Learning and Transferring IDs Representation in E-commerce

Kui Zhao, Yuechuan Li, Zhaoqian Shuai, Cheng Yang

Alibaba Machine Intelligence Technology Lab

阿里巴巴机器智能技术实验室
Alibaba Machine Intelligence Technology Lab

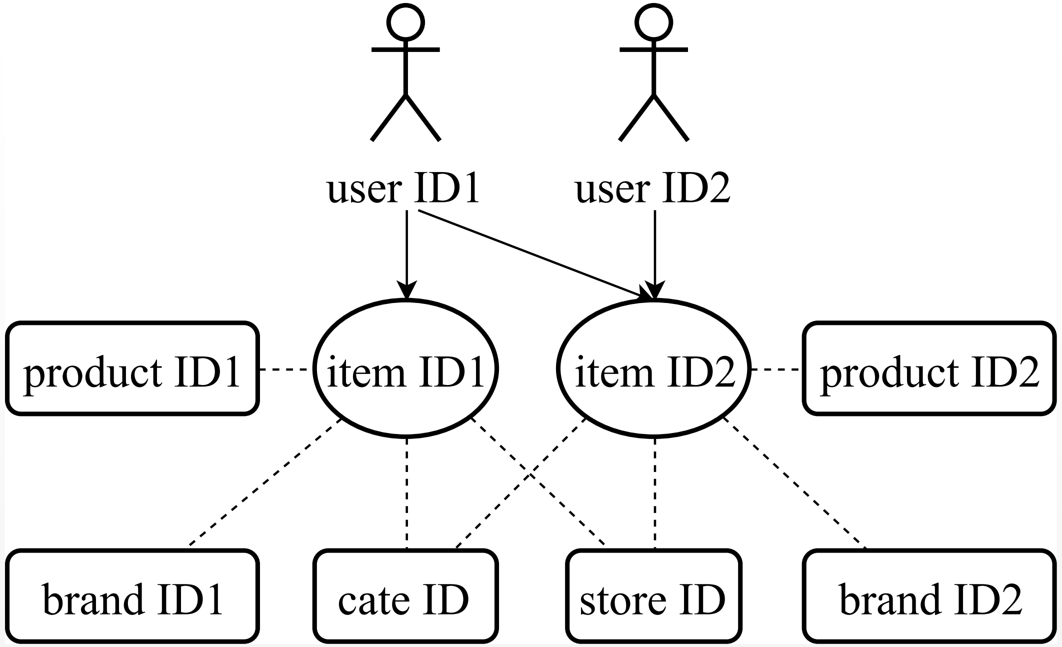
KDD

Motivation

By analyzing and modeling the collected data, intelligence techniques help us to make more informed decisions in E-commerce.

One essential step is the representation of data, especially for IDs.

We proposed an embedding-based framework to learn and transfer the representations of all IDs in E-commerce.

