How to Diversify any Personalized Recommender? A User-centric Pre-processing approach

Anonymous Author(s)*

ACM Reference Format:
Anonymous Author(s). 2024. How to Diversify any Personalized Recommender? A User-centric Pre-processing approach. In *RecSys '24: ACM*. ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 APPENDIX

1.1 Characteristics of the Pre-Processed Data

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference'24, September 2024, Washington, DC, USA
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

Table 1: Detailed statistics of the original data vs. pre-processed data with one-step and two-step variants. We have λ={0%, 1%, 2%, 5%, 10%}.

Data Sets		Pre-processing		#Users	#Click history/ #Interactions	Sparsity (%)	#Items	Features	
								#Categories	#Impressions
MIND News Subset	*Training Data*	***0***	None	1000	9368/39467	99.58	26740	17	7105
		1	1-Step	1000	9367/39467	99.58	26740	17	7105
			2-Step	1000	9364/39417	99.58	26740	17	7105
		2	1-Step	1000	9367/39766	99.58	26740	17	7105
			2-Step	1000	9364/39716	99.58	26740	17	7105
		5	1-Step	1000	9367/40383	99.57	26740	17	7105
			2-Step	1000	9365/40333	99.57	26740	17	7105
		10	1-Step	1000	9367/40387	99.57	26740	17	7105
			2-Step	1000	9364/40337	99.57	26740	17	7105
	Test Data	***None***	None	5000	15557	99.82	18723	16	7538
GoodBook Subset	*Training Data*	***0***	None	943	8477	98.77	729	31	-
		1	1-Step	943	8454	98.73	706	31	-
			2-Step	943	8354	98.75	706	31	-
		2	1-Step	943	8619	98.71	707	31	-
			2-Step	943	8499	98.73	707	31	-
		5	1-Step	943	9670	98.58	723	31	-
			2-Step	943	9290	98.64	723	31	-
		10	1-Step	943	11291	98.36	729	31	-
			2-Step	943	10271	98.51	729	31	-
	Test Data	***None***	None	943	4715	99.27	688	31	-

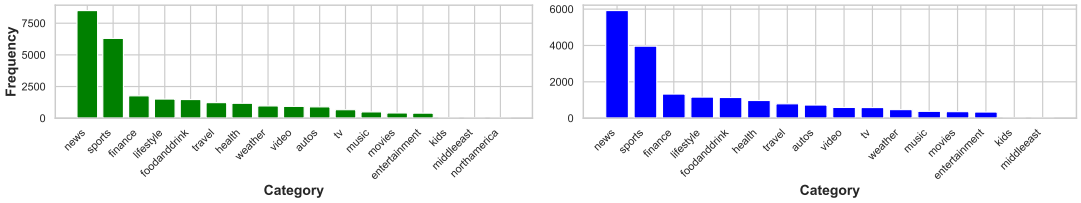


Figure 1: Categories of news in MIND Data. (Top) Categories of news in the train data. (Bottom) Categories of news in the test data.

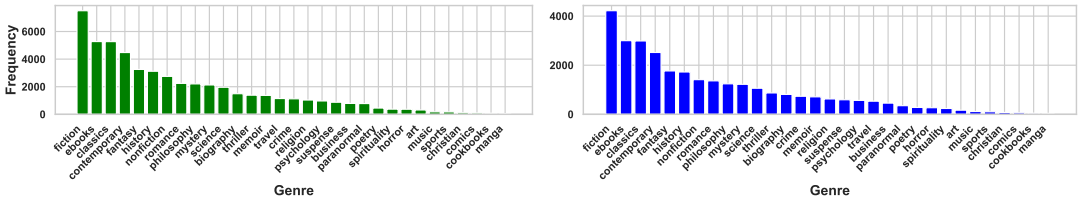
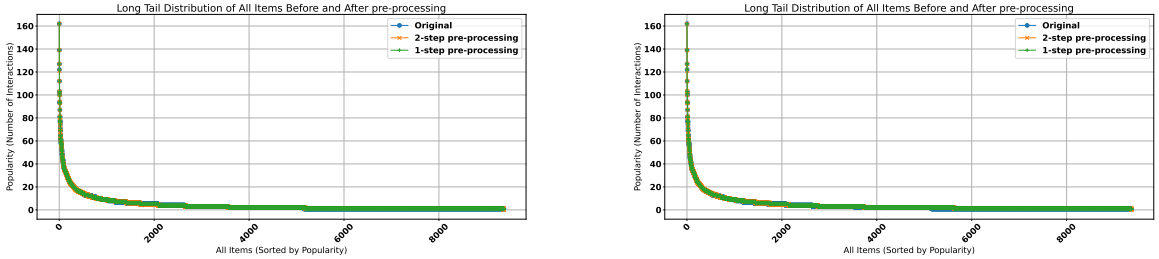
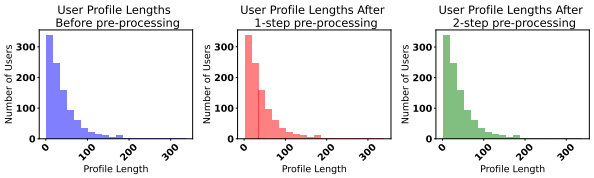


Figure 2: Genres of books in GoodBook Data. (Top) Genres of books in the train data. (Bottom) Genres of books in the test data.

