Personalized Purchase Prediction of Market Baskets with Wasserstein-Based Sequence Matching

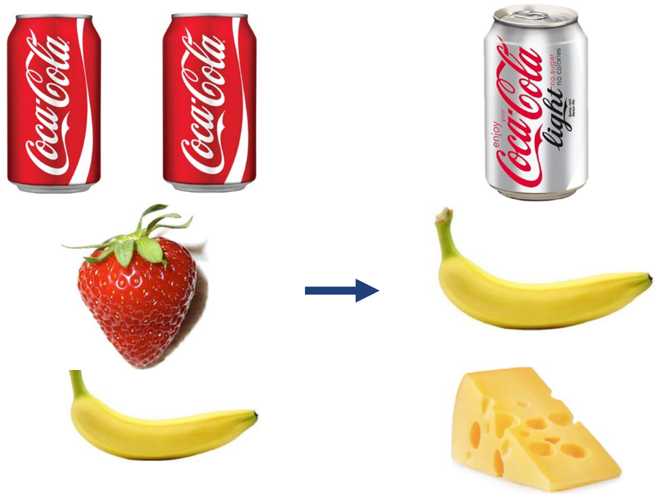
Mathias Kraus and Stefan Feuerriegel

Chair of Management Information Systems, ETH Zürich, Switzerland

Definition of market basket and purchase history

A market basket is a set of items that a customer buys.

A purchase history is a sequence of market baskets.

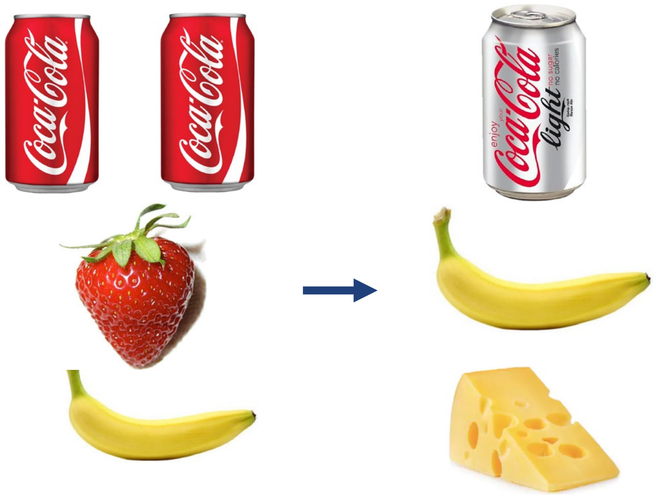


Purchase 1 → Purchase 2

time

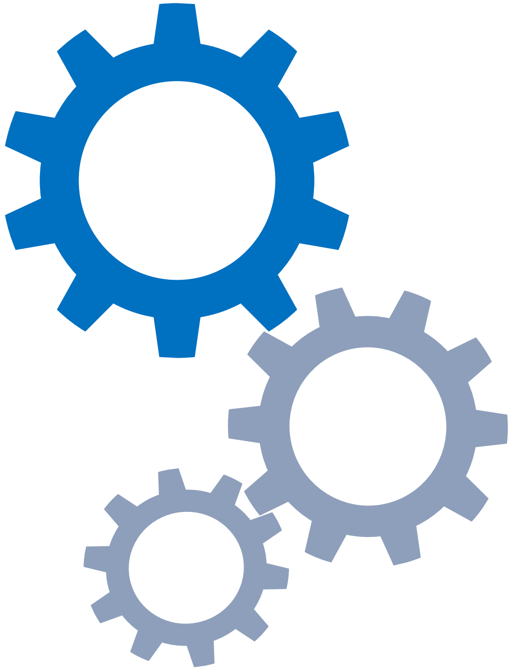
This work aims to identify similar customers to predict future purchases

Purchase history for which we want to predict the next market basket.



