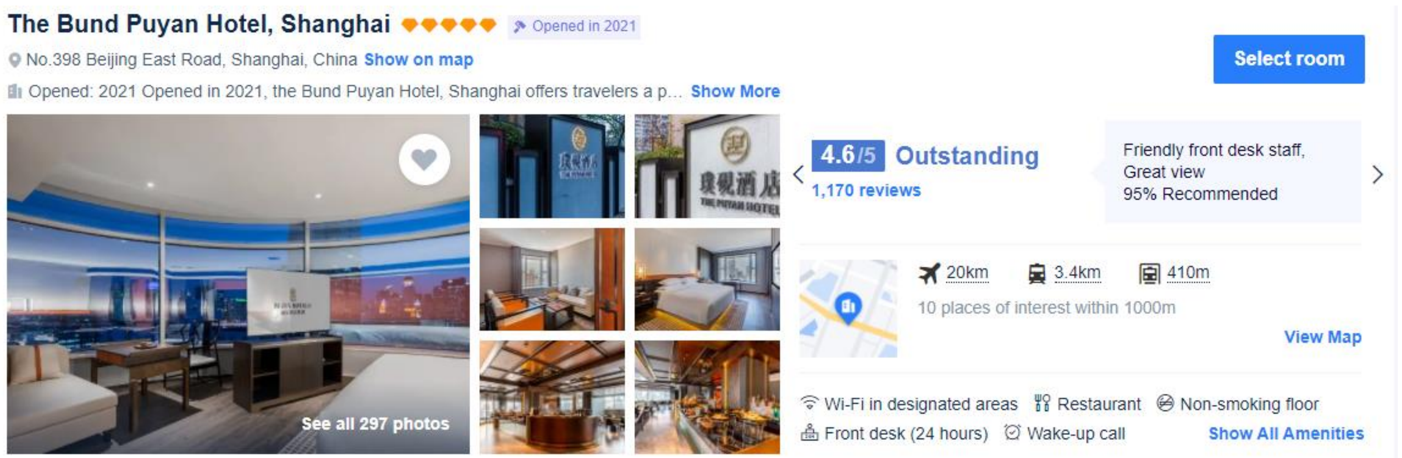
- Overview of multidomain recommender systems (Liang Hu, 25 mins)
- Multi-item domain recommender systems (Z. Y. Lai, 20 mins)
- Multi-user domain recommender systems (Qi Zhang, 20 mins)
- **Multi-data domain recommender systems (Z. Y. Lai, 20 mins)**
- Multi-spatial domain recommender systems (Qi Zhang 20 mins)
- Multi-temporal domain recommender systems (Z. Y. Lai, 20 mins)
- Multi-goal domain recommender systems (Liang Hu, 20 mins)
- Summary (Liang Hu, 5 mins)

Multi-data domain recommendation

- Modalities:
  - E.g. rating, description, photos, geo information
- Distributions:
  - Ratings have different distributions for different domains, e.g., food vs stays
  - The content of image distribution is also different, e.g., color distributions between food and stays.



Multi-data domain recommender systems

- Multi-modal modeling for multidomain recommendation
- Multi-distribution modeling for multidomain recommendation

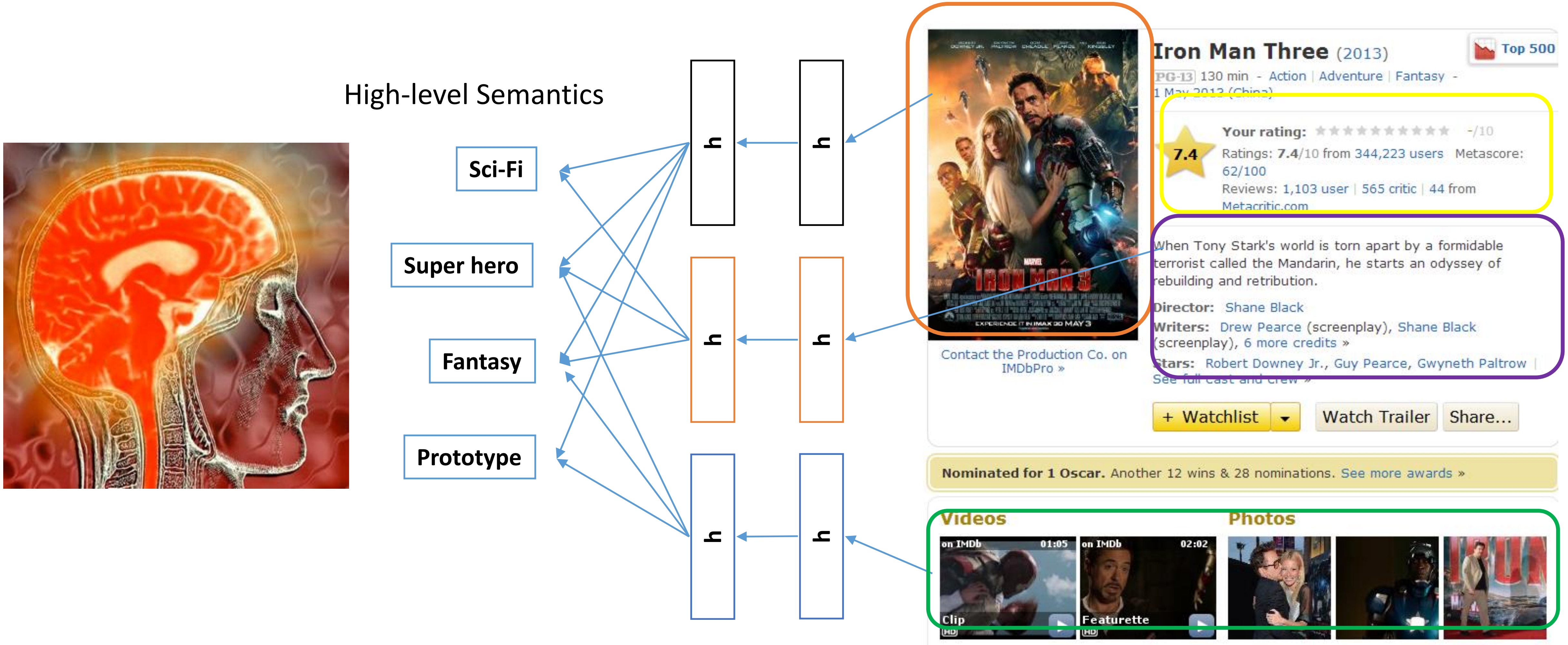
Multimodal Recommender Systems

- Traditional RSs are built on single data type.

- It may not learn sufficient information from single data due to the data incompletion and data quality.

