https://preferred.ai/recsys21-tutorial/

Multi-Modal Recommender Systems: Hands-On Exploration

Quoc-Tuan Truong

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Aghiles Salah

Rakuten Institute of Technology

Hady W. Lauw

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

PREFERRED.AI

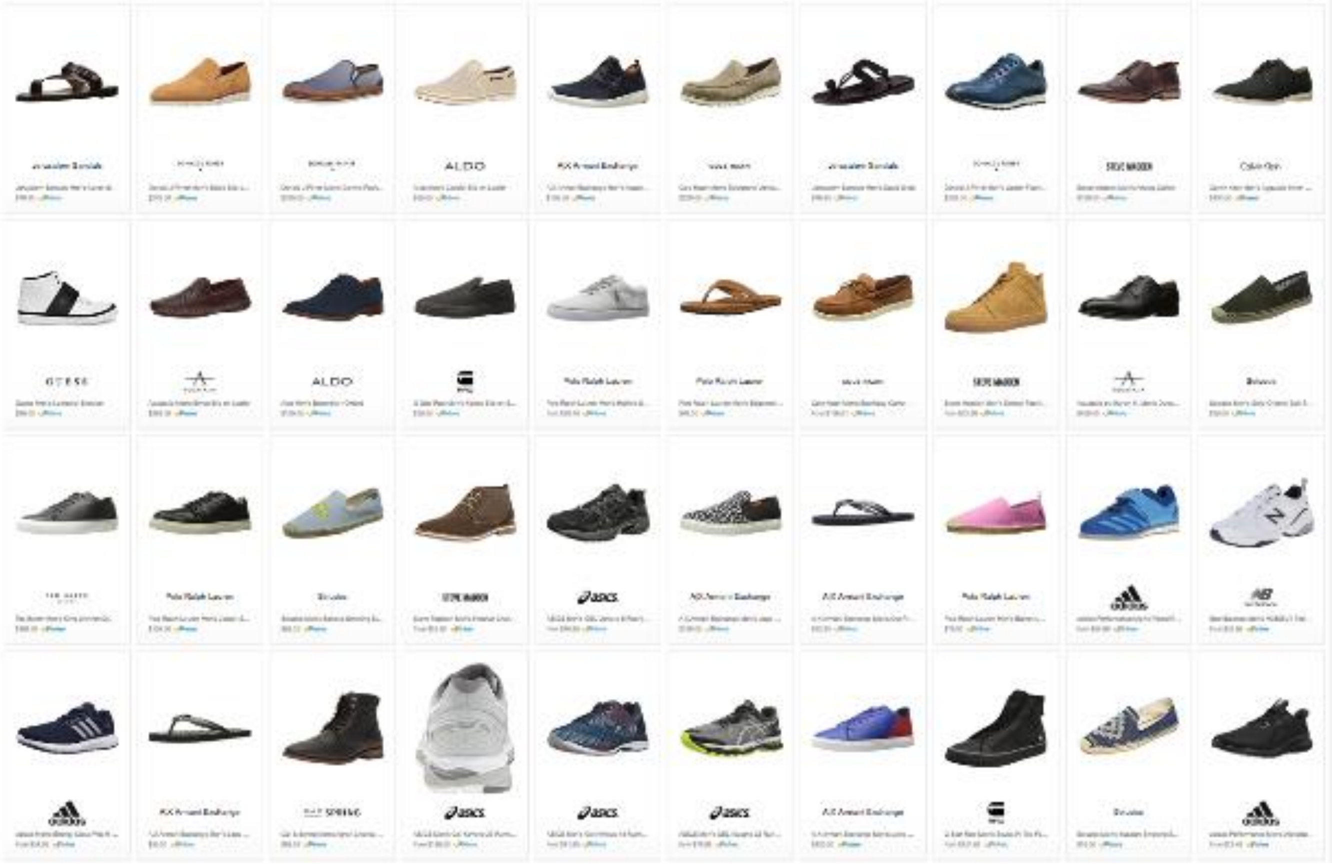
Outline

- Recommender Systems & Multi-Modality – 30 min
- Hands-on 1: Cornac – 15 min
- Text & Image Modalities – 30 min
- Q & A – 15 min
- Coffee break – 30 min
- Graph Modality – 25 min
- Hands-on 2: Multi-modal – 30 min
- Hands-on 3: Cross-modal – 20 min
- Future Directions – 5min
- Q & A – 10 min

Recommender Systems and Multi-Modality

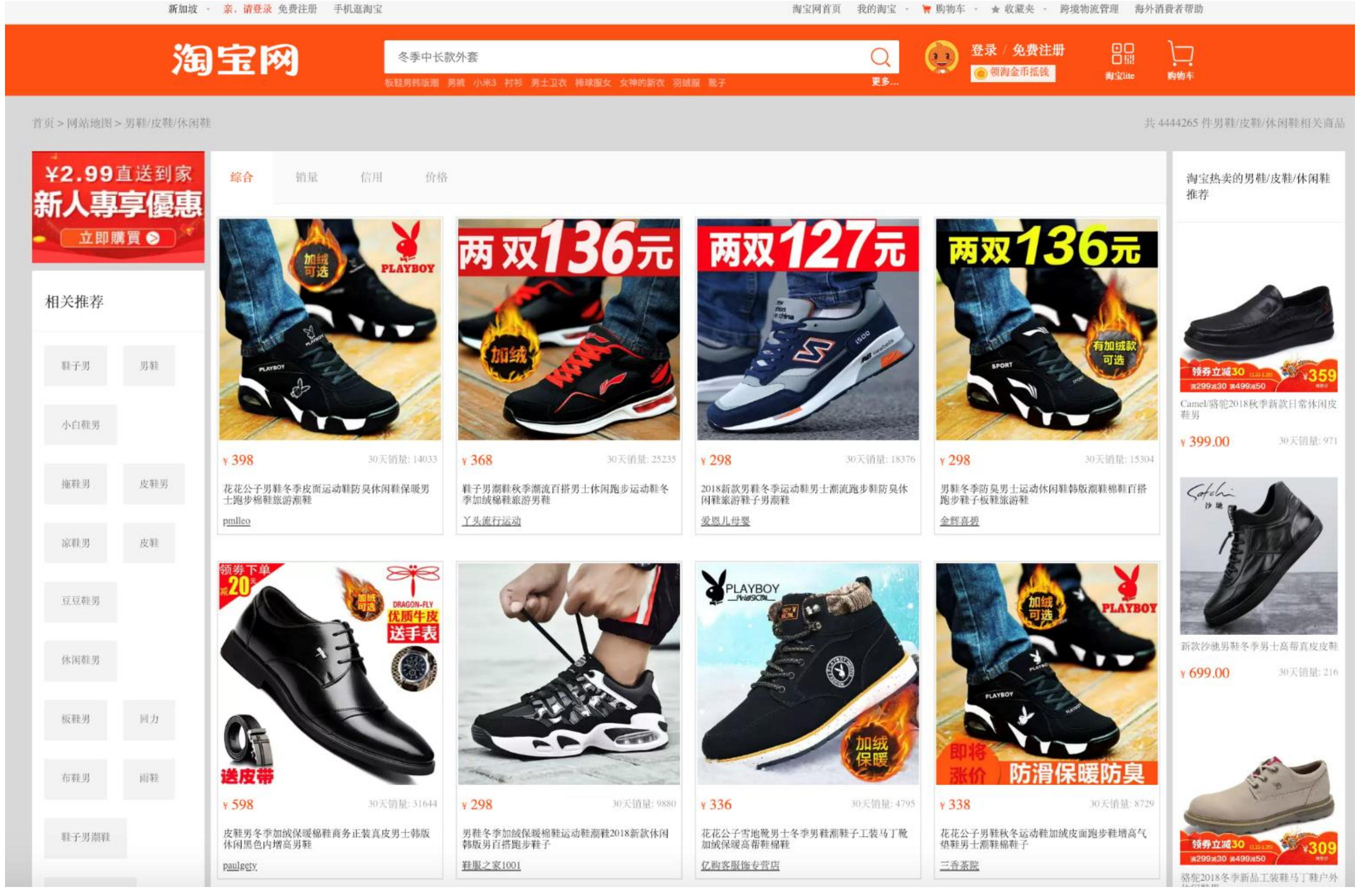
Choice Explosion

133,520 results for ”Men’s Shoes"

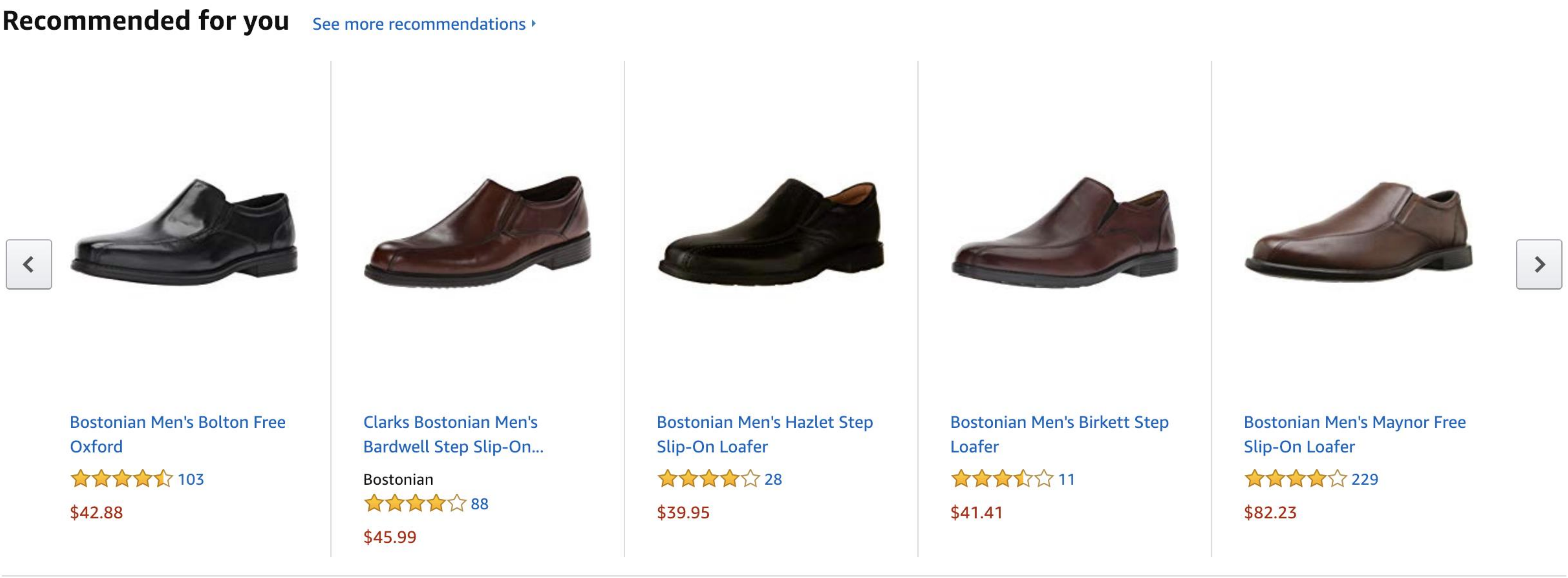


Choice Explosion

4 million results results for "男鞋/皮鞋/休闲鞋"



Solution: Recommender Systems



amazon

Where do we see a recommender system?

LAZADA
Effortless Shopping

YouTube

in

amazon

Rakuten

淘宝网
Taobao.com

last.fm
the social music revolution

Google news

tripadvisor

Spotify

Booking.com

yelp

facebook

Pinterest

foursquare

Formulation

Rating Prediction

- Given a rating dataset $\boldsymbol{R}$
  - each $r_{ui} \in \boldsymbol{R}$ indicates a rating, or the degree of preference user u has for item i
  - where $r_{ui} \notin \boldsymbol{R}$, predict rating $\hat{r}_{ui}$ as a function $f(u, i)$
- "Is the user going to like this item?"

Ranking

- Given a ranking dataset $\boldsymbol{T}$
  - each triple $t_{uij} \in \boldsymbol{T}$ indicates whether/how much u prefers i to j
  - where $t_{uij} \notin \boldsymbol{T}$, predict rating $\hat{t}_{uij}$ as a function $g(u, i, j)$
- "Which items are the user most likely to have an interest?"

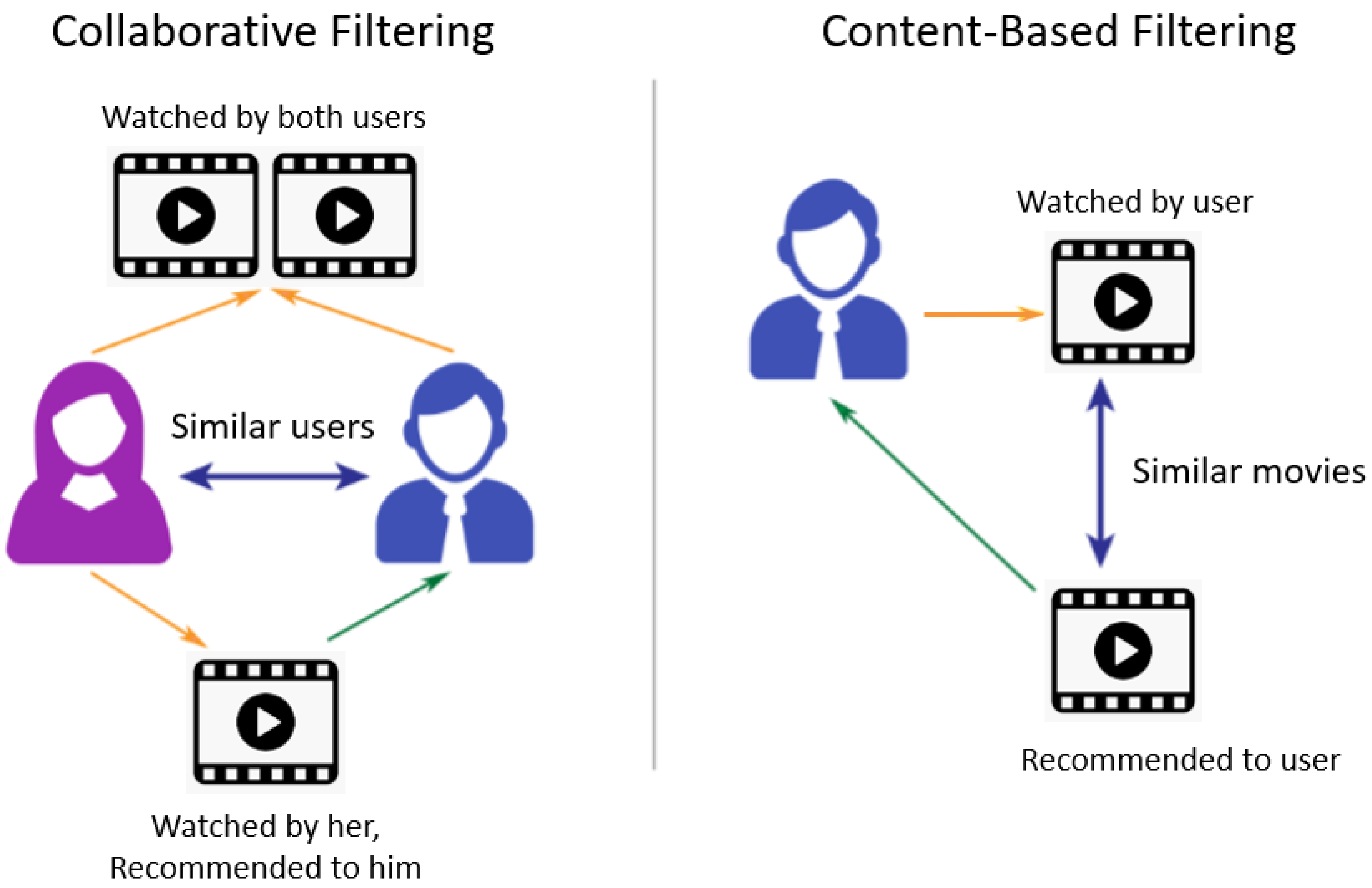
Basis for User-Item Relations

Collaborative Filtering

- A user tends to have similar consumption behavior to other 'like-minded' users

Content-Based Filtering

- A user tends to like items with similar contents to those previously consumed



https://www.kdnuggets.com/2019/11/content-based-recommender-using-natural-language-processing-nlp.html

Netflix Prize

https://bits.blogs.nytimes.com/2009/09/21/netflix-awards-1-million-prize-and-starts-a-new-contest/

Matrix Factorization

Koren, Bell, and Volinsky, "Matrix Factorization Techniques for Recommender Systems", IEEE Computer, 2009

Our winning entries consist of more than 100 different predictor sets, the majority of which are factorization models using some variants of the methods described here. Our discussions with other top teams and postings on the public contest forum indicate that these are the most popular and successful methods for predicting ratings.

~BellKor

https://datajobs.com/data-science-repo/Recommender-Systems-[Netflix].pdf

Example Rank-2 Matrix Factorization

		NERO	JULIUS CAESAR	CLEOPATRA	SLEEPLESS IN SEATTLE	PRETTY WOMAN	CASABLANCA
HISTORY	1	1	1	1	0	0	0
HISTORY	2	1	1	1	0	0	0
HISTORY	3	1	1	1	0	0	0
BOTH	4	1	1	1	1	1	1
ROMANCE	5	-1	-1	-1	1	1	1
ROMANCE	6	-1	-1	1	1	1	1
ROMANCE	7	-1	-1	-1	1	1	1

R

≈

	HISTORY	ROMANCE
1	1	0
2	1	0
3	1	0
4	1	1
5	-1	1
6	-1	1
7	-1	1

U

X

	NERO	JULIUS CAESAR	CLEOPATRA	SLEEPLESS IN SEATTLE	PRETTY WOMAN	CASABLANCA
HISTORY	1	1	1	0	0	0
ROMANCE	0	0	1	1	1	1

V^T

		NERO	JULIUS CAESAR	CLEOPATRA	SLEEPLESS IN SEATTLE	PRETTY WOMAN	CASABLANCA
HISTORY	1	0	0	0	0	0	0
HISTORY	2	0	0	0	0	0	0
HISTORY	3	0	0	0	0	0	0
BOTH	4	0	0	-1	0	0	0
ROMANCE	5	0	0	-1	0	0	0
ROMANCE	6	0	0	1	0	0	0
ROMANCE	7	0	0	-1	0	0	0

R

Residual matrix: $\boldsymbol{R} - \boldsymbol{U}\boldsymbol{V}^T$

Estimating Latent Factors

- Minimize loss function:

$$\mathcal{L}(\boldsymbol{U},\boldsymbol{V}) = \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{U}\boldsymbol{V}^T\|^2 = \frac{1}{2}\sum_{r_{ij}\in\boldsymbol{R}}\left(r_{ij} - \sum_{k=1}^{K} u_{ik}\cdot v_{jk}\right)^2$$

  - defined only over observed ratings

Regularization

- Due to sparsity, some users or items may have very few ratings
- To prevent overfitting, we can introduce regularization:

$$\mathcal{L}(\boldsymbol{U},\boldsymbol{V}|\lambda) = \frac{1}{2}\sum_{r_{ij}\in\boldsymbol{R}}\left(r_{ij} - \sum_{k=1}^{K} u_{ik}\cdot v_{jk}\right)^2 + \frac{\lambda}{2}\sum_{i=1}^{N}\|\boldsymbol{u}_i\|^2 + \frac{\lambda}{2}\sum_{j=1}^{M}\|\boldsymbol{v}_j\|^2$$

$$= \frac{1}{2}\sum_{r_{ij}\in\boldsymbol{R}}\left(r_{ij} - \sum_{k=1}^{K} u_{ik}\cdot v_{jk}\right)^2 + \frac{\lambda}{2}\sum_{i=1}^{N}\sum_{k=1}^{K}(u_{ik})^2 + \frac{\lambda}{2}\sum_{j=1}^{M}\sum_{k=1}^{K}(v_{jk})^2$$

User and Item Biases

- Users have different ranges of ratings
  - some are generous with their ratings mainly in the upper range, others are strict
- Items differ in popularity or likeability
  - some have mainly high ratings, others mainly low
- Introduce bias parameters:

$$\hat{r}_{ij} = \mu + b_{u_i} + b_{v_j} + \boldsymbol{u}_i^T \boldsymbol{v}_j$$

  - μ is the global average in the training data, to be computed
  - b_{u_i} is a bias term for a specific user, to be learnt
  - b_{v_j} is a bias term for a specific item, to be learn
- Loss function:

$$\mathcal{L}(\boldsymbol{U}, \boldsymbol{V}|\lambda) = \frac{1}{2} \sum_{r_{ij} \in \boldsymbol{R}} \left(r_{ij} - (\mu + b_{u_i} + b_{v_j} + \boldsymbol{u}_i^T \boldsymbol{v}_j) \right)^2 + \frac{\lambda}{2} \sum_{i=1}^{N} \left(\|\boldsymbol{u}_i\|^2 + \left(b_{u_i}\right)^2 \right) + \frac{\lambda}{2} \sum_{j=1}^{M} \left(\|\boldsymbol{v}_j\|^2 + \left(b_{v_j}\right)^2 \right)$$

Explicit vs. Implicit Feedback

Explicit

- Stated clearly and readily observable
- Ratings, thumbs up/down
- Some notion of positive/negative

Implicit

- Preferences are unclear
  - suggested, not directly expressed
- Noisy
  - clicks, viewing time
  - observed values may have very small or wide ranges
- Absence of negative signals
  - only positive examples observed

