Hierarchical Item Inconsistency Signal Learning for Sequence Denoising in Sequential Recommendation

Chi Zhang
Harbin Engineering University
Harbin, China
zhangchi20@hrbeu.edu.cn

Yantong Du
Harbin Engineering University
Harbin, China
duyantong94@hrbeu.edu.cn

Xiangyu Zhao*
City University of Hong Kong
Hong Kong
xianzhao@cityu.edu.hk

Qilong Han*
Harbin Engineering University
Harbin, China
hanqilong@hrbeu.edu.cn

Rui Chen*
Harbin Engineering University
Harbin, China
ruichen@hrbeu.edu.cn

Li Li
University of Delaware
Newark, United States
lilee@udel.edu

CIKM 2022
Code: https://github.com/zc-97/HSD.

Reported by liang li

Motivation

Motivation:

- Historical sequences typically contain some inherent noise (e.g., accidental interactions).

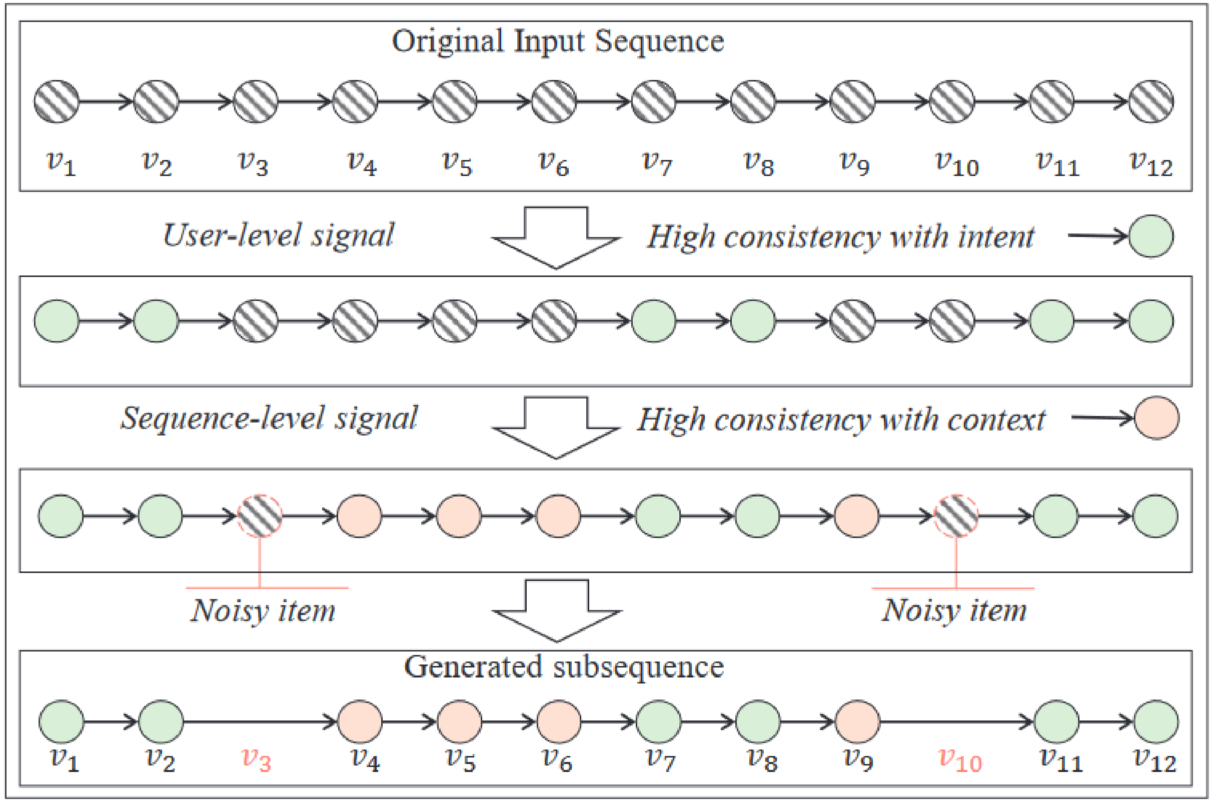

Figure 1: An illustration of hierarchical item inconsistency signals for sequence denoising.