Purchase 1 Purchase 2

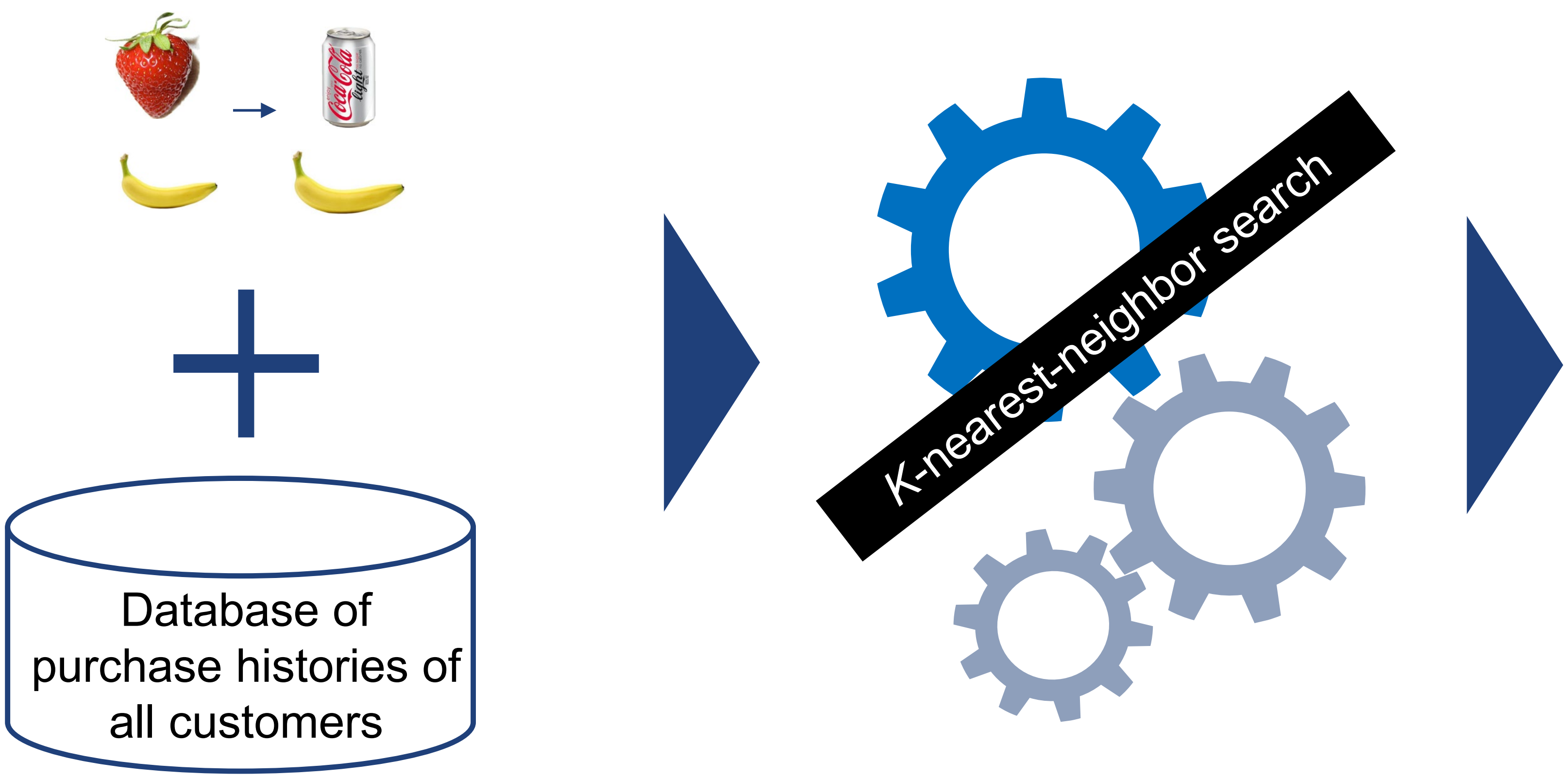
time

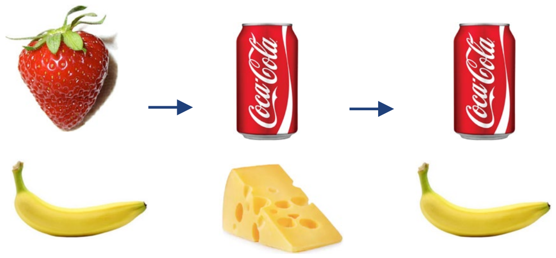


Next purchase

This task is naturally difficult, as very similar products, like coca-cola and coca-cola light are two different products.