- Joint learning multiple data types, e.g. attributes, text description and images, can obtain more comprehensive information.

Human are Joint Thinking with Related Data

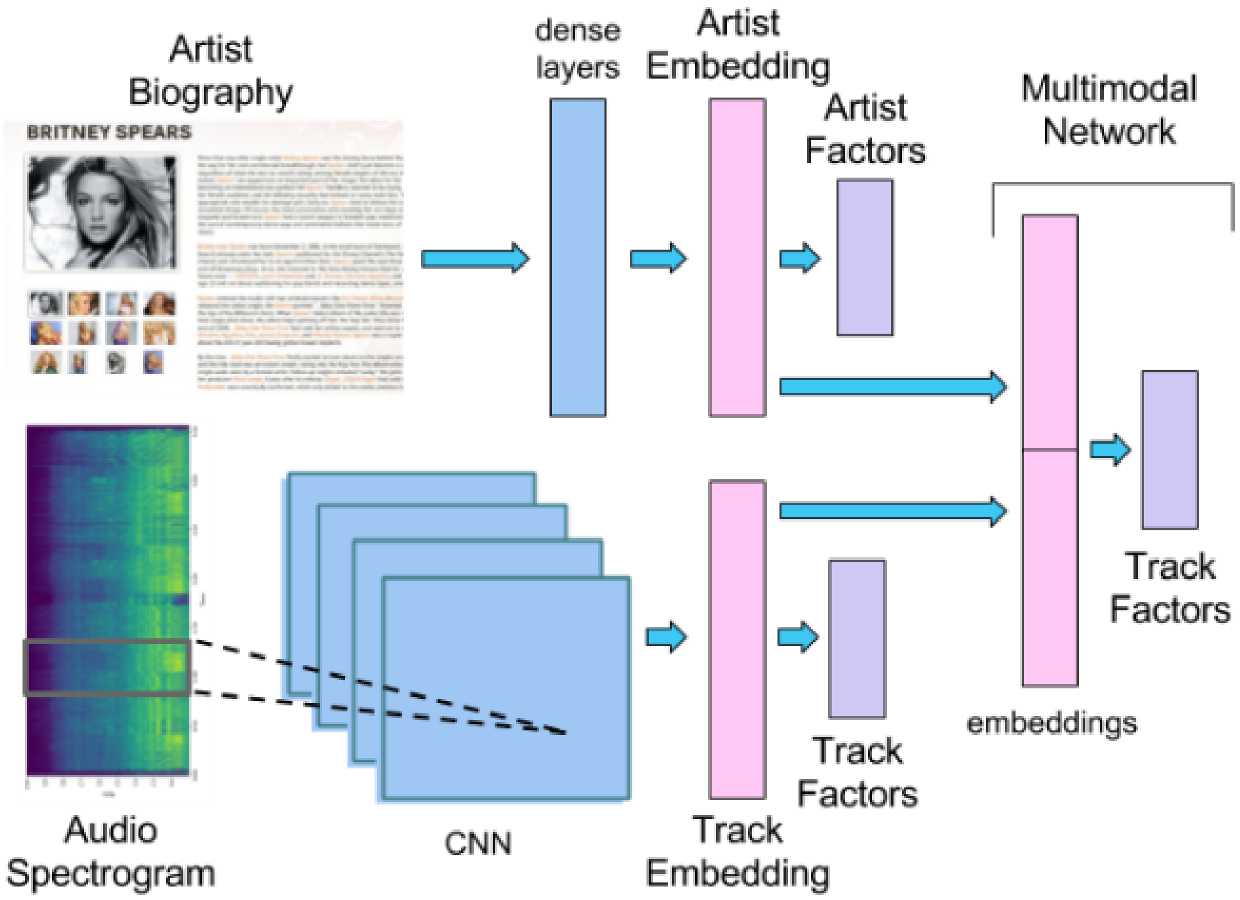


Multimodal Learning

- 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

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.

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

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.

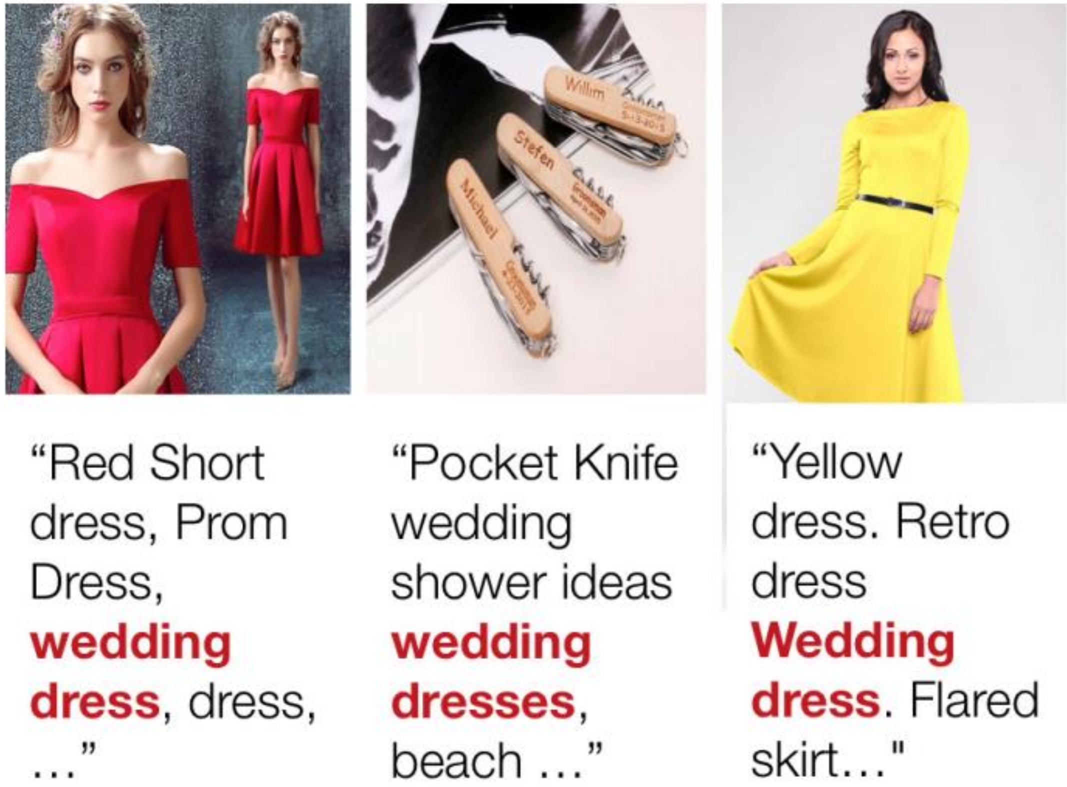
Table 2: Song Recommendation Results

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.