Problem Statement

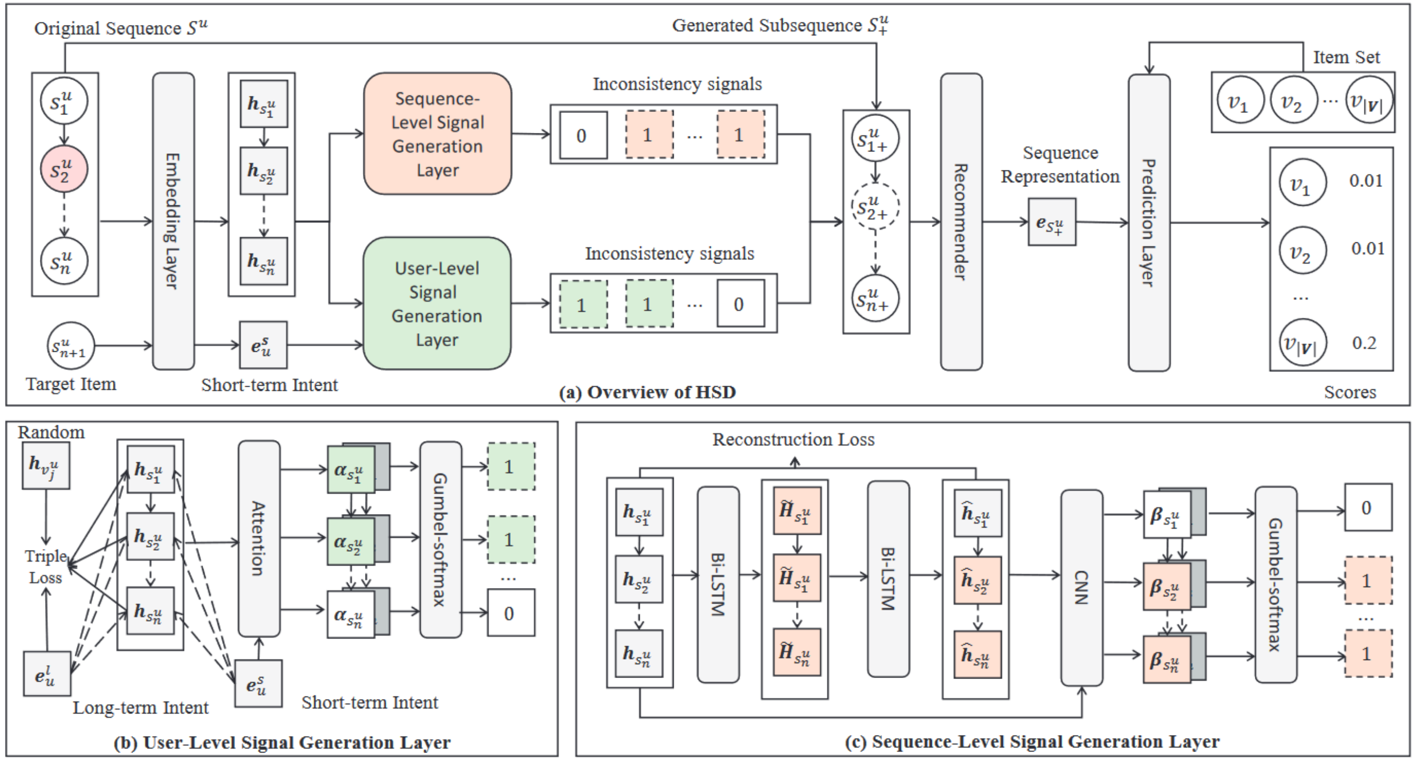

Figure 2: The overall architecture of the proposed HSD model.