Our approach uses simple *K*-nearest-neighbor search to find similar customers





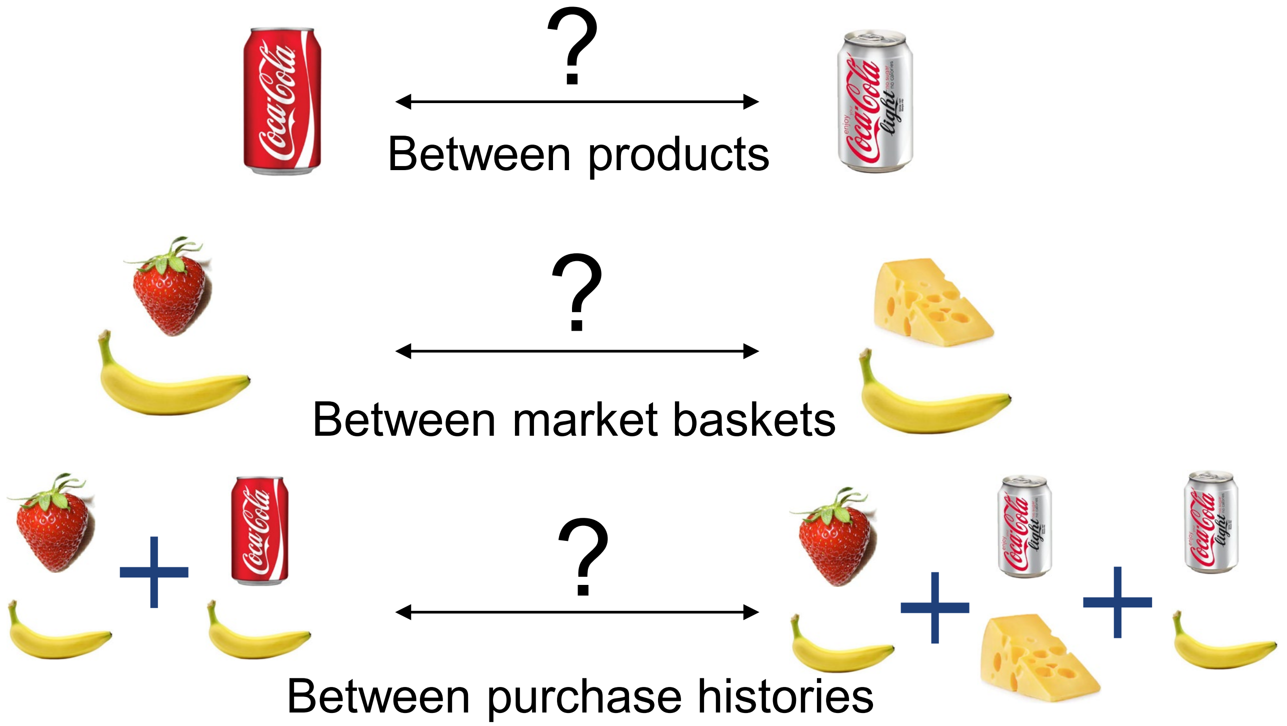
Most similar purchase history that we found in the database. Use this for prediction.

What is the distance between two purchase histories which we use for KNN?