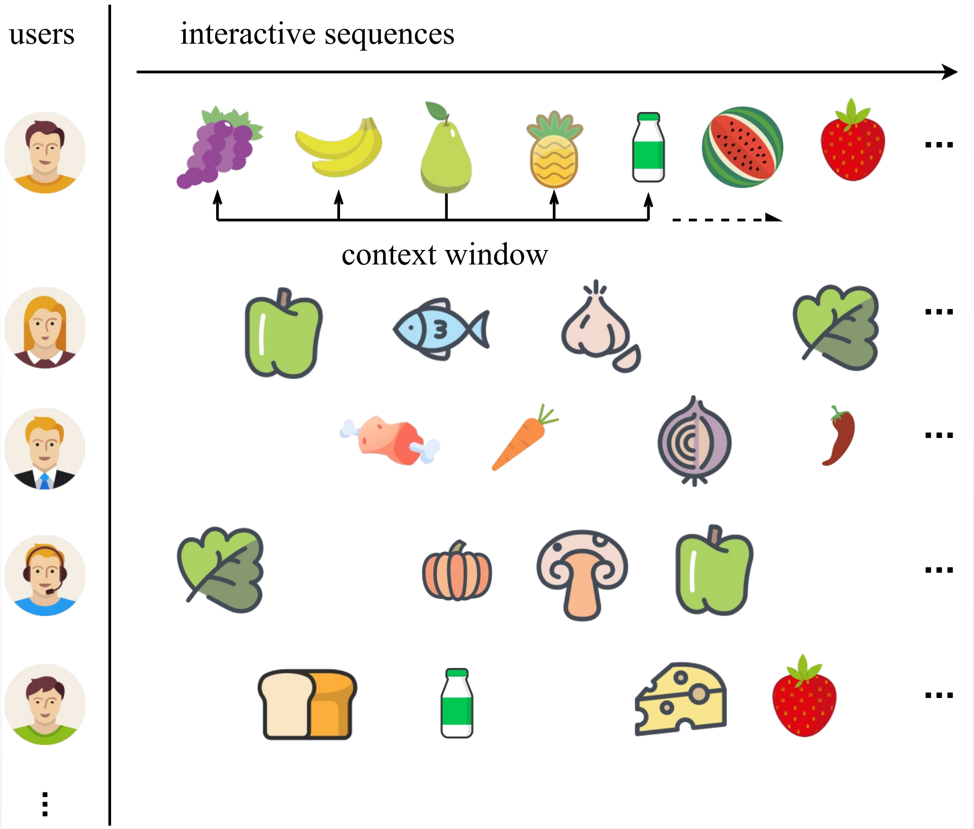


Learning IDs representation

As the implicit feedbacks of users, massive item ID sequences can be easily collected from interactive sessions.

By regarding them as "documents", item2vec embeds item IDs into low-dimensional vectors by modeling their co-occurrence in interactive sequences.

These vectors can be used to boost the accuracy of recommendation and search.



Learning IDs representation: item2vec

Given the target item ID, skip-gram is to find the useful representation by predicting the surrounding item IDs:

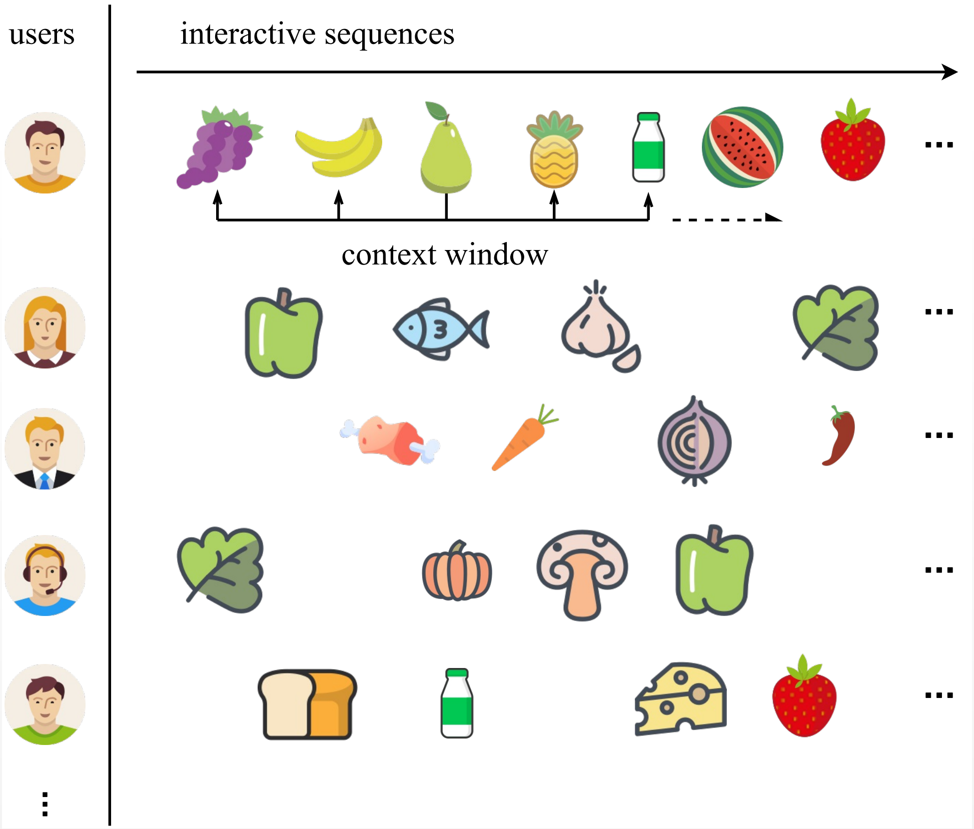
$$\mathcal{J} = \frac{1}{N}\sum_{n=1}^{N}\sum_{-C\le j\le C}^{1\le n+j\le N,\, j\ne 0} \log p(\mathrm{item}_{n+j}|\mathrm{item}_n)$$

where

$$p(\mathrm{item}_j|\mathrm{item}_i) = \frac{\exp(\mathbf{e}_j'^{\mathrm{T}}\mathbf{e}_i)}{\sum_{d=1}^{D}\exp(\mathbf{e}_d'^{\mathrm{T}}\mathbf{e}_i)}$$