Weighted Matrix Factorization (WMF)

Hu, Koren, and Volinsky, " Collaborative Filtering for Implicit Feedback Datasets", ICDM 2009

- Adoption $\hat{p}_{ij} = \boldsymbol{u}_i^T \boldsymbol{v}_j$ by user i on item j
  - $p_{\text{ij}} = \begin{cases} 1, \text{if } r_{ij} \geq \text{threshold} \\ 0, \ \text{otherwise} \end{cases}$
- Confidence
  - $c_{\text{ij}} = \begin{cases} a, \text{if } r_{ij} \geq \text{threshold} \\ b, \ \text{otherwise} \end{cases}$

  Customarily $b \ll a$ as we have less confidence on non-observations
- Loss function

$$\mathcal{L}(\boldsymbol{U}, \boldsymbol{V} | \lambda) = \frac{1}{2} \sum_{i,j} c_{ij} \left(p_{ij} - \boldsymbol{u}_i^T \boldsymbol{v}_j \right)^2 + \frac{\lambda}{2} \sum_{i=1}^{N} \|\boldsymbol{u}_i\|^2 + \frac{\lambda}{2} \sum_{j=1}^{M} \|\boldsymbol{v}_j\|^2$$

Bayesian Personalized Ranking (BPR)

Rendle, Freudenthaler, Gantner and Schmidt-Thieme, "BPR: Bayesian Personalized Ranking from Implicit Feedback", UAI 2009

- Ordinal triple

$$j >_i l \begin{cases} 1, \text{if } r_{ij} \in \boldsymbol{R}^+ \wedge r_{il} \in \boldsymbol{R}^- \\ 0, \text{if } r_{ij} \in \boldsymbol{R}^- \wedge r_{il} \in \boldsymbol{R}^+ \\ \text{otherwise unspecified} \end{cases}$$

- Triple probability as sigmoid function

$$P(j >_i l) = \text{sigmoid}(\boldsymbol{u}_i^T \boldsymbol{v}_j - \boldsymbol{u}_i^T \boldsymbol{v}_l)$$

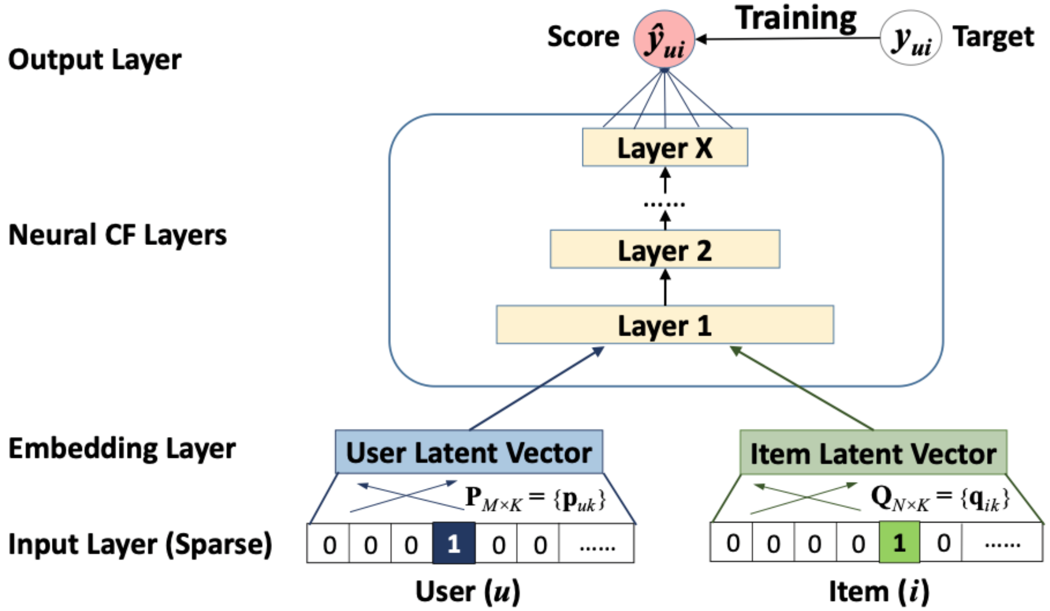
- Maximize the joint probability across all triples $\prod_{(j>_i l)\in \boldsymbol{S}} P(j >_i l)$

- With regularization, minimize regularized negative log-likelihood function:

$$\mathcal{L}(\boldsymbol{U},\boldsymbol{V}|\lambda) = \sum_{(j>_i l)\in \boldsymbol{S}} \ln(1 + \exp\{-(\boldsymbol{u}_i^T \boldsymbol{v}_j - \boldsymbol{u}_i^T \boldsymbol{v}_l)\}) + \frac{\lambda}{2}\sum_{i=1}^{N} \|\boldsymbol{u}_i\|^2 + \frac{\lambda}{2}\sum_{j=1}^{M} \|\boldsymbol{v}_j\|^2$$

Neural Collaborative Filtering

He et al., "Neural collaborative filtering", WWW, 2017

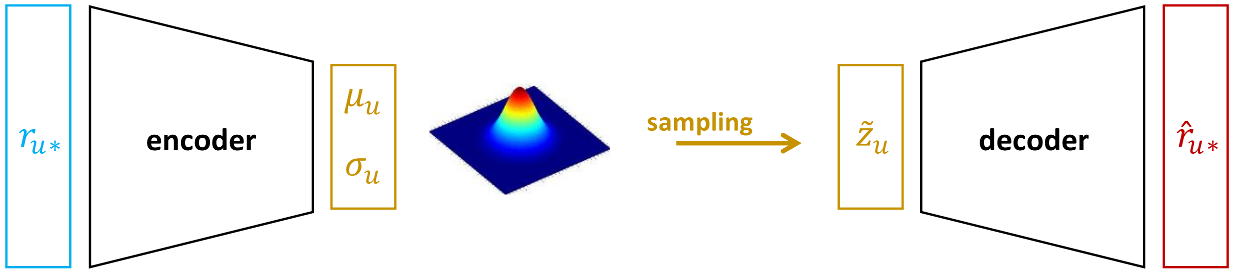


- NeuMF unifies the strengths of MF and MLP in modeling user preference
  - MF captures interaction via inner product (simple yet effective)
  - MLP is more capable of capturing complex user intention

(Bilateral) Variational Auto-Encoder

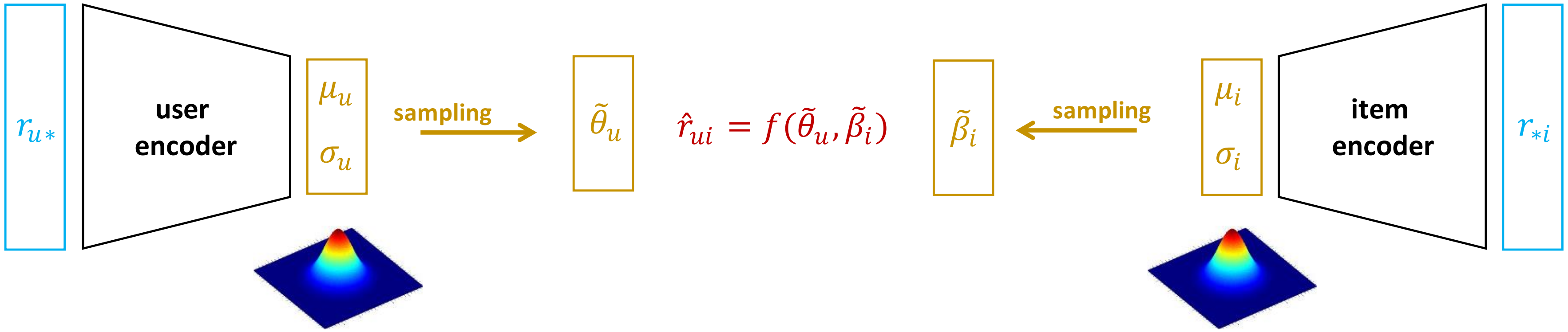
Liang, Krishnan, Hoffman, and Jebara, "Variational autoencoders for collaborative filtering", WWW, 2018.

VAE



Truong, Salah, and Lauw, "Bilateral Variational Autoencoder for Collaborative Filtering", WSDM, 2021.

BiVAE

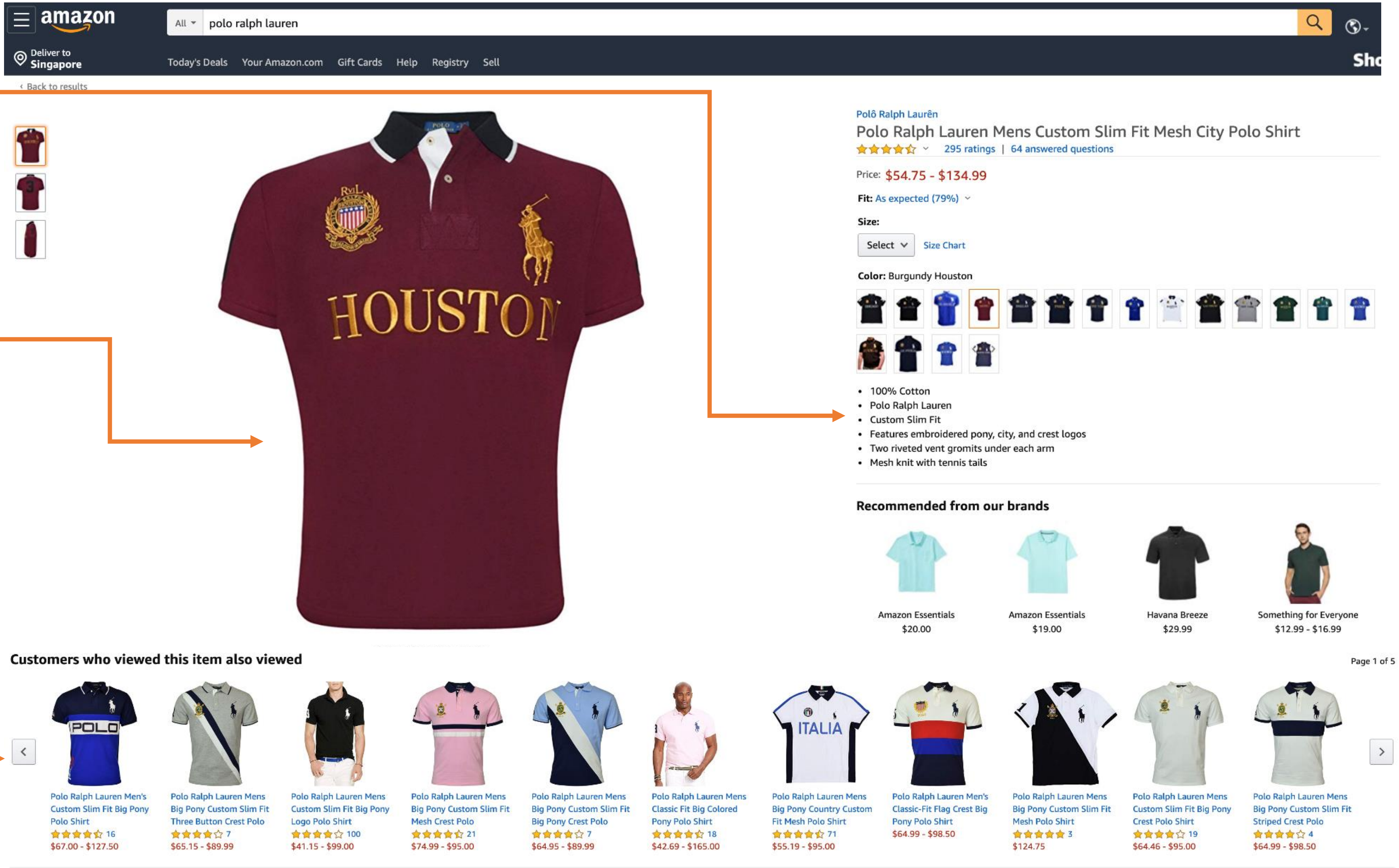


- Respect the two-way nature of dyadic data
- Can accommodate auxiliary data from both sides

Product description

Product image

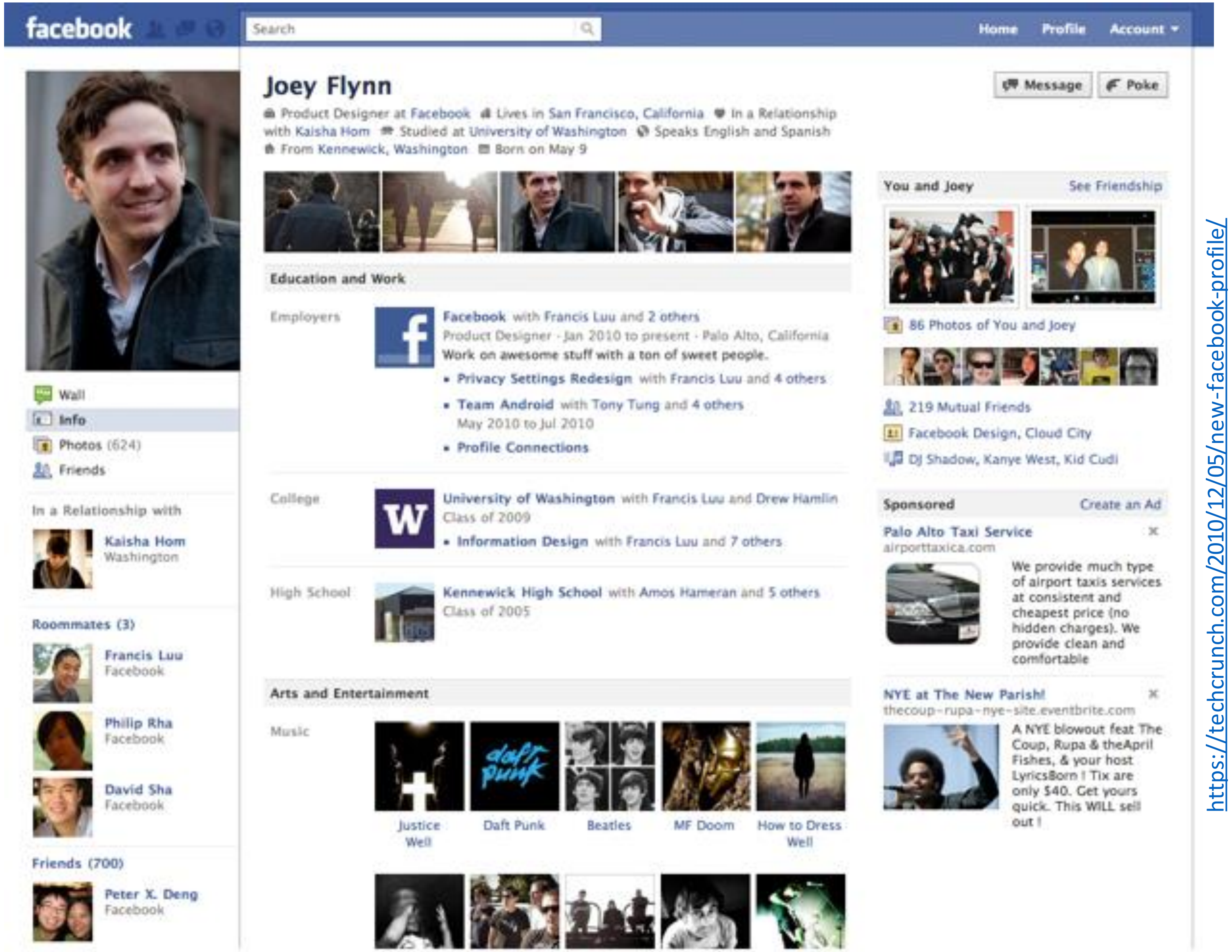
Related products



Music Genome Project by Pandora

> *A given song is represented by a vector containing values for approximately 450 "genes" (analogous to trait-determining genes for organisms in the field of genetics). Each gene corresponds to a characteristic of the music, for example, gender of lead vocalist, prevalent use of groove, level of distortion on the electric guitar, type of background vocals, etc. Rock and pop songs have 150 genes, rap songs have 350, and jazz songs have approximately 400. Other genres of music, such as world and classical music, have 300–450 genes. The system depends on a sufficient number of genes to render useful results. Each gene is assigned a number between 0 and 5, in half-integer increments.*

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



https://techcrunch.com/2010/12/05/new-facebook-profile/

Text Document as a Vector of Terms

- Each item is a "document"
- Features are vocabulary words
- TF: Term Frequency

$$\mathrm{tf}(t,d) = 0.5 + 0.5\frac{f_{t,d}}{\max\{f_{t',d}: t' \in d\}}$$

  - where $f_{t,d}$ is frequency of term t in document d

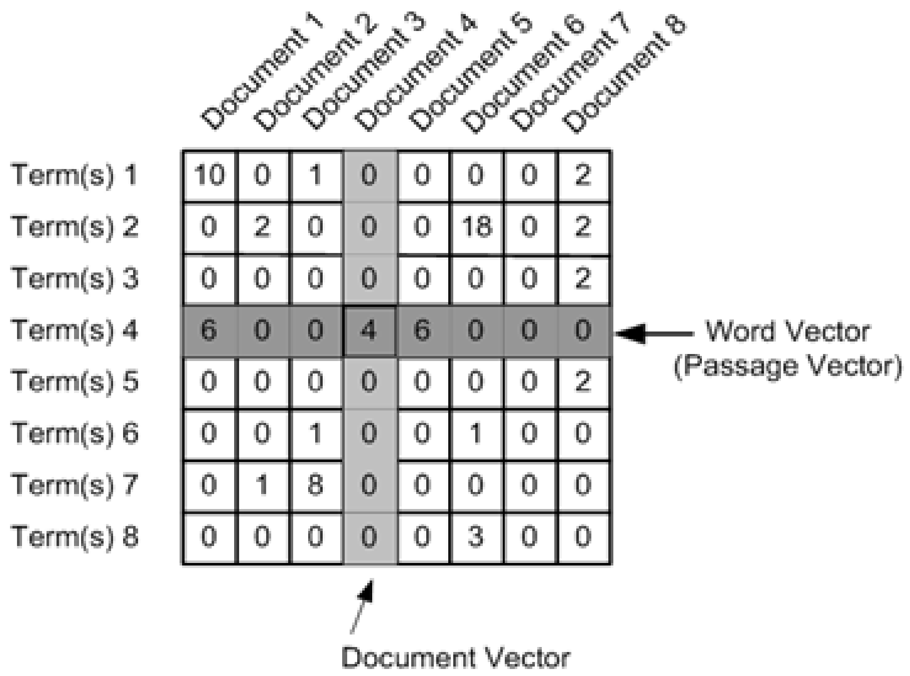
- IDF: Inverse Document Frequency

$$\mathrm{idf}(t,D) = \log\frac{N}{|\{d \in D: t \in d\}|}$$

  - where N is the size of corpus D

- TFIDF

$$\mathrm{tfidf}(t,d,D) = \mathrm{tf}(t,d) \times \mathrm{idf}(t,D)$$



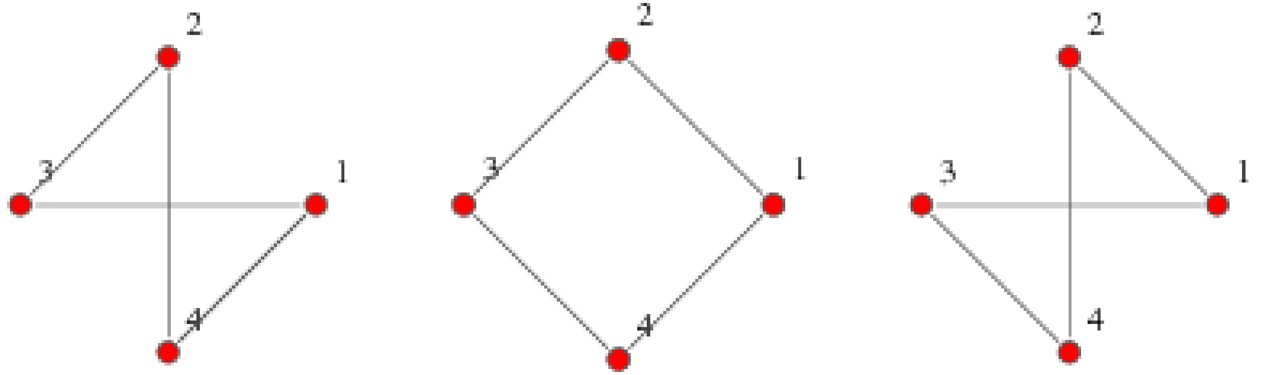
	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	Document 8
Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

https://medium.com/@ranasinghiitkgp/featurization-of-text-data-bow-tf-idf-avgw2v-tfidf-weighted-w2v-7a6c62e8b097

PREFERRED.AI

Graph Vertex as a Vector of Edges

- Directedness
  - Undirected: just links
  - Directed: outgoing or incoming links
- Weight
  - Unweighted: binary connectivity
  - Weighted: similarity values



$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

https://mathworld.wolfram.com/AdjacencyMatrix.html

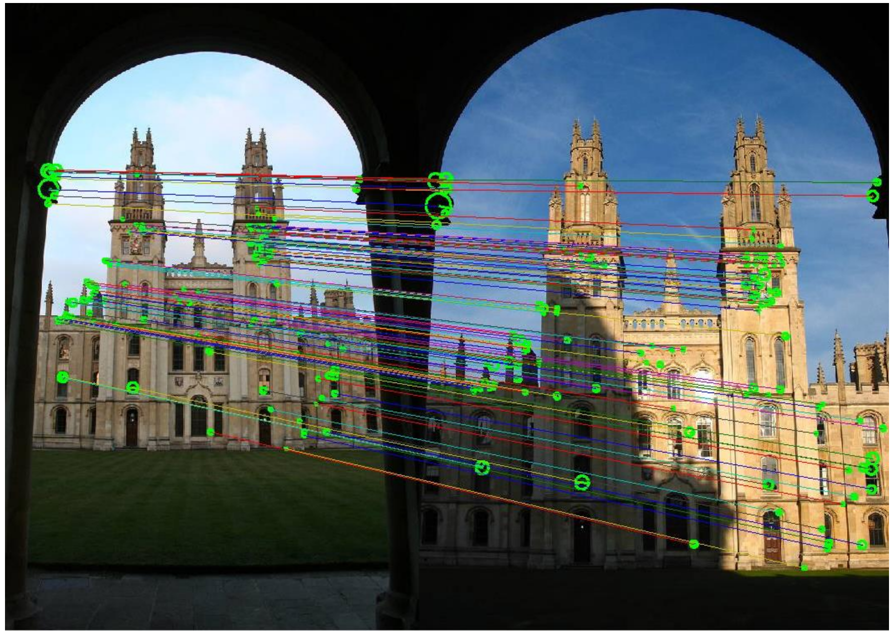
Image as a Vector of Image Features

Pixels

- An image is made of pixels
- Each pixel is described by RGB values
- An image of size $m \times n$ can be represented by a vector of length $3mn$

SIFT

- Scale-Invariant Feature Transform
- Extracting features that are robust against translation, scaling, rotation, illumination, etc.



https://blog.ekbana.com/extracting-invariant-features-from-images-using-sift-for-key-point-matching-675f818ce199

Automatic Feature Extraction

- Also known as representation learning
- Text
  - Word vectors
  - Recurrent neural networks
- Image
  - Convolutional neural networks
- Graph
  - Graph embeddings
  - Graph convolution networks
- Many are based on deep learning

Collaborative vs. Content-Based Filtering

Collaborative Filtering

- Presumes that behavior drives consumption
- Greater capacity for personalization, especially for matrix factorization with user latent vectors

Content-Based Filtering

- Presumes that content drives consumption
- Caters to cold-start items

Can we do both?