$$\mathcal{S} = \{S^{u_1}, S^{u_2}, \cdots, S^{u_{|\mathcal{U}|}}\}$$

$$\mathcal{U} = \{u_1, u_2, \cdots, u_{|\mathcal{U}|}\}$$

$$\mathcal{V} = \{v_1, v_2, \cdots, v_{|\mathcal{V}|}\}$$

$$S^u = [s_1^u, s_2^u, \cdots, s_n^u]$$

$$S_+^u = [s_{1+}^u, s_{2+}^u, \cdots, s_{n+}^u]$$

$$|S_+^u| \leq |S^u|$$

Method

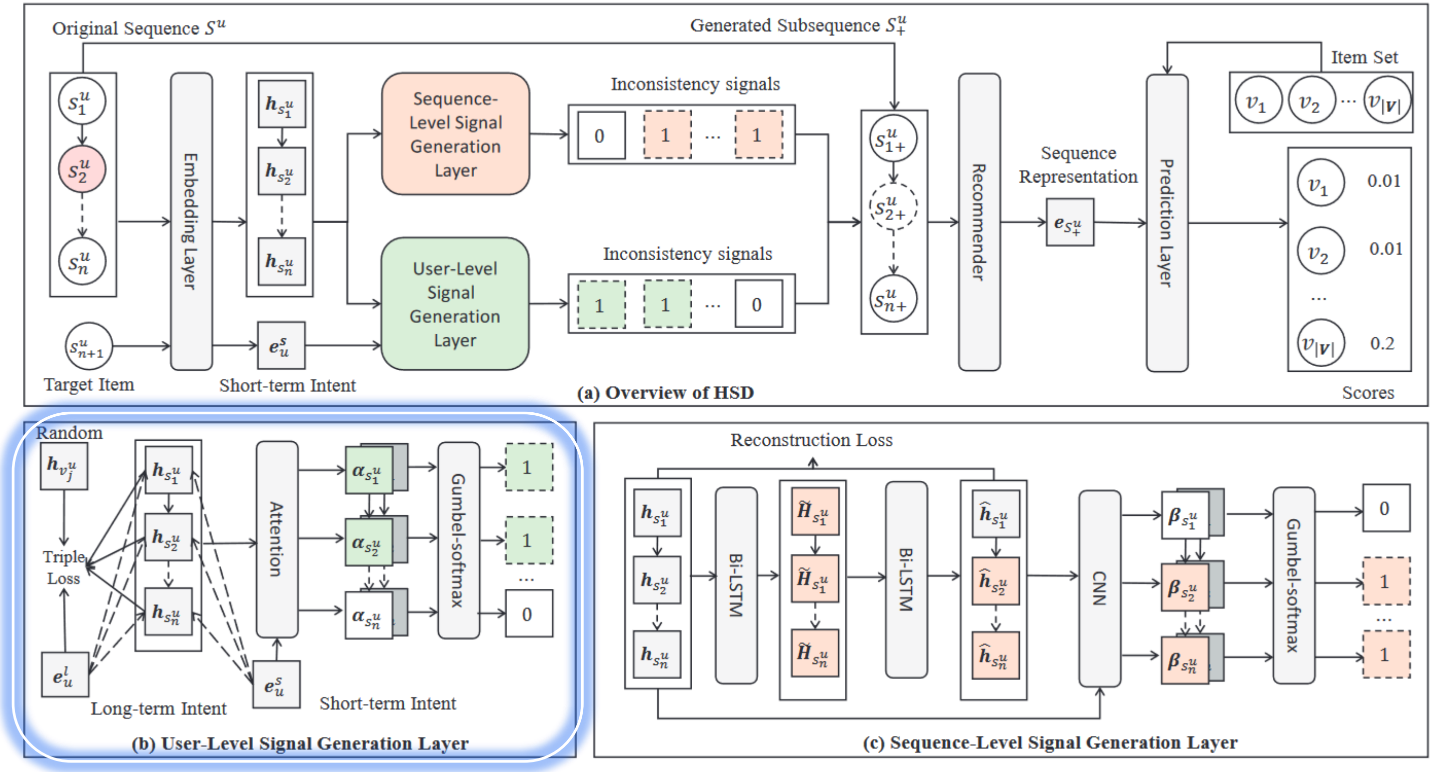

Figure 2: The overall architecture of the proposed HSD model.

$$\boldsymbol{h}_{v_i} = \boldsymbol{x}_{v_i}\boldsymbol{W}_h, \tag{1}$$

$$\boldsymbol{e}_u^s = \boldsymbol{h}_{s_{n+1}^u}, \boldsymbol{e}_u^l = \boldsymbol{h}_u, \tag{2}$$

$$\boldsymbol{\alpha}_{s_i^u} = \sigma(\tanh(\boldsymbol{h}_{s_i^u}\boldsymbol{W}_1 + \boldsymbol{e}_u^s\boldsymbol{W}_2)\boldsymbol{W}_3), \tag{3}$$

$$\boldsymbol{a}_{s_i^u} = \text{Gumbel-softmax}(\boldsymbol{\alpha}_{s_i^u}, \tau),$$
$$\boldsymbol{a}_{s_i^u} = \frac{\exp(\log(\boldsymbol{\alpha}_{s_i^u,j}) + g_j)/\tau}{\sum_{j=0}^{1}\exp(\log(\boldsymbol{\alpha}_{s_i^u,j}) + g_j)/\tau}, \tag{4}$$

$$\mathcal{L}_u = \frac{1}{n}\sum_{s_i^u \in S^u}\mathcal{L}(u, s_i^u, v_j^u),$$
$$\mathcal{L}(u, s_i^u, v_j^u) = -\log(\sigma(\text{sim}(\boldsymbol{e}_u^l, \boldsymbol{h}_{s_i^u}) - \text{sim}(\boldsymbol{e}_u^l, \boldsymbol{h}_{v_j^u}))), \tag{5}$$