It can be replaced by negative-sampling:

$$p(\mathrm{item}_j|\mathrm{item}_i) = \sigma(\mathbf{e}_j'^{\mathrm{T}}\mathbf{e}_i)\prod_{s=1}^{S}\sigma(-\mathbf{e}_s'^{\mathrm{T}}\mathbf{e}_i)$$



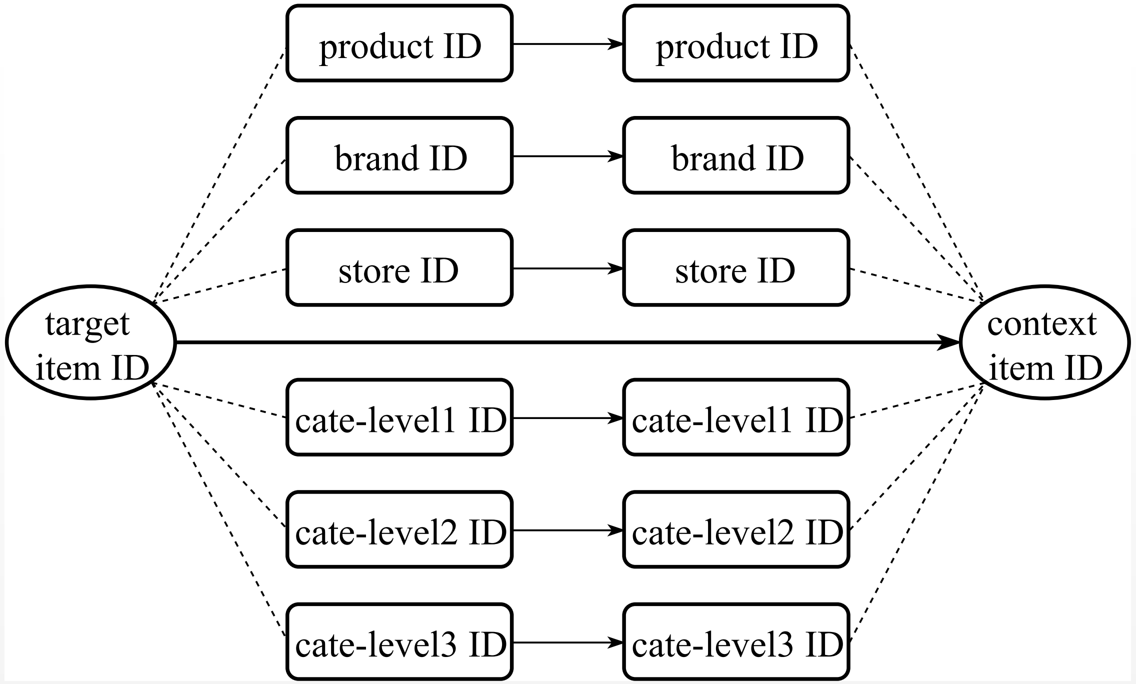
Learning IDs representation: jointly embedding

By exploring structural connections between the item and its attribute IDs, their representations can be learned jointly.

Firstly, the co-occurrence of item IDs also implicates the co-occurrence of corresponding attribute IDs:

$$p\left(\mathrm{IDs}(\mathrm{item}_j)|\mathrm{IDs}(\mathrm{item}_i)\right)$$
$$=\sigma\left(\sum_{k=1}^{K}(w_{jk}\mathbf{e}'_{jk})^{\mathrm{T}}(w_{ik}\mathbf{e}_{ik})\right)\prod_{s=1}^{S}\sigma\left(-\sum_{k=1}^{K}(w_{sk}\mathbf{e}'_{sk})^{\mathrm{T}}(w_{ik}\mathbf{e}_{ik})\right)$$

where w_{ik} is the weight of $id_k(item_i)$.