Yes, since matrix factorization or deep learning algorithms are highly extensible

Multimodality

- Key idea: learn from multiple modalities at the same time

- First modality: preference feedback
  - Explicit feedback
  - Implicit feedback

- Second (and third, etc.) modality
  - Item content in the form of text, graph, or image
  - Alternatively, user content

Hands-on #1: Cornac

https://github.com/PreferredAI/tutorials/blob/master/multimodal-recsys/01_getting_started.ipynb

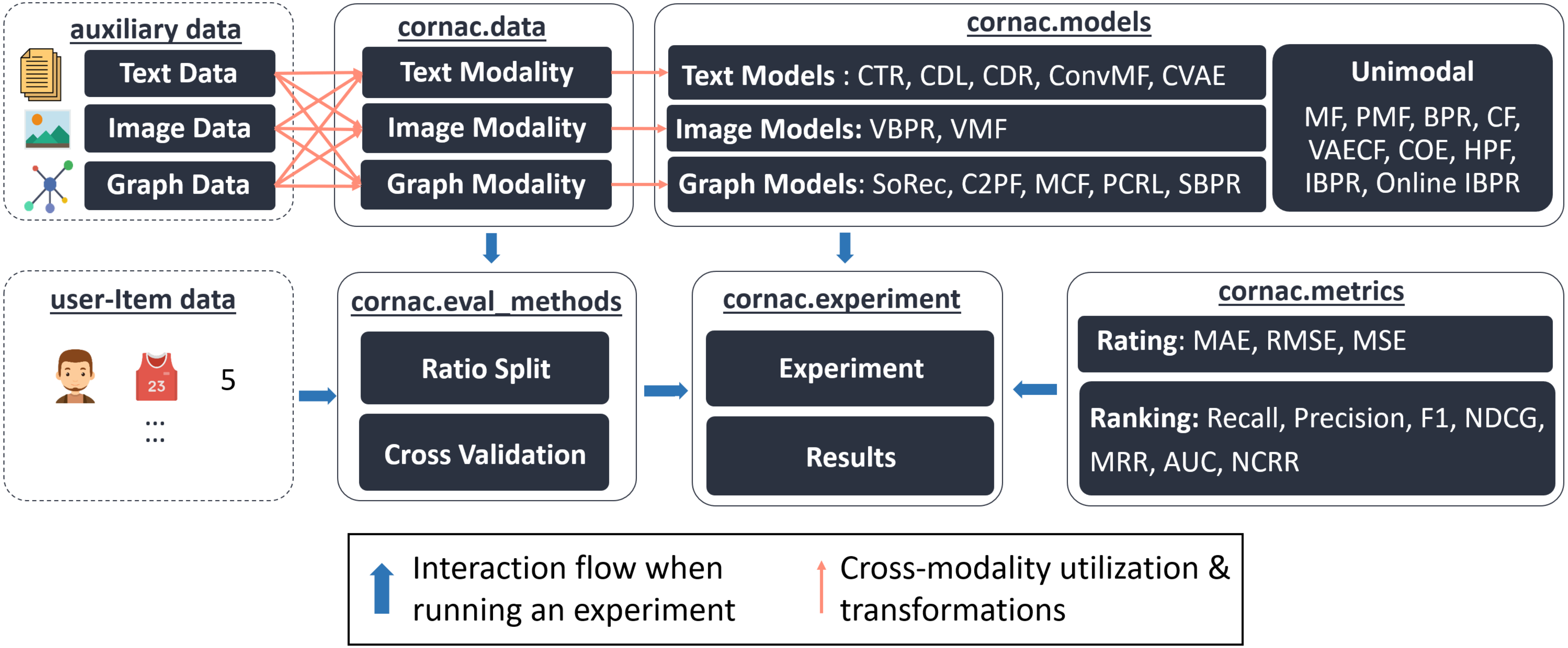
Cornac

- Python-based recommender framework

- An open-source project by Preferred.AI
  - https://cornac.preferred.ai
  - Go to our GitHub and star it

- Publication at Journal of Machine Learning Research:
  - https://jmlr.org/papers/v21/19-805.html

Overview



Fast **experimentation**, **exploration**, and **comparisons**

Cornac Key Features

- **Multimodality Support**
  - Reading, transforming, formatting and representing different types of data
  - Convenient development of new models
  - Broadening the use cases of existing models
- **Scalability**
  - A collection of Iterators for easy stochastic optimization
  - Harnessing the Python ecosystem, e.g., NumPy and Scipy, for efficient operations
  - Leverage Cython to achieve C/C++ performance
- **Accessibility & Reproducibility**
  - Open-access to a rich collection of **models** (**>40**)
  - Straightforward usage of real-world benchmark datasets
  - Full control over random number generators

Experiment oriented

Data Loading → Evaluation Method (Data Splitting) → Models → Metrics → Experiment

Text Modality

Text-Modality

Text Modeling	MF or PMF or FM	WMF	BPR	NCF
Term Vector	SVDFeature			
Matrix Factorization	CMF			
Topic Model	HFT	CTR	CTRank	
Auto-Encoder	AutoSVD	CDL, CVAE	CDR	
CNN	ConvMF, DeepCoNN	DeepMusic		
RNN, LSTM				MRG, NRT

References

- CTR: Wang, C., & Blei, D. M. (2011, August). Collaborative topic modeling for recommending scientific articles. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 448-456).
- CDL: Wang, H., Wang, N., & Yeung, D. Y. (2015, August). Collaborative deep learning for recommender systems. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1235-1244).
- ConvMF: Kim, D., Park, C., Oh, J., Lee, S., & Yu, H. (2016, September). Convolutional matrix factorization for document context-aware recommendation. In Proceedings of the 10th ACM conference on recommender systems (pp. 233-240).
- CDR: Ying, H., Chen, L., Xiong, Y., & Wu, J. (2016, April). Collaborative deep ranking: A hybrid pair-wise recommendation algorithm with implicit feedback. In Pacific-asia conference on knowledge discovery and data mining (pp. 555-567). Springer, Cham.
- CVAE: Li, X., & She, J. (2017, August). Collaborative variational autoencoder for recommender systems. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 305-314).
- CTRank: Yao, W., He, J., Wang, H., Zhang, Y., & Cao, J. (2015, February). Collaborative topic ranking: Leveraging item meta-data for sparsity reduction. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 29, No. 1).
- DeepMusic: Van Den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. In Neural Information Processing Systems Conference (NIPS 2013) (Vol. 26). Neural Information Processing Systems Foundation (NIPS).
- HFT: McAuley, J., & Leskovec, J. (2013, October). Hidden factors and hidden topics: understanding rating dimensions with review text. In Proceedings of the 7th ACM conference on Recommender systems (pp. 165-172).
- CMF: 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).
- SVDFeature: Chen, T., Zhang, W., Lu, Q., Chen, K., Zheng, Z., & Yu, Y. (2012). SVDFeature: a toolkit for feature-based collaborative filtering. The Journal of Machine Learning Research, 13(1), 3619-3622.
- libFM: Rendle, S. (2012). Factorization machines with libfm. ACM Transactions on Intelligent Systems and Technology (TIST), 3(3), 1-22.
- AutoSVD: Zhang, S., Yao, L., & Xu, X. (2017, August). Autosvd++ an efficient hybrid collaborative filtering model via contractive auto-encoders. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 957-960).
- DeepCoNN: Zheng, L., Noroozi, V., & Yu, P. S. (2017, February). Joint deep modeling of users and items using reviews for recommendation. In Proceedings of the tenth ACM international conference on web search and data mining (pp. 425-434).
- MRG: Truong, Q. T., & Lauw, H. (2019, May). Multimodal review generation for recommender systems. In The World Wide Web Conference (pp. 1864-1874).
- NRT: Li, P., Wang, Z., Ren, Z., Bing, L., & Lam, W. (2017, August). Neural rating regression with abstractive tips generation for recommendation. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 345-354).

SVDFeature

Chen, Zhang, Lu, Chen, Zheng, & Yu, "SVDFeature: a toolkit for feature-based collaborative filtering", JMLR, 2012.

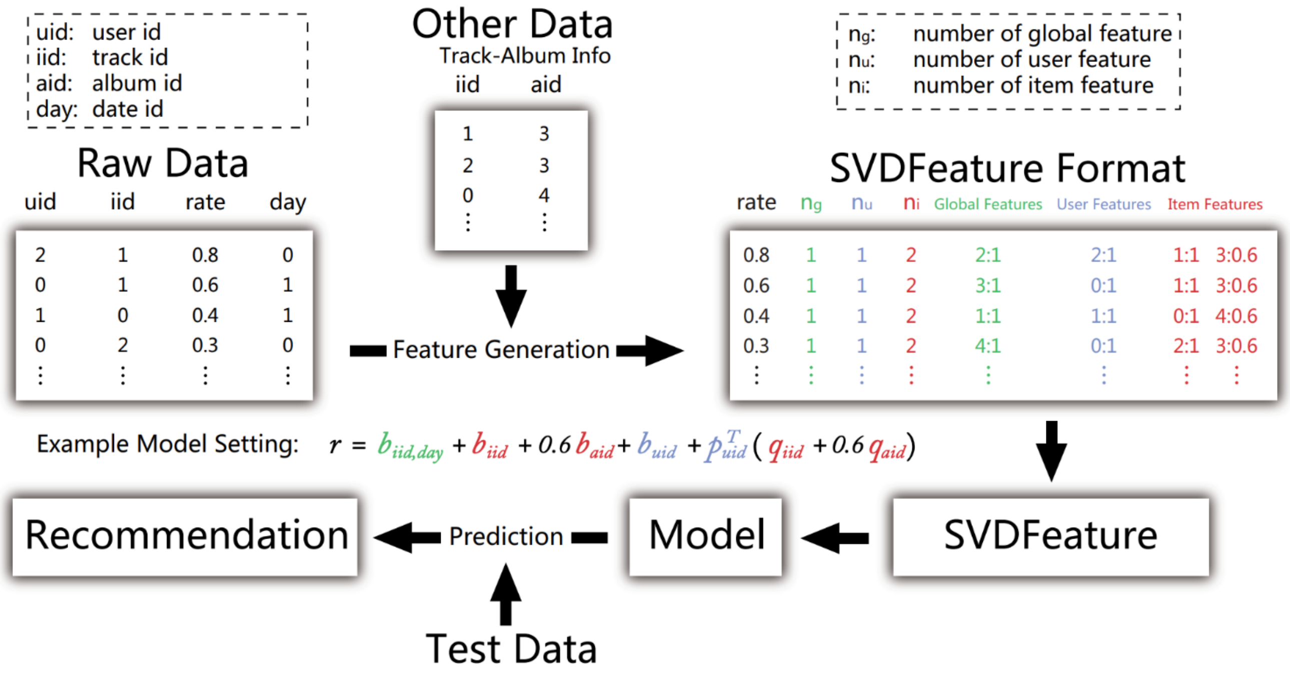


Figure 1: Usage flow example of SVDFeature

Collective Matrix Factorization

Singh and Gordon, "Relational learning via collective matrix factorization", KDD, 2008.

- Let $\boldsymbol{R} \in \mathbb{R}^{N \times M}$ be a sparse rating matrix for N users and M items
  - $r_{ij} \in \boldsymbol{R}$ is an observed rating by user i on item j
  - For each user i, a real-valued vector $\boldsymbol{u}_i \in \mathbb{R}^K$. For all users, collectively $\boldsymbol{U} \in \mathbb{R}^{N \times K}$.
  - For each item j, a real-valued vector $\boldsymbol{v}_j \in \mathbb{R}^K$. For all items, collectively $\boldsymbol{V} \in \mathbb{R}^{M \times K}$.
- Let $\boldsymbol{D} \in \mathbb{R}^{M \times L}$ be a sparse matrix for M items and L words
  - $d_{jl} \in \boldsymbol{D}$ is an observed occurrence/importance of word l in item j
  - For each word l, a real-valued vector $\boldsymbol{z}_l \in \mathbb{R}^K$. For all items, collectively $\boldsymbol{Z} \in \mathbb{R}^{M \times L}$.
- Overall loss function (with regularization)

$$\mathcal{L}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Z} | \lambda) = \frac{1}{2} \sum_{r_{ij} \in \boldsymbol{R}} \left(r_{ij} - \sum_{k=1}^{K} u_{ik} \cdot v_{jk} \right)^2 + \frac{1}{2} \sum_{d_{jl} \in \boldsymbol{D}} \left(d_{jl} - \sum_{k=1}^{K} v_{jk} \cdot z_{lk} \right)^2 + \frac{\lambda}{2} \sum_{i=1}^{N} \|\boldsymbol{u}_i\|^2 + \frac{\lambda}{2} \sum_{j=1}^{M} \|\boldsymbol{v}_j\|^2 + \frac{\lambda}{2} \sum_{l=1}^{L} \|\boldsymbol{z}_l\|^2$$

  - and λ is the regularization weight

Topic Model

Blei, Ng, and Jordan, "Latent Dirichlet Allocation", JMLR 2003.

- Text Document

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

- Topics

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

- A document of many words may discuss a relatively small number of "topics"
- A topic can be described by a series of words that frequently co-occur with one another

Collaborative Topic Regression (CTR)

Wang and Blei, "Collaborative Topic Modeling for Recommending Scientific Articles", KDD 2011.

- Combines weighted matrix factorization and topic model (LDA)
- Item j has a document that describes it, with a distribution $\boldsymbol{\theta}_j$ over K topics
- Item j also has K-dimensional latent vector $\boldsymbol{v}_j$ (in matrix factorization sense)

$$\hat{p}_{ij} = \boldsymbol{u}_i^T \boldsymbol{v}_j$$

- Intuition: two items with similar topics would behave similarly
- Assume that $\boldsymbol{v}_j$ is drawn from a Normal distribution with mean $\boldsymbol{\theta}_j$

$$\boldsymbol{v}_j \sim \mathcal{N}\left(\boldsymbol{\theta}_j, \lambda^{-1}\mathbf{I}\right)$$

  - standard deviation λ_v^{-1} will eventually translate into regularization coefficient
- Equivalently:

$$\boldsymbol{v}_j = \boldsymbol{\theta}_j + \boldsymbol{\epsilon}_j$$
$$\boldsymbol{\epsilon}_j \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{I})$$

PREFERRED.AI

CTR: Generative Process

- Generative process:
  - For each user i
    - Draw latent vector $\boldsymbol{u}_i \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{I})$
  - For each item j
    - Draw topic proportions $\boldsymbol{\theta}_j \sim \mathrm{Dirichlet}(\alpha)$
    - Draw item offset $\boldsymbol{\epsilon}_j \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{I})$
    - For the n^{th} word in the article:
      - Draw topic assignment $z_{jn} \sim \mathrm{Multinomial}(\boldsymbol{\theta}_j)$
      - Draw word $w_{jn} \sim \mathrm{Multinomial}(\beta_{jn})$
  - For each user-item pair (i, j)
    - Draw adoption $p_{ij} \sim \mathcal{N}(\boldsymbol{u}_i^T \boldsymbol{v}_j, c_{ij}^{-1})$

- Adoption:
  - Binary adoption or
  - Valued adoption

$$p_{ij} = \begin{cases} r_{ij}, \text{if } r_{ij} \geq \text{threshold} \\ 0, \text{otherwise} \end{cases}$$

- Confidence:

$$c_{ij} = \begin{cases} a, \text{if } r_{ij} \geq \text{threshold} \\ b, \text{otherwise} \end{cases}$$

  - More confidence on positive adoption, i.e., $a \gg b$

PREFERRED.AI

Learning

- Minimizes the negative log-likelihood of ratings and item descriptions

$$\mathcal{L}(\boldsymbol{U},\boldsymbol{V},\boldsymbol{\theta},\boldsymbol{\beta}|\lambda) = \frac{1}{2}\sum_{i,j} c_{ij}\left(p_{ij} - \sum_{k=1}^{K} u_{ik}\cdot v_{jk}\right)^2 - \sum_{j}\sum_{n}\log\left(\sum_{k=1}^{K}\theta_{jk}\cdot\beta_{k,w_{jn}}\right) + \frac{\lambda}{2}\sum_{i=1}^{N}\|\boldsymbol{u}_i\|^2 + \frac{\lambda}{2}\sum_{j=1}^{M}(\boldsymbol{v}_j - \boldsymbol{\theta}_j)^T(\boldsymbol{v}_j - \boldsymbol{\theta}_j)$$

- Iteratively:
  - Optimize for user and item latent vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_j$ based on the current $\boldsymbol{\theta}_j$
  - Optimize for topic proportions $\boldsymbol{\theta}_j$ based on the current vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_j$ and topic words $\boldsymbol{\beta}_k$
  - Optimize for topic words $\boldsymbol{\beta}_k$ based on the current topic proportions $\boldsymbol{\theta}_j$
- Prediction
  - For existing items: $\hat{p}_{ij} = \boldsymbol{u}_i^T\boldsymbol{v}_j$
  - For new items: $\hat{p}_{ij} = \boldsymbol{u}_i^T\boldsymbol{\theta}_j$

Generalizing to Unseen Items

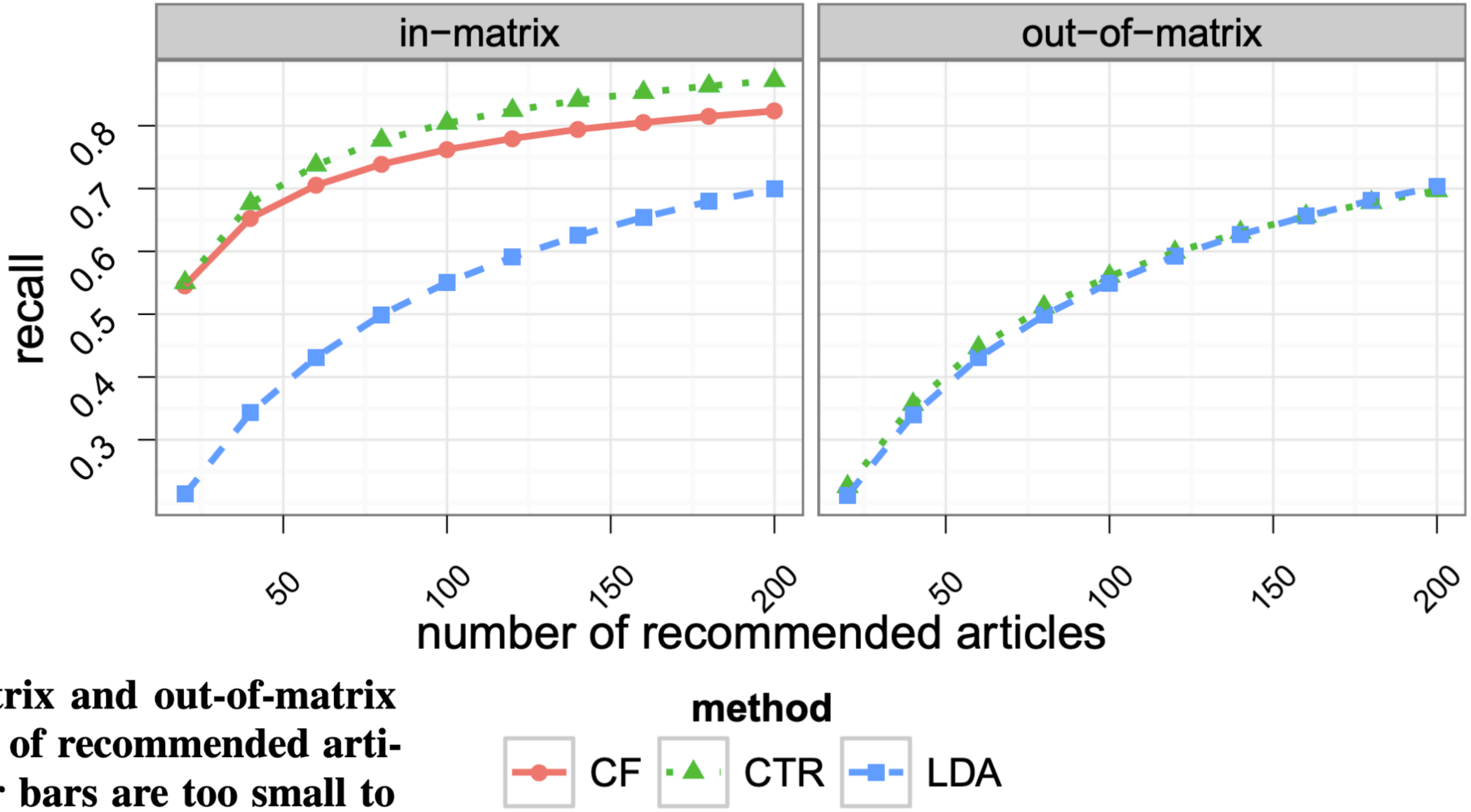


Figure 3: Recall comparison on in-matrix and out-of-matrix prediction tasks by varying the number of recommended articles. For CTR, we set $\lambda_v = 100$. Error bars are too small to show. The maximum expected recall for random recommendation is about 6%. CF can not do out-of-matrix prediction. CTR performs best.

