications.
- The **interactional view** assumes that the user behavior is induced by an underlying context, but that the context itself is not necessarily observable.

Adomavicius, G. and Tuzhilin, A., 2015. Context-aware recommender systems. In *Recommender systems handbook* (pp. 191-226). Springer US.

In short

- Context is any factor (observable or not observable) leading to user behavior

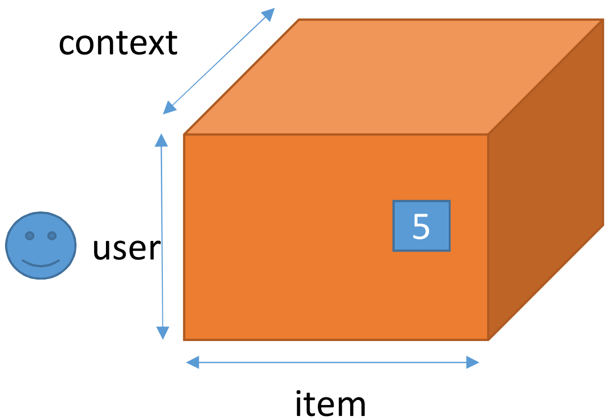
Context-aware Recommender Systems

- Rating mapping without context
  - $User \times Item \rightarrow R$

- Rating mapping with context
  - $User \times Item \times C_1 \times C_2 \times \cdots \rightarrow R$

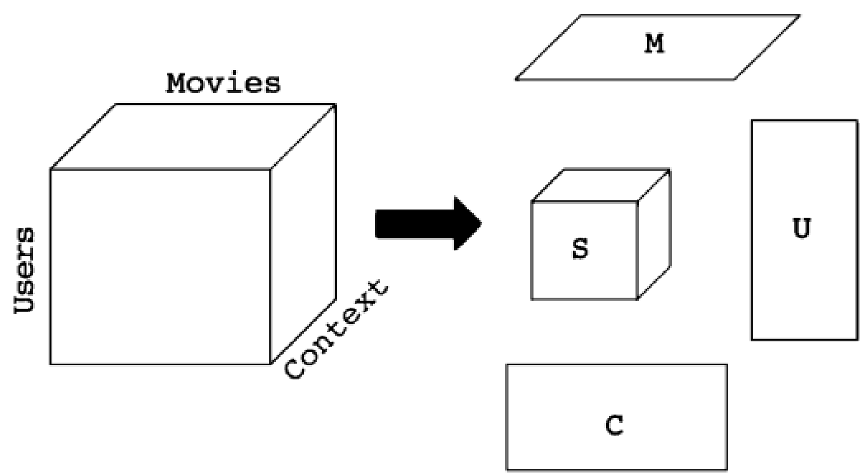
Represent context in higher dimensions

- Rating mapping
  - R(u, i, c) = 5



Tensor factorization model

- 3-dimensional tensor over <User, Movie, Context>



Karatzoglou, A., et al. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Recsys*, 79-86, 2010.

Fast context-aware recommendations with factorization machines

- The idea behind FMs is to model interactions between features using factorized parameters. The FM model has the ability to the estimate all interactions between features even with extreme sparse data.

- FM models all interactions between pairs of variables with the target (2nd-order), including nested ones (1st-order), by using factorized interaction parameters

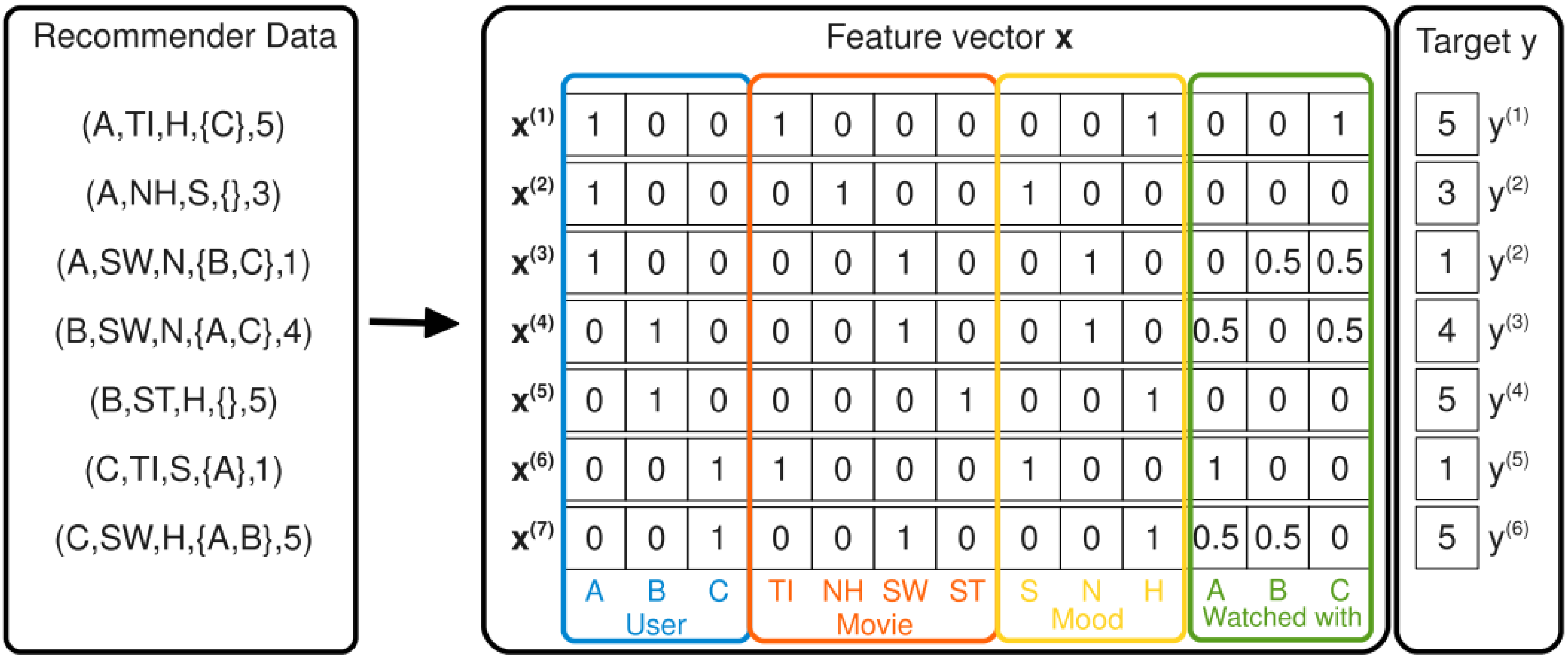
$$\hat{y}(\mathbf{x}) := \boxed{w_0 + \sum_{i=1}^{n} w_i\, x_i} + \boxed{\sum_{i=1}^{n} \sum_{j=i+1}^{n} \hat{w}_{i,j}\, x_i\, x_j}$$

where $\widehat{w}_{i,j}$ are the factorized interaction parameters between pairs:

$$\hat{w}_{i,j} := \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \sum_{f=1}^{k} v_{i,f} \cdot v_{j,f}$$

Rendle, S., Gantner, Z., Freudenthaler, C., and Schmidt-Thieme, L. Fast context-aware recommendations with factorization machines. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 635-644, 2011.

Representing context data as features



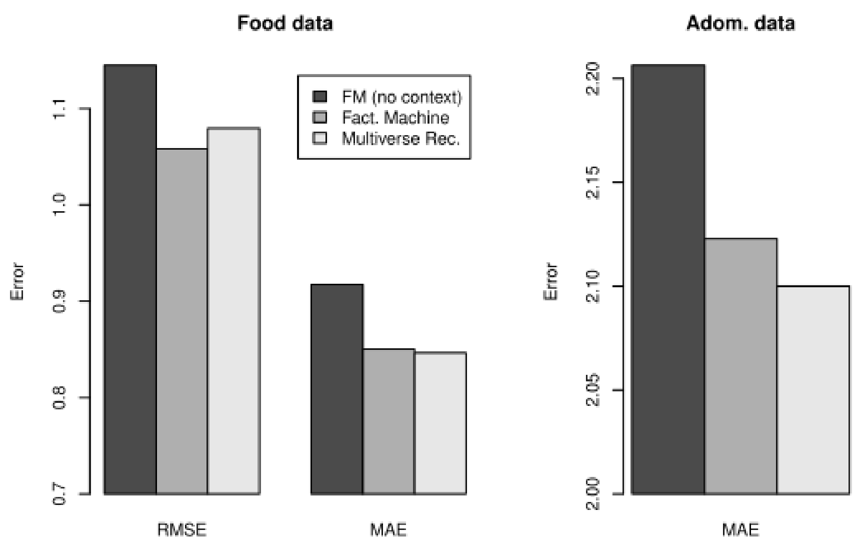
Here in the feature vector x, the first three values indicate the user, the next four ones the movie, the next three ones the mood and the last three ones the other users a movie has been watched with.

Datasets

- Adom dataset
  - 1524 rating events (1 to 15 stars) for movies with five context variables about companion, the weekday and other time information

- Food dataset
  - 6360 ratings (1 to 5 stars) by 212 users for 20 menu items with two context variables:
    - One context variable indicates whether the user is hungry or not.
    - The other one indicates how hungry the user is.

With/without context

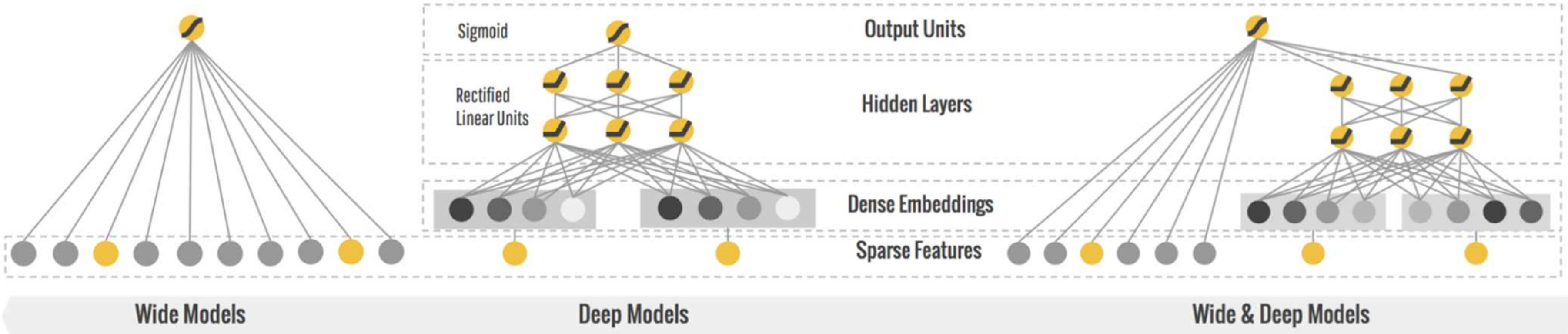


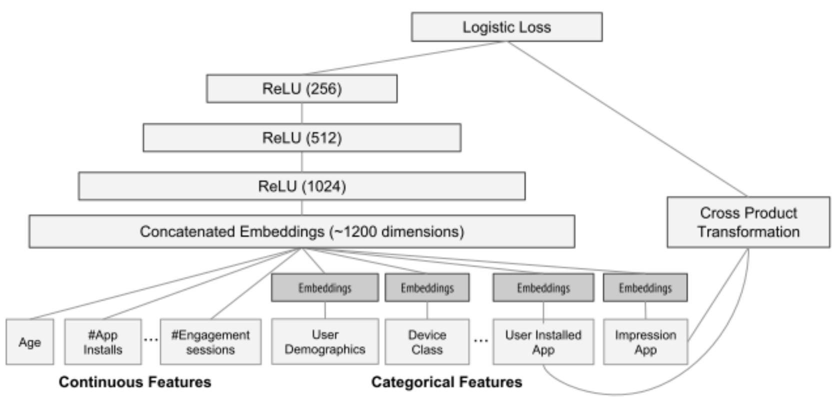
The context-aware methods Multiverse Recommendation and context-aware Factorization Machine benefit from incorporating the context-information into the rating prediction.