We build up the distance between two purchase histories

We need to define distances to apply *K*-nearest-neighbor search between purchase histories.

We make use of multiple methods to build a distance between purchase histories.



- Product embeddings + cosine simularity
- Wasserstein distance
- Dynamic Time Warping

1. Cosine similarity between two product embeddings defines the distance between two products

Our approach builds upon product embeddings:

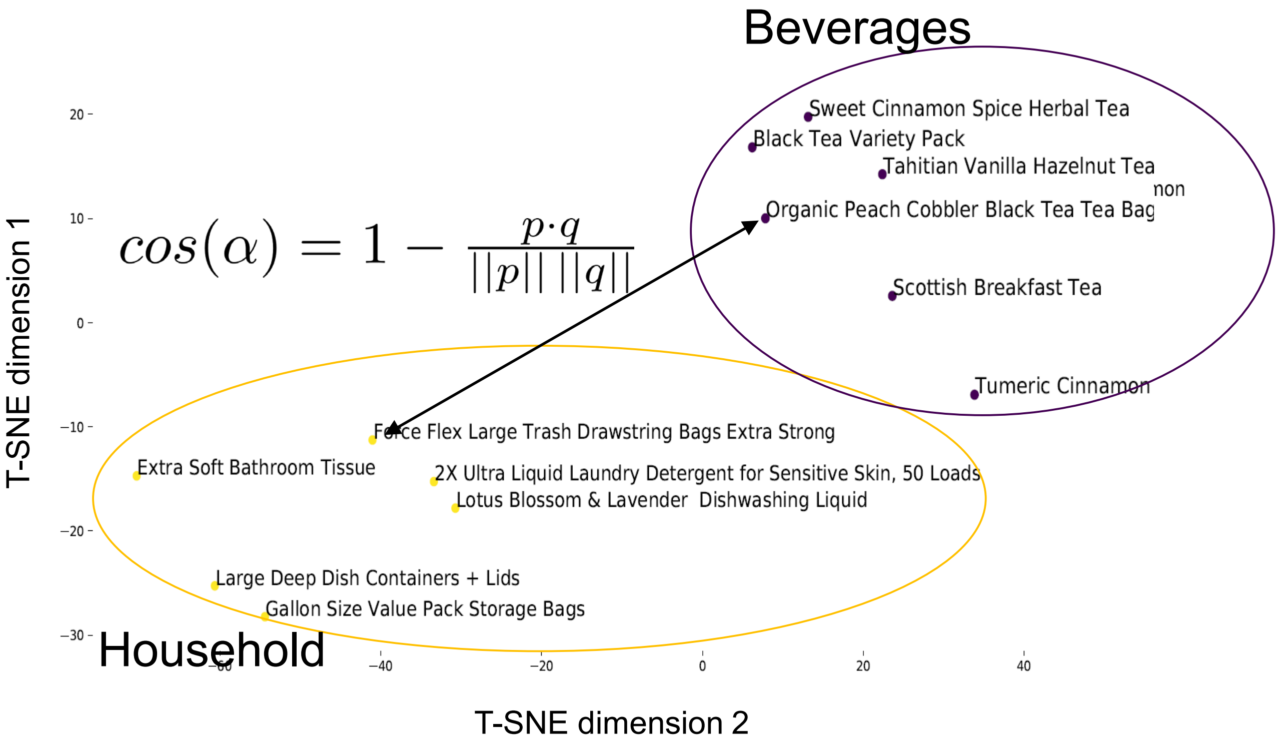
- We obtain a multi-dimensional vector for each product
- Similar products are close to each other such as substitutes like red and white wine

Formally, we maximize the log probability

$$\sum_{p \in b_c^i} \sum_{\substack{q \in b_c^i \\ q \neq p}} \log \Pr(p \mid q)$$

where b_c^i represents a set of items, and $\Pr(p \mid q)$ denotes the softmax function.