Stacked Denoising Autoencoder (SDAE)

Wang, Wang, & Yeung et al., “Collaborative deep learning for recommender systems ", KDD 2016.

- Let the content for item j be a vector $\boldsymbol{x}_j \in \mathbb{R}^S$ where S is the vocabulary
  - each element of the vector indicates the ‘importance’ (e.g., tfidf) of word in document j
  - document can be plot of movie, abstract of paper, etc.
- Let the corrupted content be a vector $\boldsymbol{x}_j^0 \in \mathbb{R}^S$
  - with some probability an element of $\boldsymbol{x}_j$ is set to zero
- Encode the corrupted version and decode into the original (clean) version

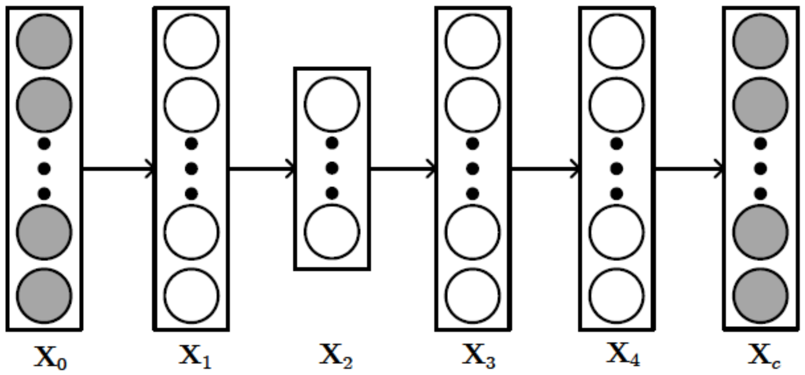
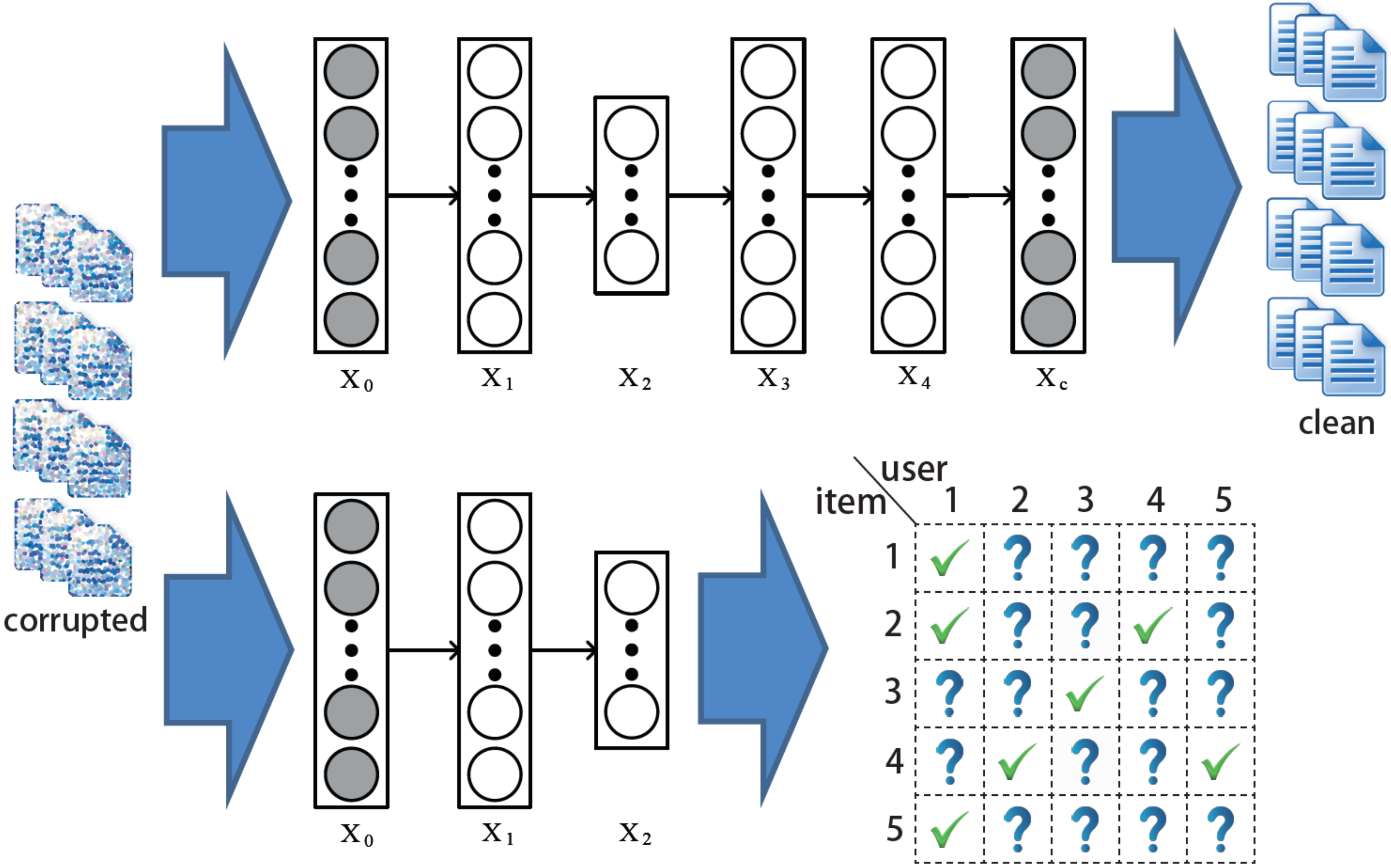


Figure 2: A 2-layer SDAE with $L = 4$.

Collaborative Deep Learning (CDL)



CDL: Generative Process

- Generative process:
  - For each user i
    - Draw latent vector $\boldsymbol{u}_i \sim \mathcal{N}(\mathbf{0}, \lambda_u^{-1}\mathbf{I})$
  - For each item j
    - Put its corrupted content $\boldsymbol{x}_j^0$ through SDAE
    - Draw clean content $\boldsymbol{x}_j \sim \mathcal{N}(\boldsymbol{x}_L, \lambda_n^{-1}\mathbf{I})$
      - Use middle representation $\boldsymbol{x}_j^{L/2}$
    - Draw item offset $\boldsymbol{\epsilon}_j \sim \mathcal{N}(\mathbf{0}, \lambda_v^{-1}\mathbf{I})$
    - Set item latent vector $\boldsymbol{v}_j$ to be
      $\boldsymbol{v}_j = \boldsymbol{\epsilon}_j + \boldsymbol{x}_j^{L/2}$
  - For each user-item pair (i, j)
    - Draw adoption $p_{ij} \sim \mathcal{N}(\boldsymbol{u}_i^T \boldsymbol{v}_j, c_{ij}^{-1})$

- Confidence:

$$c_{ij} = \begin{cases} a, \text{if } r_{ij} \geq 0 \\ b, \text{otherwise} \end{cases}$$

  - More confidence on positive adoption, i.e., $a \gg b$

CDL: Loss Function

$$\mathcal{L} =$$

$$\frac{\lambda_u}{2}\sum_i \|\boldsymbol{u}_i\|^2 +$$

- Regularize user vectors

$$\frac{\lambda_w}{2}\sum_l (\|\boldsymbol{W}_l\|^2 + \|\boldsymbol{b}_l\|^2) +$$

- Regularize autoencoder parameters

$$\frac{\lambda_v}{2}\sum_j \left\|\boldsymbol{v}_j - \boldsymbol{x}_j^{L/2}\right\|^2 +$$

- Minimize offset so item vector is close to its middle encoding

$$\frac{\lambda_n}{2}\sum_i \|\boldsymbol{x}_j^{\boldsymbol{L}} - \boldsymbol{x}_j^c\|^2 +$$

- Ensure the autoencoder reconstruction is similar to the clean output

$$\sum_{i,j} \frac{c_{ij}}{2}\left(p_{ij} - \boldsymbol{u}_i^T \boldsymbol{v}_j\right) +$$

- Minimize error in rating prediction

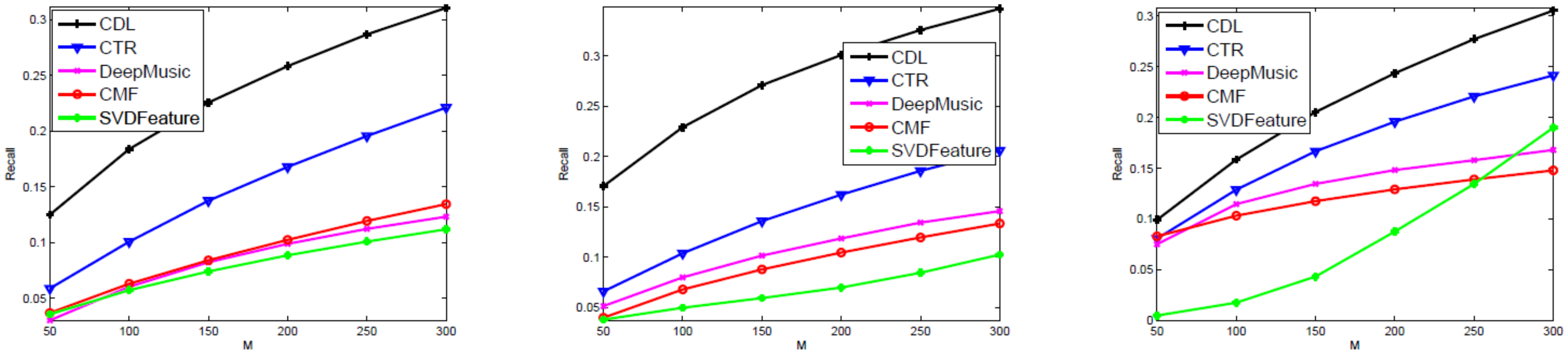


Figure 4: Performance comparison of CDL, CTR, DeepMusic, CMF, and SVDFeature based on recall@M for datasets *citeulike-a*, *citeulike-t*, and *Netflix* in the sparse setting. A 2-layer CDL is used.

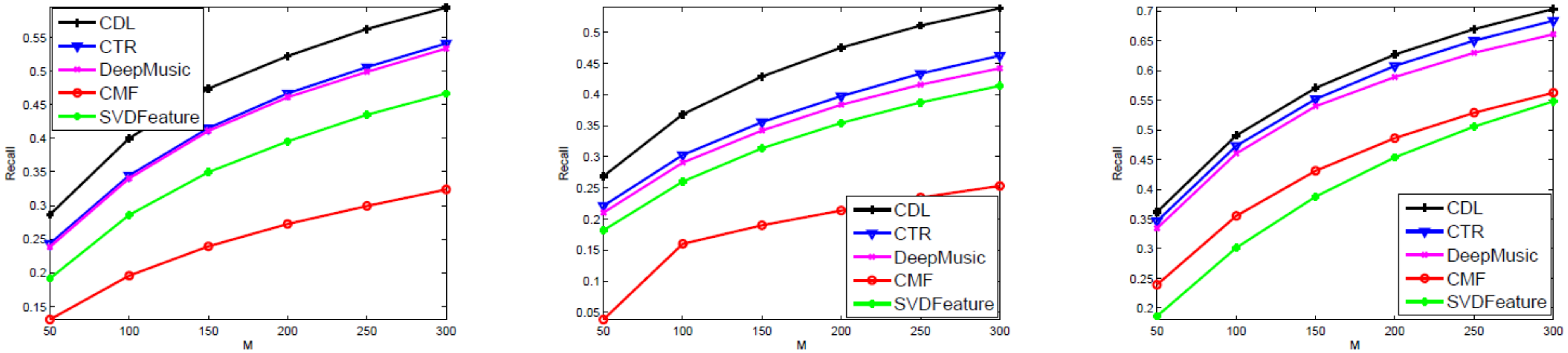
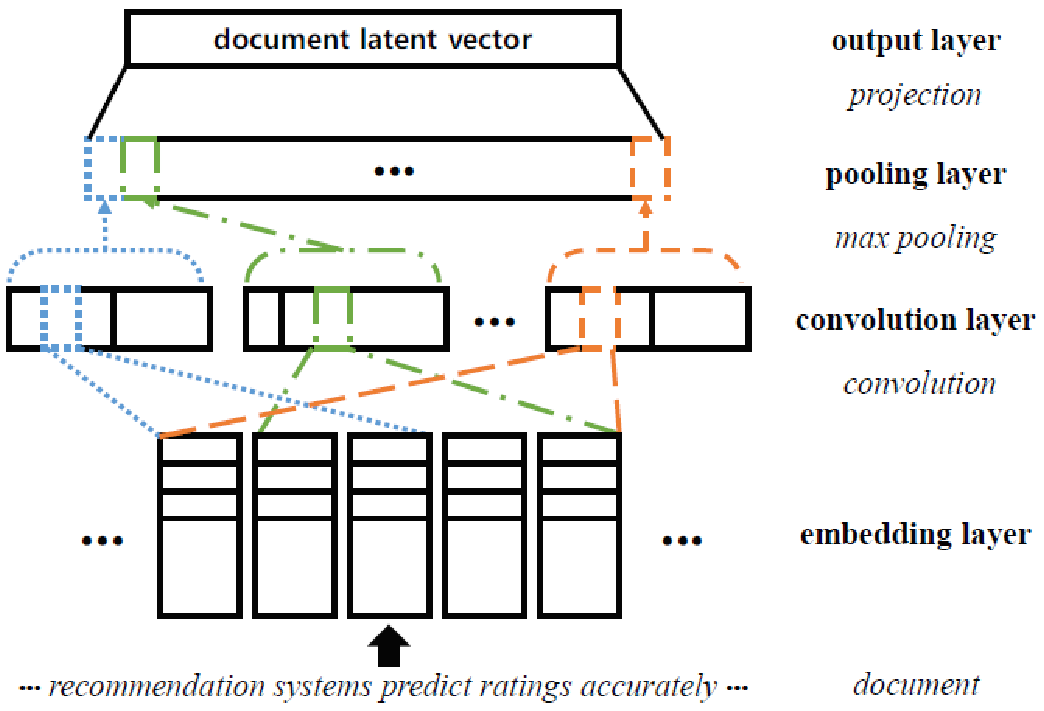


Figure 5: Performance comparison of CDL, CTR, DeepMusic, CMF, and SVDFeature based on recall@M for datasets *citeulike-a*, *citeulike-t*, and *Netflix* in the dense setting. A 2-layer CDL is used.

Convolutional Matrix Factorization (ConvMF)

Kim et al., "Convolutional matrix factorization for document context-aware recommendation ", RecSys 2016.

- Represent each product as a text document
  - e.g., product description, combine reviews into a single document
  - a term vector $\boldsymbol{x}_j \in \mathbb{R}^S$ where S is the vocabulary
- Embedding layer
  - each column initialized with p-dimensional word vector (e.g., Glove) of each word
- Derive the hidden representation using CNN



ConvMF: Generative Process

- Generative process:
  - For each user i
    - Draw latent vector $\boldsymbol{u}_i \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{I})$
  - For each item j
    - Put its content $\boldsymbol{x}_j$ through CNN parameterized by $\boldsymbol{W}$
    - Draw item offset $\boldsymbol{\epsilon}_j \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{I})$
    - Set item latent vector $\boldsymbol{v}_j$ to be
$$\boldsymbol{v}_j = \boldsymbol{\epsilon}_j + \mathrm{CNN}(\boldsymbol{W}, \boldsymbol{x}_j)$$
  - For each user-item pair (i, j)
    - Draw adoption $r_{ij} \sim \mathcal{N}\left(\boldsymbol{u}_i^T \boldsymbol{v}_j, \sigma^2\right)$

- Loss function
$$\mathcal{L} = \sum_{(i,j)\in \boldsymbol{R}} \left(r_{ij} - \boldsymbol{u}_i^T \boldsymbol{v}_j\right)^2 + \lambda_u \sum_i \|\boldsymbol{u}_i\|^2 + \lambda_v \sum_v \left\|\boldsymbol{v}_i - \mathrm{CNN}(\boldsymbol{W}, \boldsymbol{x}_j)\right\|^2 + \lambda_{\boldsymbol{W}} \sum_l \|w_l\|^2$$

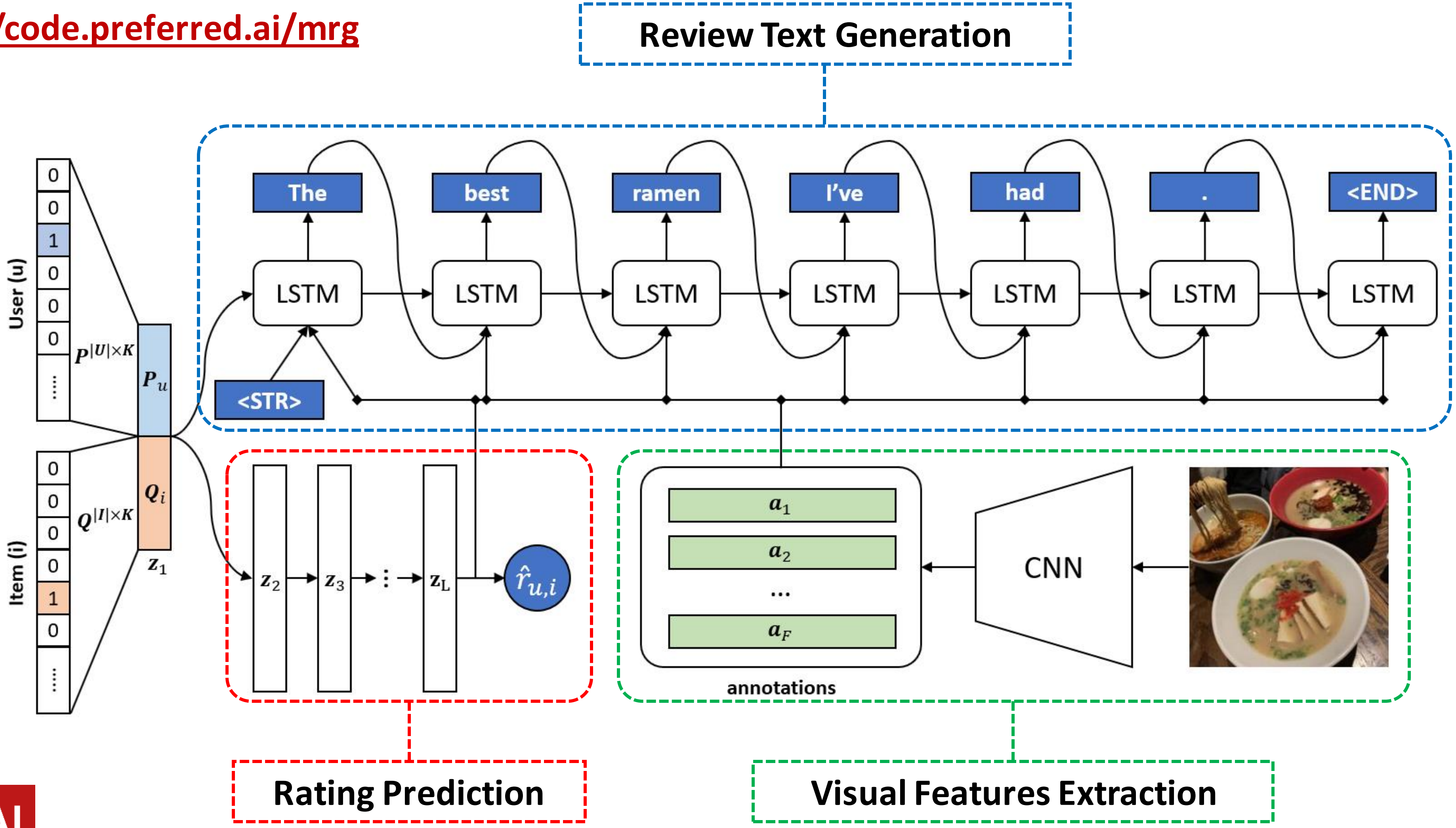
Model	Ratio of training set to the entire dataset (density)						
	20% (0.93%)	30% (1.39%)	40% (1.86%)	50% (2.32%)	60% (2.78%)	70% (3.25%)	80% (3.71%)
PMF	1.0168	0.9711	0.9497	0.9354	0.9197	0.9083	0.8971
CTR	1.0124	0.9685	0.9481	0.9337	0.9194	0.9089	0.8969
CDL	1.0044	0.9639	0.9377	0.9211	0.9068	0.8970	0.8879
ConvMF	**0.9745**	**0.9330**	**0.9063**	**0.8897**	**0.8726**	**0.8676**	**0.8531**
Improve	2.98%	3.20%	3.36%	3.41%	3.77%	3.27%	3.92%