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

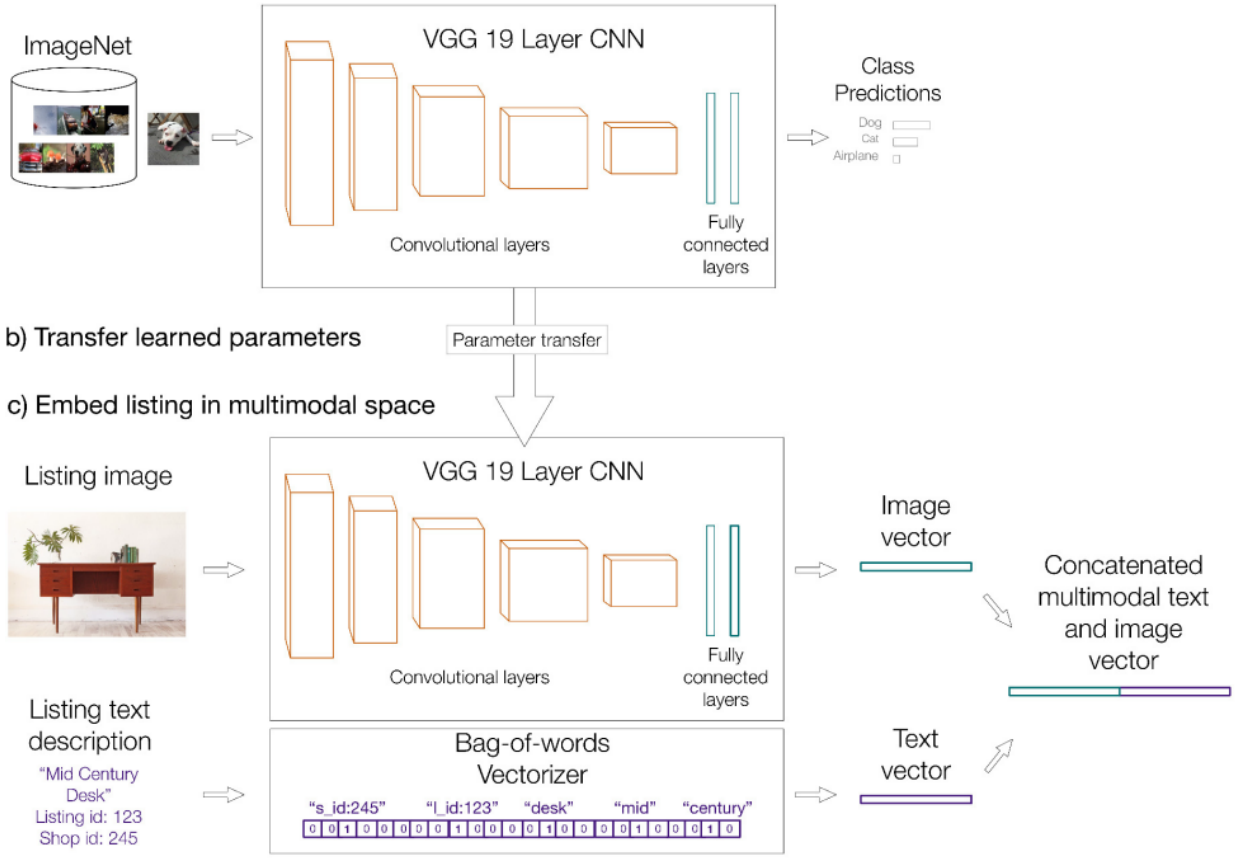
Multimodal learning for images and texts



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

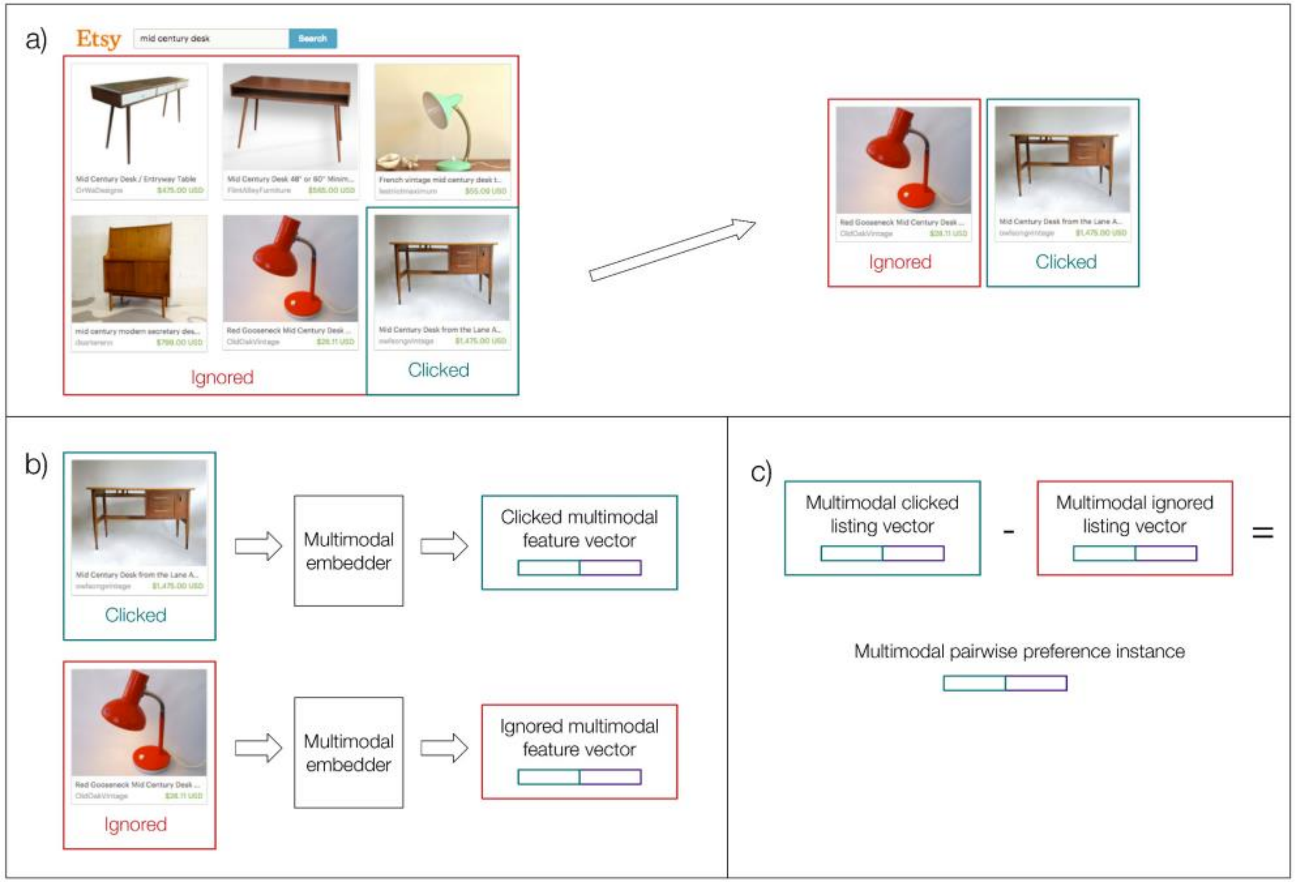
Lynch, C., Aryafar, K., and Attenberg, J. Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 541-548, 2016.