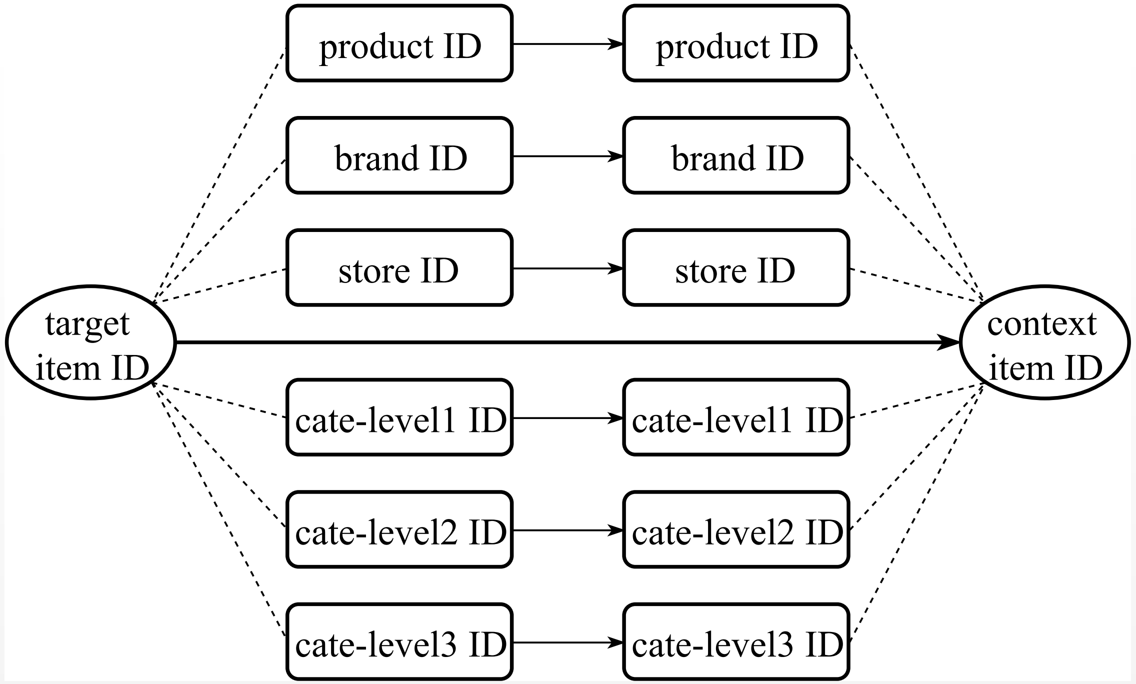
Learning IDs representation: jointly embedding

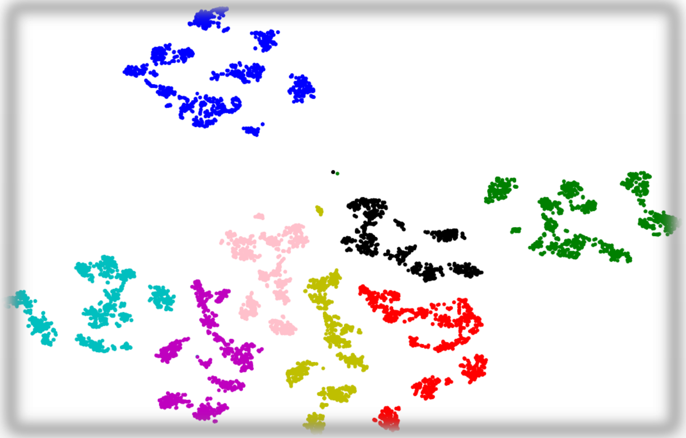
Secondly, structural connections between the item ID and attribute IDs mean constrains:

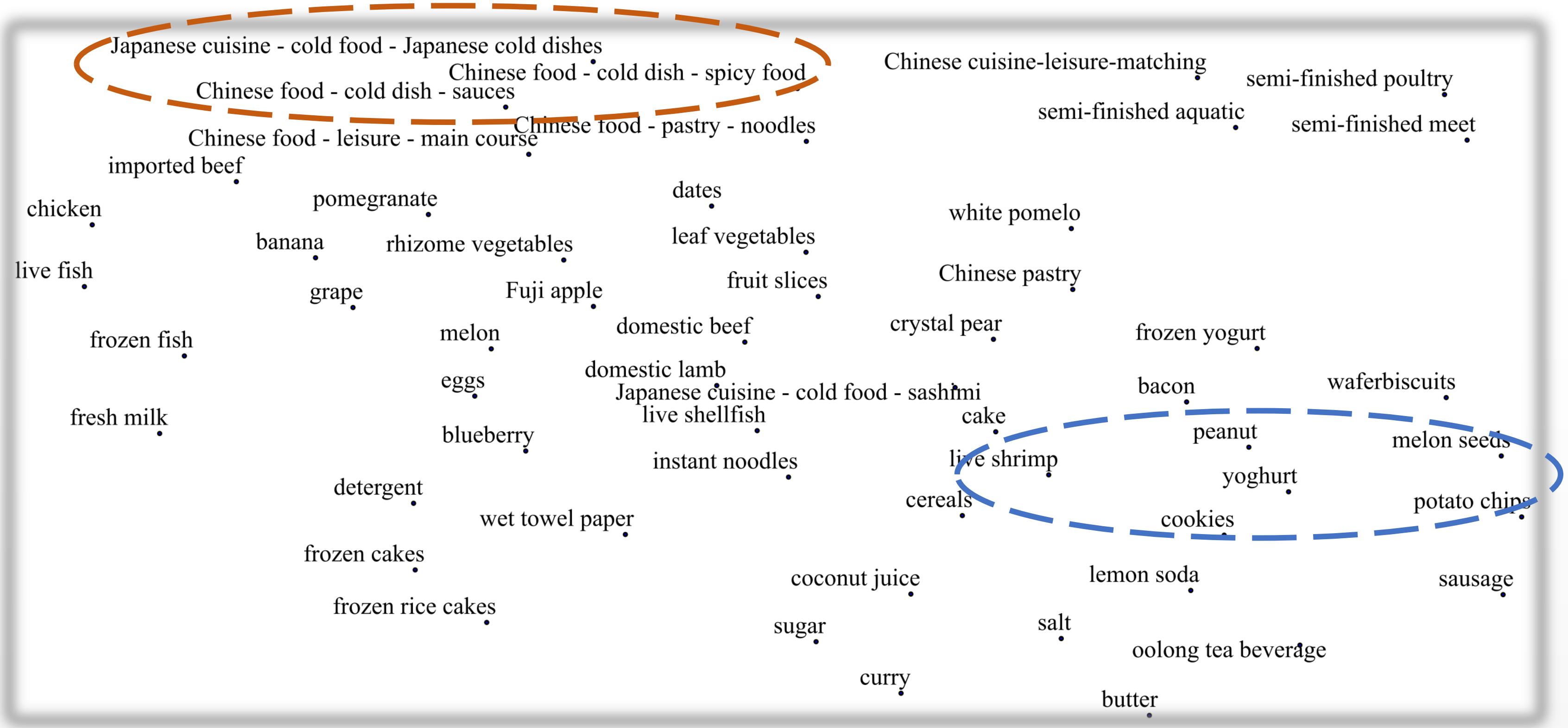
$$p(\text{item}_i|\text{IDs}(\text{item}_i)) = \sigma\left(\sum_{k=2}^{K} w_{ik}\mathbf{e}_{i1}^{T}\mathbf{M}_k\mathbf{e}_{ik}\right)$$

Then we maximize the following average log probability:

$$\mathcal{J} = \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{-C\le j\le C}^{1\le n+j\le N, j\ne 0} \log p(\text{IDs}(\text{item}_{n+j})|\text{IDs}(\text{item}_n)) + \alpha \log p(\text{item}_n|\text{IDs}(\text{item}_n)) - \beta\sum_{k=1}^{K}||\mathbf{M}_k||_2\right),$$







Visualizing vectors of
item IDs from different stores

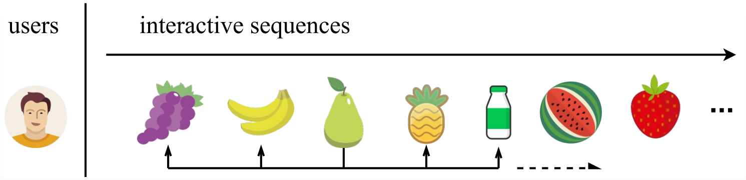
Visualizing vectors of
most popular third-level category IDs

阿里巴巴机器智能技术实验室
Alibaba Machine Intelligence Technology Lab

KDD

Learning IDs representation: user ID

Users' preferences are reflected by their interactive sequences of item IDs.