Table 4: Test RMSE over various sparseness of training data on ML-1m dataset

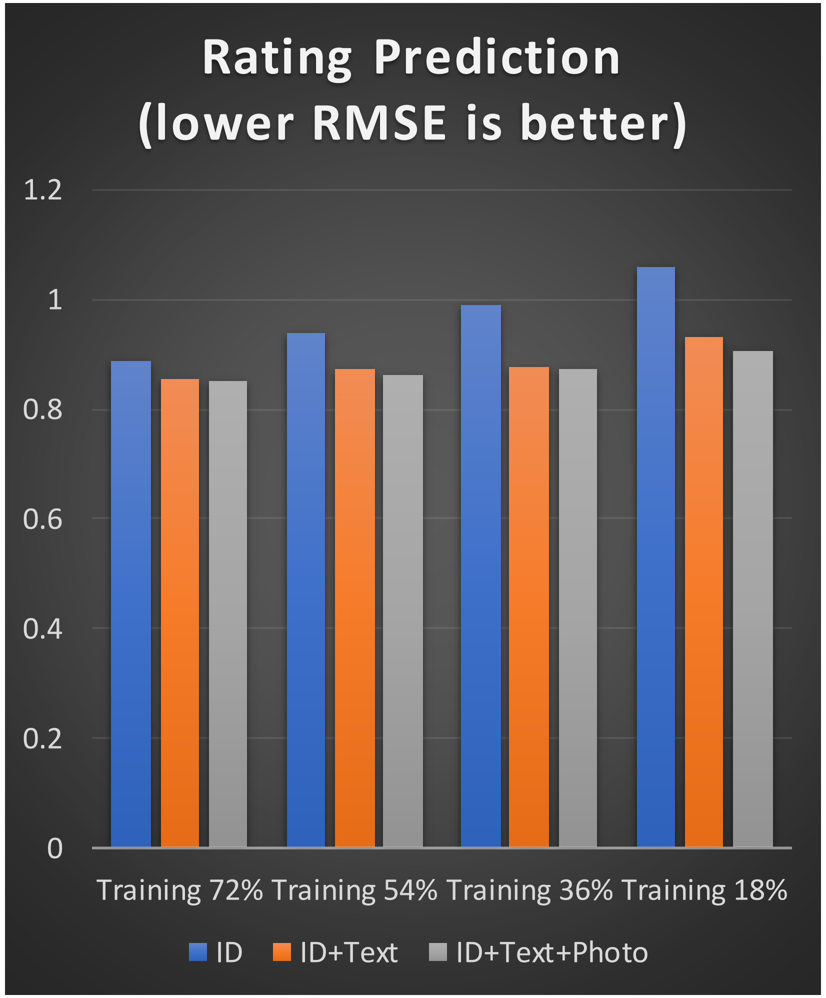
Multimodal Review Generation

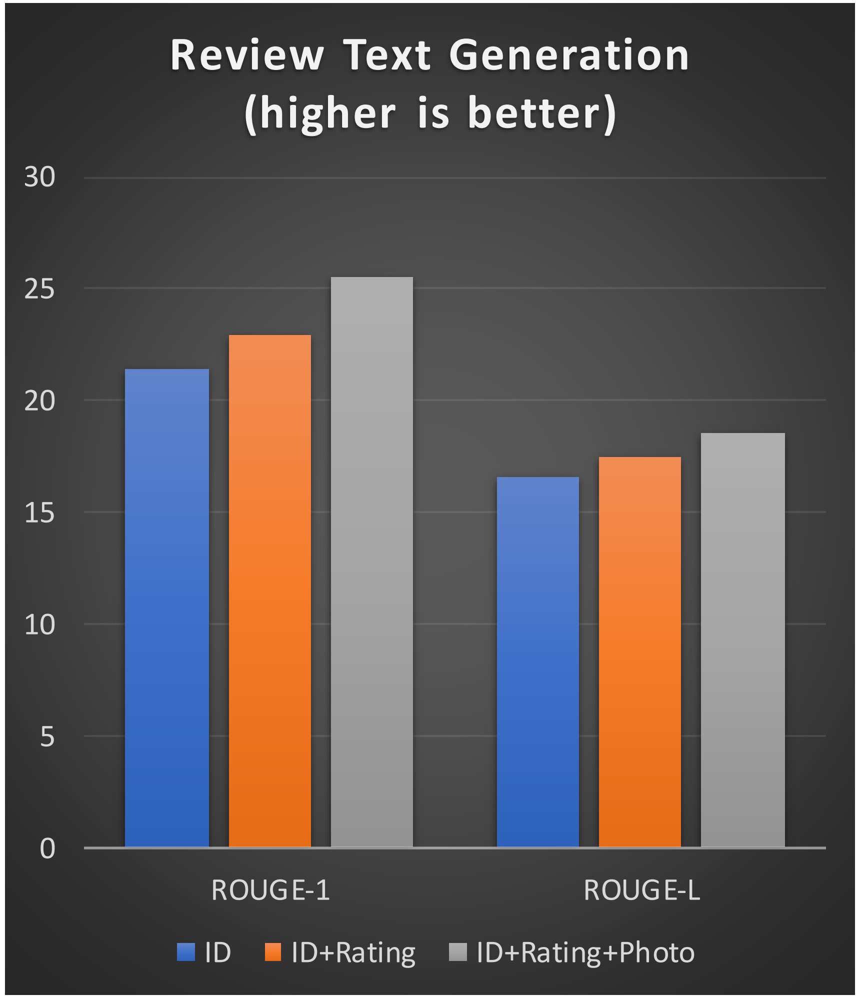
Truong and Lauw, "Multimodal Review Generation for Recommender Systems", WWW 2019.

https://code.preferred.ai/mrg



Ablation Analysis





Case Studies

	Photo	Rating	Review
Ellen "FuZe" Z.		**4.5**	**the clam chowder was good .**
		5.0	best clam chowder i 've ever had .
Young Y.		**3.4**	**the clam chowder was a bit too salty .**
		3.0	the boston clam chowder was pretty salty and i 've had lots of clam chowder before .

Cornac-Supported Text-Based Models

- Collaborative Topic Modeling (CTR)
- Hidden Factors and Hidden Topics (HFT)
- Collaborative Deep Learning (CDL)
- Collaborative Deep Ranking (CDR)
- Convolutional Matrix Factorization (ConvMF)
- …

Image Modality

Image-Modality

Image Modeling	Explicit (MF or PMF)	Implicit (BPR or WMF)
Pre-trained Embedding	VMF, VPOI	VBPR, ACF, NPR
Convolutional Neural Nets		DVBPR, CKE, CDL, JRL

References

- Chanyoung Park, Donghyun Kim, Jinoh Oh, and Hwanjo Yu. 2017. Do" Also-Viewed" Products Help User Rating Prediction?. In WWW. 1113–1122.
- Suhang Wang, Yilin Wang, Jiliang Tang, Kai Shu, Suhas Ranganath, and Huan Liu. 2017. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In WWW. 391–400.
- Ruining He and Julian McAuley. 2016. VBPR: visual bayesian personalized ranking from implicit feedback. In AAAI.
- Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, and Tat-Seng Chua. 2017. Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. In SIGIR. 335–344.
- Wei Niu, James Caverlee, and Haokai Lu. 2018. Neural personalized ranking for image recommendation. In WSDM. 423–431.
- Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian McAuley. 2017. Visually-aware fashion recommendation and design with generative image models. In ICDM. 207–216.
- Chenyi Lei, Dong Liu, Weiping Li, Zheng-Jun Zha, and Houqiang Li. 2016. Comparative deep learning of hybrid representations for image recommendations. In CVPR. 2545–2553.
- Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In SIGKDD. 353–362.
- Yongfeng Zhang, Qingyao Ai, Xu Chen, and W Bruce Croft. 2017. Joint representation learning for top-n recommendation with heterogeneous information sources. In CIKM. 1449–1458.
- Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W. L., & Leskovec, J. (2018, July). Graph convolutional neural networks for web-scale recommender systems. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 974-983.

Visual BPR

He and McAuley, "VBPR: visual bayesian personalized ranking from implicit feedback ", AAAI 2016.

- Item j has K-dimensional latent vector $\boldsymbol{v}_j$ (in matrix factorization sense)

$$\hat{r}_{ij} = \boldsymbol{u}_i^T \boldsymbol{v}_j$$

- Item j also has an image feature vector $\boldsymbol{f}_j \in \mathbb{R}^D$, which is of length $D \neq K$
  - D can be very high-dimensional
  - To reduce its dimensionality, introduce a projection matrix $\boldsymbol{E} \in \mathbb{R}^{Q \times D}$, where $Q \ll D$
  - The projection $\boldsymbol{E} \times \boldsymbol{f}_j \in \mathbb{R}^Q$ is now of length Q
- We have two representations of items
  - From collaborative filtering $\boldsymbol{v}_j$
  - From image features $\boldsymbol{E} \times \boldsymbol{f}_j$
- Intuition: user's preference is influenced by both representations

Prediction

- The preference by user i on item j is modeled as follows:

$$\hat{r}_{ij} = b_j + \boldsymbol{u}_i^T \boldsymbol{v}_j + \boldsymbol{\rho}_i^T \left(\boldsymbol{E} \times \boldsymbol{f}_j \right) + \boldsymbol{\Theta}^T \boldsymbol{f}_j$$

  - $\boldsymbol{f}_j \in \mathbb{R}^D$ is the image feature vector (input to the model)
  - $\boldsymbol{\Theta} \in \mathbb{R}^D$ is visual bias vector (learnt)
  - $\boldsymbol{\Theta}^T \boldsymbol{f}_j$ indicates the general popularity of item with particular visual features

  - $\boldsymbol{u}_i \in \mathbb{R}^K$ is user latent vector and $\boldsymbol{v}_j \in \mathbb{R}^K$ is item latent vector (learnt)
  - $\boldsymbol{u}_i^T \boldsymbol{v}_j$ indicates preference by user i on item j based on collaborative filtering

  - $\boldsymbol{\rho}_i \in \mathbb{R}^Q$ is user visual preference and $\left(\boldsymbol{E} \times \boldsymbol{f}_j \right) \in \mathbb{R}^Q$ is item visual representation (learnt)
  - $\boldsymbol{\rho}_i^T \left(\boldsymbol{E} \times \boldsymbol{f}_j \right)$ indicates preference by user i on item j based on visual signals

Learning

- Input:
  - Positive triples $\boldsymbol{S} = \{j >_i l | r_{ij} \in \boldsymbol{R}^+ \wedge r_{il} \in \boldsymbol{R}^-\}$
  - Image features $\{\boldsymbol{f}_j\}$

- Minimize regularized negative log-likelihood function:

$$\mathcal{L}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{b}, \boldsymbol{\Theta}, \mathbf{E}, \mathbf{P} | \lambda)$$
$$= \sum_{(j >_i l) \in \boldsymbol{S}} \ln(1 + \exp\{-(\hat{r}_{ij} - \hat{r}_{il})\}) + \frac{\lambda}{2} \sum_{i=1}^{N} (\|\boldsymbol{u}_i\|^2 + \|\boldsymbol{\rho}_i\|^2) + \frac{\lambda}{2} \sum_{j=1}^{M} \left(b_j^2 + \|\boldsymbol{v}_j\|^2\right) + \frac{\lambda}{2} \|\boldsymbol{\Theta}\|^2 + \frac{\lambda}{2} \|\boldsymbol{E}\|^2$$

Experiments on Fashion Datasets

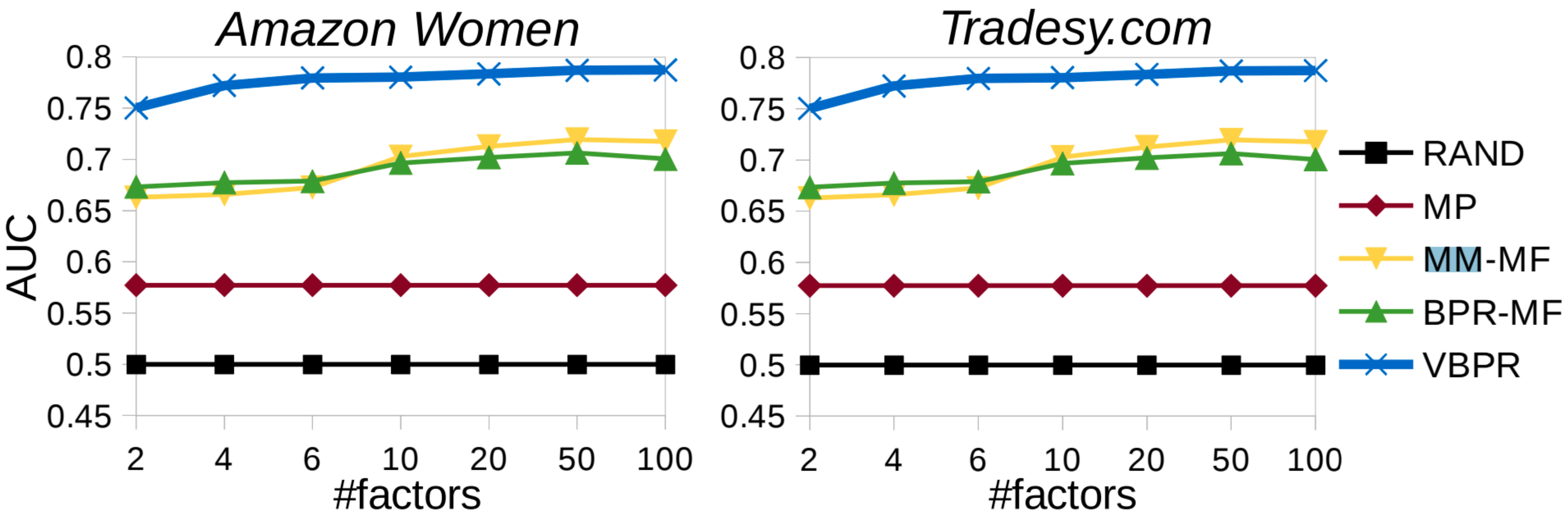


Figure 2: AUC with varying dimensions.

Figure 4: 2-D visualization (with t-SNE (?)) of the 10-D visual space learned from *Amazon Women*. All images are from the test set. For clarity, the space is discretized into a grid and for each grid cell one image is randomly selected among overlapping instances.

DVBPR: Deep Bayesian Personalized Ranking

Visually-Aware Fashion Recommendation and Design with Generative Image Models (ICDM'17)

- VBPR preference estimate: $\hat{r}_{ij} = b_j + \boldsymbol{u}_i^T \boldsymbol{v}_j + \boldsymbol{\rho}_i^T (\boldsymbol{E} \times \boldsymbol{f}_j) + \boldsymbol{\Theta}^T \boldsymbol{f}_j$

- DVBPR preference estimate: $\hat{r}_{ij} = \boldsymbol{\rho}_i^T \Phi(\boldsymbol{X}_j)$

- Pre-trained visual features and embedding matrix are replaced with a CNN network $\Phi(.)$ to extract visual features directly from the images themselves

- Item bias terms b_j and non-visual latent factors $\boldsymbol{u}_i, \boldsymbol{v}_j$ are excluded. Empirically, doing so improves performance (the remaining terms are sufficient to capture these factors implicitly).

Visual Matrix Factorization (VMF)

Do "Also-Viewed" Products Help User Rating Prediction? (WWW'17)

- VMF preference estimate is almost the same as VBPR, except bias removed:

$$\hat{r}_{ij} = \boldsymbol{u}_i^T \boldsymbol{v}_j + \boldsymbol{\rho}_i^T \left(\boldsymbol{E} \times \boldsymbol{f}_j \right)$$

- Learning via minimizing MSE instead of BPR criteria

$$\mathcal{L}(\boldsymbol{U}, \boldsymbol{V} | \lambda) = \frac{1}{2} \sum_{i,j} c_{ij} \left(p_{ij} - \hat{r}_{ij} \right)^2 + \frac{\lambda}{2} \sum_{i=1}^{N} \|\boldsymbol{u}_i\|^2 + \frac{\lambda}{2} \sum_{j=1}^{M} \left\| \boldsymbol{v}_j \right\|^2$$

Attentive Collaborative Filtering (ACF)

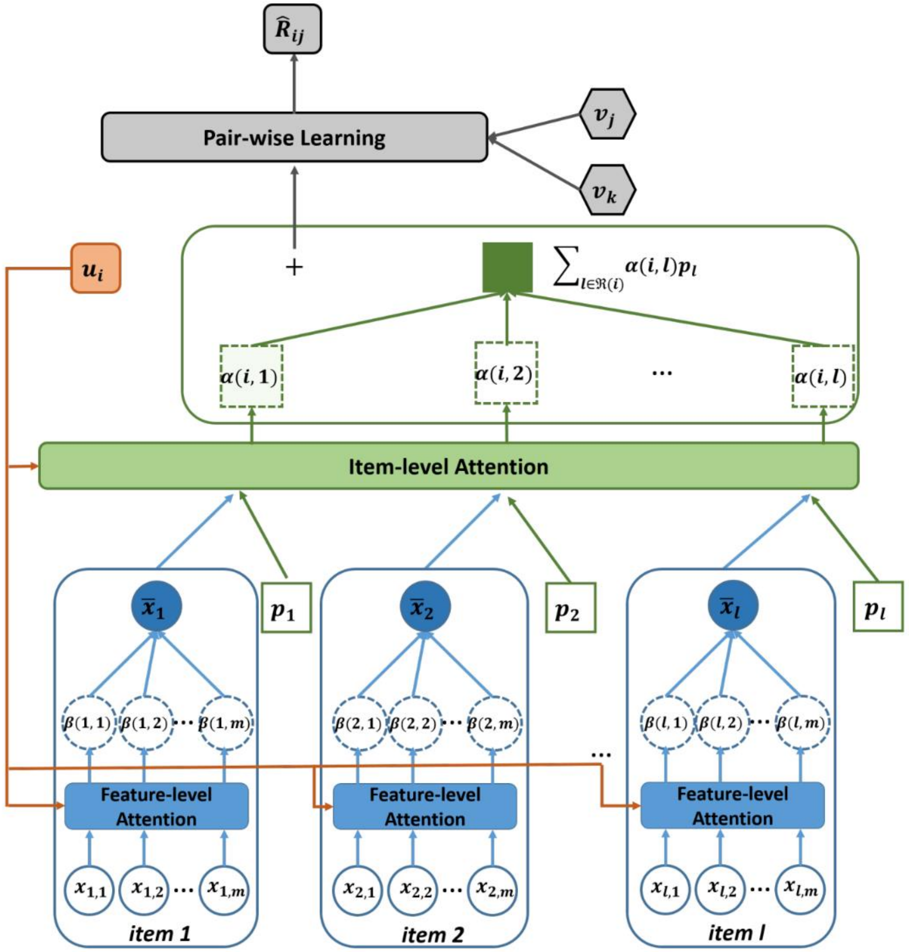
Attentive Collaborative Filtering: Multimedia Recommendation with Item- and Component-Level Attention (SIGIR'17)

- Preference with neighborhood model:

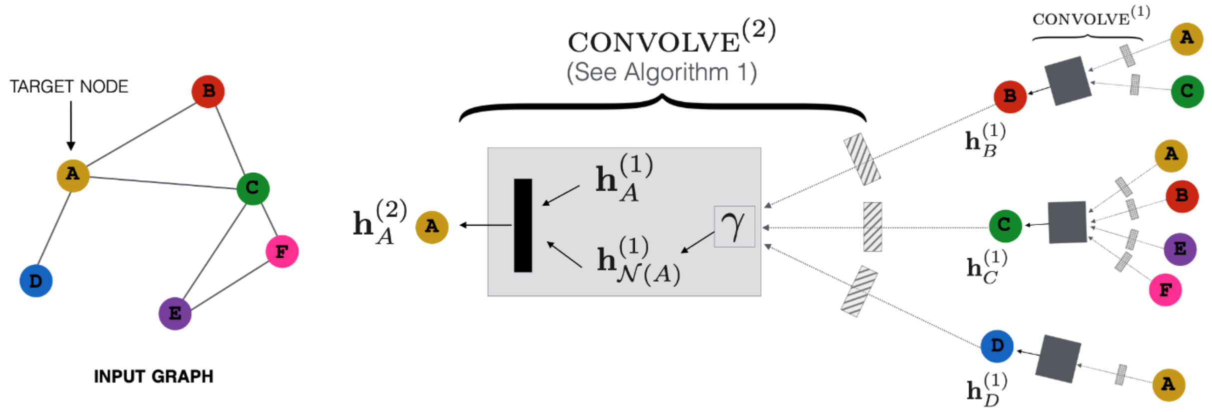
$$\hat{R}_{ij} = \overbrace{\mathbf{u}_i^T \mathbf{v}_j}^{latent\ factor\ model} + \underbrace{\sum_{l \in \mathcal{R}(i)} \alpha(i,l)\mathbf{p}_l^T \mathbf{v}_j}_{neighborhood\ model}$$