Open issues and direction

- Even though the problem of context-aware rating prediction is highly prevalent in practice, there are only a few publicly available datasets.
- Finding efficient way to capture the coupling between all context features.
- Modeling high-dimensional context features with deep models.

Wide & Deep Learning for RS



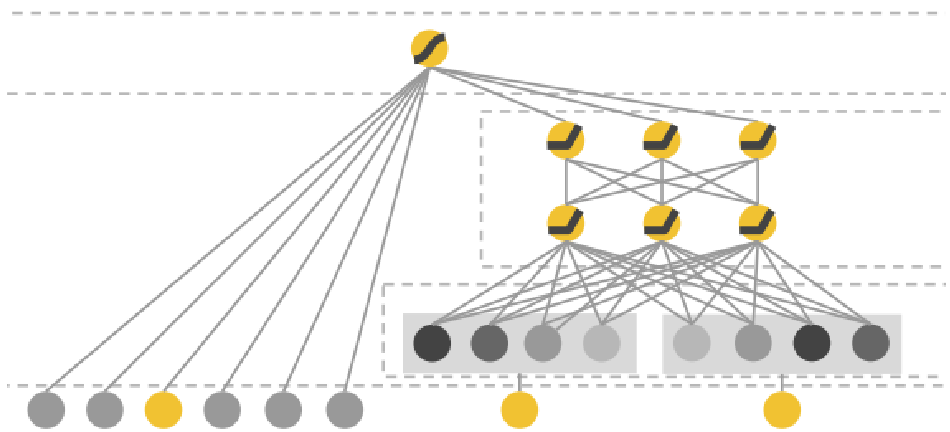


Wide & Deep model structure for apps recommendation

Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., ... & Anil, R. (2016, September). Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems* (pp. 7-10). ACM.

Wide & Deep Learning with Context Features

- Just feed all context features into the networks



All context features as input

Non-IIDness on implicit interaction

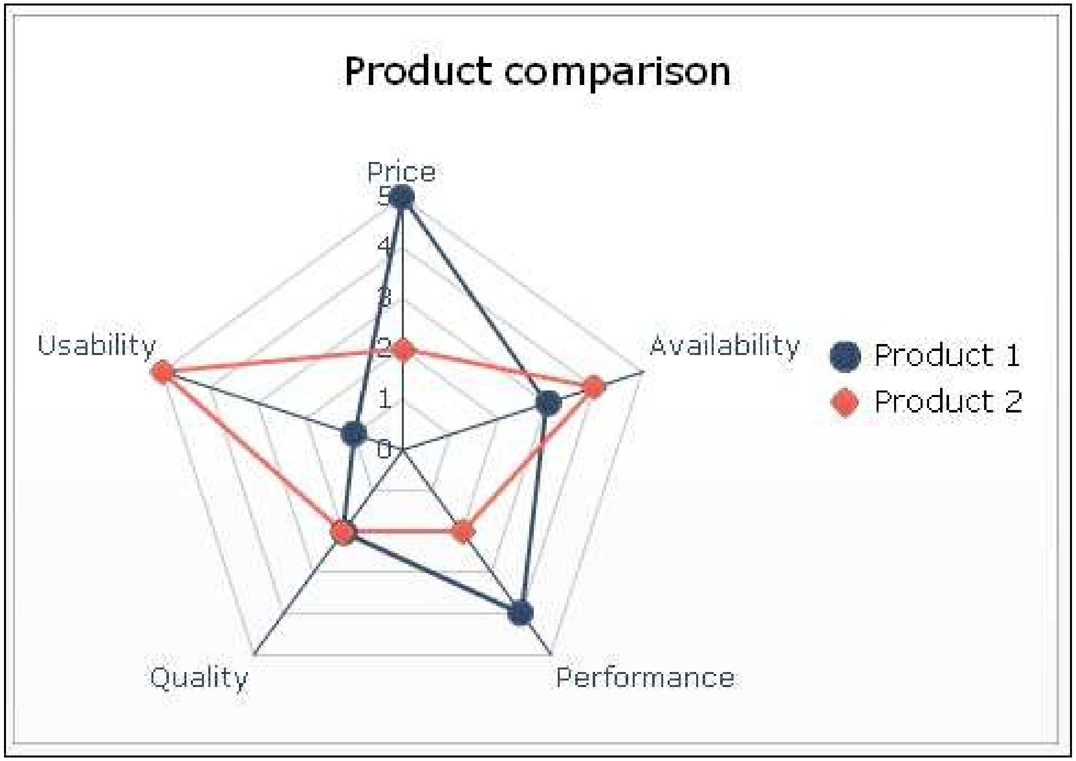
- Context-aware RS: contextual information
  - Context-aware Recommender Systems
  - factorization machines
  - Open issues and direction
- Multi-objective RS
  - Recurrent Mutual Regularization Model (RMRM)
  - Open issues and directions
- Attraction RS: subjective attention

Non-IIDness on implicit interaction

- Context-aware RS
- Multi-objective RS
- Attraction RS

Non-IID RS in practice

Rating from different perspectives

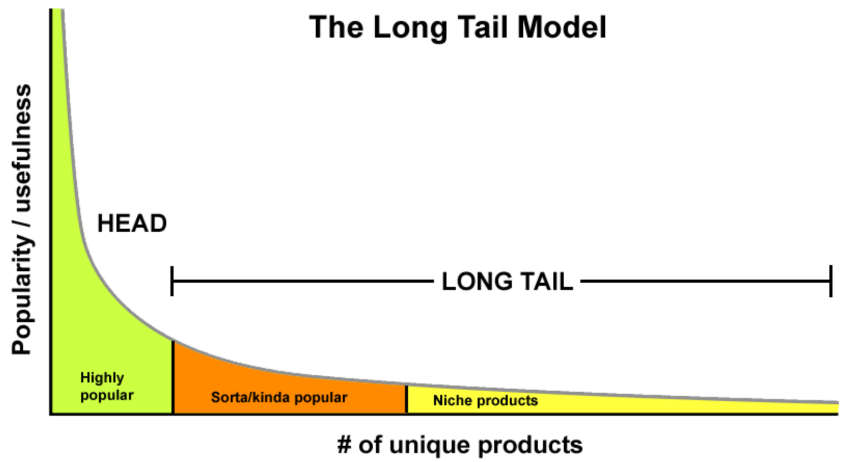


Multi-objective Recommender Systems

- Traditional RSs are built on single objective
- However, recommendations are determined by multiple aspects
  - Accuracy, diversity, novelty …
- To learn users' profile more comprehensively, we need to build new RSs to optimize multiple objectives for each aspect

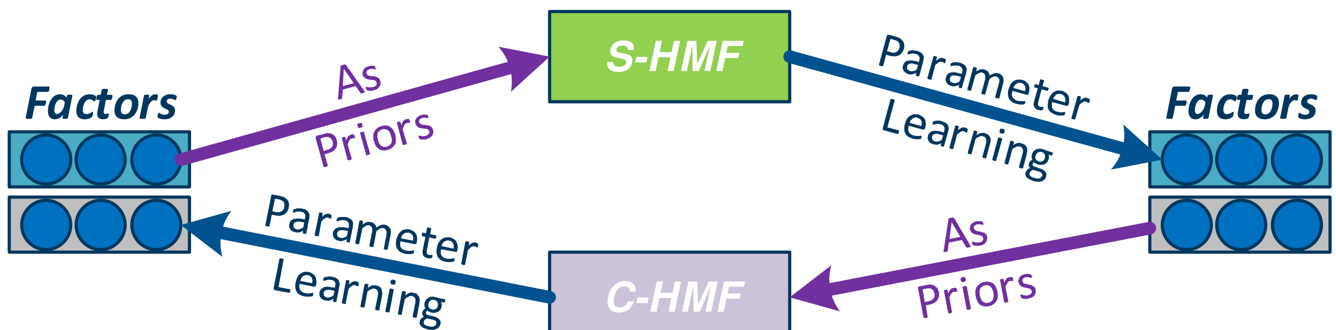
Problems for Long-tail Users/Items

- Popularity Bias
  - Short-head users and items account for the majority of data, and models tend to fit these users and items.
  - Specialty modeling is desirable
- Shilling Attack
  - Long-tail items have few data and they are more vulnerable to shilling attack.
  - Credibility modeling is desirable



RMRM : Joint Optimizing Credibility and Specialty

- Recurrent Mutual Regularization Model (RMRM) consists of two main components
  - C-HMF models user choices by emphasizing credibility
  - S-HMF models user choices by emphasizing specialty
- Each component leads to a different objective for optimization, so RMRM is a multi-objective recommenders systems



Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

Classic Probabilistic MF & Heteroscedastic MF

- $P(\boldsymbol{U}_i) = N\left(\boldsymbol{U}_i \middle| \boldsymbol{0}, \sigma_U^2 \boldsymbol{I}\right)$
- $P(\boldsymbol{V}_j) = N\left(\boldsymbol{V}_j \middle| \boldsymbol{0}, \sigma_V^2 \boldsymbol{I}\right)$
- $P\left(Y_{ij} \middle| \boldsymbol{U}_i, \boldsymbol{V}_j\right) = N\left(Y_{ij} \middle| \boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{V}_j, \sigma^2\right)$

- $P(\boldsymbol{U}_i) = N\left(\boldsymbol{U}_i \middle| \boldsymbol{\mu}_U, \sigma_U^2 \boldsymbol{I}\right)$
- $P(\boldsymbol{V}_j) = N\left(\boldsymbol{V}_j \middle| \boldsymbol{\mu}_V, \sigma_V^2 \boldsymbol{I}\right)$
- $P\left(Y_{ij} \middle| \boldsymbol{U}_i, \boldsymbol{V}_j\right) = N\left(Y_{ij} \middle| \boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{V}_j, \sigma_{ij}^2\right)$

$$P(\boldsymbol{U}, \boldsymbol{V} | \boldsymbol{Y}) \propto P(\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}) = \prod_{ij \in \boldsymbol{O}} P\left(Y_{ij} \middle| \boldsymbol{U}_i, \boldsymbol{V}_j\right) \prod_i P(\boldsymbol{U}_i) \prod_j P(\boldsymbol{V}_j)$$

- *Loss function:*
  - $-\log P\left(Y_{ij}, \boldsymbol{U}_i, \boldsymbol{V}_j\right) = \underset{\boldsymbol{U},\boldsymbol{V}}{argmin} \left[\underbrace{\boxed{\sum_{ij} \left(Y_{ij} - \boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{V}_j\right)^2}}_{unweighted\ loss} + \underbrace{\lambda_U \sum_i \|\boldsymbol{U}_i\|^2 + \lambda_V \sum_j \left\|\boldsymbol{V}_j\right\|^2}_{regularization} \right]$

- *Loss function:*
  - $-\log P\left(Y_{ij}, \boldsymbol{U}_i, \boldsymbol{V}_j\right) =$
  - $\underset{\boldsymbol{U},\boldsymbol{V}}{argmin} \left[\underbrace{\sum_{ij} w_{ij} \left(Y_{ij} - \boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{V}_j\right)^2}_{weighted\ loss} + \underbrace{\lambda_U \sum_i \|\boldsymbol{U}_i - \boldsymbol{\mu}_i\|^2 + \lambda_V \sum_j \left\|\boldsymbol{V}_j - \boldsymbol{\mu}_j\right\|^2}_{regularization} \right]$

Popularity Bias

Shilling Attack

- model variance, i.e. weight on the loss : $w_{ij} = f\left(\sigma_{ij}^{-2}\right)$

Specialty Enhancement

Popularity Bias

- S-HMF (Specialty-specific Heteroscedastic MF)
  - $\sigma_{ij}^2 = f^S(Y_{ij}) \propto {\psi_j}^{-1}$ scores the *specialty* of user choice, which tightly fits the choices over long-tail items

- Given all observed choices, the ***specialty score*** of a choice on an item j is measured by the *self-information*:
  - $\psi_j = -\,log\,\bar{p}(j|\alpha)$

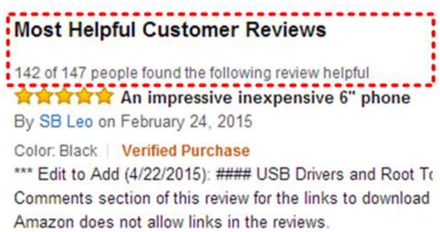
Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