Thus the representation of user IDs can be obtained by aggregating embedding vectors of interactive item IDs:

$$\text{Embedding}(u) = \frac{1}{T}\sum_{t=1}^{T} e_t$$

As the embedding vectors of user IDs should be updated frequently to reflect the latest preference, Average is chosen instead of RNN etc.

Deploying IDs representation

HEMA App: it is an O2O platform mainly providing fresh food...

Recommendation

Challenge: new items and new users

Delivery demand forecast

Challenge: depict different stores properly

Deploying IDs representation

Items Similarity

Item-item relationships

New Items

Transferring from Seen Items to Unseen Items

New Users

Transferring across Different Domains

Sales Forecast

Transferring across Different Tasks

Recommendation

Forecast

阿里巴巴机器智能技术实验室
Alibaba Machine Intelligence Technology Lab

KDD

Deploying IDs representation: recommendation

Measuring item similarity

The similarity between two item IDs can be measured by: $\cos(\mathbf{v}_i, \mathbf{v}_j) = \dfrac{\mathbf{v}_i^T \mathbf{v}_j}{||\mathbf{v}_i||_2 \cdot ||\mathbf{v}_j||_2}$

Table 1: The click recall@top-N of all methods (higher is better).

	top-N	10	20	30	40	50	60	70	80	90	100	1000
weekday	CF	2.46%	4.46%	6.07%	7.44%	8.66%	9.82%	10.88%	11.82%	12.69%	13.49%	29.83%
	ITEM2VEC	4.72%	7.46%	9.43%	11.00%	12.35%	13.53%	14.16%	14.73%	15.26%	15.74%	30.35%
weekend	CF	3.44%	6.18%	8.46%	10.39%	12.12%	13.65%	15.05%	16.32%	17.49%	18.57%	42.33%
	ITEM2VEC	6.49%	10.42%	13.29%	15.58%	17.48%	19.04%	19.81%	20.55%	21.23%	21.88%	43.34%

recall@top-N

$$= \frac{\sum_{u_i} \#\{\text{hits in top-N}\}_{u_i}}{\sum_{u_i} \#\{\text{total clicks}\}_{u_i}}$$

Table 2: The click recall@top-1000 of all methods at different popularity levels (higher is better).

popular-level	1	2	3	4	5	6	7	8	9	10
CF	22.67%	31.53%	36.81%	39.85%	43.17%	46.72%	47.22%	47.50%	45.85%	58.27%
ITEM2VEC	25.13%	40.44%	45.02%	47.14%	49.33%	51.34%	49.70%	49.72%	47.99%	48.97%

Ensemble in online system: increased by 24.0%

Deploying IDs representation: recommendation

Transferring from seen items to unseen items

New items cause cold-start problems, and approximate embedding vectors are constructed to relieve that. The basic idea is that attribute IDs connected to new item IDs usually have historical records.

Since $$p(\text{item}_i|\text{IDs}(\text{item}_i)) = \sigma\left(\sum_{k=2}^{K} w_{ik}\mathbf{e}_{i1}^T\mathbf{M}_k\mathbf{e}_{ik}\right) \propto \sum_{k=2}^{K} w_{ik}\mathbf{e}_{i1}^T\mathbf{M}_k\mathbf{e}_{ik} = \mathbf{e}_{i1}^T\left(\sum_{k=2}^{K} w_{ik}\mathbf{M}_k\mathbf{e}_{ik}\right)$$

We have $$p(\text{item}_i|\text{IDs}(\text{item}_i)) \to 1 \Rightarrow \mathbf{e}_{i1}^T\left(\sum_{k=2}^{K} w_{ik}\mathbf{M}_k\mathbf{e}_{ik}\right) \text{ is relatively large } \Rightarrow \mathbf{e}_{i1} \approx \sum_{k=2}^{K} w_{ik}\mathbf{e}_{ik}^T\mathbf{M}_k^T$$

Deploying IDs representation: recommendation

Transferring from seen items to unseen items

New items cause cold-start problems, and approximate embedding vectors are constructed to relieve that. The basic idea is that attribute IDs connected to new item IDs usually have historical records.

