Personalized Item Frequency for Next Basket Recommendation

Sebastiaan Dijkstra, Giray Keskin, Bram Veenman

University of Amsterdam

Next basket Recommendation

Definition

- Set of items $I = \{i_1, i_2, ..., i_m\}$
- Users with basket history $B_u = \{b_1, b_2, ..., b_t\}$, $b_i \subseteq I$
- Goal: Predict next basket b_{t+1}

Summary of contributions

Reproducibility test

- Verify original results
- Set baseline for other experiments

Extension experiments

- Evaluate performance in new domains and datasets
- Alternative distance metrics
- Try density-based clustering algorithms

Main Objective: Assess robustness and generalizibility of TIFU-KNN
Knowledge provides nsights to advance NBR research

Reproducibility

Table: Side-by-side comparison of the results originally reported and those obtained in our replication attempt using identical hyperparameters. 'Metric' identifies the evaluation measure, 'Original' and 'Reproduced' show the respective results, and 'Difference' gives the percentual change between them. A negative 'Difference' indicates a decrease in the replicated results.

Dataset	Metric	Original	Reproduced	Difference (%)
ValuedShopper	Recall@10	0.2162	0.2216	2.5
	Recall@20	0.3028	0.3036	0.3
	NDCG@10	0.2171	0.2217	2.1
	NDCG@20	0.2589	0.2617	1.1
Instacart	Recall@10	0.3952	0.3814	-3.5
	Recall@20	0.4875	0.4857	-0.4
	NDCG@10	0.3825	0.3831	0.2
	NDCG@20	0.4384	0.4382	-0.1
Dunnhumby	Recall@10	0.2087	0.2073	-0.7
	Recall@20	0.2692	0.2675	-0.6
	NDCG@10	0.1983	0.1986	0.2
	NDCG@20	0.2302	0.2270	-1.4
TaFeng	Recall@10	0.1301	0.1291	-0.8
	Recall@20	0.1810	0.1893	4.6
	NDCG@10	0.1011	0.0983	-2.8
	NDCG@20	0.1206	0.1211	0.4

DBSCAN

- Density clustering alternative to KNN
- Clusters governed by maximum distance ϵ

HDBSCAN

- Search across range of ϵ
-

Distance Metrics

Table: Comparison of different distance metrics across the datasets using the same hyperparameters as the original paper. 'Metric' identifies the evaluation measure, while 'Euclidean', 'Cosine', and 'Manhattan' columns show the results obtained using the respective distance metrics. Numbers in **bold** represent the best-achieved results.

Dataset	Metric	Euclidean	Cosine	Manhattan
ValuedShopper	Recall@10	0.2216	**0.2232**	0.2198
	NDCG@10	0.2217	**0.2230**	0.2212
Instacart	Recall@10	0.3814	**0.3830**	0.3816
	NDCG@10	0.3831	**0.3841**	0.3828
Dunnhumby	Recall@10	0.2073	0.2063	**0.2075**
	NDCG@10	0.1986	0.1916	**0.1988**
TaFeng	Recall@10	0.1291	**0.1312**	0.1287
	NDCG@10	0.0983	**0.1003**	0.0987

Personalized Item Frequency

Capture two purchase patterns

- Repetition: users buy the same items as in the past.
- Collaboration: similar users buy similar items.

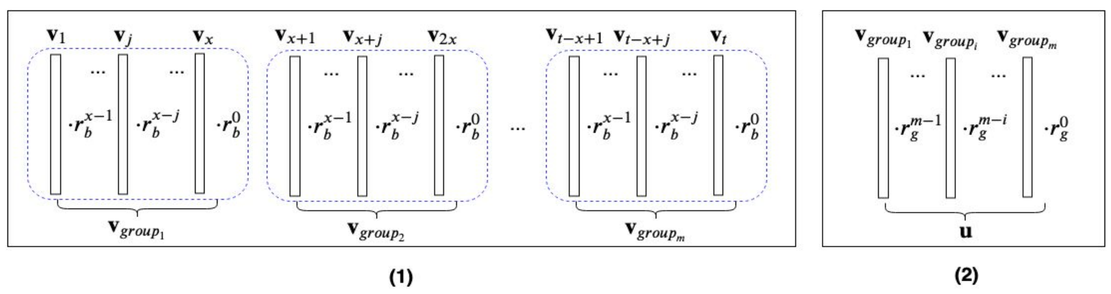
TIFU-KNN

Prediction for next basket:

$$\mathbf{P} = \alpha \cdot \mathbf{u_t} + (1 - \alpha) \cdot \mathbf{u_n}$$

- $\mathbf{u_t}$ is the target user's representation.
- $\mathbf{u_n}$ is the average of the user's k nearest neighbors.
- α balances the repetition and collaboration components.

Figure: Generating User Vector Representations



Two time-decayed weight factors

- Divide basket history into equally sized groups **(1)**
- Average over weighted groups **(2)**

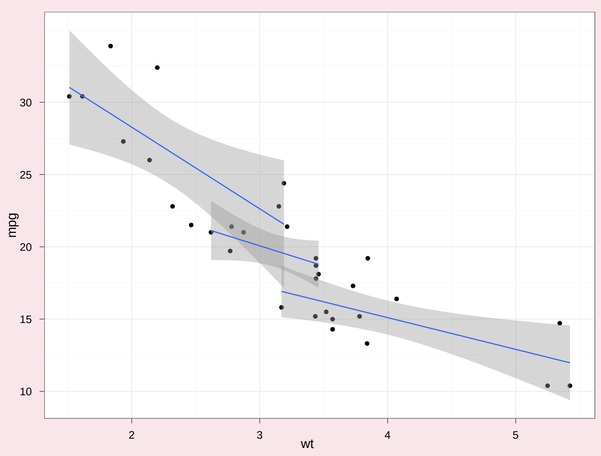
Generalizability assesment

Challenge: Data not divided into "baskets"

- Discard items with little interaction
- Group actions into time intervals

Last FM

Figure: Effect of X on Y



Explain the findings.

Conclusion

Our work includes:

- Reproduced original findings with very similar results
- Tested density based clustering algorithms
- Tried different distance metrics
- Evaluated model in different domains

References:

- Hu, Haoji, et al. "Modeling personalized item frequency information for next-basket recommendation." Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020.

UNIVERSITY OF AMSTERDAM

bram.veenman@student.uva.nl