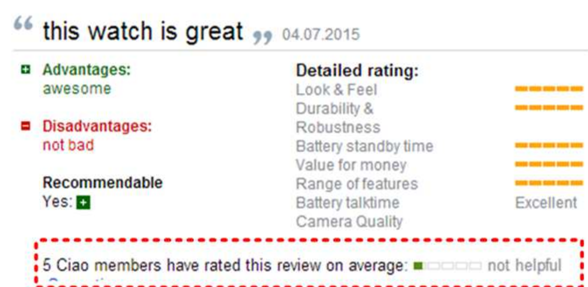
Credibility Enhancement

Shilling Attack

- C-HMF (Credibility-specific Heteroscedastic MF)
  - $\sigma_{ij}^2 = f^C(Y_{ij}) \propto \varphi_i^{-1}$ scores the *credibility* of each review
- Bayesian Reputation Modeling
  - *Reputation Score: Given the helpfulness scores h_i of a user i, the reputation score on this user is defined by:*

$$\varphi_i = \mathcal{R}(\boldsymbol{e}_i|\boldsymbol{h}_i) \stackrel{\text{def}}{=} \frac{r+\alpha}{r+s+\alpha+\beta}$$

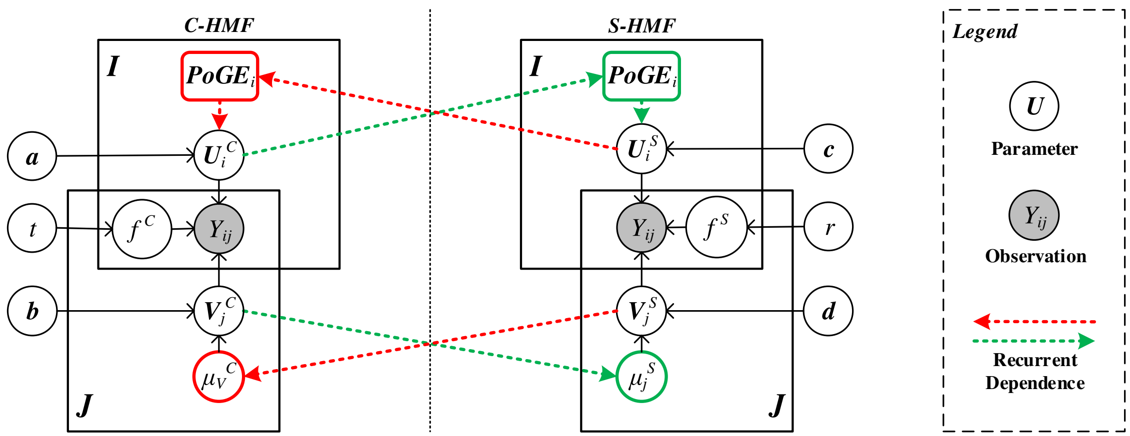




Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

Recurrent Mutual Regularization

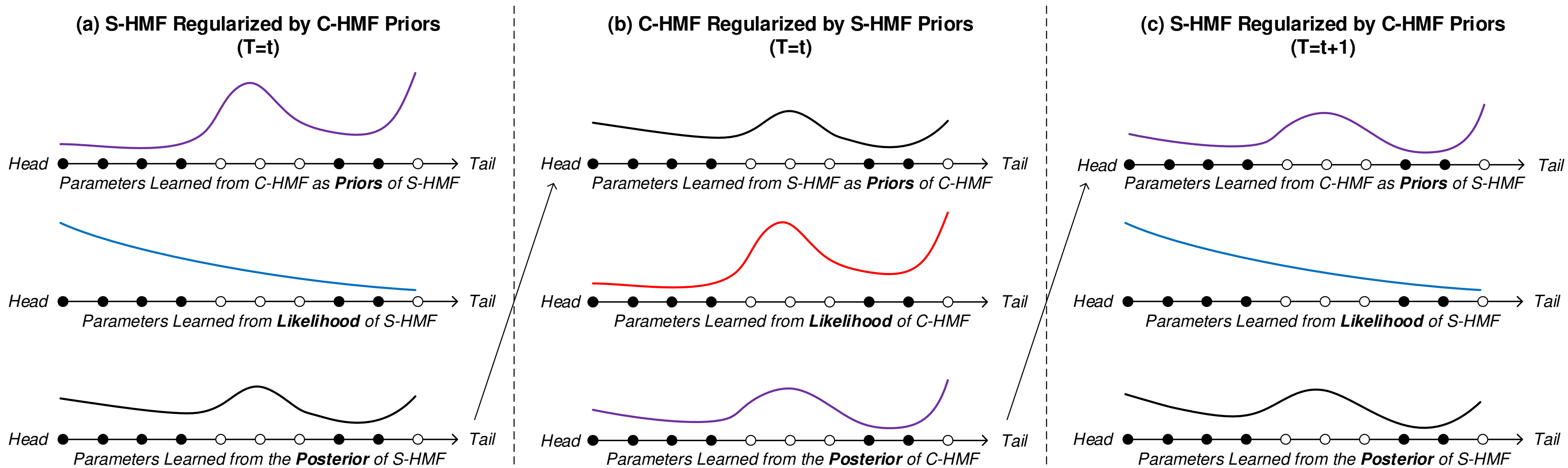
- A **recurrent mutual regularization process** couples S-HMF and C-HMF using the user and items factors learned from each other as the **empirical priors**



Graphical model of RMRM framework

Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

Demonstration of the recurrent mutual Regularization process



$$\underset{U^C,V^C}{argmin}\left[\underbrace{\sum_{ij} w_{ij}\left(Y_{ij} - U_i^{C^{\mathrm{T}}} V_j^C\right)^2}_{weighted\ loss} + \underbrace{\lambda_U \sum_i \left\|U_i^C - U_i^S\right\|^2 + \lambda_V \sum_j \left\|V_j^C - V_j^S\right\|^2}_{regularization}\right]$$

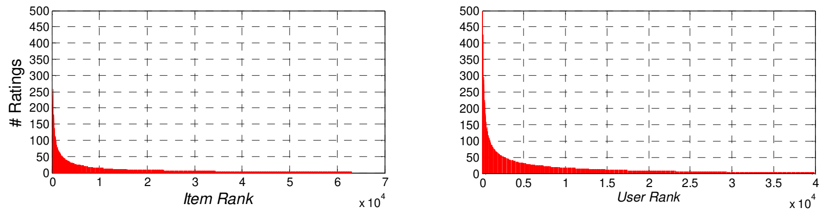
Mutual Regularization

$$\underset{U^S,V^S}{argmin}\left[\underbrace{\sum_{ij} w_{ij}\left(Y_{ij} - U_i^{S^{\mathrm{T}}} V_j^S\right)^2}_{weighted\ loss} + \underbrace{\lambda_U \sum_i \left\|U_i^S - U_i^C\right\|^2 + \lambda_V \sum_j \left\|V_j^S - V_j^C\right\|^2}_{regularization}\right]$$

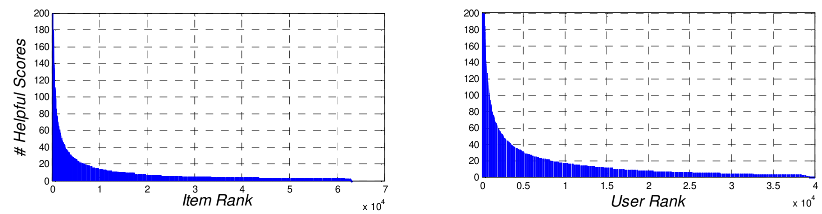
Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

Dataset: the Epinions

# users: 39,902	# items: 63,027
# trust links: 43,8965	# trusters / users: 11
max # of trusters: 1,713	# users with zero truster: 14,202
# ratings: 734,441	density: 0.029%
# ratings / users: 18	# ratings / items: 11
max # ratings of user: 1,809	max # ratings of item: 2,112



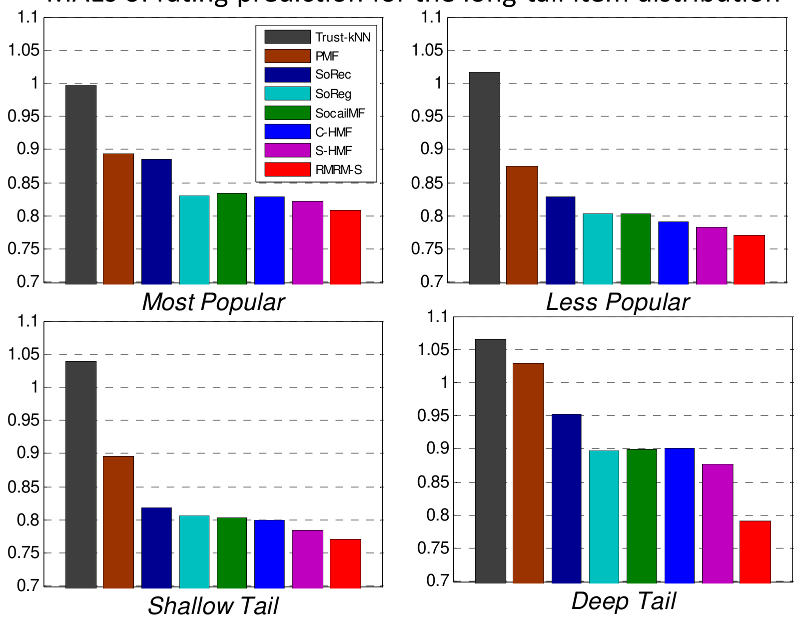
Long-tail distributions for the number of ratings of items and users (truncated from 0 to 500)



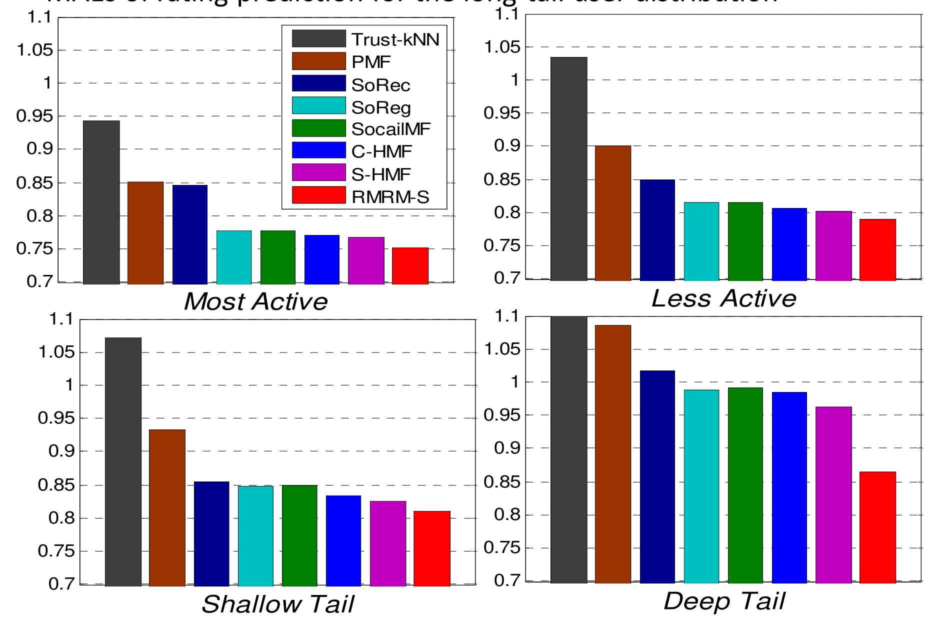
The distributions for the number of helpful scores w.r.t. items and users (truncated from 0 to 200)

