2020.10.9

Paper Sharing (Rec)

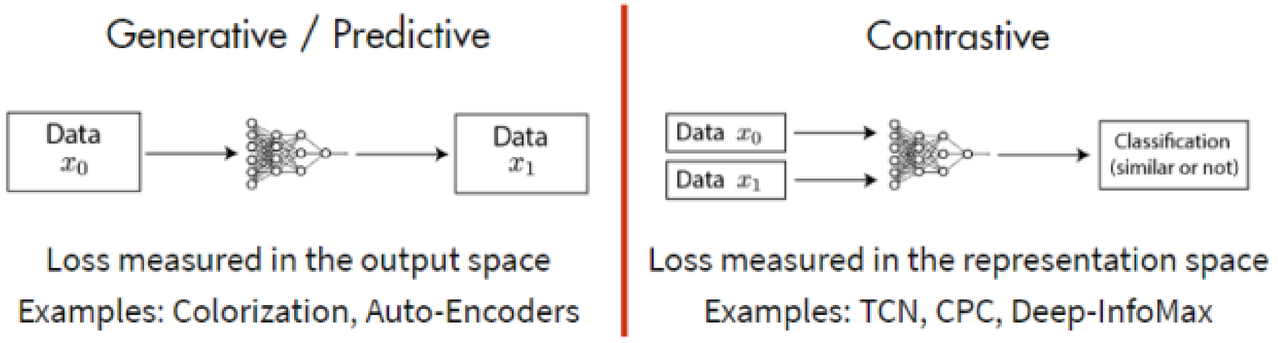
S^3-Rec:

- Self-Supervised Learning for Sequential Recommendation with Mutual Information Maximization, CIKM 2020
  - Self-Supervised – Learning from intrinsic structure of raw data

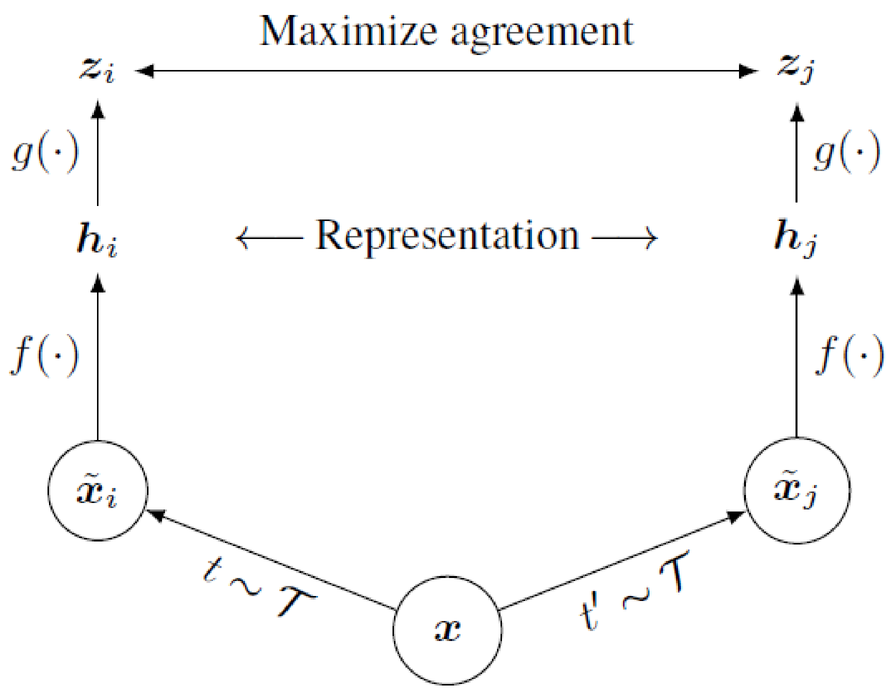
- Contrastive Learning
  - A Simple Framework for Contrastive Learning of Visual Representations, ICML 2020
  - Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere, ICML 2020

Contrastive Learning

- Self-Supervised



- SimCLR



Contrastive learning

(a) Original (b) Crop and resize (c) Crop, resize (and flip) (d) Color distort. (drop) (e) Color distort. (jitter)

(f) Rotate $\{90°, 180°, 270°\}$ (g) Cutout (h) Gaussian noise (i) Gaussian blur (j) Sobel filtering

S^3-Rec:

- Objective - Mutual Information Maximization

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

  - Given two random variables X and Y, it can be understood as how much knowing X reduces the uncertainty in Y or vice versa.
  - Intractable – find a lower bound(InfoNCE)

- Base Model (Transformer?)
  - Embedding Layer
  - Self-Attention Block (Bidirectional)
  - Prediction Layer

- 4 task
  - Associated Attribute Prediction
  - Masked Item Prediction
  - Masked Attribute Prediction
  - Segment Prediction

S^3-Rec:

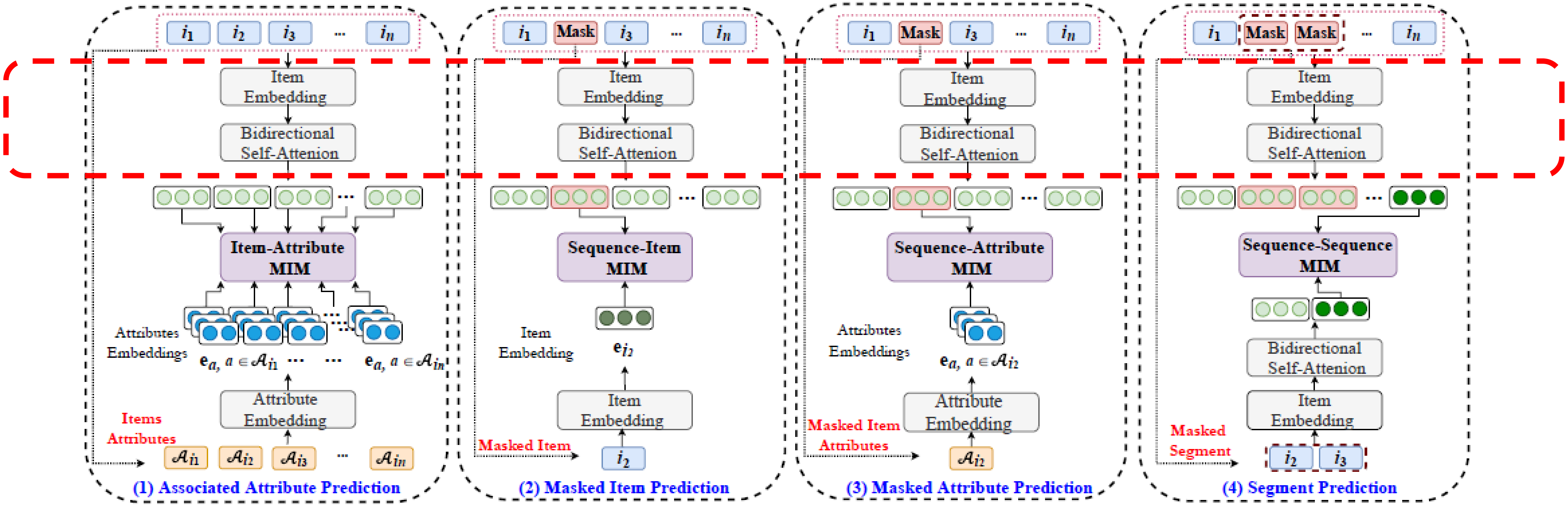


Figure 1: The overview of S^3-Rec in the pre-training stage. We assume that the user sequence is $\{i_1, \cdots, i_n\}$ and each item i is associated with several attributes $\mathcal{A}_i = \{a_1, \cdots, a_m\}$. We incorporate four self-supervised learning objectives: (1) Associated Attribute Prediction (AAP), (2) Masked Item Prediction (MIP), (3) Masked Attribute Prediction (MAP), and (4) Segment Prediction (SP). The embedding layers and bidirectional self-attention blocks are shared by the four pre-training objectives.

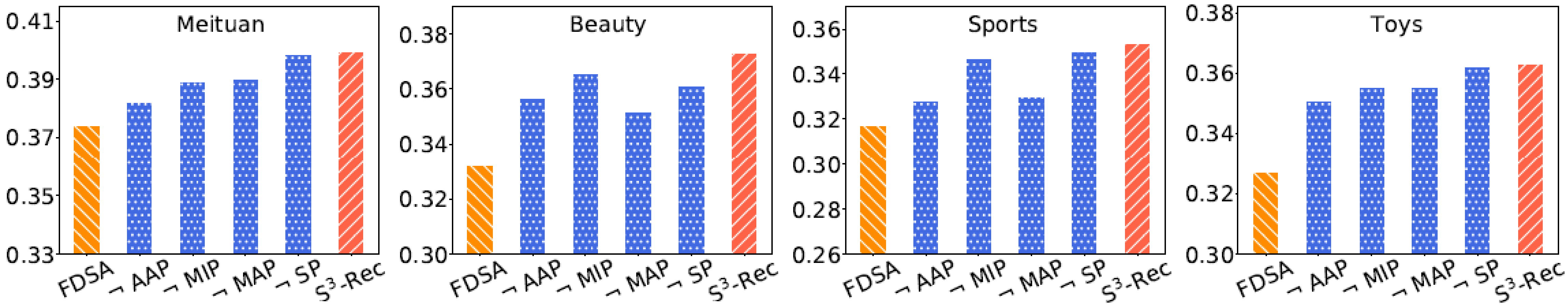
S^3-Rec:

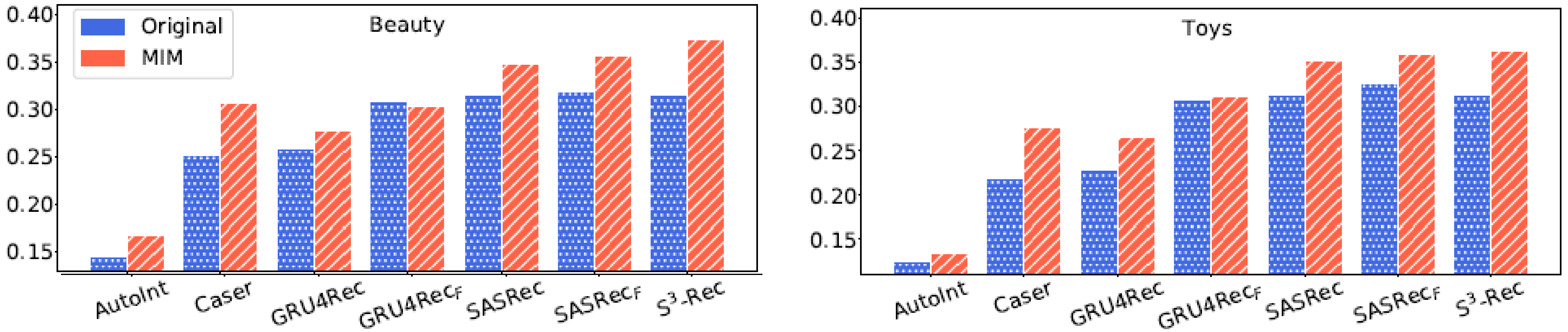
- Experiment
  - Dataset: Meituan(6 years in Beijing), Amazon, Yelp, LastFM
  - “we only keep the 5-core datasets, and filter unpopular items and inactive users with fewer than five interaction records”

Datasets	Metric	PopRec	FM	AutoInt	GRU4Rec	Caser	SASRec	BERT4Rec	HGN	GRU4Rec$_F$	SASRec$_F$	FDSA	S^3-Rec	Improv.
Meituan	HR@1	0.0946	0.1084	0.0804	0.1194	0.1368	0.1797	0.1381	0.1603	0.1436	0.1746	0.1778	**0.2040***	13.52%
	HR@5	0.2660	0.3218	0.2662	0.3382	0.3812	0.4524	0.3985	0.4110	0.3799	0.4386	0.4595	**0.4925***	7.18%
	NDCG@5	0.1813	0.2170	0.1739	0.2303	0.2619	0.3207	0.2713	0.2887	0.2639	0.3098	0.3236	**0.3527***	8.99%
	HR@10	0.3863	0.4709	0.4077	0.4881	0.5267	0.6053	0.5514	0.5573	0.5378	0.5962	0.6164	**0.6368***	3.31%
	NDCG@10	0.2200	0.2651	0.2194	0.2787	0.3090	0.3700	0.3208	0.3359	0.3149	0.3607	0.3743	**0.3994***	6.71%
	MRR	0.1923	0.2242	0.1854	0.2359	0.2617	0.3146	0.2689	0.2863	0.2666	0.3064	0.3167	**0.3421***	8.02%
Yelp	HR@1	0.0801	0.0624	0.0731	0.2053	0.2188	0.2375	0.2405	0.2428	0.2293	0.2301	0.2198	**0.2591***	6.71%
	HR@5	0.2415	0.2036	0.2249	0.5437	0.5111	0.5745	0.5976	0.5768	0.5858	0.5937	0.5728	**0.6085***	1.82%
	NDCG@5	0.1622	0.1333	0.1501	0.3784	0.3696	0.4113	0.4252	0.4162	0.4137	0.4178	0.4014	**0.4401***	3.50%
	HR@10	0.3609	0.3153	0.3367	0.7265	0.6661	0.7373	0.7597	0.7411	0.7574	0.7706	0.7555	**0.7725**	0.25%
	NDCG@10	0.2007	0.1692	0.1860	0.4375	0.4198	0.4642	0.4778	0.4695	0.4694	0.4751	0.4607	**0.4934***	3.26%
	MRR	0.1740	0.1470	0.1616	0.3630	0.3595	0.3927	0.4026	0.3988	0.3929	0.3962	0.3834	**0.4190***	4.07%

S^3-Rec

- Ablation study





Recent reading

- Deconstructing the Filter Bubble: User Decision-Making and Recommender Systems
- Debiasing Item-to-Item Recommendations With Small Annotated Datasets
  - ~~Few-shot~~ => In its design, IPS learning only requires a small annotated dataset
- Unbiased Ad Click Prediction for Position-aware Advertising Systems
  - Imputation model for unrevealed

- Bias from observation
  - Inverse Propensity Score(IPS) - $1/P(O=1|U,V)$
  - $P(C,U,V,O=1) = P(C|U,V,O=1)\,P(U,V,O=1)$

    $\qquad\qquad\qquad = P(C|U,V,O=1)\,P(U,V)\,P(O=1|U,V)$

  $P(O=1|U,V)$ is not uniform