$$cos(\alpha) = 1 - \frac{p \cdot q}{||p|| \, ||q||}$$



2. Wasserstein distance defines the distance between two market baskets

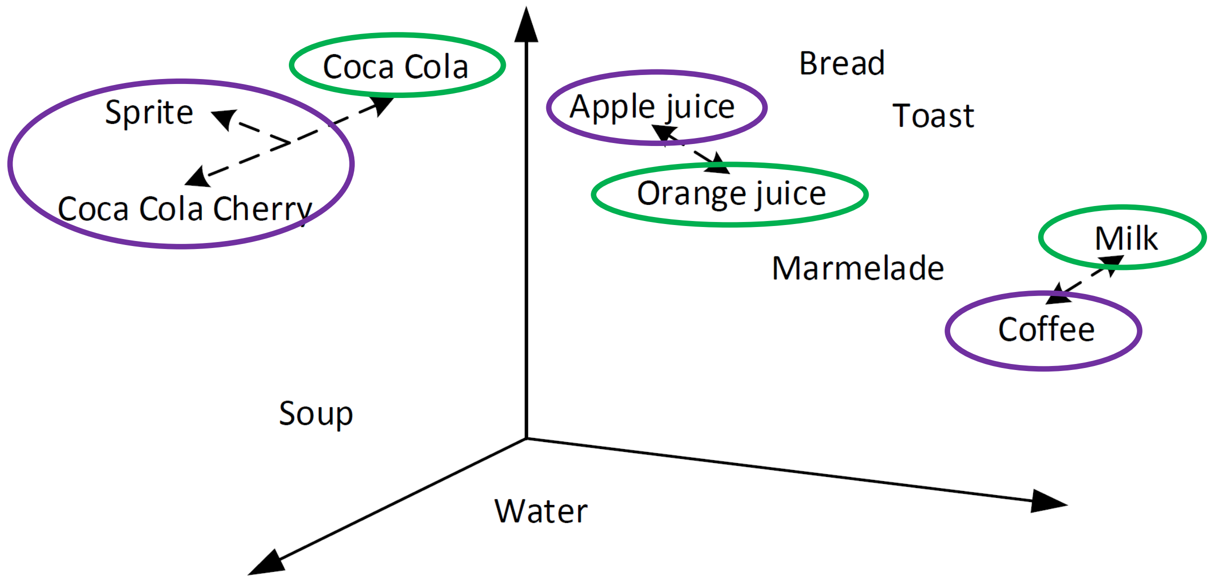
The Wasserstein distance measures distance between two sets of products

- It measures the minimum amount of distance the embedded products of one set A need to 'travel' to reach the embedded products of set B
- Previously utilized in NLP as distance measure between documents

The Wasserstein distance of order *p* is defined as

$$d_{\mathrm{W}}^{(p)}(X,Y) \stackrel{\text{def}}{=} \min_{C} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}\, d(x_i, y_j)^p \right)^{\frac{1}{p}}$$

where *d* is a metric and *C* is an *m x n* transportation matrix.



Market basket A
• Coca Cola • Orange juice • Milk

Market basket B
• Apple juice • Sprite • Coca Cola Cherry • Coffee

3. Dynamic Time Warping defines the distance between two sequences of market baskets (i.e., purchase histories)

We utilize a customized form of dynamic time warping to find similarity between (sub-) sequences of market baskets

- DTW has been shown to be a powerful distance measure between time series
- The mapping can occur non-linear
- It is a natural choice for our task as customers might make random in-between purchases, which can be skipped by DTW

We compute the DTW matrix recursively via

$$D_{ij} = d_{\mathrm{W}}^{(p)}(b_c^i, b_d^j) + \min\{D_{i,j-1}, D_{i-1,j}, D_{i-1,j-1}\}$$
$$D_{i,0} = D_{0,0} = 0, \quad \text{and}$$
$$D_{0,j} = \infty.$$