Rating Prediction on Long-tail Distributed Items and Users

MAEs of rating prediction for the long-tail item distribution



MAEs of rating prediction for the long-tail user distribution

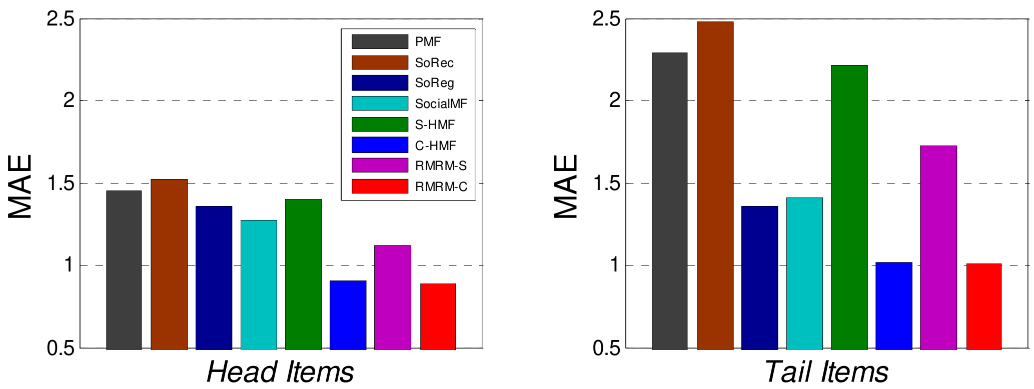


Most Popular: The items in the headmost 5% of the distribution
Less Popular: The items in the 5~20% interval of the distribution
Shallow Tail: The items in the 20~50% interval of the distribution
Deep Tail: The items in the endmost 50% of the distribution

Most Active: The users in the headmost 5% of the distribution
Less Active: The users in the 5~20% interval of the distribution
Shallow Tail: The users in the 20~50% interval of the distribution
Deep Tail: The endmost 50% users of the distribution of the distribution

Shilling Attack Simulation

- To simulate such an environment
  - We created 1,000 virtual spam users to conduct the attack
  - We selected 100 items from the head (0%~20%) and the tail (20%~100%) as the attack targets.
- Nuke attack in the case of the average attack model



Open issues and directions

- How to integrate the impacts from multiple objectives?
  - Different users may pay attention to different objectives
  - The importance of objectives are often dependent on the context

- Modeling with game theory to find equilibria over multiple objectives

- Applying multi-objective optimization methods in RS
  - Multiple-criteria decision analysis
  - Multidisciplinary design optimization

Non-IIDness on implicit interaction

- Context-aware RS: contextual information
  - Context-aware Recommender Systems
  - factorization machines
  - Open issues and direction
- Multi-objective RS
  - Recurrent Mutual Regularization Model (RMRM)
  - Open issues and directions
- Attraction RS: subjective attention

Non-IIDness on implicit interaction

- Context-aware RS
- Multi-objective RS
- Attraction RS

Non-IID RS in practice

Why modeling attraction?

- First, the attraction is the ***highlights*** that largely lead to a person's selection and decision.
- For example,
  - We often cannot recite a whole poem but we can always recall some impressive sentences;
  - We may not remember a whole song but we can hum some touching lyrics.
  - These highlights make a person to be attracted by the poem or the song.

Why modeling attraction?

- Second, the attraction is a ***subjective feeling*** which is often different from person to person.
- For example,
  - Readers in Go community may be attracted by the target problem, i.e., Go playing, of this scientific paper while readers in AI community may be attracted by the technical methods.

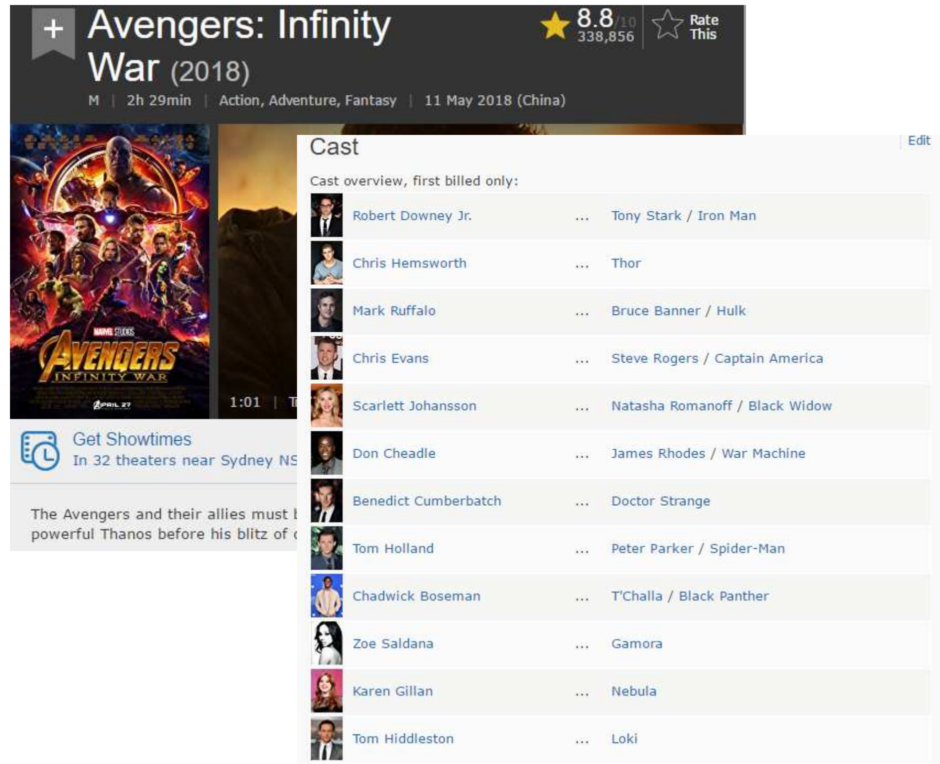
Example: Attraction on Movies

- The internet movie has accounted for the major traffic in new media age.

- In particular, the ***story*** and the ***cast members***, e.g., actors, directors and writers, are two most important aspects of a movie to attract audience.

  - A person may be caught by some attractive words by the story of a movie. Only a few sentences of the core plot instead of all sentences actually attract a user.

  - Cast members of a movie are another very important factor to attract users.

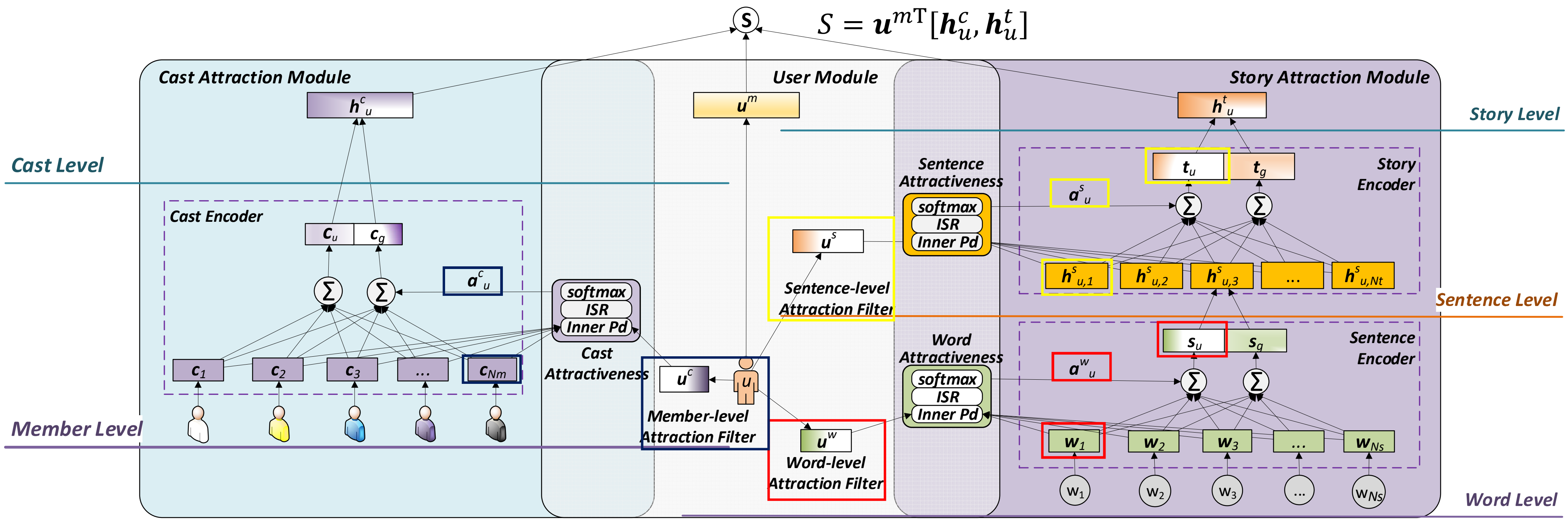


Hu, L., Jian, S., Cao, L., Chen. Q. Interpretable Recommendation via Attraction Modeling: Learning Multilevel Attractiveness over Multimodal Movie Contents. IJCAI2018

Multimodal and Multilevel Attraction Model

- One multilevel neural model on the movie story to capture
  - Word-level attraction: e.g. some character, some place
  - Sentence-level attraction: e.g. some core plot
  - Story-level attraction: e.g. like the movie to what extent
- The other multilevel neural model on the cast to capture
  - Member-level attraction: e.g. a fan of some actor
  - Cast-level attraction: e.g. attracted by the movie to what extent

Model Architecture



$$a_u^{c_i} = softmax\left(isr\left(\boldsymbol{u}^{c\mathrm{T}}\boldsymbol{c}_i\right)\right) \qquad \boldsymbol{c}_u = \sum a_u^{c_i}\boldsymbol{c}_i$$

$$a_u^{w_i} = softmax\left(isr\left(\boldsymbol{u}^{w\mathrm{T}}\boldsymbol{w}_i\right)\right) \qquad \boldsymbol{s}_u = \sum a_u^{w_i}\boldsymbol{w}_i$$

$$a_u^{s_i} = softmax\left(isr\left(\boldsymbol{u}^{s\mathrm{T}}\boldsymbol{h}_i^s\right)\right) \qquad \boldsymbol{t}_u = \sum a_u^{s_i}\boldsymbol{h}_i^s$$

Experiments

- The experiments are conducted on the real-world movie watch dataset MovieLens 1M. The model is evaluated from three aspects:
  - Recommendation accuracy
  - New movie recommendation
  - Interpretation of attraction on movies

Datasets

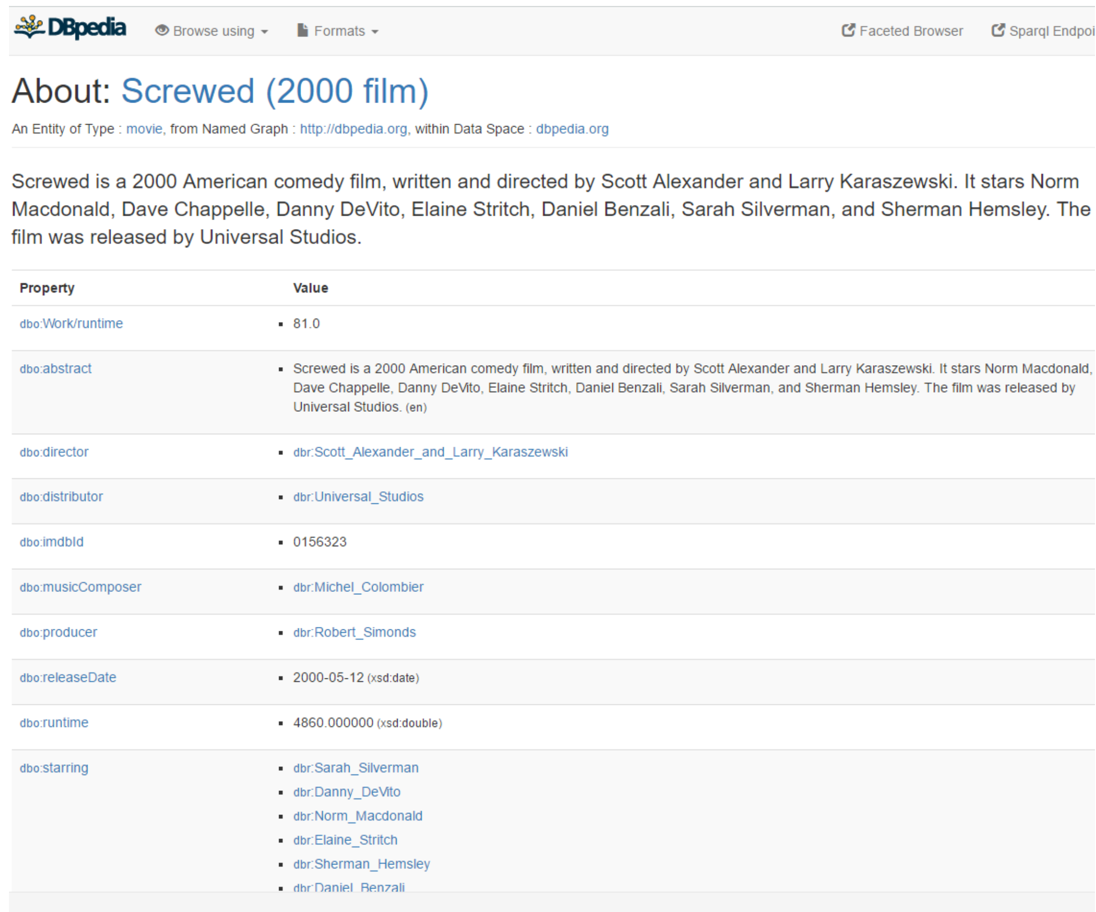
- We collect user watch records from the MovieLens 1M dataset.
  - https://grouplens.org/datasets/movielens/1m/