Transferring Parameters of A CNN to The Task of Multimodal Embedding



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.

From search logs to multimodal pairwise classification instances



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.

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/

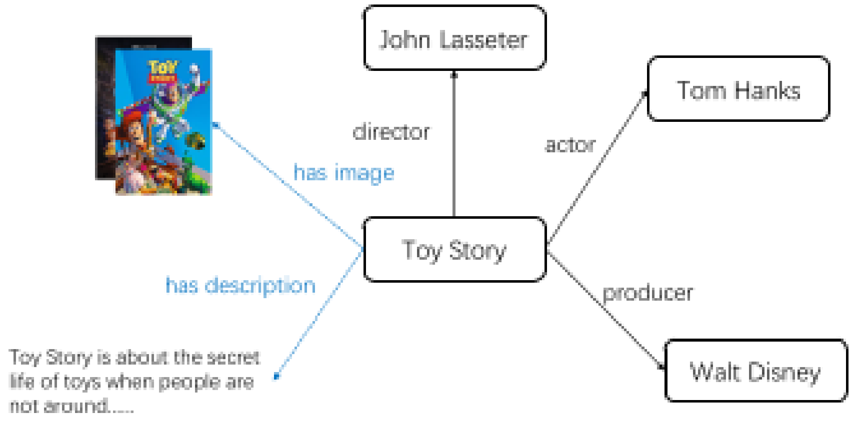
Image information can help disentangle different listings considered similar by a text model

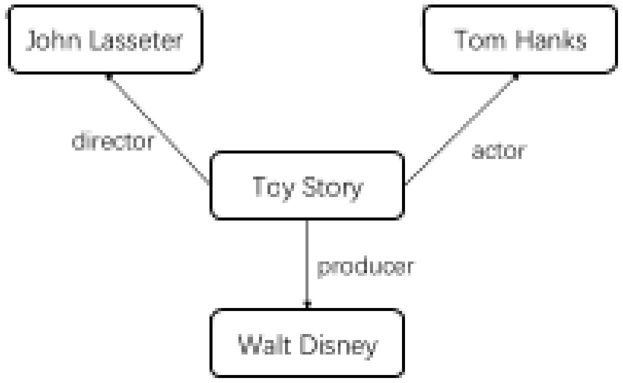
(a) (b) (c)

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.

Visualizing ranking changing by incorporating image information

Knowledge Graphs

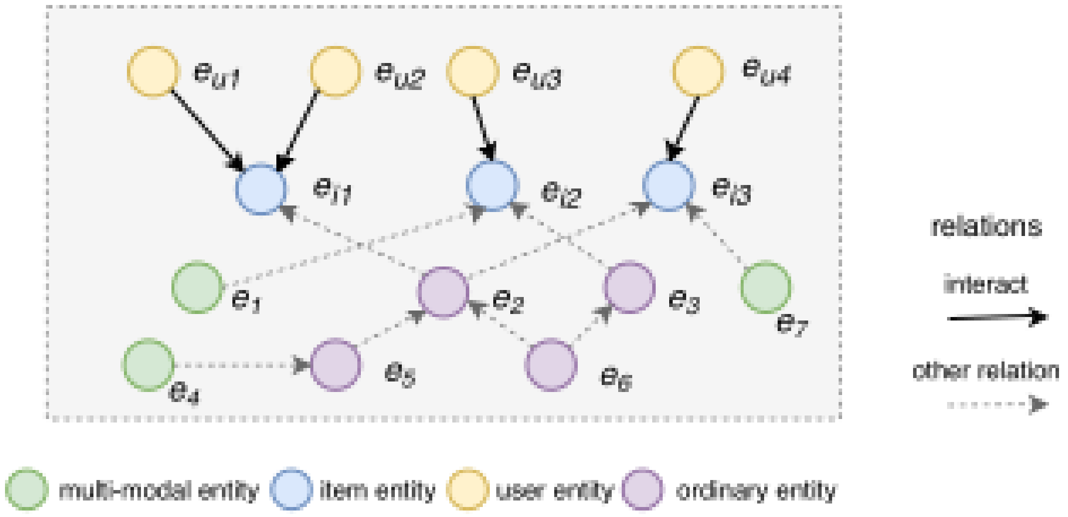




- **Problem to be solved:** how to integrate other sources of relationship information between user / items into CDR?
- **In this paper:** knowledge graphs are specific types of graphs that define relationships between nodes
  - Such relationships contain additional information apart from existence of link

Rui Sun, Xuezhi Cao, Yan Zhao, Junchen Wan, Kun Zhou, Fuzheng Zhang, Zhongyuan Wang, and Kai Zheng. 2020. Multi-modal Knowledge Graphs for Recommender Systems. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management(CIKM '20). Association for Computing Machinery, New York, NY, USA, 1405–1414.