- ACF optimizes BPR objective:

$$\arg\min_{\mathbf{U},\mathbf{V},\mathbf{P},\Theta} \sum_{(i,j,k)\in\mathcal{R}_B} -\ln\sigma\left\{\left(\mathbf{u}_i + \sum_{l\in\mathcal{R}(i)} \alpha(i,l)\mathbf{p}_l\right)^T \mathbf{v}_j - \left(\mathbf{u}_i + \sum_{l\in\mathcal{R}(i)} \alpha(i,l)\mathbf{p}_l\right)^T \mathbf{v}_k\right\} + \lambda(||\mathbf{U}||^2 + ||\mathbf{V}||^2 + ||\mathbf{P}||^2),$$

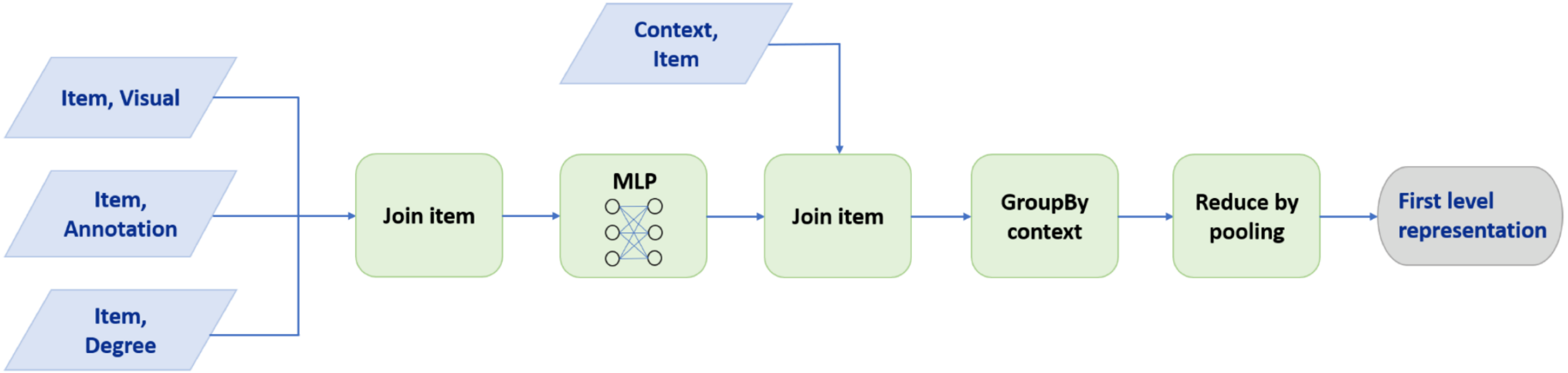


PinSage



- Learning node embedding with GNN

- Node embedding includes visual features



- Recommendation by nearest-neighbor lookup in the learned embedding space

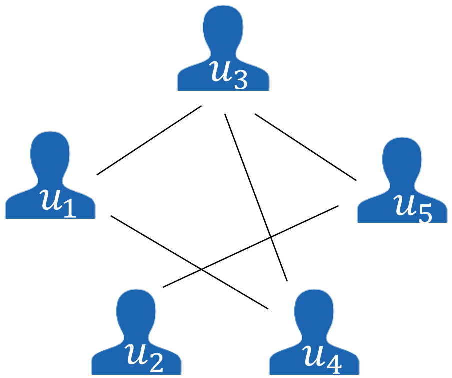
Graph Convolutional Neural Networks for Web-Scale Recommender Systems (KDD'18)

Cornac-Supported Visual-Based Models

- Visual Bayesian Personalized Ranking (VBPR)
- Visual Matrix Factorization (VMF)
- Adversarial Multimedia Recommendation (AMR)
- …

Graph Modality

User-Side: Social Relationships

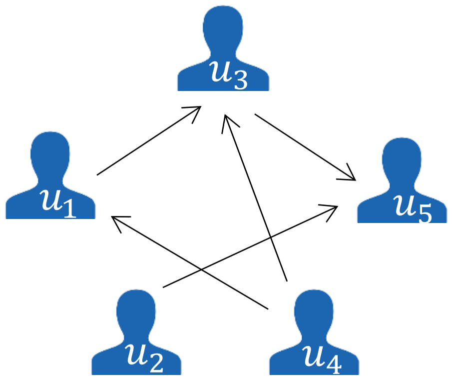


Undirected (e.g., friendship)

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

Adjacency Matrix, Symmetric

User-Side: Social Relationships



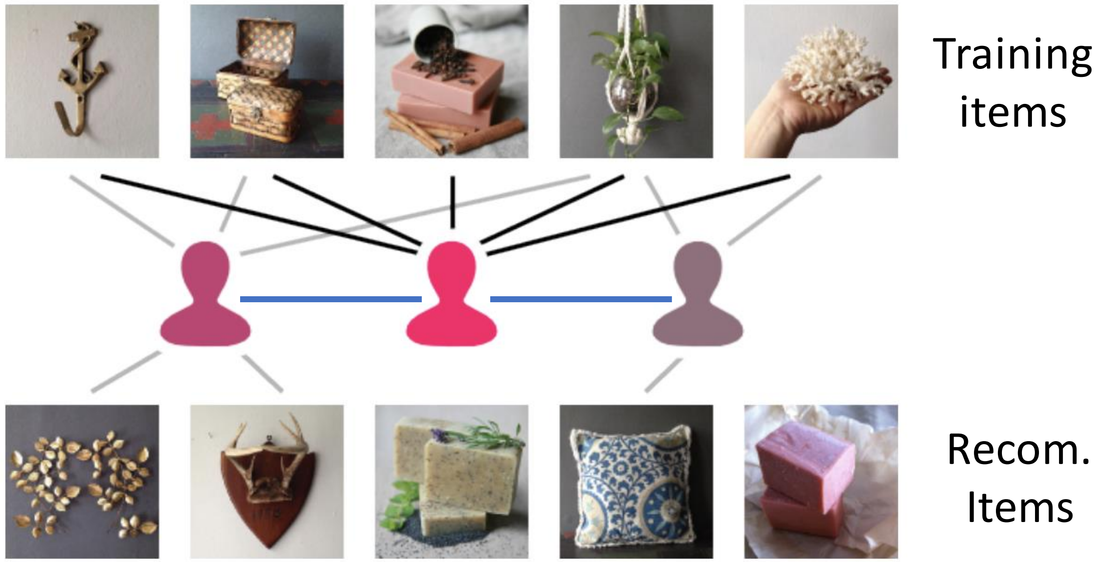
Directed (e.g., follow, trust relationships)

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

Adjacency Matrix, Asymmetric

Social Collaborative Filtering: Intuition

- Our consumption behavior is biased by our social connections
- Knowing a user's connections and what her friends like should help better predict her preferences
- Two signals driving user preferences
  - Collaborative, capturing general user preferences
  - Social, capturing influence from friends
- Social Collaborative Filtering aims to capture user preferences from both signals



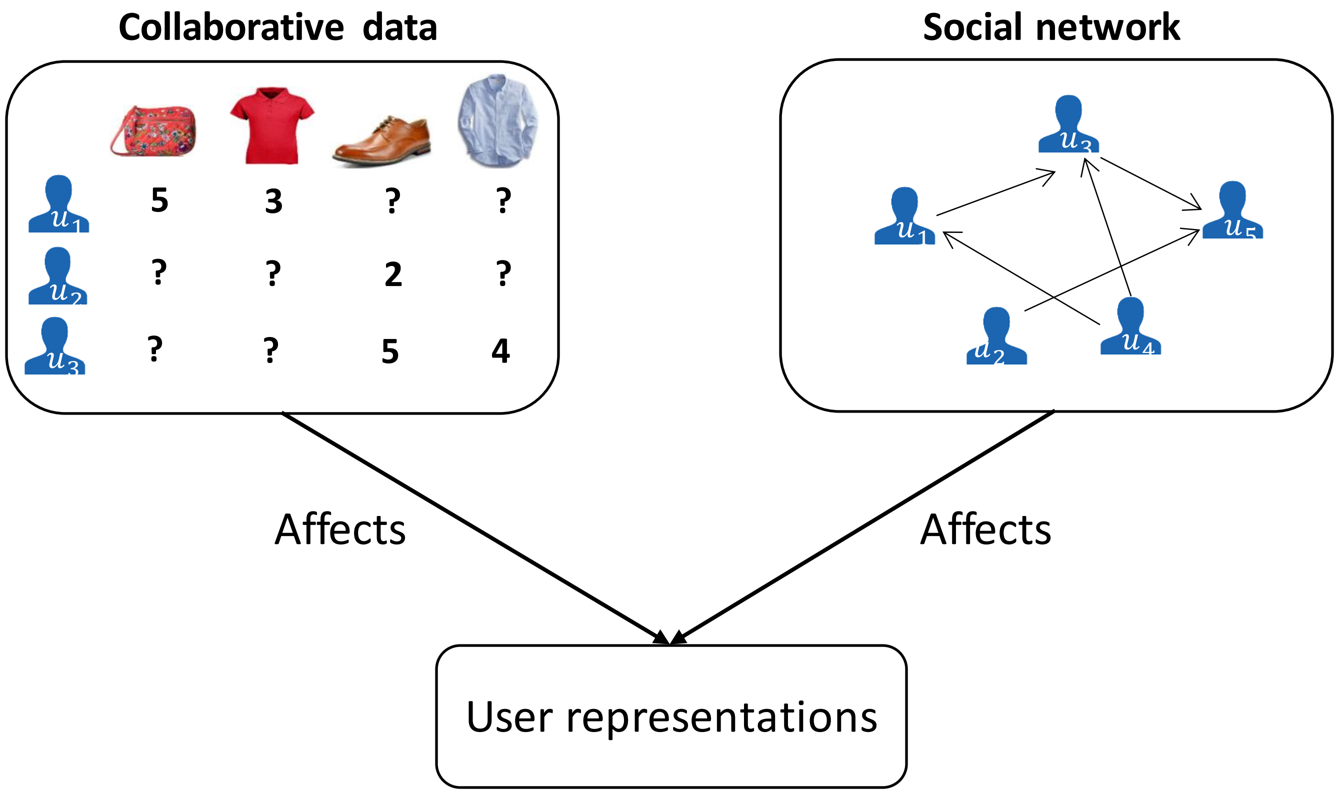
Target user (center) along with her training items (top), recommendations (bottom), and her friends (left and right users).

Social Collaborative Model: Main Families

- **Feature-based**
  - Derive user representations from Social and Collaborative Filtering (CF) data
  - Existing methods differ mainly in the models used for CF and Social data
- **Regularization-based**
  - Regularize the latent space of the CF model using the social network
  - Users who are socially connected are encouraged to have similar latent representations
- **Social network-aware architectures**
  - Embed the social network into the CF model's architecture
  - Learning is driven by the collaborative signal only

Feature-Based Approaches: the big picture

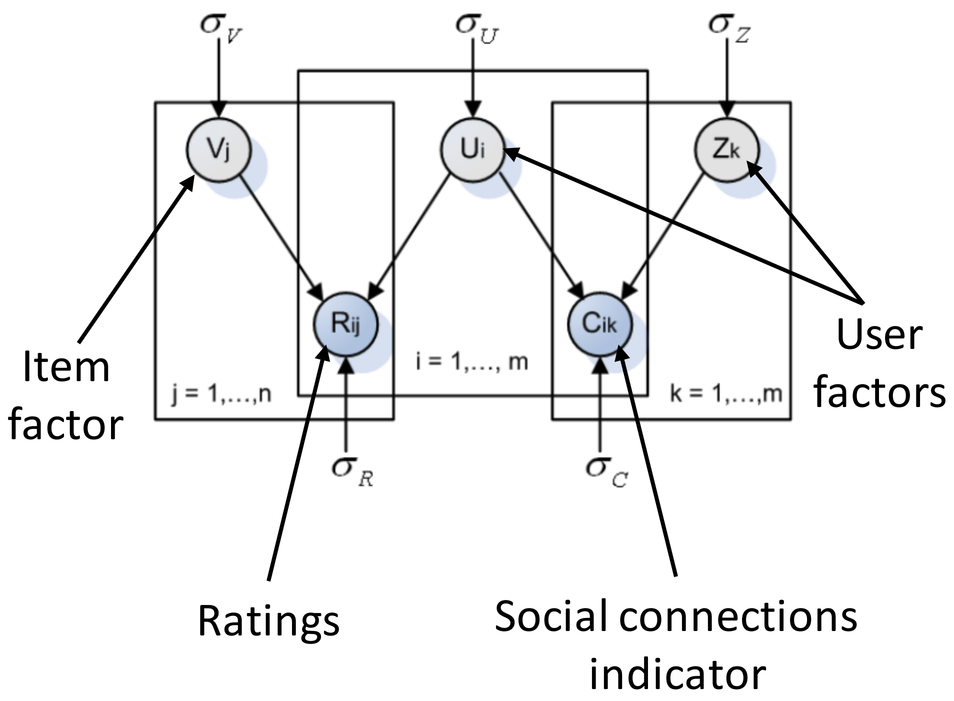
Any adequate models can be used for the two signals, namely Collaborative and Social



SoRec Model

Ma, Yang, Lyu, and King, "SoRec: Social Recommendation Using Probabilistic Matrix Factorization", CIKM 2009.

SoRec graphical model



- Modeling assumptions

$$\boldsymbol{U}_i \sim N(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}) \qquad \boldsymbol{C}_{ik} | \boldsymbol{U}, \boldsymbol{Z} \sim N(g(\boldsymbol{U}_i^\top \boldsymbol{Z}_k), \sigma_C^2)$$

$$\boldsymbol{Z}_k \sim N(\boldsymbol{0}, \sigma_z^2 \boldsymbol{I}) \qquad \boldsymbol{R}_{ij} | \boldsymbol{U}, \boldsymbol{V} \sim N(g(\boldsymbol{U}_i^\top \boldsymbol{V}_j), \sigma_R^2)$$

$$V_j \sim N(\boldsymbol{0}, \sigma_v^2 \boldsymbol{I})$$

- Loss function

$$\sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij}^{R} \left(r_{ij} - g\left(\boldsymbol{U}_i^\top \boldsymbol{V}_j \right) \right)^2 + \lambda_C \sum_{i=1}^{m} \sum_{k=1}^{m} I_{ik}^{C} \left(C_{ik} - g\left(\boldsymbol{U}_i^\top \boldsymbol{Z}_k \right) \right)^2 + \ Reg.$$

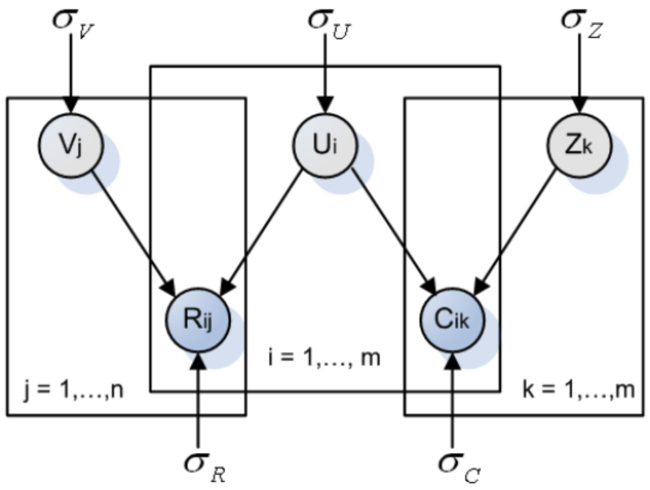
- Gradient with respect to $\boldsymbol{U_i}$ components

$$\frac{d\mathcal{L}}{du_{ik}} = \sum_{j:\, r_{ij} \in \boldsymbol{R}} -v_{jk} \cdot \left(r_{ij} - \boldsymbol{U}_i^\top \boldsymbol{V}_j \right) + \lambda_C \sum_{h:\, g_{ih} \in \boldsymbol{G}} -z_{hk} \cdot \left(c_{ik} - \boldsymbol{U}_i^\top \boldsymbol{Z}_k \right) + \lambda \cdot u_{ik}$$

SoRec Model: Some properties

Ma, Yang, Lyu, and King, "SoRec: Social Recommendation Using Probabilistic Matrix Factorization", CIKM 2009.

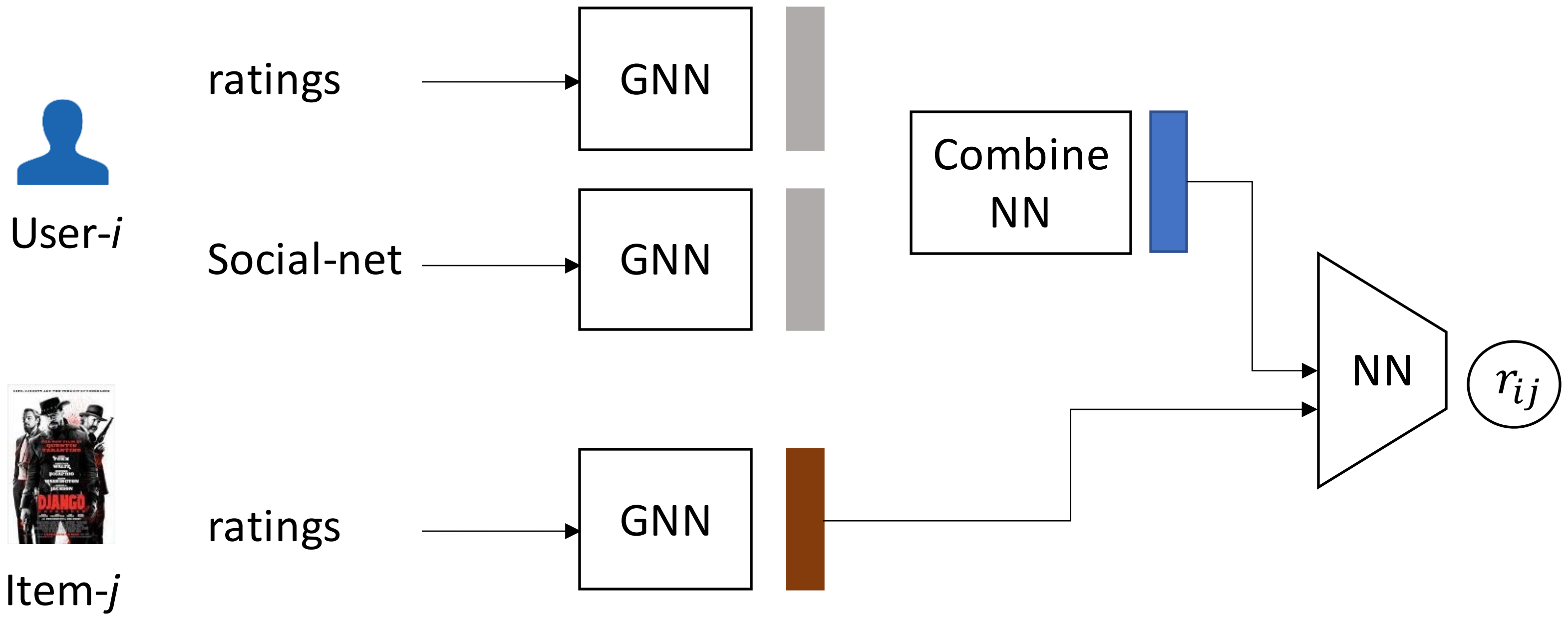
- Simple and intuitive
- Equivalent to applying Probabilistic Matrix Factorization (PMF) on the concatenation of the rating and adjacency matrices $[\boldsymbol{R}, \boldsymbol{C}]$
- Efficient and scalable training with SGD
- Improves upon PMF in recommendation
- Can predict both ratings and social connections
- Learning can be dominated by the social-net



GraphRec

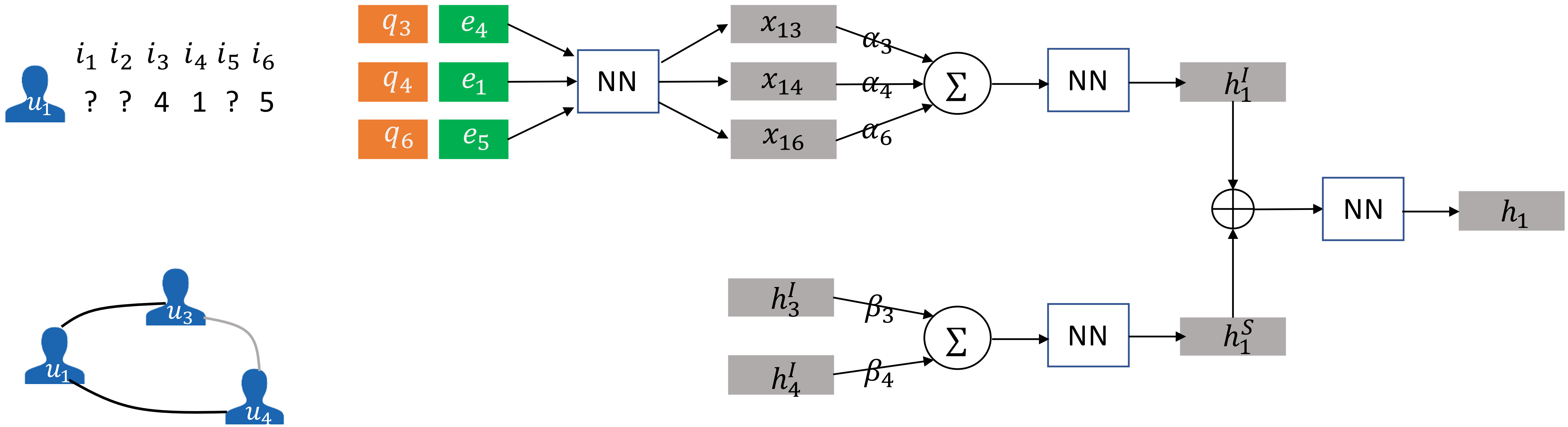
Fan, Wenqi, et al. "Graph neural networks for social recommendation." *WWW*. 2019.