- Story and cast data are provided the mapping from MovieLens ID to DBPedia URI
  - https://github.com/sisinflab/LODrecsys-datasets/tree/master/Movielens1M

Augment information from DBPedia

- SPARQL Interface

```
PREFIX
movie:http://dbpedia.org/resource/Screwed_(2000_film)
select ?abstract ?director ?writer ?starring
 { movie: dbo:abstract ?abstract.
  optional { movie: dbo:director ?director }
  optional { movie: dbo:writer ?writer }
  optional { movie: dbo:starring ?starring }
 FILTER (langMatches(lang(?abstract),"en")) }
```



Statistics of the Enriched Dataset

# movies:	3,900	# users:	6,040
# watch record:	1,000,209	# cast:	9,398
movie story vocabulary	22,582	# sentences per story	10.2
# cast members per movie	6.44	# plays per cast	2.10

Table 1: Statistics of content-enriched MovieLens dataset

Training and Testing Sets

- **Released movie recommendation**: we randomly held out 20% user watch records as the testing set, and the remainder were served as the training set.
- **New movie recommendation:** we randomly selected 10% movies and held out all their watch records from the dataset, and the remainder of 90% movies and their watch records were used for training.
- For each hold-out test sample in above two testing sets, we randomly draw ten noisy samples to test whether the testing methods can rank the true sample at a top position out of noisy samples.

Comparison Methods

- ***CENTROID***: We create user profiles using the centroid of all word embedding vectors from the users' movie stories. Then, we rank recommendations by the similarity between the user profile and the controid of word embedding vectors of movie story.
- ***CTR***: Collaborative topic regression performs user regression over the latent topic distribution of movie stories learned from LDA.
- ***CWER***: Similar to CTR, we create the collaborative word embedding user regression (CWER) to perform regression over the centroid word embedding vector of each movie story initialized by GloVe embeddings.
- ***MLAM***: This is the full multilevel attraction model proposed in this paper.
- ***MLAM-S***: This is the single-modal version MLAM that only has the story attraction module.
- ***MLAM-C***: This is the single-modal version MLAM that only has the cast attraction module.

Ranking Performance

- Recommendation accuracy on released movies and new movies

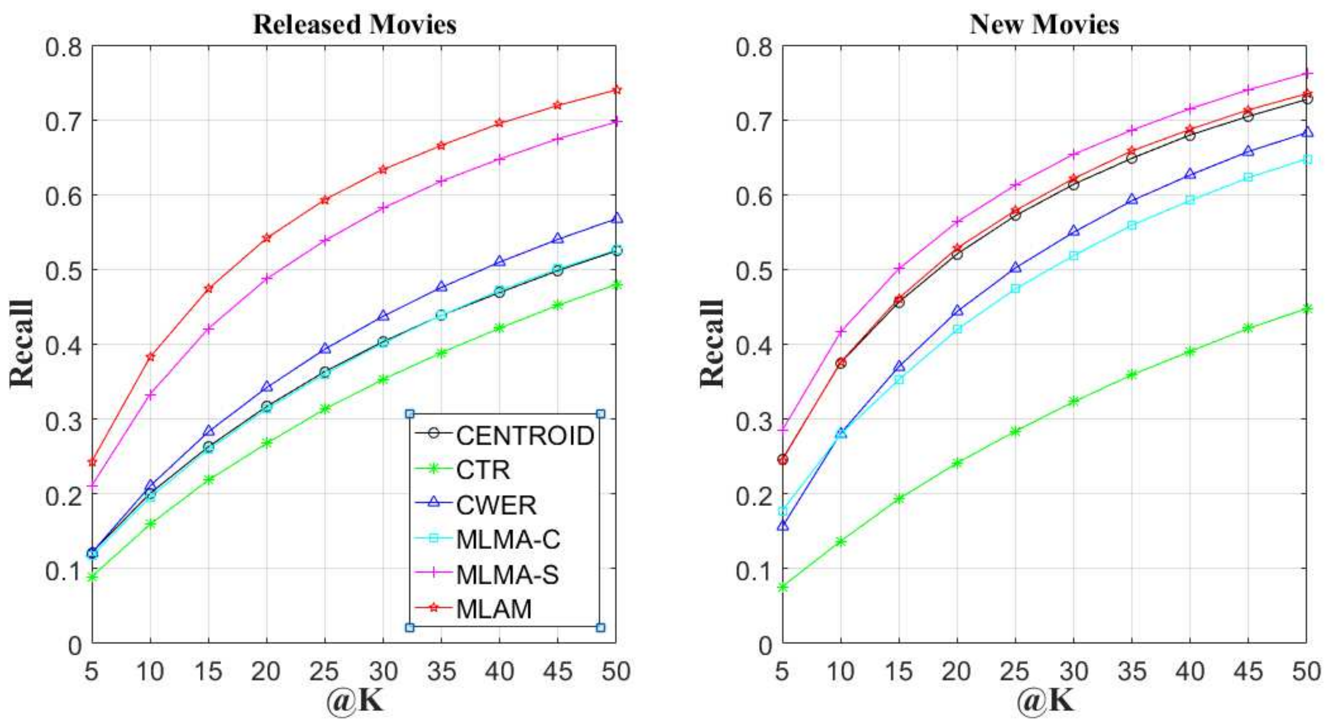
Method	MAP@5	MAP@20	MRR@5	MRR@20
CENTROID	0.1738	0.1481	0.0763	0.0958
CTR	0.1226	0.1069	0.0514	0.0692
CWER	0.1666	0.1580	0.0798	0.1089
MLAM-C	**0.4243**	**0.3963**	**0.2118**	**0.2398**
MLAM-S	0.3816	0.3451	0.1822	0.2093
MLAM	**0.4252**	**0.3997**	**0.2187**	**0.2464**

Table 2: Ranking performance on released movies (80% training)

Method	MAP@5	MAP@20	MRR@5	MRR@20
CENTROID	0.2381	0.2409	0.1623	0.1900
CTR	0.1056	0.1374	0.0798	0.1089
CWER	0.1971	0.2346	0.1461	0.1801
MLAM-C	0.1817	0.1664	0.1132	0.1370
MLAM-S	**0.3001**	**0.3059**	**0.2091**	**0.2371**
MLAM	**0.2573**	**0.2671**	**0.1794**	**0.2090**

Table 3: Ranking performance on new movies (90% training)

Recall on Release Movies and New Movies



Visualization and Interpretation

User 156	*Sentence level attractiveness*	**Election is a 1999 American comedy-drama film directed and written by Alexander Payne and adapted by him and Jim Taylor from Tom Perrotta's 1998 novel of the same title.** The plot revolves around a high school election and satirizes both suburban high school life and politics. **The film stars Matthew Broderick as Jim McAllister, a popular high school social studies teacher in suburban Omaha, Nebraska, and Reese Witherspoon as Tracy Flick, around the time of the school's student body election. When Tracy qualifies to run for class president, McAllister believes she does not deserve the title and tries his best to stop her from winning. Election opened to acclaim from critics, who praised its writing and direction.** The film received an Academy Award nomination for Best Adapted Screenplay, a Golden Globe nomination for Witherspoon in the Best Actress category, and the Independent Spirit Award for Best Film in 1999.
	Word level attractiveness	Election is a 1999 American **comedy-drama** film directed and written by Alexander Payne and adapted by him and Jim Taylor from Tom Perrotta's 1998 novel of the same title.
	Cast member attractiveness	**Alexander Payne**, Reese Witherspoon, Matthew Broderick, Jim Taylor
User 2163	*Sentence level attractiveness*	Election is a 1999 American comedy-drama film directed and written by Alexander Payne and adapted by him and Jim Taylor from Tom Perrotta's 1998 novel of the same title. The plot revolves around a high school election and satirizes both suburban high school life and politics. The film stars Matthew Broderick as Jim McAllister, a popular high school social studies teacher in suburban Omaha, Nebraska, and Reese Witherspoon as Tracy Flick, around the time of the school's student body election. When Tracy qualifies to run for class president, McAllister believes she does not deserve the title and tries his best to stop her from winning. Election opened to acclaim from critics, who praised its writing and direction. **The film received an Academy Award nomination for Best Adapted Screenplay, a Golden Globe nomination for Witherspoon in the Best Actress category, and the Independent Spirit Award for Best Film in 1999.**
	Word level attractiveness	The film received an Academy **Award** nomination for **Best** Adapted Screenplay, a Golden Globe nomination for Witherspoon in the **Best** Actress category, and the Independent Spirit **Award** for **Best** Film in 1999
	Cast member attractiveness	Alexander Payne, **Reese Witherspoon**, Matthew Broderick, Jim Taylor

Statistical attractiveness on movie ***Election (1999)*** w.r.t. sentences, words in the most attractive sentences and cast members. The larger size and deeper color of font denote the larger attractiveness weight is assigned.

Open issues and directions

- More advanced approaches involving Psychology, Neuroscience, Brain science, are demanded to precisely model attraction.

- Attraction modeling on more data types as well as text, e.g. images, videos, audios.

- Attraction is quite subjective, which changes with context
  - Incorporating contextual information for modeling context-aware attraction is more preferable

Non-IID RS in practice

- Non-IID RS in practice
  - The evolution of recommendation in Netflix
  - System architecture for recommendation
  - The process of launching new recommendation algorithms

Non-IID RS in practice

Non-IID RS in practice

- Non-IID RS in practice
  - The evolution of recommendation in Netflix
  - System architecture for recommendation
  - The process of launching new recommendation algorithms

