

Experiments

Dataset

- Amazon, Yelp, GoodReads
- Filtered out noisy data from Yelp and GoodReads datasets by maintaining only user who made at least 10 interactions and item that were involved to at least 5 interactions.

TABLE I: Data Statistics. Int. denotes user-item interactions.

Data	# Users	# Items	# Int.(M)	Avg. Int. per user	Period
Tools	16,472	10,177	0.133	7.7	Nov 1999 - Jul 2014
Toys	19,153	11,865	0.165	8.3	Jul 2000 - Jul 2014
Cell Phones	27,372	10,279	0.190	6.5	Feb 2001 - Jul 2014
Clothing	38,651	22,974	0.274	6.6	Mar 2003 - Jul 2014
Sports	34,974	18,294	0.291	7.9	Mar 2002 - Jul 2014
Health	37,842	18,358	0.339	8.4	Dec 2000 - Jul 2014
Kindle	67,193	58,110	0.935	12.7	Mar 2000 - Jul 2014
CDs	74,926	64,342	1.093	14.4	Nov 1997 - Jul 2014
Movies	122,923	49,976	1.688	13.3	Nov 1997 - Jul 2014
Yelp	47,906	78,734	2.304	47.2	Oct 2004 - Nov 2018
GoodReads	58,003	45,330	2.791	47.5	Feb 2001 - Nov 2017

Experiments

Evaluation Protocol

- Used user-item interactions in the least one month as test data, and the others as training data, then set the interactions in the latest one month in the training data as validation data.
- Removed users and items, which don't appear in the training data, from the validation and test data.
- As metrics, adopt hit ratio ($H@k$) and normalized discounted cumulative gain ($N@k$) to evaluate the ranking performance.
- Due to space limitation, report results with $k = 10$, ran system 5 times and averaged results.