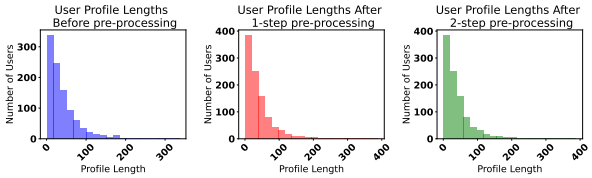


(a) 1% pre-processing **(b) 50% pre-processing**

Figure 3: Comparing items' long tail distribution on `MIND News data` for 1% (1-step and 2-step) pre-processing vs. 50% (1-step and 2-step) pre-processing.

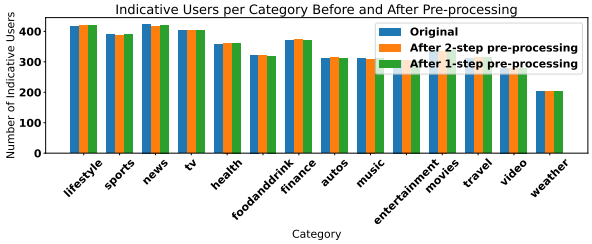


(a) 1% pre-processing

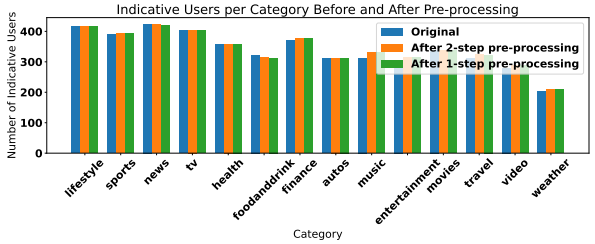


(b) 10% pre-processing

Figure 4: Comparing users' profile length on `MIND News data` for 1% (1-step and 2-step) pre-processing vs. 50% (1-step and 2-step) pre-processing.

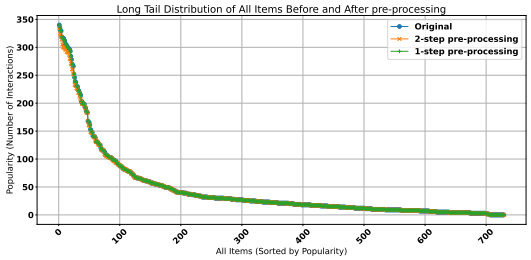


(a) 1% pre-processing

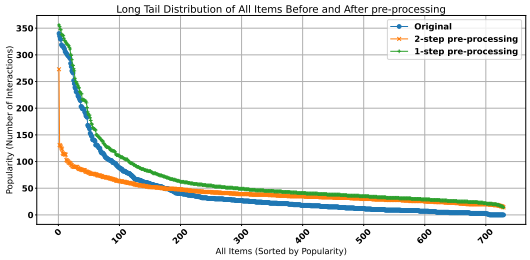


(b) 50% pre-processing

Figure 5: Comparing indicative users per category on `MIND News data` for 1% (1-step and 2-step) pre-processing vs. 50% (1-step and 2-step) pre-processing.

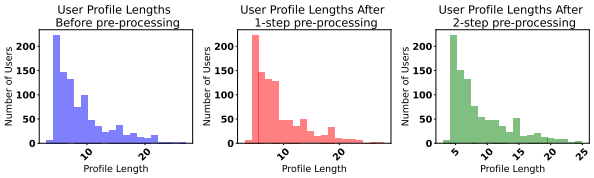


(a) 1% pre-processing

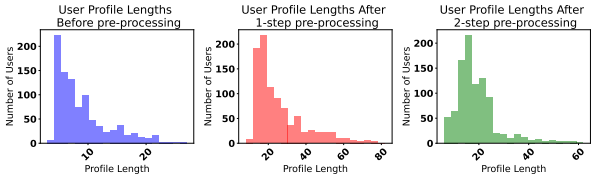


(b) 50% pre-processing

Figure 6: Comparing items' long tail distribution on `GoodBook data` for 1% (1-step and 2-step) pre-processing vs. 50% (1-step and 2-step) pre-processing.

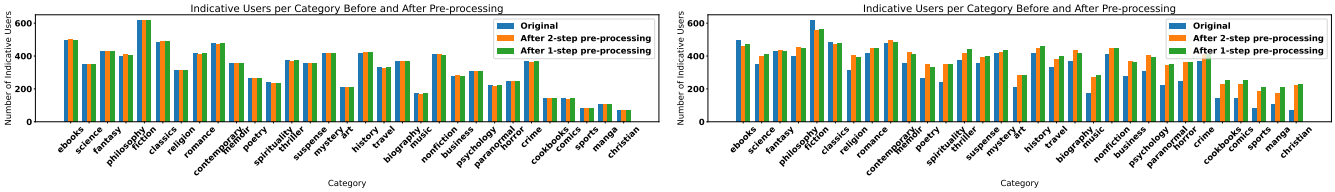


(a) 1% pre-processing



(b) 50% pre-processing

Figure 7: Comparing users' profile length on `GoodBook data` for 1% (1-step and 2-step) pre-processing vs. 50% (1-step and 2-step) pre-processing.



(a) 1% pre-processing

(b) 50% pre-processing

Figure 8: Comparing indicative users per category on `GoodBook data` for 1% (1-step and 2-step) pre-processing vs. 50% (1-step and 2-step) pre-processing.