Non-IID RS in practice

Recommendation in Netflix

https://medium.com/netflix-techblog

Recommending for the World

#AlgorithmsEverywhere

by Yves Raimond and Justin Basilico

https://medium.com/netflix-techblog/recommending-for-the-world-8da8cbcf051b

Evolution of Netflix

Personalized Recommendation

2006

2016

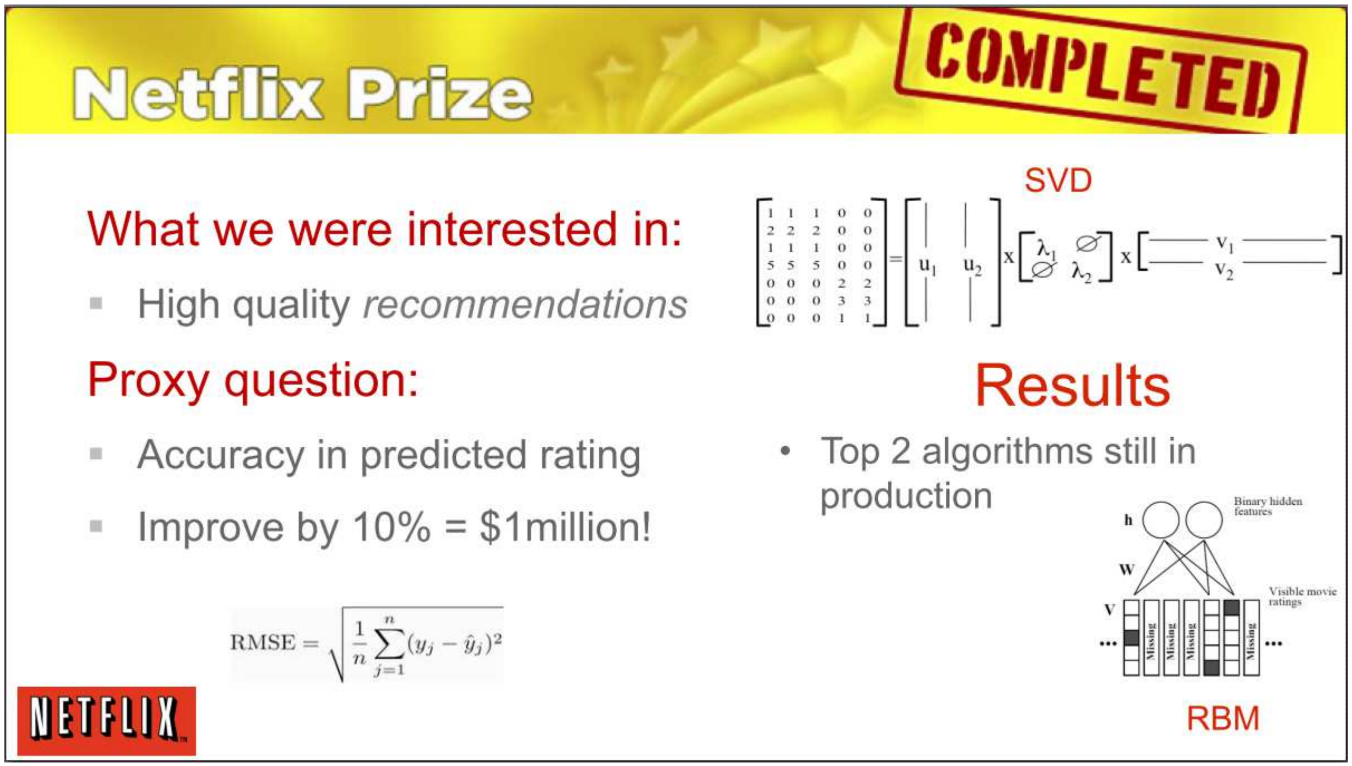
Netflix Prize

- In October 2006, Netflix as a service peddling discs of movie and TV show, announced “The Netflix Prize”

- The Netflix Prize was an open competition for the best collaborative filtering algorithm to predict user ratings for films

- The mission: Make the company's recommendation engine 10% more accurate

Netflix Prize in 2012



https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429

MF wins Netflix Prize (2012)

SVD for Rating Prediction

- User factor vectors $p_u \in \Re^f$ and item-factors vector $q_v \in \Re^f$
- Baseline $b_{uv} = \mu + b_u + b_v$ (user & item deviation from average)
- Predict rating as $r'_{uv} = b_{uv} + p_u^T q_v$
- SVD++ (Koren et. Al) asymmetric variation w. implicit feedback

$$r'_{uv} = b_{uv} + q_v^T \left(|R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} (r_{uj} - b_{uj}) x_j + |N(u)|^{-\frac{1}{2}} \sum_{j \in N(u)} y_j \right)$$

- Where
  - $q_v, x_v, y_v \in \Re^f$ are three item factor vectors
  - Users are not parametrized, but rather represented by:
    - R(u): items rated by user u
    - N(u): items for which the user has given implicit preference (e.g. rated vs. not rated)

NETFLIX

RBM wins Netflix Prize (2012)

RBM for the Netflix Prize

Restricted Boltzmann Machines for Collaborative Filtering

Ruslan Salakhutdinov RSALAKHU@CS.TORONTO.EDU
Andriy Mnih AMNIH@CS.TORONTO.EDU
Geoffrey Hinton HINTON@CS.TORONTO.EDU
University of Toronto, 6 King's College Rd., Toronto, Ontario M5S 3G4, Canada

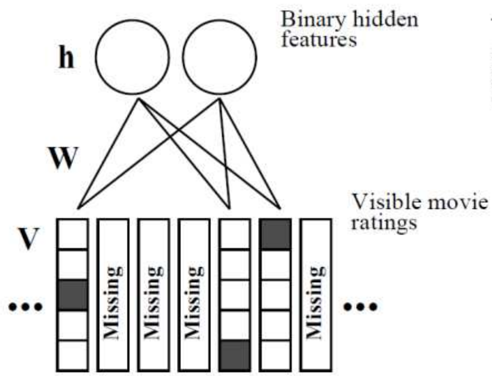


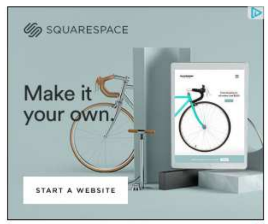
Figure 1. A restricted Boltzmann machine with binary hidden units and softmax visible units. For each user, the RBM only includes softmax units for the movies that user has rated. In addition to the symmetric weights between each hidden unit and each of the $K = 5$ values of a softmax unit, there are 5 biases for each softmax unit and one for each hidden unit. When modeling user ratings with an RBM that has Gaussian hidden units, the top layer is composed of linear units with Gaussian noise.

Most watch on Netflix comes from Recommendation (2013)

TECH 08/01/2013 08:00 am ET | Updated Aug 01, 2013

Netflix Launches Profiles, Finally Realizing How People Really Watch Movies On It

By Timothy Stenovec



TRENDING

- Donald Trump Stayed On The Golf Course As Hawaii Panicked
- Projector Lights Up Trump's D.C. Hotel With 'Shithole' And Poop Emojis
- Bill Murray Slays As The 'Bannon Cannon' On 'Saturday Night Live'
- Liam Neeson Calls The #MeToo Movement A 'Bit Of A Witch Hunt'

For years, people who share Netflix accounts have befuddled the streaming service's recommendation engine, the tool that in theory is supposed to use what you've watched before to suggest movies, documentaries and TV shows you'd like. But your kids may stream Disney movies and Sesame Street, and you may binge on episodes of "House of Cards" and "Breaking Bad," leading Netflix to suggest movies and TV shows that may not appeal to anyone in your household.

In an attempt to fix this, Netflix today begins rolling out profiles, a free feature that allows any of the company's 37 million subscribers to create up to five different profiles on one account. Each profile will be treated like its own account, so recommendations will be more aligned with a single person's interests.

Figuring out what people want to watch is key to Netflix's success. In an increasingly competitive streaming environment, where Hulu Plus and Amazon Prime Instant Video ink their own deals for exclusive and original content, Netflix needs not only to continue to attract new subscribers, but also keep existing ones happy. One way the company can do that — and keep people from ditching its service for a competitor — is by suggesting content that subscribers will like.

Introducing profiles is a move to combat "churn," the number of people who sign up and then quit paying the $7.99 monthly fee if they feel like it's not valuable, said Mike McGuire, a vice president at Gartner, the technology research firm.

"When you're in the subscription business, churn is your worst enemy," said McGuire. "If there's not something else they're surfacing that meets your interest beyond what you initially dialed in for, then you're out."

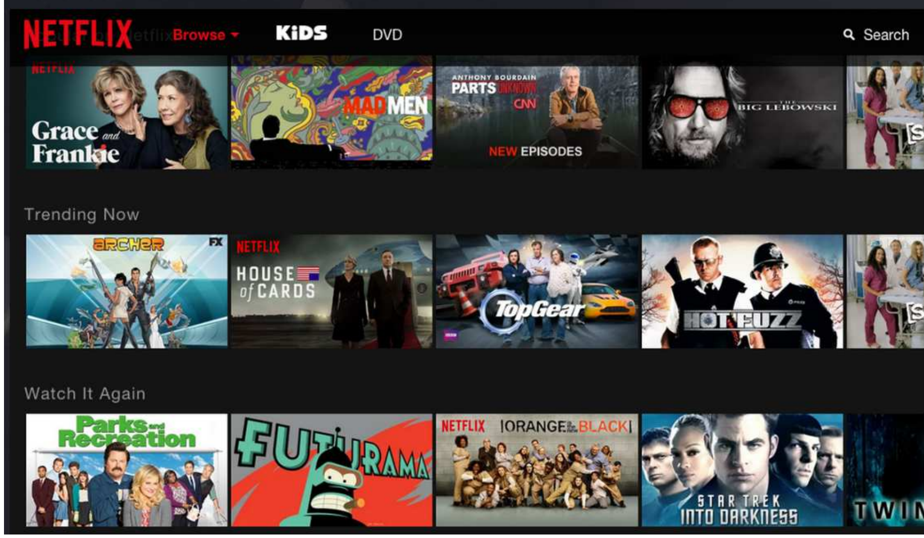
About 75 percent to 80 percent of what people watch on Netflix comes from what Netflix recommends, not from what people search for, said Yellin.

https://www.huffingtonpost.com/entry/netflix-profiles_n_3685876

Netflix has applied deep learning for recommendation

Deep Learning with Tony Jebara, Director of ML Research at Netflix

By **Sophie Curtis**, Marketing Director - RE•WORK
May 26, 2016



Tell us more about your work as Director of Machine Learning Research at Netflix.
At Netflix we are inventing the future of Internet television and helping members across the world find videos to watch and enjoy. We help them make a selection from a catalog of thousands of titles. But we need to tailor recommendations to each user and each session within seconds and within a menu of 10 to 20 visible options. Achieving this relies on our recommendation system which is really an ecosystem of many machine learning algorithms that operate together. We are constantly working on improving these algorithms. We are also leveraging machine learning across all parts of Netflix: from deciding which new titles to add to our catalog to finding ways to more efficiently stream videos across the Internet.

What do you feel are the leading factors enabling recent advancements and uptake of deep learning?
The recent adoption of deep learning has been enabled by the confluence of several factors. Of course, bigger data-sets and more computational power have been essential. But, they triggered something more important: the freedom to try bigger and deeper models. Recent theoretical research by Choromanska et al shows that models with more layers and more neurons per layer become resilient to the local optima that plague the optimization landscape. Bigger models find more reliable solutions during learning while small ones get stuck in bad solutions. So, it's been a chain reaction: bigger computation led to bigger data, then to bigger models, then to better optima, and finally to better performance.

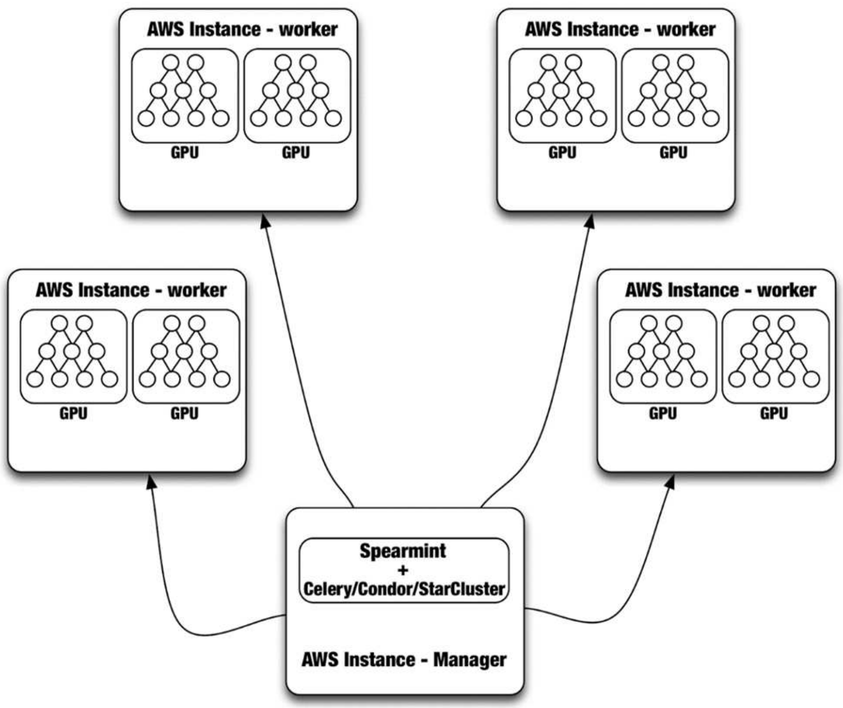
What are your thoughts on the recent surge of media interest surrounding deep learning?
The media interest certainly adds to the excitement. The uptake in press and articles has often revolved on deep learning shattering AI milestones in areas such as game-playing, computer vision and so on. But more practical progress is happening in business and commercial fronts where deep learning and machine learning are permeating almost every component of the workplace. So, beyond exciting milestones in the media (such as beating the grandmaster of Go), we are seeing a sustained ground swell in deep learning at companies all over the world.