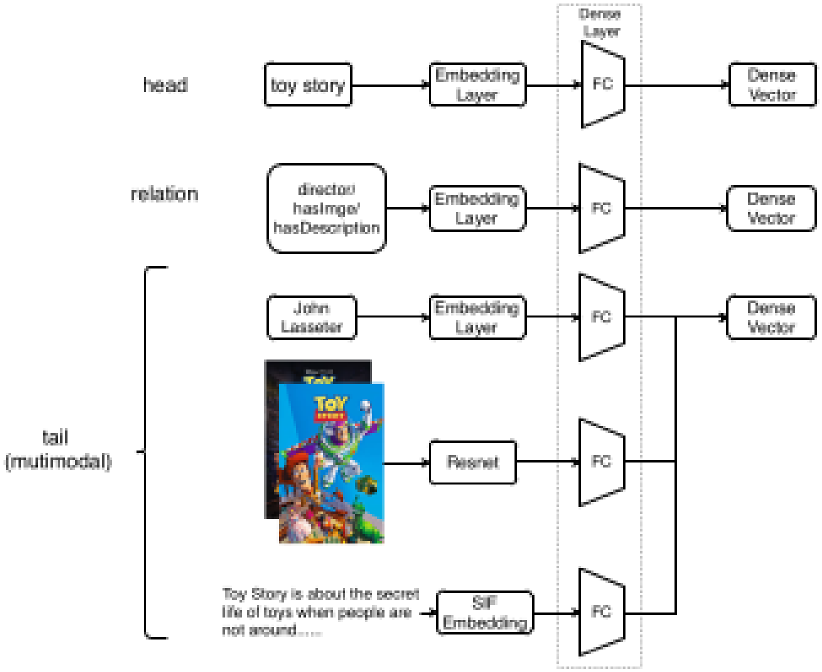
Knowledge Graphs

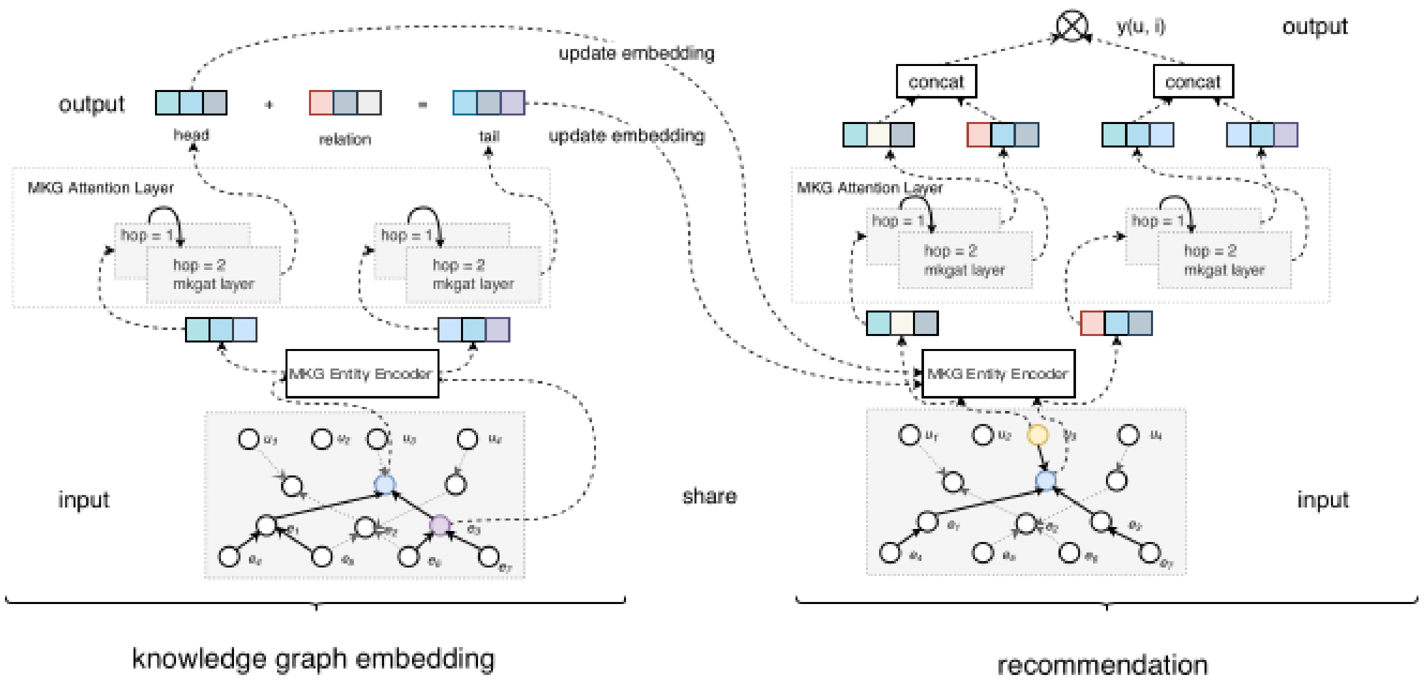


- KGs are directed graph defining characteristics of a central node; MMKGs are KGs which incorporate MM data as 1st class citizens (nodes and edges);
- Collaborative KGs are generalized KGs incorporating user behavior and item knowledge as unified relational graph;
- Task description:
  - Input: collaborative knowledge graph including user-item graph and multimodal knowledge graph;
  - Output: prediction function for probability of user adopting an item.

Rui Sun, Xuezhi Cao, Yan Zhao, Junchen Wan, Kun Zhou, Fuzheng Zhang, Zhongyuan Wang, and Kai Zheng. 2020. Multi-modal Knowledge Graphs for Recommender Systems. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management(CIKM '20). Association for Computing Machinery, New York, NY, USA, 1405–1414.

Knowledge Graphs





- framework consists of two main parts: MMKG Embedding module and Recommendation module;
  - MMKG Embedding Module: consists of graph *entity encoder* and *attention layer;*
    - use different encoders for different modalities (graph structure, images, text) (KGs have distinct structure in form, e.g., (*ToyStory, DirectorOf, John Lasseter*));
    - encoders output multiple encodings in form of dense vectors.
    - attention layer: composed of *propagation* and *aggregation* layers
  - Recommendation Module:

Rui Sun, Xuezhi Cao, Yan Zhao, Junchen Wan, Kun Zhou, Fuzheng Zhang, Zhongyuan Wang, and Kai Zheng. 2020. Multi-modal Knowledge Graphs for Recommender Systems. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management(CIKM '20). Association for Computing Machinery, New York, NY, USA, 1405–1414.

Knowledge Graphs

Models	MovieLens		Dianping	
	recall	ndcg	recall	ndcg
NFM	0.3591	0.4698	0.1163	0.0724
CKE	0.3600	0.4723	0.1321	0.0895
KGAT	0.3778	0.4827	0.1522	0.1301
MMGCN	0.3966	0.5023	0.1424	0.1255
MKGAT	**0.4134**	**0.5181**	**0.1646**	**0.1433**
%Improv.	4.2%	3.1%	8.1%	10.1%

Models	KGAT		MKGAT	
	recall	ndcg	recall	ndcg
base	0.1522	0.1301	0.1542	0.1341
base + text	0.1544	0.1343	0.1589	0.1389
%Improv.	1.5%	3.2%	**3.1%**	**3.5%**
base + image	0.1572	0.1352	0.1612	0.1396
%Improv.	3.3%	3.9%	**4.5%**	**4.1%**
base + text + image	0.1598	0.1361	0.1646	0.1433
%Improv.	4.9%	4.6%	**6.7%**	**6.8%**

- baseline models include neural factorization machines (NFM), KG methods (CKE,KGAT) and multimodal methods (MMGCN);
- baselines are not optimized to handle MM data, and its graphical relational information renders it better than naïve CF methods;
- MM methods coupled with relational information is indeed powerful framework for CDR tasks.

