CURE4Rec: A Benchmark for Recommendation Unlearning with Deeper Influence

Chaochao Chen[1], Jiaming Zhang[1], Yizhao Zhang[1], Li Zhang[1], Lingjuan Lyu[2], Yuyuan Li[3,1*], Biao Gong[3], Chenggang Yan[3]

Zhejiang University[1], Sony AI[2], Hangzhou Dianzi University[3]

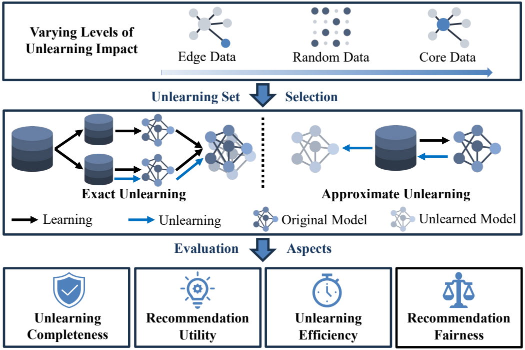
Table of Contents

Recommendation Unlearning

Machine unlearning aims to eliminate the memory of specific data, serving purposes such as privacy protection and erasing data biases. Recommendation unlearning aims to eliminate the influence of target data within the recommender system.

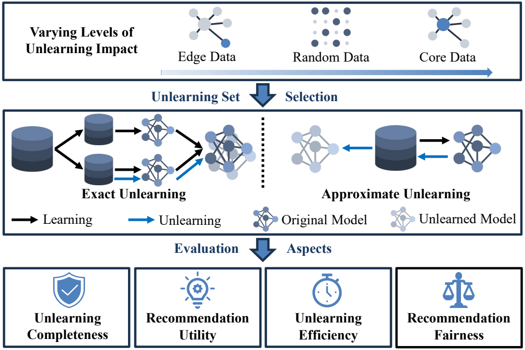
- *Exact Unlearning (EU)* aims to completely eliminate the influence of target data on the model.
- *Approximate Unlearning (AU)* achieves unlearning through direct parameter manipulation, avoiding the significant time cost of retraining.

Overview



CURE4Rec: a comprehensive benchmark specifically designed to evaluate recommendation unlearning methods.

Unlearning Sets Selection Strategies



Based on the level of collaboration, we select core data, edge data, and random data to construct unlearning sets respectively.

Unlearning Evaluation Aspects

1. **Unlearning Completeness:** To what extent does the unleaning method achieve authorized unlearning?
2. **Recommendation Utility:** How much does unlearning cause harm to the knowledge acquired from the remaining data?
3. **Unlearning Efficiency:** How fast is the time required for unlearning?
4. **Recommendation Fairness:** How significantly does unlearning affect fairness?

Overview

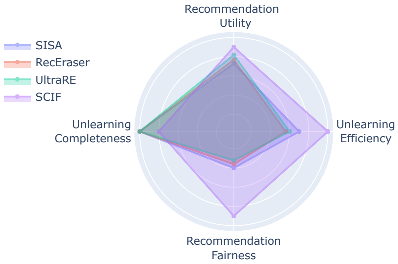


Figure: A visualized evaluation overview of recommendation unlearning methods in four aspects (↑), where the result is the normalized average outcome obtained across all models and datasets.

Results for Approximate Unlearning

		ML-100K				ML-1M				ADM			
		NDCG@20	HR@20	MIO	A-IGF	NDCG@20	HR@20	MIO	A-IGF	NDCG@20	HR@20	MIO	A-IGF
Learn		0.3215	0.3415	0.722	-0.0450	0.2144	0.2112	0.741	-0.042	0.0277	0.0578	0.756	0.0167
Retrain	Core	0.3187	0.3295	0.540	-0.0184	0.2196	0.2174	0.544	-0.0188	0.0221	0.0446	0.555	0.0053
	Random	0.2872	0.3353	0.538	-0.0403	0.2124	0.2108	0.547	-0.0507	0.0252	0.0519	0.556	0.0141
	Edge	0.3091	0.3140	0.536	-0.0430	0.2148	0.2051	0.546	-0.0518	0.0272	0.0554	0.556	0.0164
SCIF	Core	0.2483	0.2382	0.561	-0.0322	0.1865	0.1629	0.569	-0.0213	0.0194	0.0398	0.571	0.0094
	Random	0.2699	0.2617	0.563	-0.0268	0.1922	0.1785	0.571	-0.0311	0.0227	0.0461	0.575	0.0106
	Edge	0.2894	0.3012	0.601	-0.0375	0.2031	0.1811	0.623	-0.0191	0.0245	0.0502	0.579	0.0103

Table: Results in terms of unlearning completeness (MIO accuracy - approaching 0.5), recommendation utility (NDCG and HR ↑), and recommendation fairness (A-IGF - approaching Retrain) for the approximate recommendation unlearning method, where `Learn` denotes the results before unlearning. Core, random, and edge respectively refer to the selection of the unlearning sets as core data, random data, and edge data.