- Uses GNNs to learn user/item representations from user-item preferences and social network data

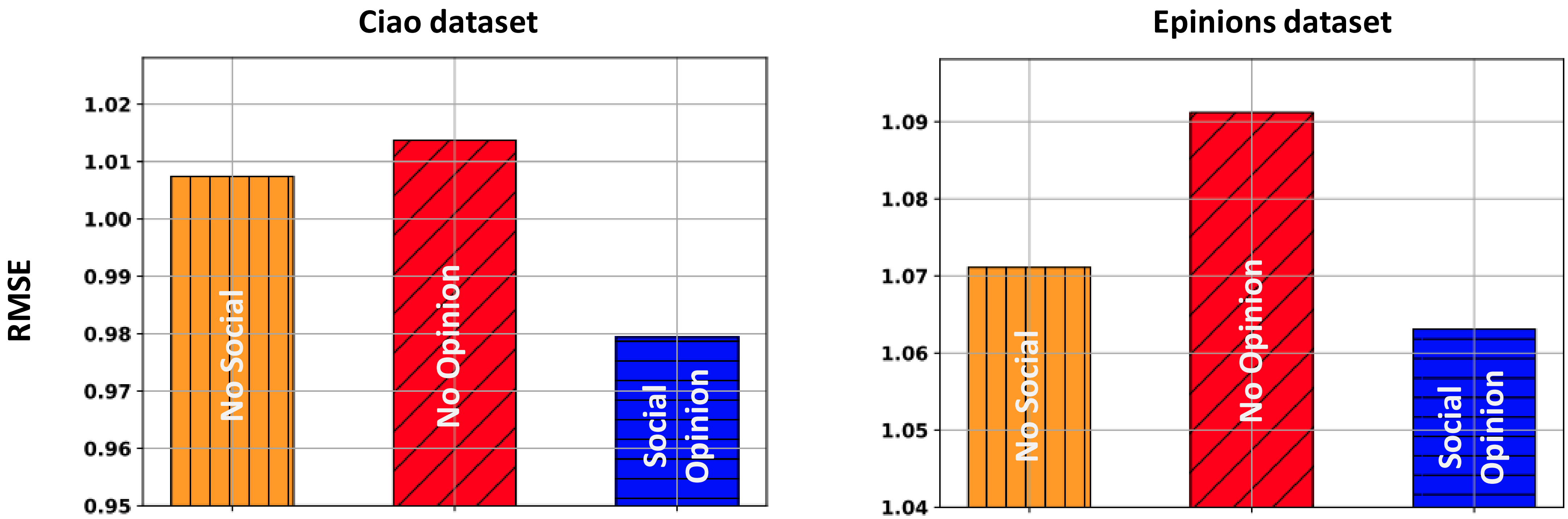


GNN-based user representation

Fan, Wenqi, et al. "Graph neural networks for social recommendation." *WWW*. 2019.



GraphRec: Results



RMSE: Root Mean Square Error

Effect of Social Network and User Opinions

Regularization-based methods

- **Core idea:** bring the latent representations of socially connected users closer to each other
- Act at the objective function level by introducing a suitable regularization term

$$\textit{Objective} = \textit{Collaborative-Objective} + \underbrace{\lambda}_{\text{Regularization paramter}} \times \underbrace{\underline{\textit{Social-Aware Regularization}}}_{\text{Pull closer the representations of connected users}}$$

SoReg: Social Regularized Matrix Factorization

Ma, Hao, et al. "Recommender systems with social regularization." *WSDM*. 2011.

- SoReg's objective function

$$\min_{U,V} \mathcal{L}_1(R,U,V) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij} - U_i^T V_j)^2$$

- Collaborative MF term

$$+ \frac{\alpha}{2}\sum_{i=1}^{m} \|U_i - \frac{\sum_{f\in\mathcal{F}^+(i)} Sim(i,f) \times U_f}{\sum_{f\in\mathcal{F}^+(i)} Sim(i,f)}\|_F^2$$

- Average-based regularization

Set of friends of user i

$$+ \frac{\beta}{2}\sum_{i=1}^{m}\sum_{f\in\mathcal{F}^+(i)} Sim(i,f)\|U_i - U_f\|_F^2$$

- Individual-based regularization

Social Mixture of Von Mises-Fisher distributions (Social-movMF)

Salah, A., & Nadif, M. 2017. Social regularized von Mises–Fisher mixture model for item recommendation. *Data Min. Knowl. Discov. 31*(5), 1218-1241.

- MovMF social-regularized log-likelihood:

$$L_r(\Theta;\mathbf{X},\mathcal{T}) = L(\Theta;\mathbf{X}) - \lambda R(\mathcal{T})$$

$$L(\Theta;\mathbf{X}) = \sum_i \log\left(\sum_h \alpha_h f_h(\mathbf{x}_i|\boldsymbol{\mu}_h,\kappa_h)\right)$$

$$R(\mathcal{T}) = \frac{1}{2}\sum_h\sum_i\sum_j \tau_{ij}(\tilde{z}_{ih} - \tilde{z}_{jh})^2$$

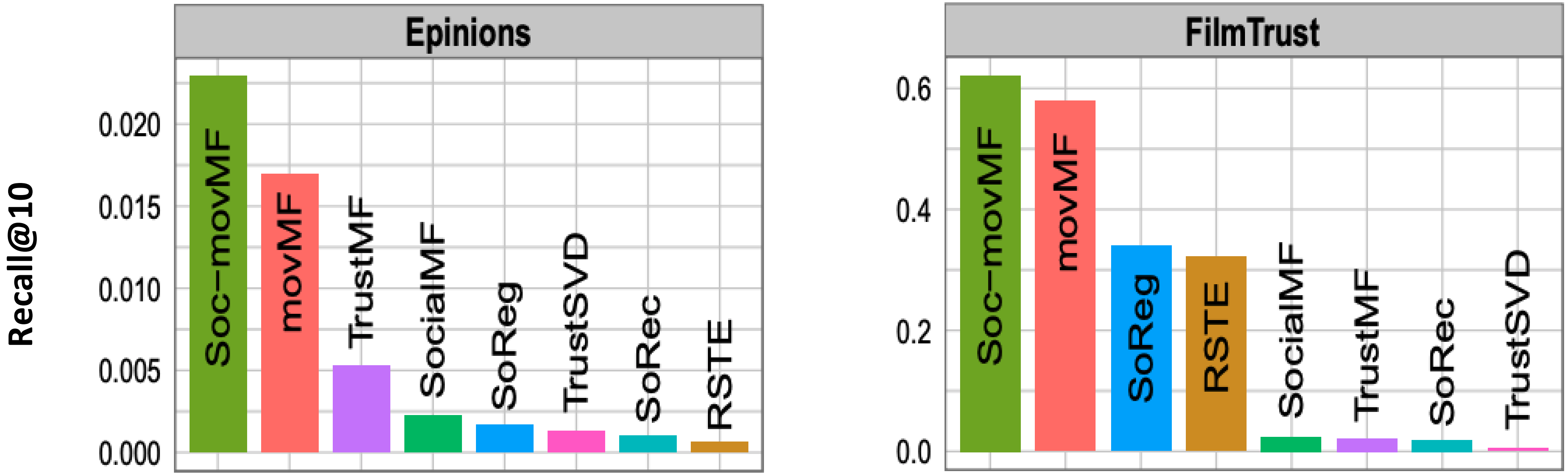
Friendship indicator

Cluster membership probabilities

- Regularized log-likelihood
- h cluster index, μ centroid, κ concentration
  $\boldsymbol{x}_i$ preferences of user i, α cluster proportion
- Smooth the posterior probabilities $\tilde{z}_{ih}$ based on the social network
- Missing rating prediction,

$$\hat{\mathbf{x}}_i = \frac{\sum_h \tilde{z}_{ih}\boldsymbol{\mu}_h}{\|\sum_h \tilde{z}_{ih}\boldsymbol{\mu}_h\|}.$$

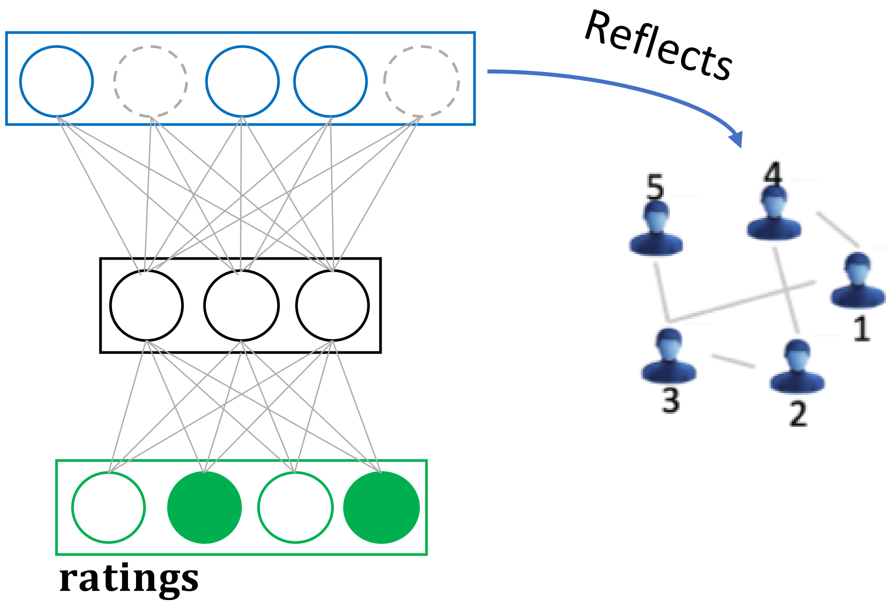
Social-movMF vs other Models on Epinions and FilmTrust datasets



Integrating social information may not be enough, the base model of the rating data is also important.

Social Network-Aware Architecture

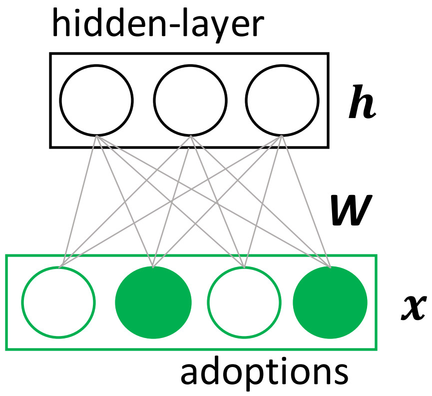
The social network is reflected in the collaborative model's architecture



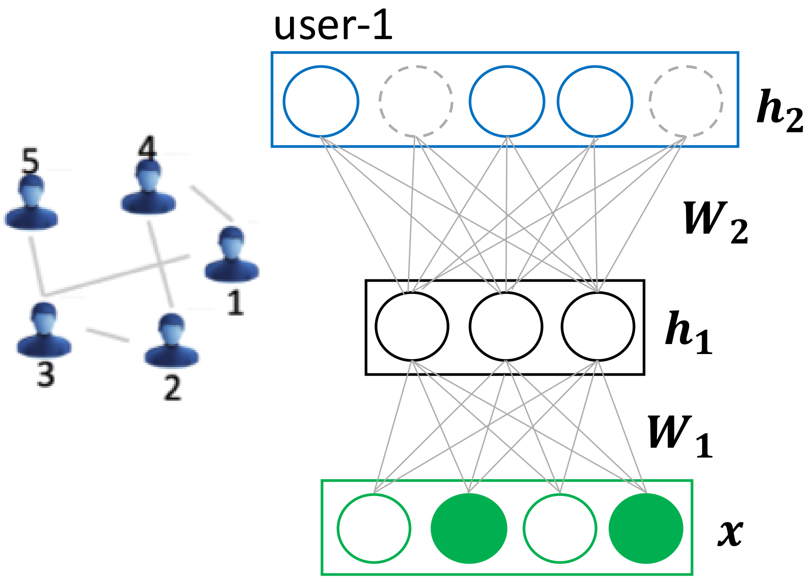
Social RBM-Deep

RBM: Restricted Boltzmann Machine

Nguyen, Trong T., and Hady W. Lauw. "Representation learning for homophilic preferences." *RecSys*. 2016.



Vanilla RBM

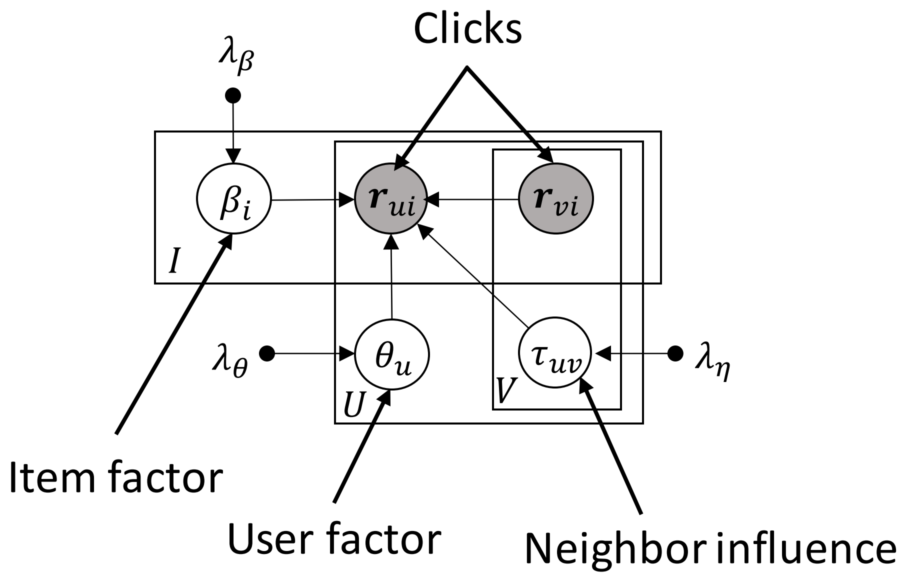


SocialRBM-Deep

- The top layer $\boldsymbol{h}_2$ has U hidden units, corresponding to U users
- User-1 has connections to users 3 and 4, thus the hidden units h_2^1, h_2^3 and h_2^4 are available for encoding user-1 adoptions

Social Poisson Factorization (SPF)

Chaney, A. J., Blei, D. M., & Eliassi-Rad, T. A probabilistic model for using social networks in personalized item recommendation. RecSys. 2015. (pp. 43-50).

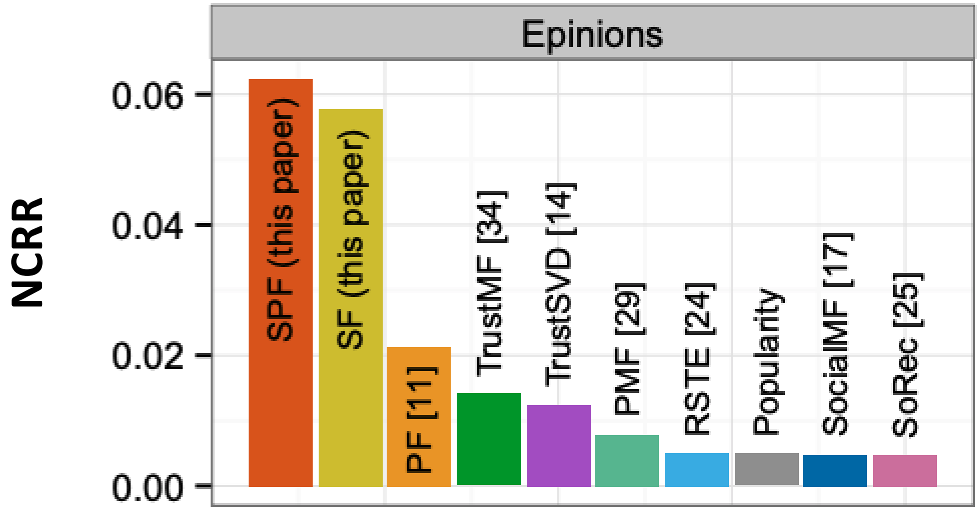


- Two signal driving each user's clicks
  - Latent preference for items captured by β_i and θ_i
  - The latent influence of her friends represented by τ_{uv} and r_{vi}
- The model is specified conditionally

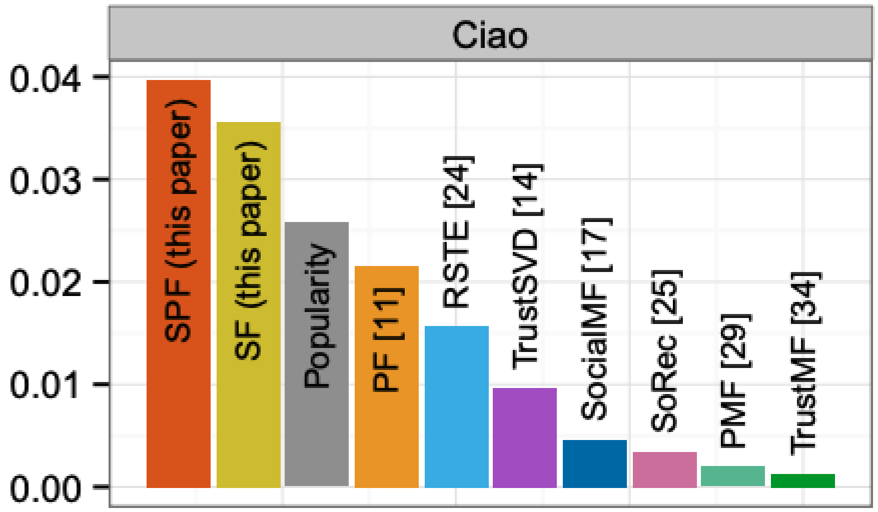
$$r_{ui} \mid r_{-u,i} \sim \text{Poisson}\left(\theta_u^\top \beta_i + \sum_{v \in N(u)} \tau_{uv} r_{vi}\right)$$

- This is an improper model, but offer strong performances
- Scalable fitting to data with Variational Inference

SPF vs other Models on Epinions and Ciao



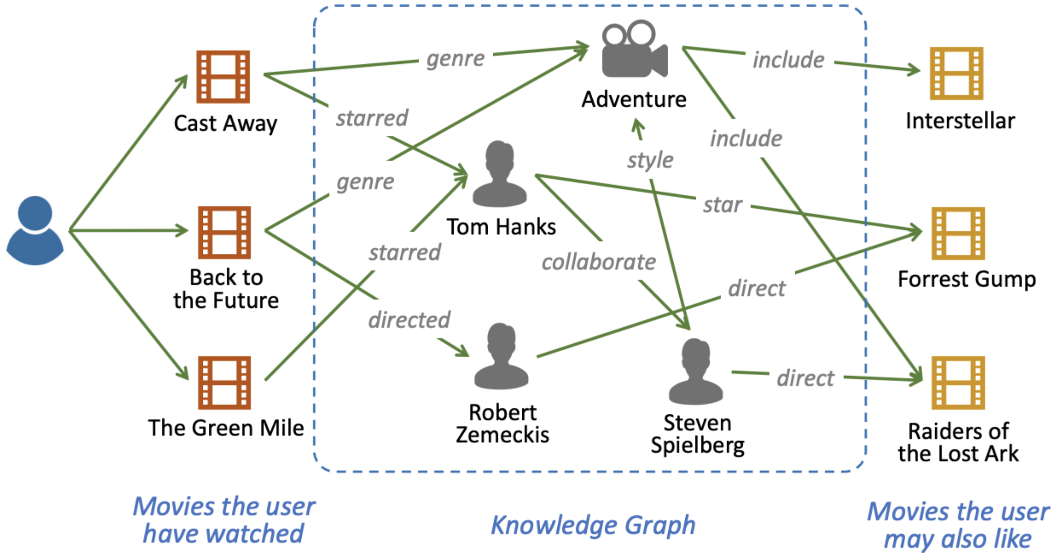
NCRR: normalized cumulative reciprocal rank



- The base model of the clicks is important
- Popularity-based recommendation is a strong baseline

Item-Side

- We are interested in ***item-relatedness***



Knowledge Graph

Browsed/Bought Together

Why is Item-Relatedness Important

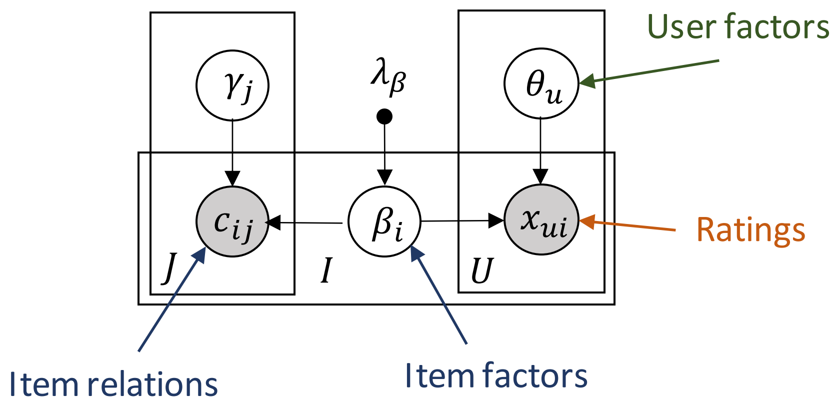
- We rarely consume items, e.g., pair of jeans, with identical or very similar features to those we have already experienced.
- We tend to consume items which
  - can complement each other, e.g., shirt and matching pair of jeans,
  - are alternatives to each other with some different features, e.g., two shirts of the same style but different colors.
- Aspects that are beyond feature similarities, which difficult to capture with the text or image modalities

Substitutes & Complements

Matrix Co-Factorization (MCF)

Park, C., Kim, D., Oh, J., & Yu, H. Do" Also-Viewed" Products Help User Rating Prediction?. WWW. 2017. (pp. 1113-1122).