Method

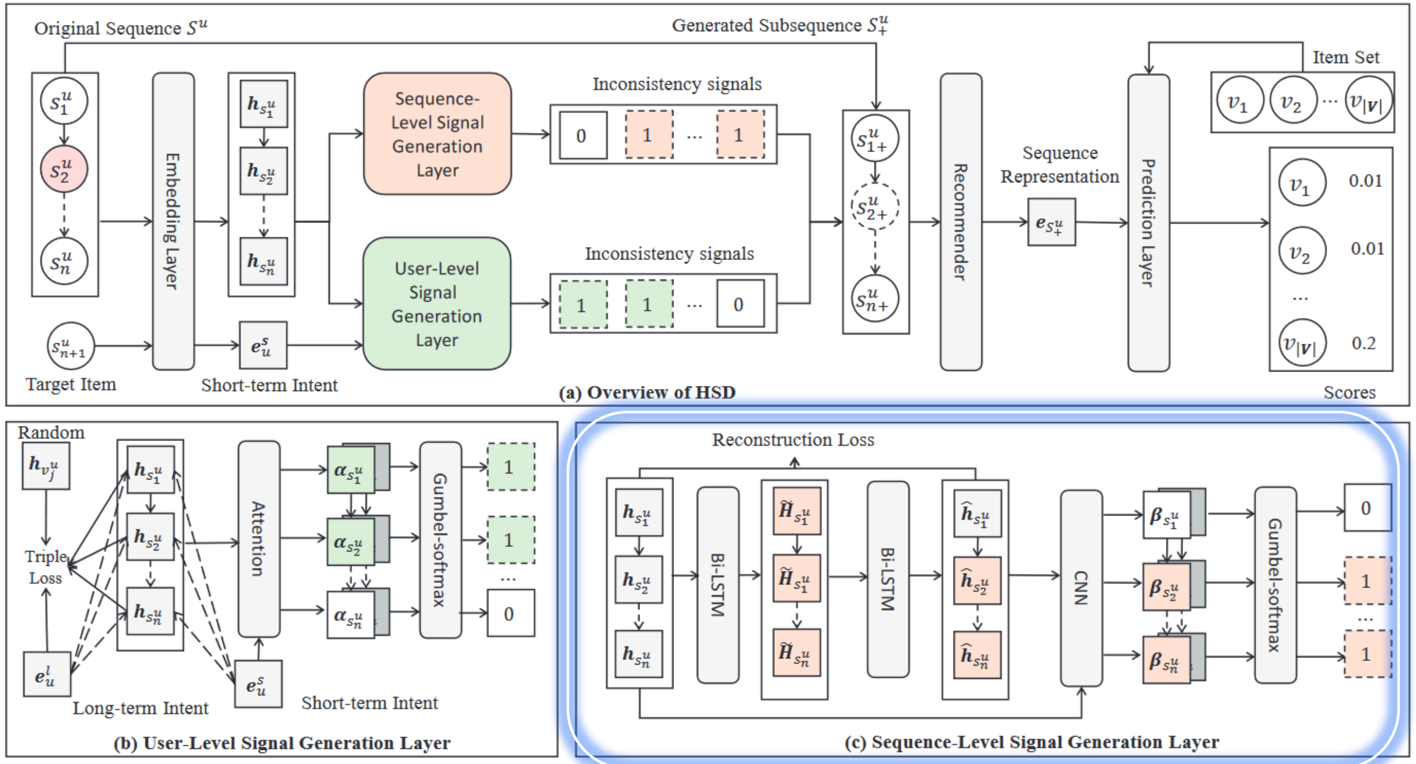

Figure 2: The overall architecture of the proposed HSD model.

$$\tilde{H}^L_{S^u}, \tilde{H}^R_{S^u} = \text{Bi-LSTM}(H_{S^u}, \Theta_1),$$
$$\tilde{H}_{S^u} = \tilde{H}^L_{S^u} + \tilde{H}^R_{S^u},$$
$$\hat{H}^L_{S^u}, \hat{H}^R_{S^u} = \text{Bi-LSTM}(\tilde{H}_{S^u}, \Theta_1),$$
$$\hat{H}_{S^u} = \hat{H}^L_{S^u} + \hat{H}^R_{S^u}, \tag{6}$$

$$\mathcal{L}_{S^u} = \frac{1}{n}\sum_{j=1}^{n}(\hat{h}_{s^u_i} - h_{s^u_i})^2. \tag{7}$$

$$c_{s^u_i} = \text{Conv}^i([\hat{h}_{s^u_i} \| h_{s^u_i}], \Theta^i_2),$$
$$\beta_{s^u_i} = \sigma(c_{s^u_i} W_4),$$
$$b_{s^u_i} = \text{Gumbel-softmax}(\beta_{s^u_i}, \tau), \tag{8}$$

$$b_{s^u_i} = \frac{\exp(\log(\beta_{s^u_i,j}) + g_j)/\tau}{\sum_{j=0}^{1}\exp(\log(\beta_{s^u_i,j}) + g_j)/\tau}, \tag{9}$$