The distance between the sequences is defined by $min_i D_{in}$ where n is the length of the sequence.

	b_d^1	b_d^2	b_d^3	b_d^4	b_d^5
b_c^3	0.8	0.5	0.5	0.9	0.2
b_c^2	0.7	0.2	0.1	0.1	0.3
b_c^1	0.4	0.3	0.6	0.3	0.7

Wasserstein distance between Baskets b_c^1 and b_d^2

Distance between customer c and d is 0.3 + 0.1 + 0.1 + 0.2 = 0.7

Based on the nearest neighbors, we make the prediction of the next market basket

- Given a purchase history as input, we compute the next market basket for a customer *c*, by considering all purchase histories of all *k* customers.
- We solve

$$(i^*, i_s^*, i_e^*) = \underset{\substack{1 \le i_s \le i_e \\ i=1,\ldots,k}}{\arg\min}\, d_{\mathrm{SDTW}}(B_c, B_i[i_s : i_e])$$

in order to find i^*, the closest customer within the complete customer base.

The prediction for the next market basket is simply the next basket of the most similar customer i^*

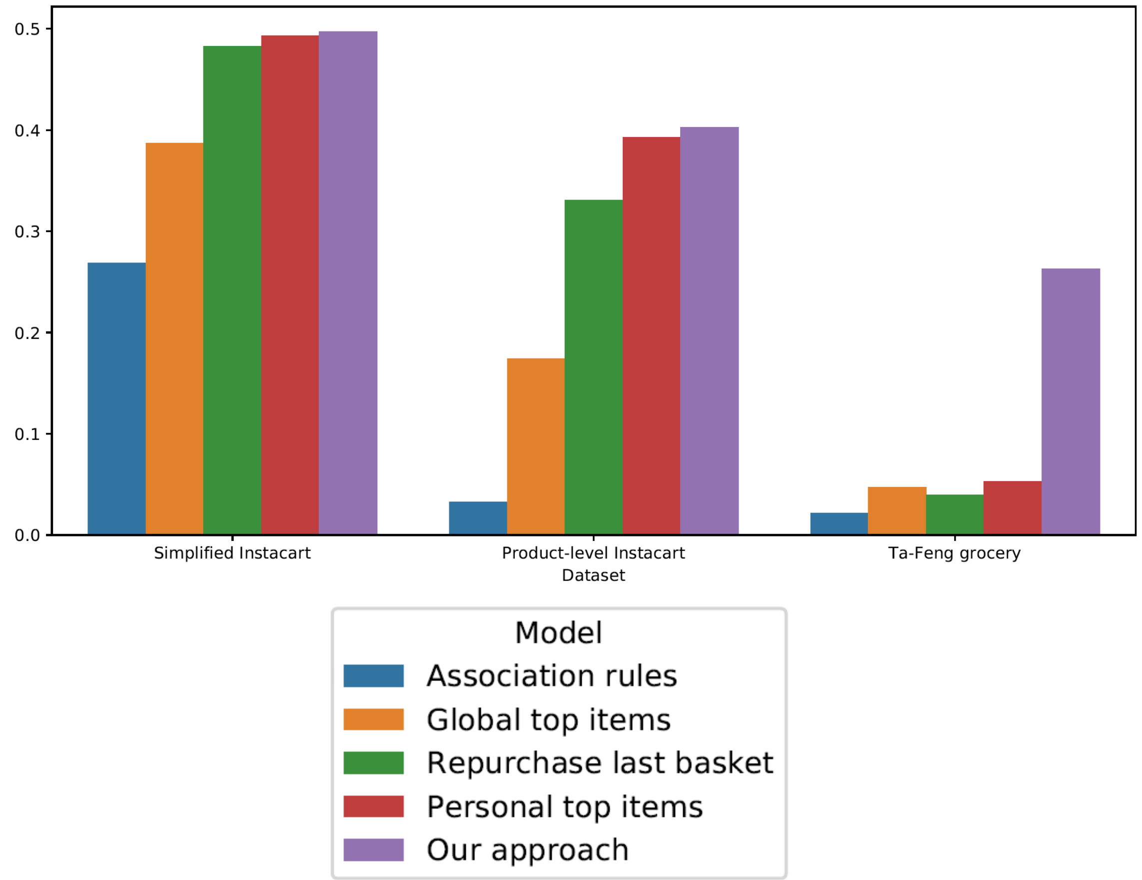
$$B_{i^*}[i_e^* + 1]$$

+

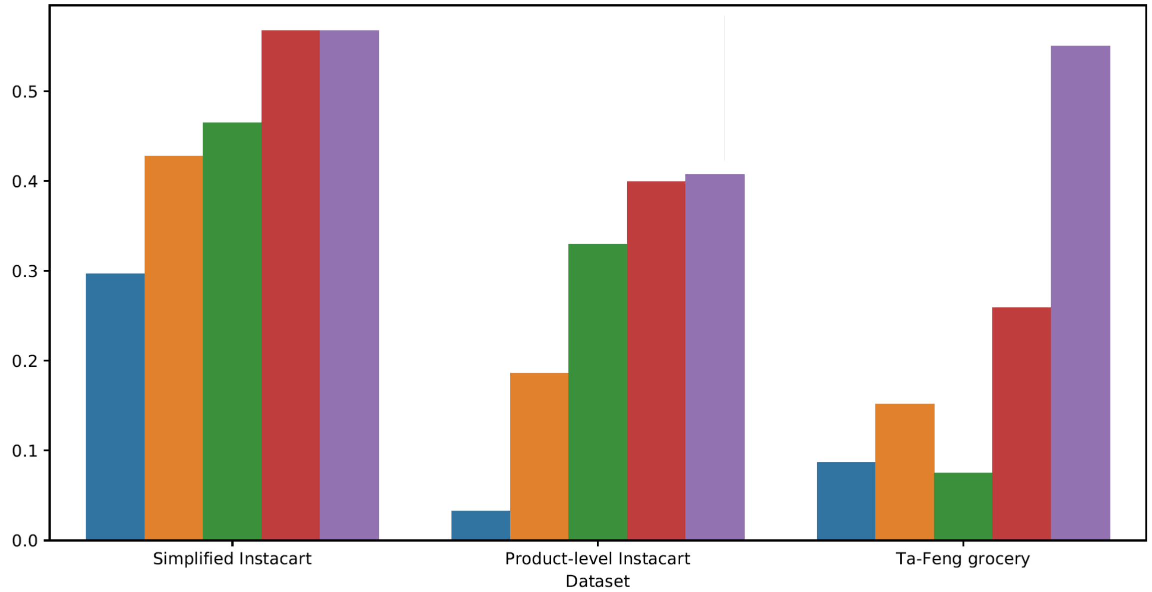
- Our approach has the advantage of returning a distance to the most similar customer.
- We use this distance for a fallback if the distance is above a threshold $\mathcal{T}$.

We experiment with three publicly available datasets

F1 Score



Jaccard Coefficient



Dataset	Customers	Baskets	Products
Simplified Instacart	65710	1634548	134
Product-level Instacart	27139	603457	500
Ta-Feng grocery dataset	9451	172086	500

Conclusion

Contribution

- Our novel prediction algorithm can learn hidden structures among products.
- It can draw upon the complete database of trajectories during prediction time and thus leverages cross-customer knowledge.
- This work provides the first combination of dynamic time warping for subsequence matching and the Wasserstein distance.

Potentials

This approach is not limited to prediction of next market baskets

- Based on the Wasserstein distance, we can provide customers with recommendations of similar baskets (i.e., ‘You seem to like baking apple pie, what about a blueberry pie?’)
- The distance between customers can be utilized for clustering algorithms to identify similar groups of customers

Thank you