- Analogous to SoRec
- Can be implemented by concatenating $X = (x_{ui})$ and $C = (c_{ij})$ row-wise
- Can be used for item relatedness prediction

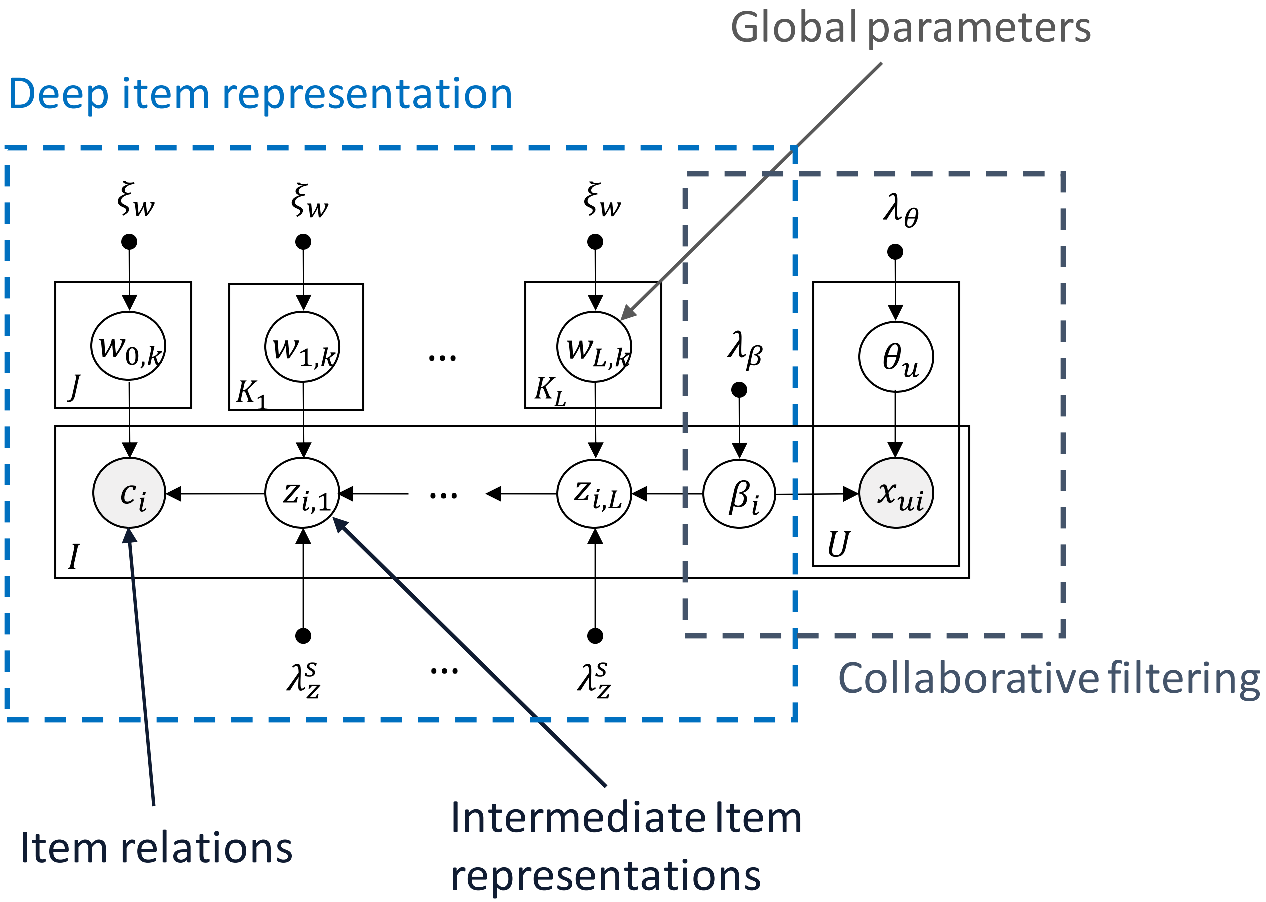


MCF in plate notations

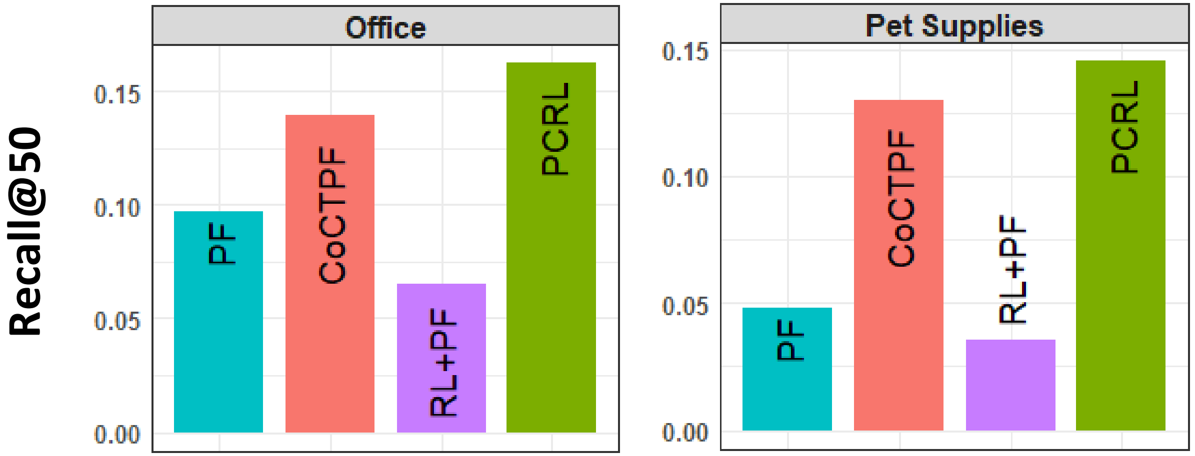
Probabilistic Collaborative Representation learning (PCRL)

Salah, A., & Lauw, H. W. Probabilistic collaborative representation learning for personalized item recommendation. UAI. 2018.

- Preferences guide representation learning
- The item graph helps in predicting preferences
- The deep part of PCRL is robust thanks to its probabilistic nature



Results: item recommendation on Amazon data



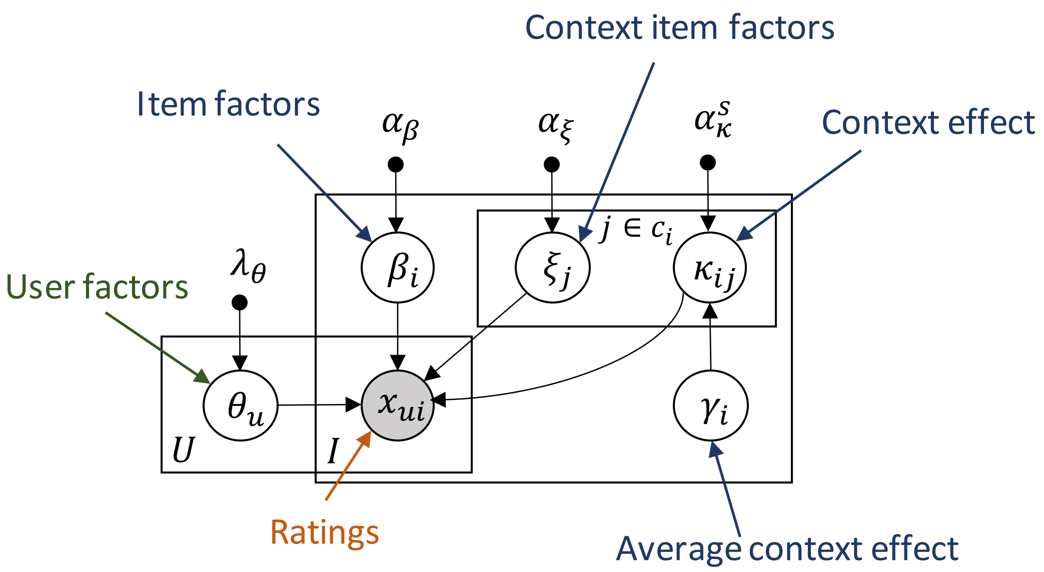
PF: Poisson Factorization
CoCTPF: Content Only Collaborative Topic PF. Equivalent to MCF with a Gamma-Poisson model
RL+PF: Representation Learning + PF. Like PCRL but uses disjoint learning.

Main findings

- Item relationships boost recommendation performance
- PCRL's Deep structure is useful
- Joint learning/modeling is important

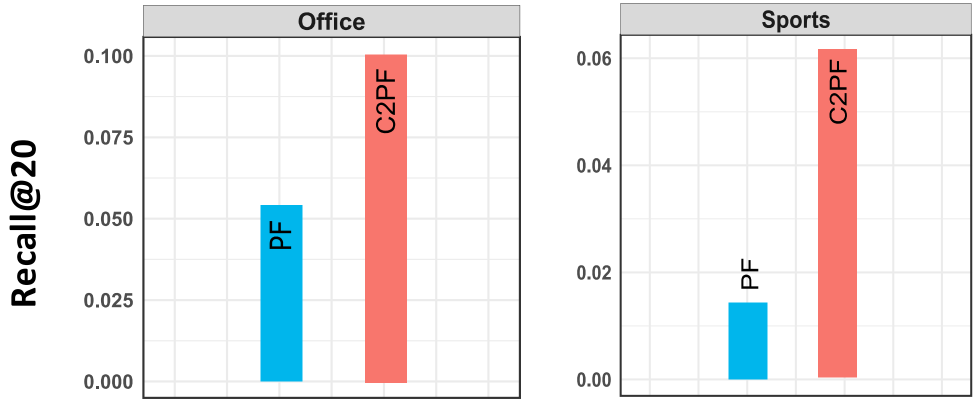
Collaborative Context PF (C2PF)

Salah, Aghiles, and Hady W. Lauw. "A bayesian latent variable model of user preferences with item context." IJCAI, 2018.



C2PF in plate notations

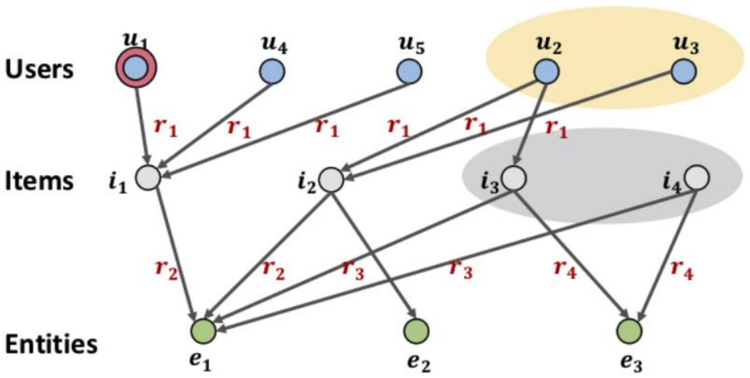
- Side-Information embedded into model's architecture
- Leaning driven by rating signals only
- Strong improvements over Co-Factorization on Amazon



C2PF vs PF on Amazon data

Knowledge Graph Attention Network (KGAT)

Wang, X., He, X., Cao, Y., Liu, M., & Chua, T. S. Kgat: Knowledge graph attention network for recommendation. *SIGKDD.* 2019 (pp. 950-958).



Collaborative Knowledge Graph (CKG)
User-Item graph ∪ Knowledge Graph

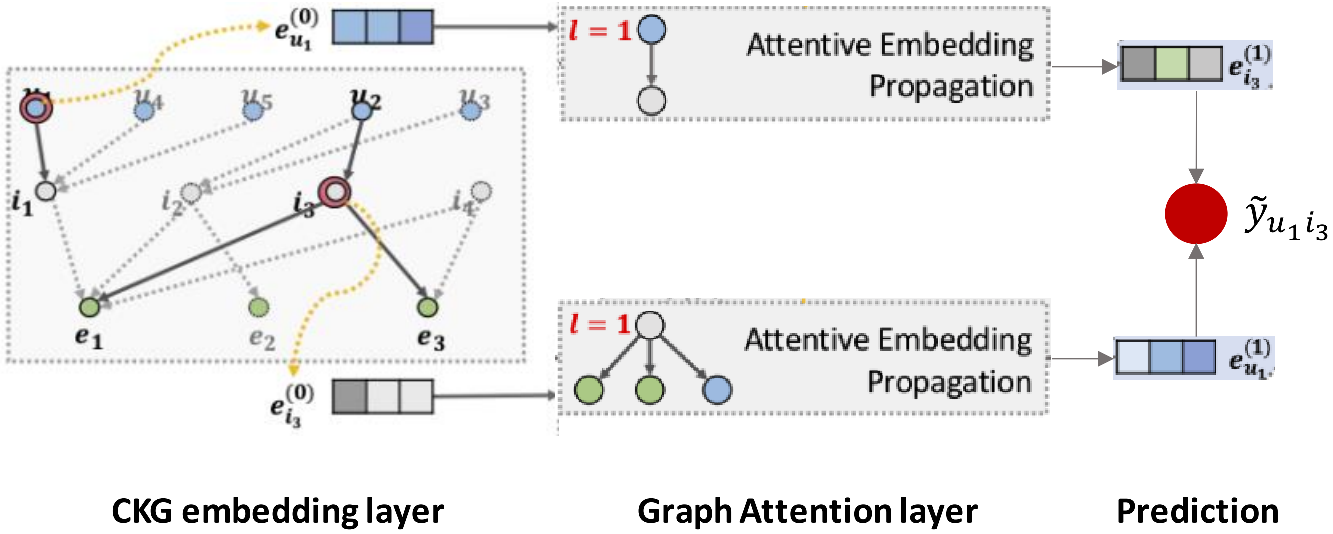


Illustration of KGAT model

- Embed entities and relations
- Graph Attention based representation of entities
- Predict user-item interactions

Cornac-Supported Graph-Based Models

- User network:
  - Social Recommendation using PMF (SoRec)
  - Social Bayesian Personalized Ranking (SBPR)
  - …
- Item network:
  - Collaborative Context Poisson Factorization (C2PF)
  - Probabilistic Collaborative Representation Learning (PCRL)
  - Matrix Co-Factorization (MCF)
  - …

More on Graph-related Recommendation

- Social Recommendation (Users)
  - Deep Social Collaborative Filtering (RecSys'19)
  - Deep Adversarial Social Recommendation (IJCAI'19)
  - Graph Neural Network for Social Recommendation (WWW'19)
  - A Neural Influence Diffusion Model for Social Recommendation (SIGIR'19)
  - A Graph Neural Network Framework for Social Recommendations (TKDE'20)
- Knowledge-graph-aware Recommendation (Items)
  - Leveraging Meta-path based Context for Top-N Recommendation with A Neural Co-Attention Model (KDD'18)
  - Knowledge Graph Convolutional Networks for Recommender Systems with Label Smoothness Regularization (KDD'19 and WWW'19)
  - RippleNet: Propagating User Preferences on the Knowledge Graph for Recommender Systems (CIKM'18)

Hands-on #2: Multi-Modal

https://github.com/PreferredAI/tutorials/blob/master/multimodal-recsys/02_multimodality.ipynb

Hands-on #3: Cross-Modal

https://github.com/PreferredAI/tutorials/blob/master/multimodal-recsys/03_cross_modal.ipynb

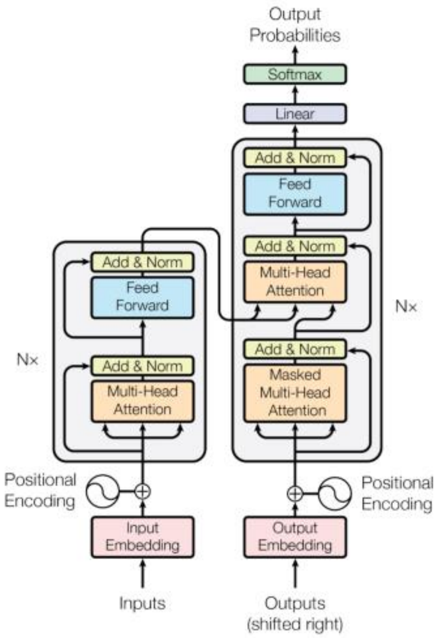
Summary

- Multi-modal recommender systems help to improve performance
- Cross-modal utilization is not only feasible, but positively beneficial and worthy of further exploration
- Cornac is a recommender library that fully supports multi-modality

Future Directions

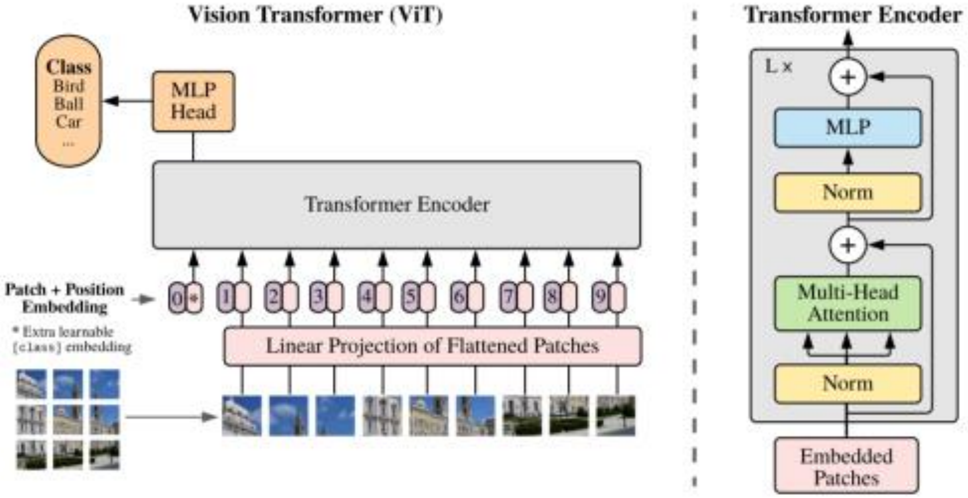
Newer ways to represent a particular modality

Transformers for Text



https://proceedings.neurips.cc/paper/2017

Visual Transformers for Images



https://arxiv.org/abs/2010.11929

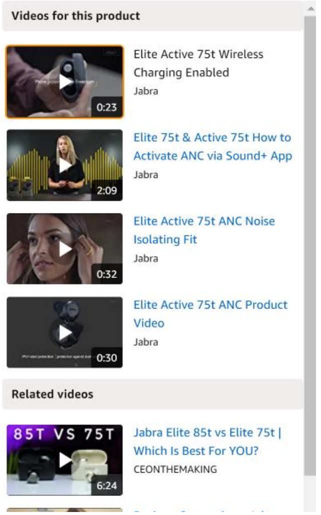
Future Directions

New or under-explored modalities

Videos

Elite Active 75t Wireless Charging Enabled

Jabra



Sound

Music Genome Project

From Wikipedia, the free encyclopedia

The **Music Genome Project** is an effort to "capture the essence of music at the most fundamental level" using various attributes to describe songs and mathematics to connect them together into an interactive map. The Music Genome Project cover 5 music genres: Pop/Rock, Hip-Hop/Electronica, Jazz, World Music, and Classical.

Future Directions

Inter-play of various modalities

Modality-agnostic models

- Development of models that do not specify particular modalities

- Experimental evaluation of models that consider different modalities

Modality-integrated models

- Development of models that work with two or more modalities simultaneously

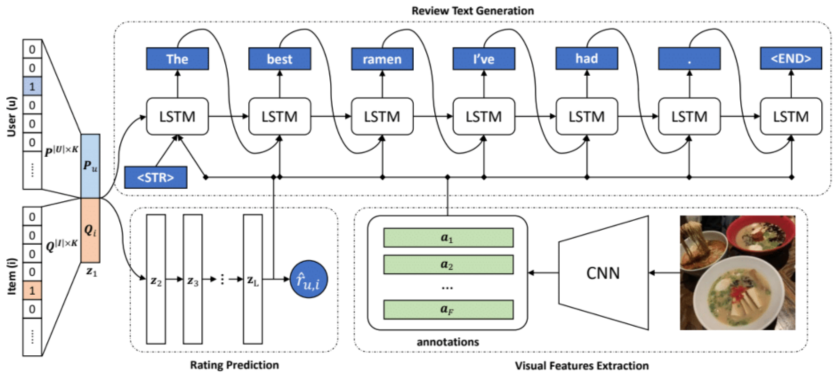


Figure 2: Overall Architecture of Multimodal Review Generation (MRG) model