Rui Sun, Xuezhi Cao, Yan Zhao, Junchen Wan, Kun Zhou, Fuzheng Zhang, Zhongyuan Wang, and Kai Zheng. 2020. Multi-modal Knowledge Graphs for Recommender Systems. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management(CIKM '20). Association for Computing Machinery, New York, NY, USA, 1405–1414.

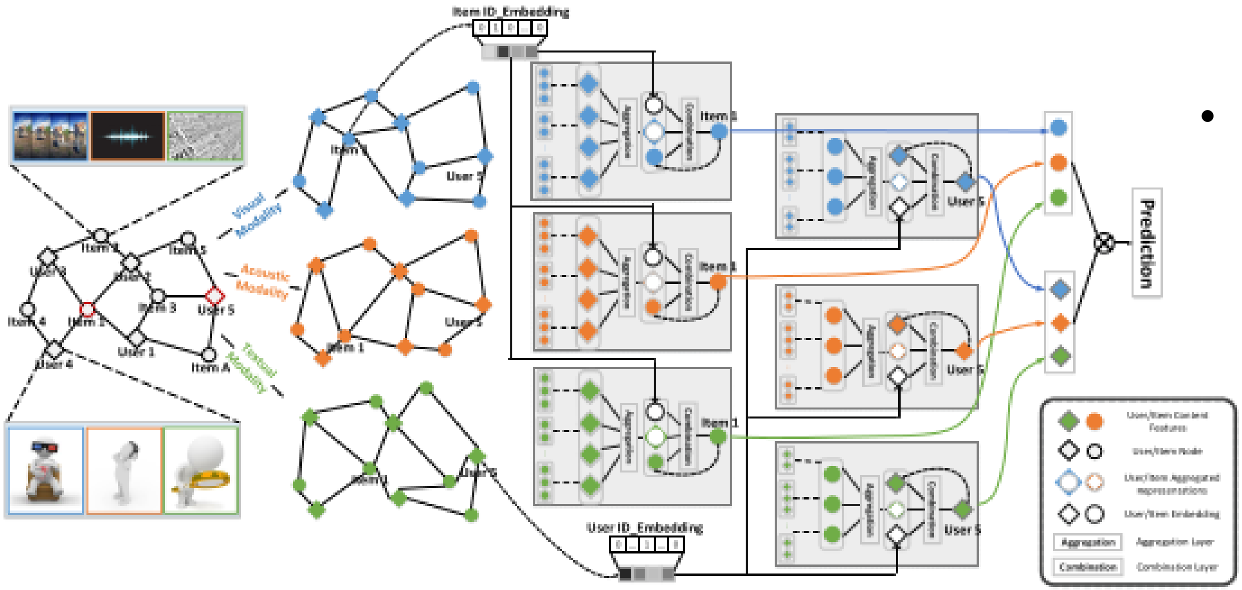
Multi-modal Graph Convolution Network for CDR

- **Problem to be solved:** how to model modality-specific user preferences?

- **In this paper:** construct user-item bipartite graph on each modality.
  - Within each modality user-item interactions aggregated;
  - Aggregated modality embeddings integrated with user representations.

Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and Tat-Seng Chua. 2019. MMGCN: Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video. In Proceedings of the 27th ACM International Conference on Multimedia (MM '19). Association for Computing Machinery, New York, NY, USA, 1437–1445.

Multi-modal Graph Convolution Network for CDR



- use information propagation of GCNs to encode higher-order connectivity between users; construct separate graphs for each modality instead of unifying these;
- Architecture consists of graph inputs, aggregation, combination and prediction layers;
  - Graphs: each modality is treated separately, as a bipartite user-item graph;
  - Aggregation: aggregates historical interaction and user group data; employs either mean or max agg. functions.
  - Combination: needs to account for multiple modalities; projects modality information u_m onto same latent space as user ID embedding u_{id}. Here u_{id} is the bridge between modalities.
  - Prediction: Outputs probability of item-user connections.

Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and Tat-Seng Chua. 2019. MMGCN: Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video. In Proceedings of the 27th ACM International Conference on Multimedia (MM '19). Association for Computing Machinery, New York, NY, USA, 1437–1445.

Multi-modal Graph Convolution Network for CDR

Model	Kwai			Tiktok			MovieLens		
	Precision	Recall	NDCG	Precision	Recall	NDCG	Precision	Recall	NDCG
VBPR	0.2673	0.3386	0.1988	0.0972	0.4878	0.3136	**0.1172**	**0.4724**	**0.2852**
ACF	0.2559	0.3248	0.1874	0.8734	0.4429	0.2867	0.1078	0.4304	0.2589
GraphSAGE	0.2718	0.3412	0.2013	0.1028	0.4972	0.3210	0.1132	0.4532	0.2647
NGCF	**0.2789**	**0.3463**	**0.2058**	**0.1065**	**0.5008**	**0.3226**	0.1156	0.4626	0.2732
MMGCN	**0.3057***	**0.3996***	**0.2298***	**0.1164***	**0.5520***	**0.3423***	**0.1215***	**0.5138***	**0.3062***
%Improv.	**9.61%**	**15.59%**	**11.66%**	**9.03%**	**10.23%**	**6.11%**	**3.67%**	**8.76%**	**7.36%**

- Datasets:
  - Tiktok: micro-video sharing site, 3-15 second videos
  - Kwai: micro-video sharing site, similar to Tiktok
  - MovieLens
- Baselines are distinguished by MF methods (VBPR), attention-based frameworks (ACF) and graphical approaches (GraphSAGE, NGCF)
  - Graph structure outperforms naïve MF methods due to structural information;
  - Other methods were generally unable to capture structural information as well as MMGCN.

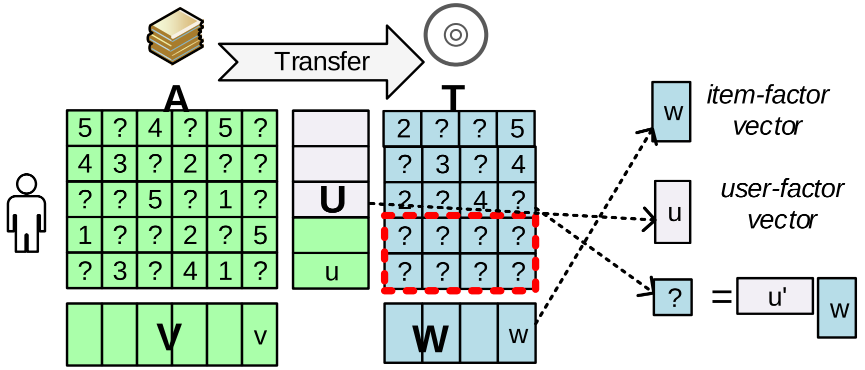
Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and Tat-Seng Chua. 2019. MMGCN: Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video. In Proceedings of the 27th ACM International Conference on Multimedia (MM '19). Association for Computing Machinery, New York, NY, USA, 1437–1445.