How can larger corporations working on deep learning ensure that their work benefits others within this field?
Corporations are increasingly part of the conversation and now have a much stronger presence at leading conferences in machine learning and deep learning. They are not only providing funding and exhibit booths at the events but are also contributing papers and organizing workshops. Some are even releasing open-source software and systems (such as Google's TensorFlow) which broaden the reach of deep learning to anyone in the world who wants to get involved.

What present or potential future applications of deep learning excite you most?
I'm excited to see how deep learning can help in recommendation, personalization and search. So far, we've seen deep learning solve tasks that humans are already good at, such as vision, speech recognition, gaming or natural language processing. But humans are notoriously bad at recommending content or items to their friends. We think aspirationally rather than realistically; we recommend a high-brow documentary that sounds intellectual rather than what our friends really would rather watch. I'm excited to see how deep learning can anticipate these bias and preferences and help us each optimize our entertainment, our disposable time and our everyday life in general.

https://www.re-work.co/blog/deep-learning-tony-jebara-machine-learning-research-netflix

Distributed Neural Networks with GPUs in the AWS Cloud



- Implementing bleeding edge solutions to train large-scale Neural Networks using GPUs
- The cost and the complexity might be overwhelming if doing it in own custom infrastructure.
- Levering the public AWS cloud with the customization and use of the instance resources.

https://medium.com/netflix-techblog/distributed-neural-networks-with-gpus-in-the-aws-cloud-ccf71e82056b

Vectorflow:
a neural network library optimized for sparse data

Scope

NETFLIX

Lines of code*:

*: git ls-files | xargs cat | wc -l

TensorFlow	PYTORCH	mxnet	Chainer	vectorflow
1,888k	252k	2,322k	110k	**6k**

vectorflow

- 0.05 dev (I spend 5% of my time on it)
- offers a minimal DAG with backprop for feed-forward nets
- sparse data as first class citizen
- arbitrary loss function
- extremely fast on CPU
    - 0 memory allocation
    - lock-free inter-core parallelism
    - LLVM intrinsics for dense ops SIMD vectorization

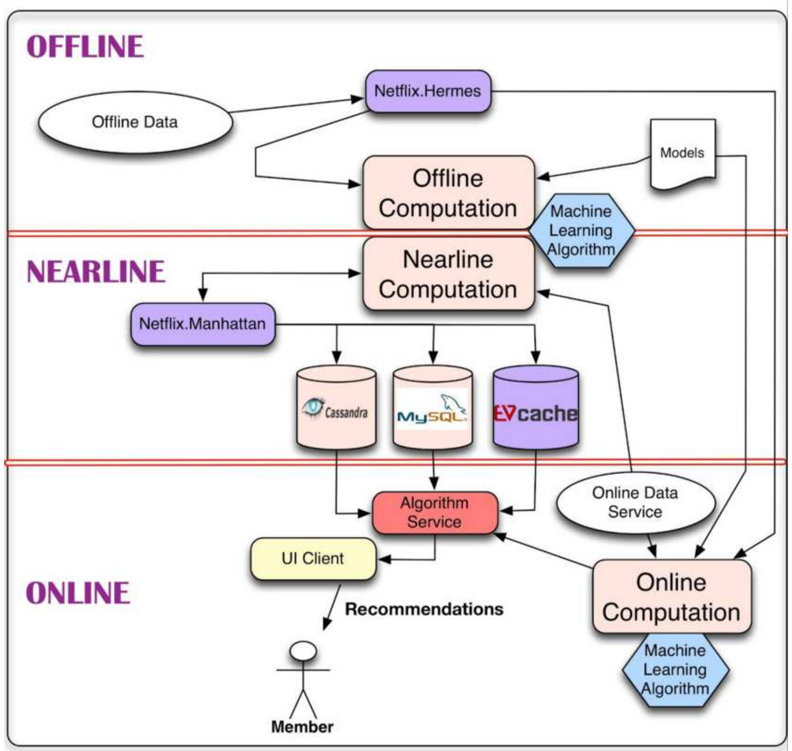
https://github.com/Netflix/vectorflow

Non-IID RS in practice

- Non-IID RS in practice
  - The evolution of recommendation in Netflix
  - System architecture for recommendation
  - The process of launching new recommendation algorithms

Non-IID RS in practice

System diagram for personalized recommendation (2013)



- Offline jobs: **model training** and **batch computation of intermediate or final results**.
- **Nearline computation** is an intermediate compromise between these two modes in which we can perform online-like computations, but do not require them to be served in real-time.
- **Online computation** responds better to recent events and user interaction, and responds to requests in real-time.

https://medium.com/netflix-techblog/system-architectures-for-personalization-and-recommendation-e081aa94b5d8

Non-IID RS in practice

- Non-IID RS in practice
  - The evolution of recommendation in Netflix
  - System architecture for recommendation
  - The process of launching new recommendation algorithms

Non-IID RS in practice

Incomplete list of methods in machine learning for personalization

- Linear regression
- Logistic regression
- Elastic nets
- Singular Value Decomposition
- Restricted Boltzmann Machines
- Markov Chains
- Latent Dirichlet Allocation
- Association Rules
- Gradient Boosted Decision Trees
- Random Forests
- Clustering techniques from the simple k-means to novel graphical approaches
- Matrix factorization

https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-2-d9b96aa399f5

Algorithms selection and validation in Netflix

- When we test something, we want to understand why it failed or succeeded.
- So, how does this work in practice?
- It is a slight variation over the traditional scientific process called A/B testing (or bucket testing):

The process of A/B testing in Netflix

1. **Start with a hypothesis**

   Algorithm/feature/design X will increase member engagement with service and *ultimately member retention*

2. **Design a test**

   Develop a solution or prototype. Ideal execution can be 2X as effective as a prototype, but not 10X.

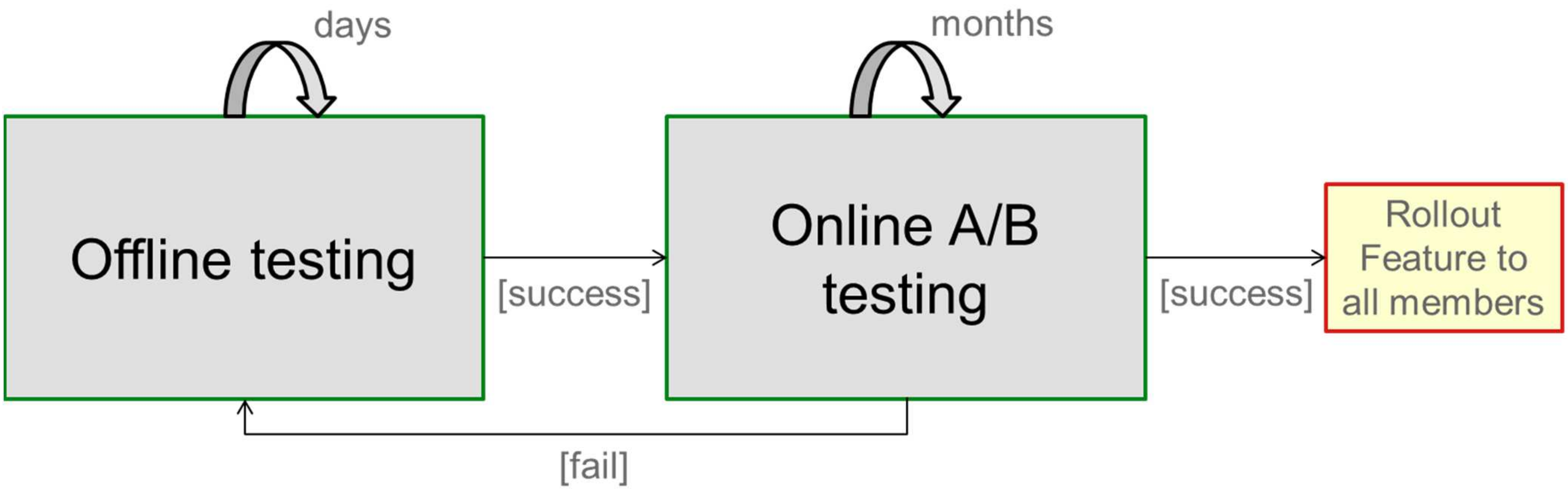
3. **Execute the test**

4. **Let data speak for itself**
   - When executing A/B tests, Netflix track many different metrics.
   - Tests usually have thousands of members and anywhere from 2 to 20 cells exploring variations of a base idea.
   - The key advantage of A/B tests is that they allow decisions to be data-driven.

https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-2-d9b96aa399f5

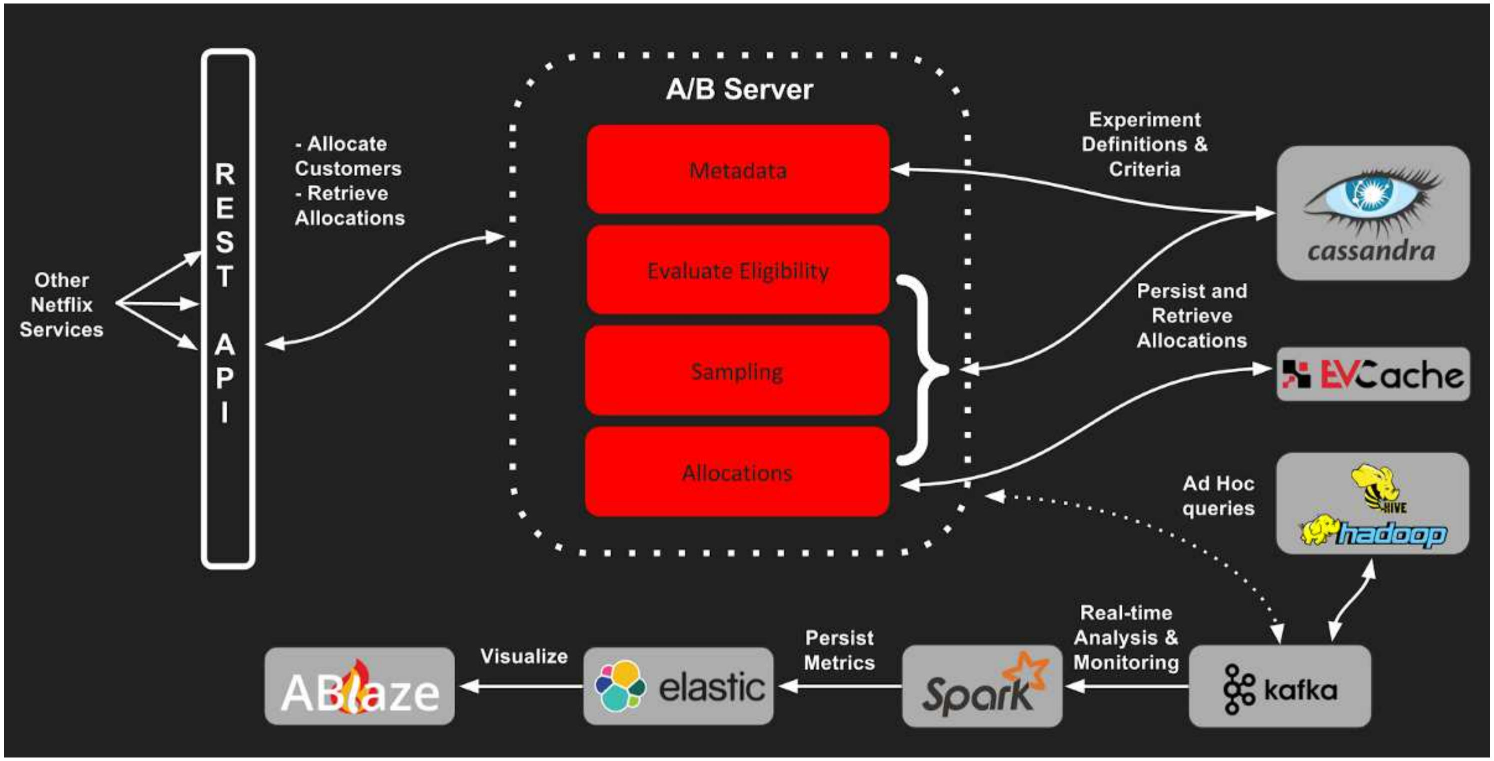
Offline/online Testing

The offline testing cycle is a step to test and optimize algorithms prior to performing online A/B testing