Method

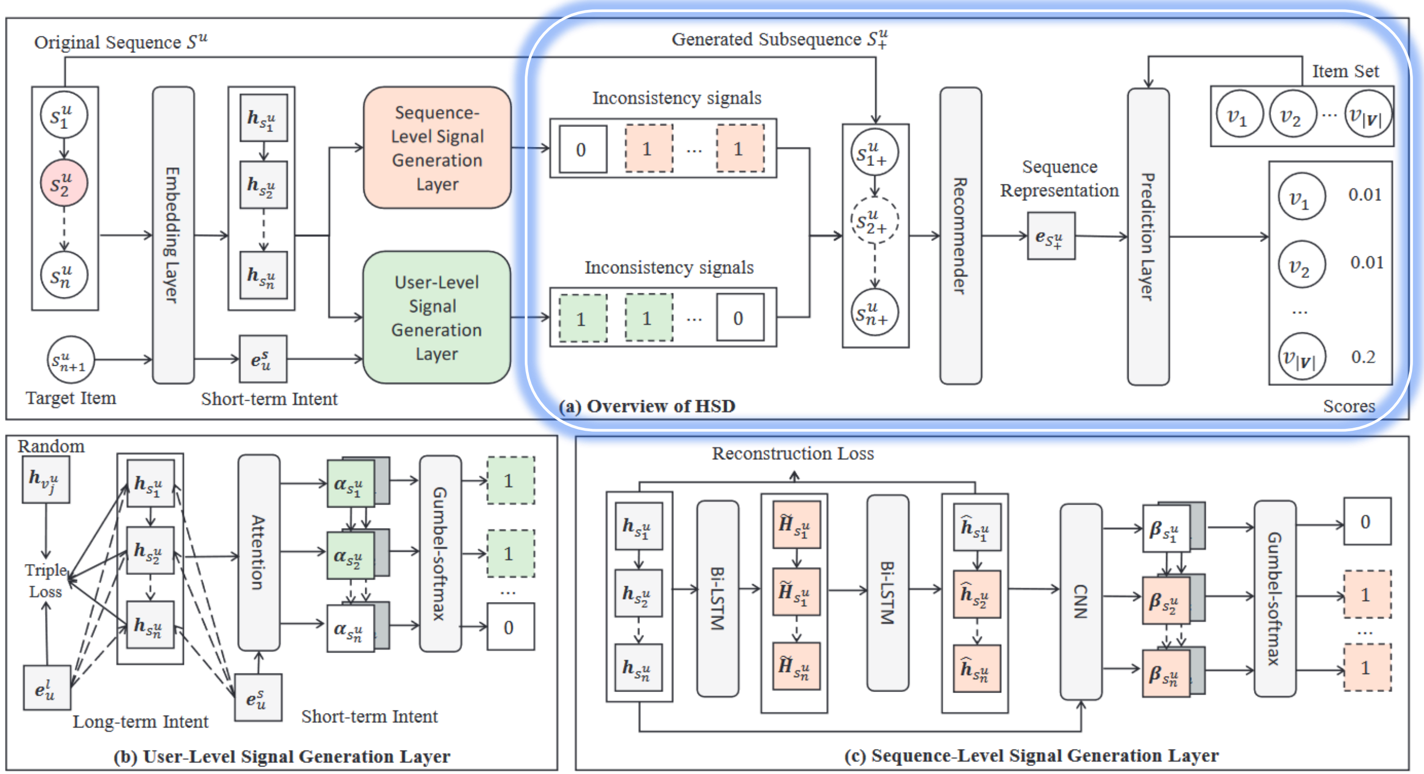

Figure 2: The overall architecture of the proposed HSD model.

$$r_{s_i^u} = a_{s_i^u} b_{s_i^u},$$
$$H_+^u = [r_{s_1^u} \boldsymbol{h}_{s_1^u}, r_{s_2^u} \boldsymbol{h}_{s_2^u}, \cdots, r_{s_n^u} \boldsymbol{h}_{s_n^u}], \qquad (10)$$

$$\boldsymbol{e}_{S_+^u} = \mathrm{F}(H_+^u), \qquad (11)$$

$$z_i = \boldsymbol{e}_{S_+^u} \boldsymbol{h}_{v_i}^{\top}. \qquad (12)$$

$$\hat{y}_i = \frac{\exp(z_i)}{\sum_{v_j \in \mathcal{V}} \exp(z_j)}. \qquad (13)$$

$$\mathcal{L}(\mathrm{y}, \hat{\mathrm{y}}) = -\sum_{i=1}^{|\mathcal{V}|} \left(y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right), \qquad (14)$$

$$Loss = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \left(\mathcal{L}(\mathrm{y}, \hat{\mathrm{y}}) + \mathcal{L}_u + \mathcal{L}_{S^u}\right) + \lambda \|\Theta\|_2^2, \qquad (15)$$

Experiments

Table 1: Statistics of the datasets

Dataset	# Sequences	# users	# Items	# Sparsity
ML-100k	99,287	944	1,350	92.21%
ML-1M	999,611	6,041	3,417	95.16%
Beauty	198,502	22,364	12,102	99.93%
Sports	296,337	35,599	18,358	99.95%
Yelp	317,078	30,495	20,062	99.95%

Experiments

Table 2: Experimental results of different sequential recommendation methods with (w) or without (w/o) HSD on the five datasets. The best results are boldfaced, and Imprv indicates the average improvements over all metrics. All improvements are statistically significant (i.e., two-sided t-tests with $p < 0.05$).