Multi-data domain recommender systems

- Multi-modal modeling for multidomain recommendation
- Multi-distribution modeling for multidomain recommendation

MF based Transfer Learning

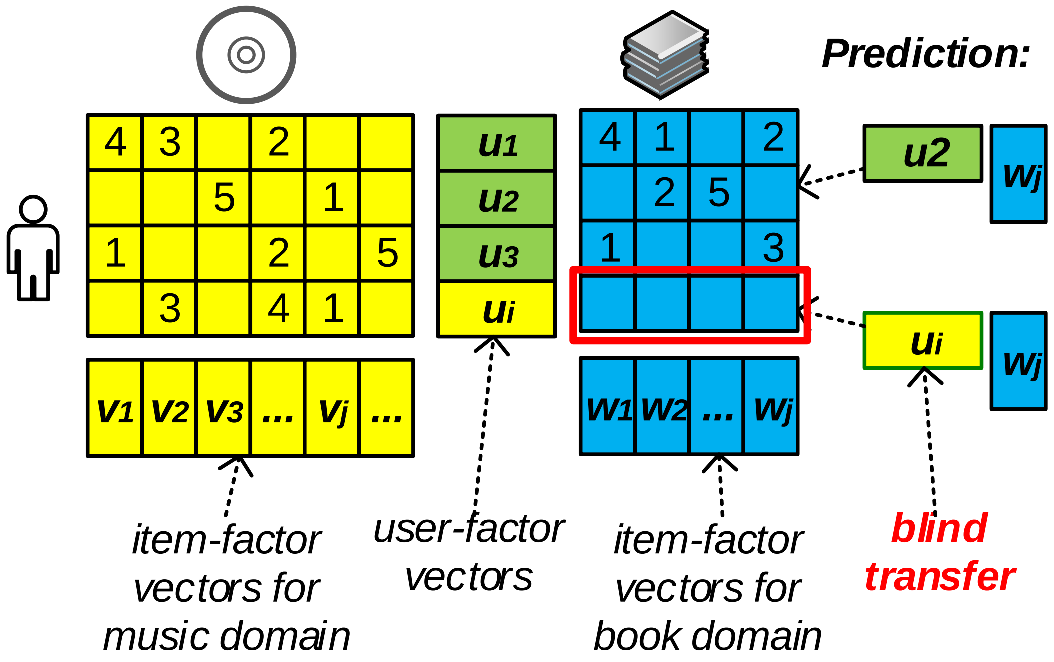
- Transfer the knowledge learned from the auxiliary domain to the target domain
  - The user-factor vectors are *co-determined* by the feedback in auxiliary and target domains



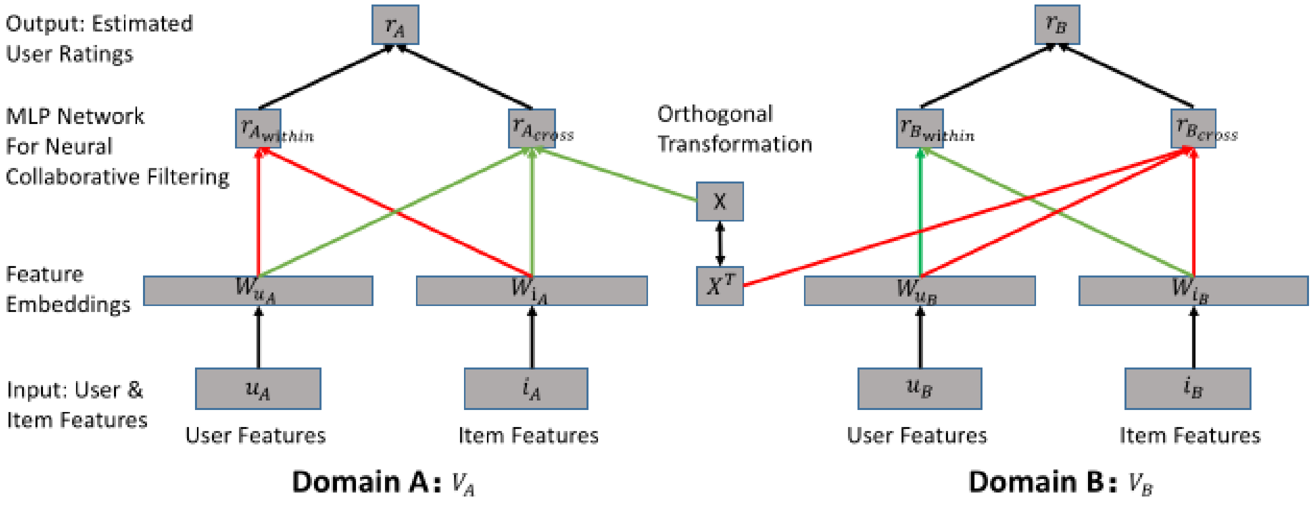
Pan, W., Xiang, E. W., Liu, N. N., & Yang, Q. (2010, July). Transfer Learning in Collaborative Filtering for Sparsity Reduction. In *AAAI* (Vol. 10, pp. 230-235).

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.



Deep Dual Transfer Cross Domain Recommendation



- **Problem to be solved:** how to effectively transfer learned knowledge from one domain to another?
  - Needs to include not only user-item information, but also latent and complex relationships.
- **In this paper:** latent orthogonal embedding approach to transfer learned latents across domains

Li, Pan, and Alexander Tuzhilin. "DDTCDR: Deep Dual Transfer Cross Domain Recommendation." Proceedings of the 13th International Conference on Web Search and Data Mining, 2020, pp. 331–339.