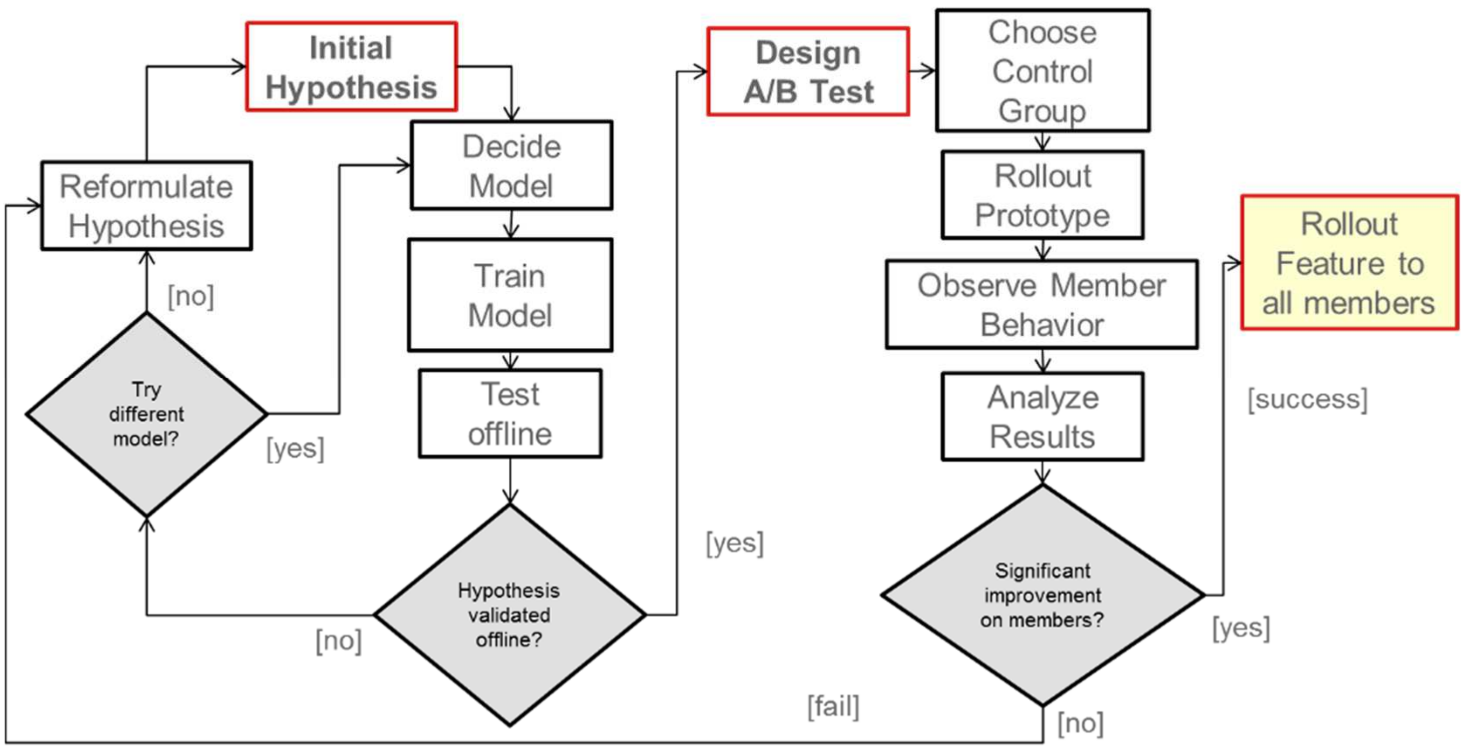
https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-2-d9b96aa399f5

The experimentation platform for A/B testing



https://medium.com/netflix-techblog/its-all-a-bout-testing-the-netflix-experimentation-platform-4e1ca458c15

The process of rolling out feature



https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-2-d9b96aa399f5

Innovation Cycle: Top10 Marathon

- 10-week effort to quickly test dozens of algorithmic ideas related to improving Top10 row

https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-2-d9b96aa399f5

The lesson learned from Netflix

- How to seamlessly integrate personalized recommendation into real business
- How to bridge the gap between the algorithms in papers to the real systems.
- How to design and build large-scale and real-time recommender systems
- How to adopt a scientific process to select and validate algorithmic ideas

https://medium.com/netflix-techblog

Conclusion in one word

Data complexity is the main challenge in RS

Non-IID and RS

Overview
Challenges
Non-IIDness in RS

Representations is the foundation of RS

Data Representation

Attributes

- Text
- Rating table
- Image
- Sequence

Each choice is not only made by oneself

Non-IIDness on users

- Social RS
- Group RS

Items in different domains have resonance

Non-IIDness on items

- Cross-domain RS
- Session-based RS

Choices are made with unseen interaction

Non-IIDness on implicit interaction

- Context-aware RS
- Multi-objective RS
- Attraction RS

Real-world RS is more than algorithms

Non-IID RS in practice

Words for taking home

- Get insight into the ubiquitous non-IIDness in modern RSs due to the data complexity
- Try to link modern AI techniques to represent heterogeneities and couplings in complex data
- Practice building novel non-IID RSs with state-of-the-art machine learning approaches
- Customize and deploy real-world RSs from new ideas with practical methodology

References

- Kantor, P. B. (2015). *Recommender systems handbook*. F. Ricci, L. Rokach, & B. Shapira (Eds.). Berlin, Germany:: Springer.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. http://www.deeplearningbook.org/
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning (Information Science and Statistics): Springer-Verlag New York, Inc.
- Cao, L. (2016). Non-IID Recommender Systems: A Review and Framework of Recommendation Paradigm Shifting. *Engineering*, *2*(2), 212-224.
- Pan, W., Xiang, E. W., Liu, N. N., & Yang, Q. (2010, July). Transfer Learning in Collaborative Filtering for Sparsity Reduction. In AAAI (Vol. 10, pp. 230-235).
- Singh, A. P., & Gordon, G. J. (2008, August). Relational learning via collective matrix factorization. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 650-658). ACM.
- Marlin, B.M., Zemel, R.S., Roweis, S., and Slaney, M. Collaborative filtering and the missing at random assumption. In Proceeding 23rd Conference on Uncertainty in Artificial Intelligence, 2007.
- Hu, Y., Koren, Y., and Volinsky, C. Collaborative Filtering for Implicit Feedback Datasets. In Eighth IEEE International Conference on Data Mining, 263-272, 2008.
- Elkahky, A.M., Song, Y., and He, X. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In Proceedings of the 24th International Conference on World Wide Web, 278-288, 2015.
- Kim, T., Cha, M., Kim, H., Lee, J., & Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. arXiv preprint arXiv:1703.05192.
- Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., & Yang, D. (2016). Learning Informative Priors from Heterogeneous Domains to Improve Recommendation in Cold-Start User Domains. ACM Transactions on Information Systems (TOIS), 35(2), 13.
- Jian, S, Hu, L, Cao, L & Lu, K. AAAI 2018. Metric-based Auto-Instructor for Learning Mixed Data Representation
- Mikolov, T., Corrado, G., Chen, K., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space.

References

- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality
- Ma, H., Yang, H., Lyu, M. R., & King, I. (2008, October). Sorec: social recommendation using probabilistic matrix factorization. In Proceedings of the 17th ACM conference on Information and knowledge management (pp. 931-940). ACM.
- Jamali, M., & Ester, M. (2010, September). A matrix factorization technique with trust propagation for recommendation in social networks. In Proceedings of the fourth ACM conference on Recommender systems (pp. 135-142). ACM.
- Ma, H., Zhou, D., Liu, C., Lyu, M. R., & King, I. (2011, February). Recommender systems with social regularization. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 287-296). ACM.
- Wang, X., He, X., Nie, L., & Chua, T. S. (2017). Item Silk Road: Recommending Items from Information Domains to Social Users. arXiv preprint arXiv:1706.03205.
- Oramas, S., Nieto, O., Sordo, M., & Serra, X. (2017). A deep multimodal approach for cold-start music recommendation. arXiv preprint arXiv:1706.09739.
- Lynch, C., Aryafar, K., & Attenberg, J. (2016, August). Images don't lie: Transferring deep visual semantic features to large-scale multimodal learning to rank. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 541-548). ACM.
- Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., & Wang, J. (2017). Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. *ACM Transactions on Information Systems (TOIS)*, *35*(3), 25.
- Karatzoglou, A., Amatriain, X., Baltrunas, L., & Oliver, N. (2010, September). Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 79-86). ACM.
- Rendle, S., Gantner, Z., Freudenthaler, C., & Schmidt-Thieme, L. (2011, July). Fast context-aware recommendations with factorization machines. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 635-644). ACM.
- Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., ... & Anil, R. (2016, September). *Wide & deep learning for recommender systems*. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems* (pp. 7-10). ACM.
- Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... & Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In *ICML 2015*.

References

- Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Rendle, S., Freudenthaler, C., and Schmidt-Thieme , L. (2010, August). *Factorizing Personalized Markov Chains for Next-Basket Recommendation*. WWW2010.
- Hidasi, B., Karatzoglou,A., Baltrunas, L., and Tikk, D. (2016, May). *Session-based Recommendations with Recurrent Neural Networks*. ICLR2016.
- Gravity R, B., Quadrana, M., Karatzoglou, A., and Tikk, D. (2016 August). *Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations*. RecSys'2016.
- Hu, L., Cao, L., Wang, S., Xu, G., Cao, J., & Gu, Z. (2017, January). Diversifying Personalized Recommendation with User-session Context. In *IJCAI International Joint Conference on Artificial Intelligence*.
- Wang, S., Hu, L., & Cao, L. (2017, September). *Perceiving the Next Choice with Comprehensive Transaction Embeddings for Online Recommendation*. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*(pp. 285-302). Springer, Cham.
- Wang, S., Hu, L., & Cao, L. (2018, February). *Attention-based Transactional Context Embeddings for Next-Item Recommendation*. AAAI2018
- Loyola, P., Liu, C., and Hirate, Y. (2017 August). *Modeling User Session and Intent with an Attention-based Encoder-Decoder Architecture*. RecSys'2017.
- Lu, Q., Yang, D., Chen, T., Zhang, W., & Yu, Y. (2011, October). Informative household recommendation with feature-based matrix factorization. In *proceedings of the 2nd Challenge on Context-Aware Movie Recommendation* (pp. 15-22). ACM.
- Hu, L., Cao, J., Xu, G., Cao, L., Gu, Z., & Cao, W. (2014, July). Deep Modeling of Group Preferences for Group-Based Recommendation. In *AAAI* (Vol. 14, pp. 1861-1867).
- Hu, L., Jian, S., Cao, L., & Chen, Q., Interpretable Recommendation via Attraction Modeling: \\Learning Multilevel Attractiveness over Multimodal Movie Contents, IJCAI 2018
- Masthoff, J. (2015). Group recommender systems: aggregation, satisfaction and group attributes. In *Recommender Systems Handbook* (pp. 743-776). Springer US.
- Netflix Tech Blog: https://medium.com/netflix-techblog

Thank you for attention

- Papers & Slides:
  - https://sites.google.com/view/lianghu/
  - http://www-staff.it.uts.edu.au/~lbcao/
  - https://jiansonglei.github.io/
  - http://www.datasciences.org/

- Comments & collaborations:

  - rainmilk@gmail.com
  - longbing.cao@uts.edu.au
  - jiansonglei@163.com