Datasets	Metric	GRU4Rec		NARM		STAMP		Caser		SASRec		BERT4Rec	
		w/o	w	w/o	w	w/o	w	w/o	w	w/o	w	w/o	w
ML-100k	HR@5	0.0191	**0.0276**	0.0180	**0.0223**	0.0201	**0.0255**	0.0212	**0.0255**	0.0191	**0.0371**	0.0191	**0.0339**
	HR@10	0.0286	**0.0488**	0.0403	**0.0403**	0.0392	**0.0467**	0.0339	**0.0477**	0.0371	**0.0583**	0.0414	**0.0732**
	HR@20	0.0594	**0.0742**	0.0657	**0.0785**	0.0700	**0.0870**	0.0679	**0.0742**	0.0764	**0.1018**	0.0912	**0.1294**
	N@5	0.0104	**0.0173**	0.0132	**0.0144**	0.0115	**0.0151**	0.0113	**0.0154**	0.0114	**0.0225**	0.0117	**0.0178**
	N@10	0.0134	**0.0241**	**0.0202**	0.0201	0.0176	**0.0217**	0.0153	**0.0224**	0.0172	**0.0292**	0.0189	**0.0305**
	N@20	0.0212	**0.0306**	0.0267	**0.0299**	0.0253	**0.0320**	0.0238	**0.0291**	0.0270	**0.0401**	0.0315	**0.0447**
	MRR	0.0109	**0.0185**	0.0162	**0.0169**	0.0132	**0.0171**	0.0119	**0.0168**	0.0139	**0.0234**	0.0157	**0.0218**
	Imprv.		**63.91%**		**7.82%**		**26.78%**		**35.74%**		**71.88%**		**52.68%**
ML-1M	HR@5	0.0194	**0.0260**	0.0151	**0.0260**	0.0232	**0.0301**	0.0104	**0.0252**	0.0397	**0.0500**	0.0224	**0.0477**
	HR@10	0.0373	**0.0427**	0.0349	**0.0417**	0.0440	**0.0530**	0.0215	**0.0424**	0.0666	**0.0858**	0.0495	**0.0886**
	HR@20	0.0690	**0.0745**	0.0591	**0.0677**	0.0677	**0.0899**	0.0589	**0.0725**	0.1007	**0.1326**	0.0980	**0.1399**
	N@5	0.0135	**0.0152**	0.0080	**0.0162**	0.0150	**0.0190**	0.0063	**0.0161**	0.0207	**0.0282**	0.0132	**0.0297**
	N@10	0.0190	**0.0206**	0.0144	**0.0212**	0.0218	**0.0264**	0.0099	**0.0216**	0.0294	**0.0397**	0.0218	**0.0429**
	N@20	0.0270	**0.0286**	0.0205	**0.0277**	0.0278	**0.0357**	0.0194	**0.0291**	0.0379	**0.0515**	0.0339	**0.0558**
	MRR	0.0159	**0.0161**	0.0100	**0.0168**	0.0168	**0.0209**	0.0091	**0.0173**	0.0203	**0.0290**	0.0169	**0.0328**
	Imprv.		**9.68%**		**62.42%**		**25.63%**		**108.61%**		**36.26%**		**95.82%**
Beauty	HR@5	0.0077	**0.0131**	0.0120	**0.0187**	0.0080	**0.0201**	0.0072	**0.0106**	0.0242	**0.0267**	0.0060	**0.0261**
	HR@10	0.0135	**0.0229**	0.0209	**0.0311**	0.0135	**0.0325**	0.0133	**0.0183**	0.0386	**0.0449**	0.0127	**0.0447**
	HR@20	0.0256	**0.0378**	0.0367	**0.0495**	0.0231	**0.0505**	0.0235	**0.0290**	0.0561	**0.0674**	0.0204	**0.0683**
	N@5	0.0045	**0.0082**	0.0071	**0.0114**	0.0046	**0.0128**	0.0044	**0.0064**	0.0129	**0.0147**	0.0037	**0.0147**
	N@10	0.0064	**0.0113**	0.0099	**0.0153**	0.0064	**0.0168**	0.0064	**0.0089**	0.0175	**0.0205**	0.0059	**0.0207**
	N@20	0.0094	**0.0150**	0.0139	**0.0199**	0.0088	**0.0213**	0.0090	**0.0116**	0.0219	**0.0261**	0.0078	**0.0266**
	MRR	0.0051	**0.0088**	0.0077	**0.0118**	0.0049	**0.0133**	0.0051	**0.0068**	0.0122	**0.0146**	0.0044	**0.0151**
	Imprv.		**71.71%**		**52.54%**		**161.44%**		**38.06%**		**17.05%**		**262.99%**
Sports	HR@5	0.0064	**0.0097**	0.0090	**0.0096**	0.0071	**0.0097**	0.0069	**0.0077**	0.0113	**0.0131**	0.0055	**0.0135**
	HR@10	0.0114	**0.0152**	0.0138	**0.0169**	0.0123	**0.0155**	0.0115	**0.0129**	0.0175	**0.0224**	0.0104	**0.0224**
	HR@20	0.0183	**0.0230**	0.0223	**0.0282**	0.0182	**0.0252**	0.0178	**0.0214**	0.0268	**0.0348**	0.0167	**0.0365**
	N@5	0.0035	**0.0067**	0.0058	**0.0063**	0.0046	**0.0065**	0.0046	**0.0049**	0.0059	**0.0071**	0.0036	**0.0072**
	N@10	0.0051	**0.0085**	0.0073	**0.0087**	0.0062	**0.0084**	0.0061	**0.0065**	0.0079	**0.0100**	0.0051	**0.0101**
	N@20	0.0068	**0.0105**	0.0094	**0.0115**	0.0077	**0.0108**	0.0077	**0.0086**	0.0102	**0.0132**	0.0067	**0.0137**
	MRR	0.0036	**0.0070**	0.0059	**0.0070**	0.0048	**0.0069**	0.0049	**0.0052**	0.0055	**0.0071**	0.0040	**0.0073**
	Imprv.		**71.53%**		**17.12%**		**39.05%**		**8.85%**		**25.80%**		**102.39%**
yelp	HR@5	0.0057	**0.0104**	0.0113	**0.0180**	0.0060	**0.0147**	0.0045	**0.0169**	0.0293	**0.0334**	0.0087	**0.0292**
	HR@10	0.0102	**0.0180**	0.0187	**0.0248**	0.0099	**0.0216**	0.0084	**0.0198**	0.0352	**0.0446**	0.0159	**0.0408**
	HR@20	0.0184	**0.0317**	0.0315	**0.0349**	0.0161	**0.0345**	0.0146	**0.0251**	0.0439	**0.0613**	0.0273	**0.0593**
	N@5	0.0034	**0.0066**	0.0075	**0.0142**	0.0038	**0.0107**	0.0028	**0.0151**	0.0251	**0.0255**	0.0054	**0.0223**
	N@10	0.0048	**0.0090**	0.0099	**0.0164**	0.0051	**0.0129**	0.0040	**0.0160**	0.0270	**0.0291**	0.0077	**0.0260**
	N@20	0.0068	**0.0124**	0.0131	**0.0189**	0.0066	**0.0162**	0.0055	**0.0173**	0.0292	**0.0333**	0.0105	**0.0307**
	MRR	0.0037	**0.0072**	0.0081	**0.0145**	0.0040	**0.0112**	0.0031	**0.0152**	0.0250	**0.0255**	0.0060	**0.0228**
	Imprv.		**87.83%**		**64.52%**		**158.53%**		**315.36%**		**11.33%**		**246.14%**

Experiments

Table 3: The experimental comparison between HSD with the best/worst base model and the state-of-the-art denoising methods on the five datasets. The best results are boldfaced, the second-best results are underlined. Imprv indicates the average improvements between the best and worst ones over the baselines. All improvements are statistically significant (i.e., two-sided t-test with $p < 0.05$) over the best baselines.