Recommendation Utility(Exact Unlearning)

ML-100K		Retrain			SISA			RecEraser			UltraRE		
		Core	Random	Edge	Core	Random	Edge	Core	Random	Edge	Core	Random	Edge
WMF	NDCG@20	0.3187	0.2872	0.3091	0.2096	0.2092	0.2041	0.2285	0.2208	0.2109	0.2303	0.2354	0.2149
	HR@20	0.3295	0.3353	0.3140	0.2094	0.2049	0.1892	0.2218	0.2142	0.1979	0.2267	0.2282	0.2027
BPR	NDCG@20	0.3111	0.3003	0.3043	0.2244	0.2324	0.2298	0.2614	0.2615	0.2694	0.2708	0.2764	0.2743
	HR@20	0.3151	0.3028	0.2987	0.2203	0.2259	0.2179	0.2724	0.2658	0.2620	0.2851	0.2813	0.2695
LightGCN	NDCG@20	0.3175	0.3121	0.3101	0.1802	0.1932	0.1964	0.2856	0.2905	0.2886	0.2952	0.3069	0.3063
	HR@20	0.3250	0.3253	0.3244	0.1724	0.1907	0.1911	0.3053	0.3099	0.3121	0.3123	0.3201	0.3185

ML-1M		Retrain			SISA			RecEraser			UltraRE		
		Core	Random	Edge	Core	Random	Edge	Core	Random	Edge	Core	Random	Edge
WMF	NDCG@20	0.2196	0.2124	0.2148	0.1780	0.1639	0.1714	0.1894	0.1796	0.1838	0.1926	0.1891	0.1970
	HR@20	0.2174	0.2108	0.2051	0.1612	0.1485	0.1493	0.1731	0.1592	0.1596	0.1747	0.1680	0.1717
BPR	NDCG@20	0.2462	0.2319	0.2336	0.1545	0.1530	0.1628	0.1826	0.1660	0.1860	0.1828	0.1856	0.1913
	HR@20	0.2279	0.2162	0.2118	0.1353	0.1329	0.1367	0.1627	0.1450	0.1624	0.1652	0.1632	0.1651
LightGCN	NDCG@20	0.2177	0.2108	0.2147	0.1504	0.1533	0.1642	0.1864	0.1863	0.1814	0.1969	0.1867	0.1806
	HR@20	0.2138	0.2045	0.2186	0.1365	0.1323	0.1581	0.1825	0.1804	0.1818	0.1907	0.1855	0.1798

ADM		Retrain			SISA			RecEraser			UltraRE		
		Core	Random	Edge	Core	Random	Edge	Core	Random	Edge	Core	Random	Edge
WMF	NDCG@20	0.3691	0.3566	0.3556	0.2720	0.2589	0.2515	0.3373	0.3256	0.3185	0.3420	0.3334	0.3347
	HR@20	0.4071	0.3822	0.3848	0.2617	0.2492	0.2471	0.3527	0.3467	0.3203	0.3689	0.3595	0.3501
BPR	NDCG@20	0.3566	0.3453	0.3499	0.2806	0.2708	0.2757	0.3286	0.3295	0.3212	0.3325	0.3301	0.3314
	HR@20	0.3821	0.3628	0.3718	0.2745	0.2638	0.2611	0.3486	0.3406	0.3483	0.3541	0.3569	0.3608
LightGCN	NDCG@20	0.0105	0.0106	0.0096	0.0075	0.0054	0.0048	0.0084	0.0085	0.0079	0.0097	0.0088	0.0086
	HR@20	0.0221	0.0234	0.0208	0.0157	0.0112	0.0103	0.0171	0.0176	0.0154	0.0191	0.0185	0.0183

Unlearning Efficiency

Time (s)		ML-100K			ML-1M			ADM		
		WMF	BPR	LightGCN	WMF	BPR	LightGCN	WMF	BPR	LightGCN
Retrain	Core	4296	5238	4734	7748	9113	8645	3682	6998	5225
	Random	4526	5494	5044	8693	9461	10324	3972	7127	5354
	Edge	4687	5527	5274	8006	9748	10497	4127	7351	6359
SISA	Core	402	488	437	1160	1160	1523	669	1750	1009
	Random	467	586	528	1256	1265	1605	717	1842	1246
	Edge	442	504	515	1280	1292	1659	751	1902	1077
RecEraser	Core	463	582	561	1533	1568	1846	865	1892	1106
	Random	476	693	656	1654	1660	1952	912	1945	1490
	Edge	489	659	617	1736	1819	1964	965	2032	1190
UltraRE	Core	457	591	559	1507	1493	1667	819	1810	1057
	Random	482	618	645	1595	1550	1834	901	1862	1283
	Edge	466	518	666	1781	1791	1955	923	1904	1368
SCIF	Core	289	336	316	784	784	1034	453	1186	682
	Random	325	403	368	862	860	1083	497	1242	841
	Edge	316	358	359	887	877	1126	520	1282	733

Table: Results in terms of unlearning efficiency (running time in seconds $\downarrow$).

Recommendation Fairness

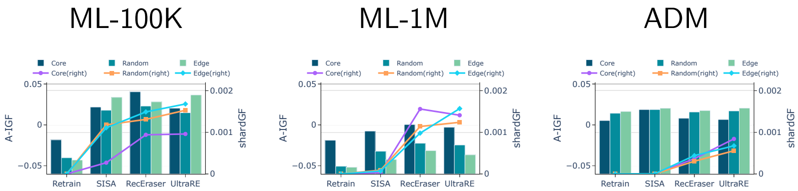


Figure: Results in terms of recommendation fairness for exact recommendation unlearning methods on WMF, where A-IGF (approaching Retrain) and shardGF (↓) evaluate the fairness of group-level and shard-level, respectively.

- The division-aggregation design of the EU approach achieves unlearning completeness while compromising other evaluation aspects.
- The AU approach outperforms the EU approach in all aspects except completeness, with less negative influence on model properties.
- Exact unlearning methods impact fairness of the original recommendation model.
- All unlearning methods exhibit significant differences in utility and efficiency across different unlearning sets.