Deep Dual Transfer Cross Domain Recommendation

Algorithm 1 Dual Neural Collaborative Filtering

Input: Domain V_A and V_B, autoencoder AE_A and AE_B, transfer rate α, learning rates γ_A and γ_B, initial recommendation models RS_A and RS_B, initial mapping function X

repeat

- Sample user-item records d_A and d_B from V_A and V_B respectively
- Unpack records d_A, d_B as user features u_A, u_B, item features i_A, i_B and ratings r_A, r_B
- Generate feature embeddings from autoencoder as $W_{u_A} = AE_A(u_A)$, $W_{u_B} = AE_B(u_B)$, $W_{i_A} = AE_A(i_A)$, $W_{i_B} = AE_B(i_B)$
- Estimate the ratings in domain A via $r'_A = (1-\alpha)RS_A(W_{u_A}, W_{i_A}) + \alpha RS_B(X * W_{u_A}, W_{i_A})$
- Estimate the ratings in domain B via $r'_B = (1-\alpha)RS_B(W_{u_B}, W_{i_B}) + \alpha RS_A(X^T * W_{u_B}, W_{i_B})$
- Compute MSE loss $\hat{r_A} = r_A - r'_A$, $\hat{r_B} = r_B - r'_B$
- Backpropogate $\hat{r_A}$, $\hat{r_B}$ and update RS_A, RS_B;
- Backpropogate orthogonal constraint on X; Orthogonalize X

until convergence

- transfer learning approach to multi-distribution recommendation;
- main intuition is that similar preferences in source will be replicated in target domain; procedure quantified via LOTs;
- Algorithm:
  - construct feature embeddings from user and item features;
  - design LOT to transfer feature embeddings across domains since OTs preserve inner product of vectors;
  - minimize MSE between actual and predicted user ratings;
  - backprop MSE losses and update recommendation models;
  - backprop orthogonal constraint and orthogonalize;

Li, Pan, and Alexander Tuzhilin. “DDTCDR: Deep Dual Transfer Cross Domain Recommendation.” Proceedings of the 13th International Conference on Web Search and Data Mining, 2020, pp. 331–339.

Deep Dual Transfer Cross Domain Recommendation

Category	Feature Group	Dimensionality	Type
User Features	Gender	2	one-hot
	Age	$\sim 10^2$	numerical
	Movie Taste	12	one-hot
	Residence	12	one-hot
	Preferred Category	9	one-hot
	Recommendation Usage	5	one-hot
	Marital Status	3	one-hot
	Personality	6	one-hot
Book Features	Category	8	one-hot
	Title	$\sim 10^5$	one-hot
	Author	$\sim 10^4$	multi-hot
	Publisher	$\sim 10^2$	one-hot
	Language	4	one-hot
	Country	4	one-hot
	Price	$\sim 10^2$	numeric
	Date	$\sim 10^3$	date
Movie Features	Genre	6	one-hot
	Title	$\sim 10^5$	one-hot
	Director	$\sim 10^3$	multi-hot
	Writer	$\sim 10^3$	multi-hot
	Runtime	$\sim 10^3$	numeric
	Country	4	one-hot
	Rating	$\sim 10^2$	numeric
	Votes	$\sim 10^4$	numeric
Music Features	Listener	$\sim 10^3$	numeric
	Playcount	$\sim 10^3$	numeric
	Artist	$\sim 10^4$	one-hot
	Album	$\sim 10^4$	one-hot
	Tag	8	one-hot
	Release	$\sim 10^3$	date
	Duration	$\sim 10^3$	numeric
	Title	$\sim 10^5$	one-hot

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

Table 4: Comparison of recommendation performance in Book-Movie Dual Recommendation: Improved Percentage versus the second best baselines

Algorithm	Book				Music			
	RMSE	MAE	Precision@5	Recall@5	RMSE	MAE	Precision@5	Recall@5
DDTCDR	$\mathbf{0.2209}^*$	$\mathbf{0.1704}^*$	$\mathbf{0.8570}^*$	$\mathbf{0.9602}^*$	$\mathbf{0.2753}^*$	$\mathbf{0.2302}^*$	$\mathbf{0.8392}^*$	$\mathbf{0.8928}^*$
Improved %	(+4.07%)	(+8.87%)	(+3.97%)	(+3.15%)	(+2.14%)	(+4.74%)	(+5.51%)	(+5.35%)
NCF	0.2315	0.1887	0.8230	0.9294	0.2828	0.2423	0.7930	0.8450
CCFNet	0.2630	0.1842	0.8150	0.9108	0.3090	0.2422	0.7902	0.8388
CDFM	0.2489	0.2155	0.8104	0.9102	0.3252	0.2463	0.7895	0.8365
CMF	0.2921	0.2478	0.8072	0.8978	0.3478	0.2698	0.7820	0.8324
CoNet	0.2307	0.1897	0.8230	0.9300	0.2801	0.2410	0.7912	0.8428

Table 5: Comparison of recommendation performance in Book-Music Dual Recommendation: Improved Percentage versus the second best baselines

Algorithm	Movie				Music			
	RMSE	MAE	Precision@5	Recall@5	RMSE	MAE	Precision@5	Recall@5
DDTCDR	$\mathbf{0.2174}^*$	$\mathbf{0.1720}^*$	$\mathbf{0.8926}^*$	$\mathbf{0.9869}^*$	$\mathbf{0.2758}^*$	$\mathbf{0.2311}^*$	$\mathbf{0.8370}^*$	$\mathbf{0.8902}^*$
Improved %	(+3.75%)	(+9.77%)	(+5.32%)	(+3.68%)	(+1.89%)	(+4.24%)	(+4.30%)	(+4.38%)
NCF	0.2276	0.1895	0.8428	0.9495	0.2828	0.2423	0.7970	0.8501
CCFNet	0.2468	0.1932	0.8398	0.9310	0.3090	0.2433	0.7952	0.8498
CDFM	0.2289	0.1895	0.8306	0.9382	0.3252	0.2467	0.7880	0.8460
CMF	0.2738	0.2293	0.8278	0.9222	0.3478	0.2698	0.7796	0.8400
CoNet	0.2302	0.1908	0.8450	0.9508	0.2811	0.2428	0.8010	0.8512

Table 6: Comparison of recommendation performance in Movie-Music Dual Recommendation: Improved Percentage versus the second best baselines

Li, Pan, and Alexander Tuzhilin. “DDTCDR: Deep Dual Transfer Cross Domain Recommendation.” Proceedings of the 13th International Conference on Web Search and Data Mining, 2020, pp. 331–339.

Next chapter

- Overview of multidomain recommender systems (Liang Hu, 25 mins)
- Multi-item domain recommender systems (Z. Y. Lai, 20 mins)
- Multi-user domain recommender systems (Qi Zhang, 20 mins)
- Multi-data domain recommender systems (Z. Y. Lai, 20 mins)
- **Multi-spatial domain recommender systems (Qi Zhang 20 mins)**
- Multi-temporal domain recommender systems (Z. Y. Lai, 20 mins)
- Multi-goal domain recommender systems (Liang Hu, 20 mins)
- Summary (Liang Hu, 5 mins)