Dataset	Model	HR@5	HR@10	HR@20	N@5	N@10	N@20	MRR
ML-100k	DSAN (*AAAI'21*)	0.0201	0.0435	0.0700	0.0115	0.0188	0.0254	0.0133
	FMLP-Rec (*WWW'22*)	0.0170	0.0477	0.0764	0.0117	0.0216	0.0288	0.0160
	HSD+NARM	0.0223	0.0403	0.0785	0.0144	0.0201	0.0299	0.0169
	HSD+BERT4Rec	**0.0339**	**0.0732**	**0.1294**	**0.0178**	**0.0305**	**0.0447**	**0.0218**
	Imprv.	40%±29%	19%±34%	36%±33%	38%±15%	17%±24%	30%±26%	21%±15%
ML-1M	DSAN	0.0098	0.0336	0.0651	0.0048	0.0122	0.0200	0.0081
	FMLP-Rec	0.0210	0.0449	0.0707	0.0120	0.0199	0.0263	0.0142
	HSD+NARM	0.0260	0.0417	0.0677	0.0162	0.0212	0.0277	0.0168
	HSD+BERT4Rec	**0.0477**	**0.0886**	**0.1399**	**0.0297**	**0.0429**	**0.0558**	**0.0328**
	Imprv.	75%±52%	45%±52%	47%±51%	91%±56%	61%±55%	59%±53%	75%±56%
Beauty	DSAN	0.0092	0.0152	0.0264	0.0058	0.0077	0.0105	0.0062
	FMLP-Rec	0.0095	0.0166	0.0284	0.0056	0.0078	0.0107	0.0060
	HSD+Caser	0.0106	0.0183	0.0290	0.0064	0.0089	0.0116	0.0068
	HSD+BERT4Rec	**0.0261**	**0.0447**	**0.0683**	**0.0147**	**0.0207**	**0.0266**	**0.0151**
	Imprv.	93%±82%	90%±80%	71%±69%	82%±72%	90%±76%	79%±70%	77%±67%
Sports	DSAN	0.0061	0.0105	0.0215	0.0042	0.0056	0.0084	0.0049
	FMLP-Rec	0.0068	0.0117	0.0180	0.0044	0.0059	0.0075	0.0046
	HSD+Caser	0.0077	0.0129	0.0214	0.0049	0.0065	0.0086	0.0052
	HSD+BERT4Rec	**0.0120**	**0.0190**	**0.0303**	**0.0078**	**0.0100**	**0.0129**	**0.0081**
	Imprv.	45%±32%	36%±26%	20%±21%	44%±33%	40%±30%	28%±26%	36%±30%
Yelp	DSAN	0.0269	0.0369	0.0541	0.0211	0.0242	0.0285	0.0216
	FMLP-Rec	0.0203	0.0294	0.0436	0.0142	0.0171	0.0207	0.0144
	HSD+GRU4Rec	0.0104	0.0180	0.0317	0.0066	0.0090	0.0124	0.0072
	HSD+BERT4Rec	**0.0292**	**0.0408**	**0.0593**	**0.0223**	**0.0260**	**0.0307**	**0.0228**
	Imprv.	-26%±35%	-20%±31%	-16%±26%	-32%±37%	-28%±35%	-24%±32%	-31%±36%

Experiments

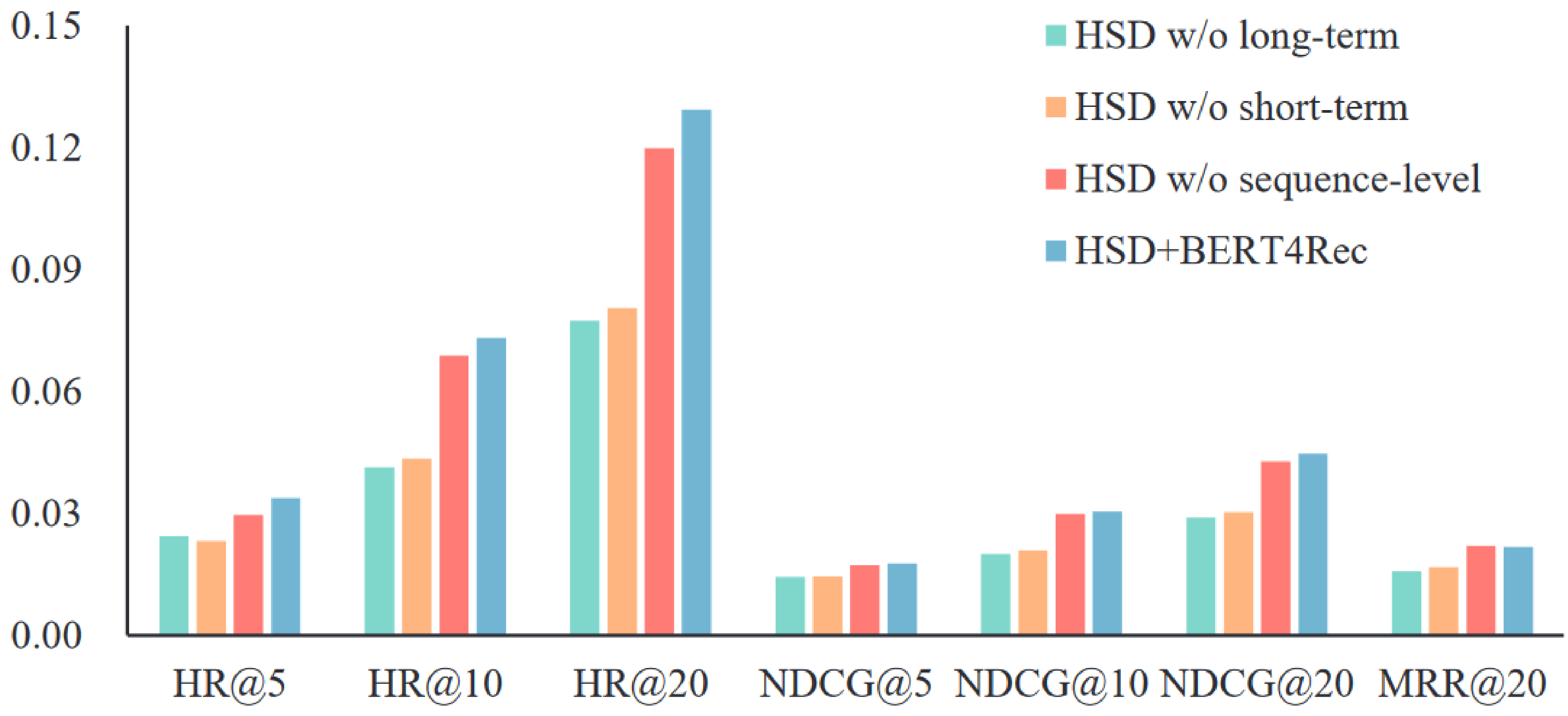

Figure 3: Impact of different inconsistency signal generation layers.

Experiments

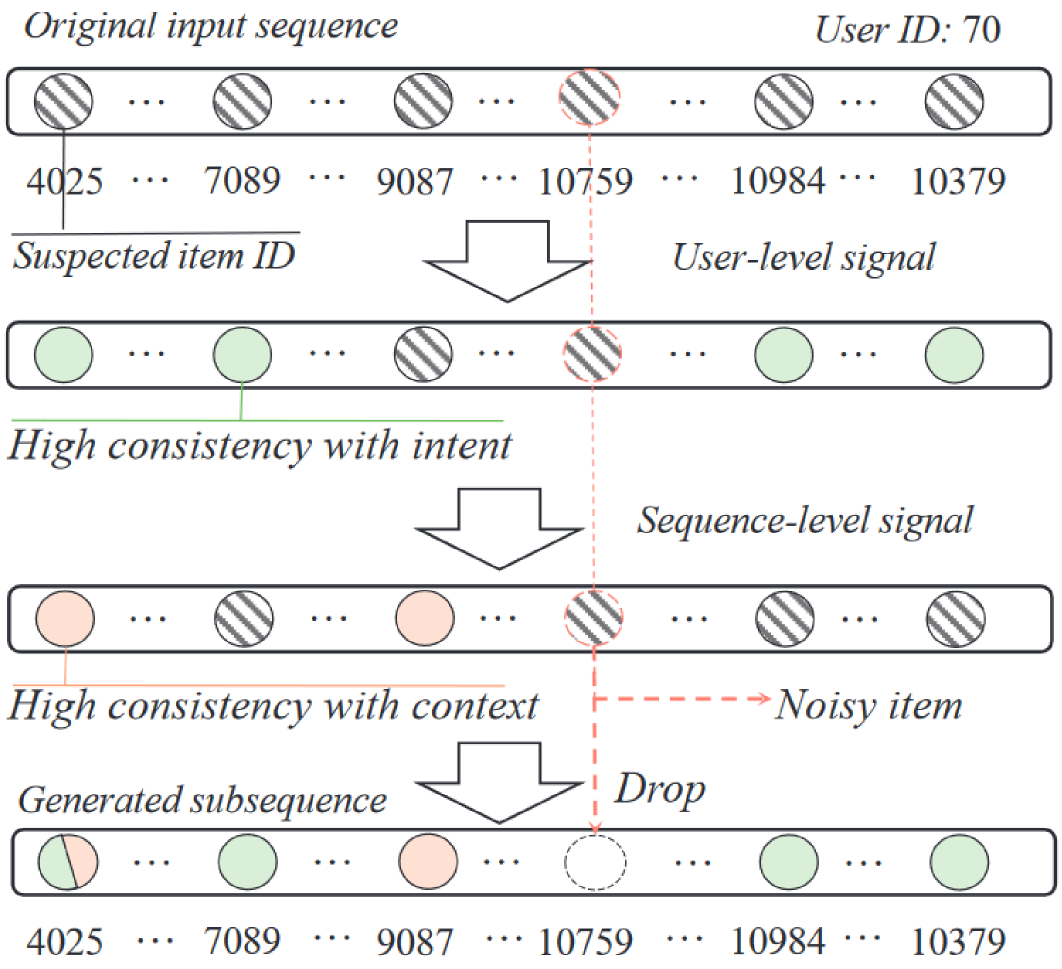

Figure 4: A case study to show how different inconsistency signals affect the denoising process.

Thanks