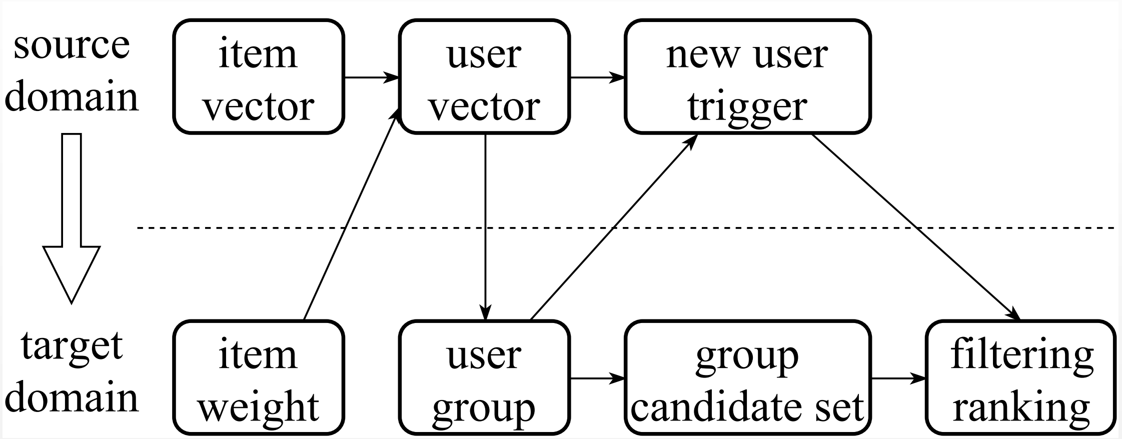
Table 3: The click recall@top-N of baselines and constructed vectors (higher is better).

	top-N	10	20	30	40	50	60	70	80	90	100	1000
weekday	RANDOM	0.01%	0.01%	0.02%	0.03%	0.04%	0.04%	0.05%	0.06%	0.07%	0.07%	0.53%
	HOT	1.60%	2.46%	3.19%	3.77%	4.36%	4.83%	5.39%	5.86%	6.39%	6.97%	27.67%
	NEW2VEC	4.56%	7.05%	8.73%	9.94%	10.88%	11.63%	11.75%	11.87%	11.99%	12.11%	16.95%
weekend	RANDOM	0.00%	0.01%	0.02%	0.03%	0.04%	0.05%	0.06%	0.07%	0.08%	0.08%	0.70%
	HOT	1.69%	2.51%	3.19%	3.79%	4.35%	4.83%	5.42%	6.00%	6.53%	7.13%	27.75%
	NEW2VEC	6.26%	9.86%	12.27%	14.02%	15.33%	16.37%	16.47%	16.56%	16.66%	16.75%	21.89%

Deploying IDs representation: recommendation

Transferring across different domains

For emerging platforms like Hema, a high proportion of users are new customers. We transfer the preference of users on long existing platforms (such as Taobao, Tmall), onto the emerging platform (such as Hema).



The items that are similar to items in the target domain should have higher weights.

$$\text{Embedding}(u) = \frac{\sum_{t=1}^{T} w_t e_t}{\sum_{t=1}^{T} w_t}$$

Naive average increased PPM by 71.4%, and the weighted average increased PPM by 141.8%

Deploying IDs representation: forecast

Transferring across different tasks

Sales forecast can help to manage the workforce, such as guiding us to pre-order appropriate number of delivery staff. By taking embedding vectors of store IDs as a part of input, it helps to depict different stores efficiently and improve the accuracy of forecast.

Table 4: The RMAE scores of different methods in forecasting delivery demand (lower is better).

dataset	day 1	day 2	day 3
HISTORY	43.23%	40.75%	34.26%
HISTORY with ONE-HOT	42.57%	39.00%	34.57%
HISTORY with VEC	40.95%	33.75%	33.02%

$$\mathrm{RMAE} = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{\sum_{i=1}^{N} y_i}$$

Conclusion

The unordered discrete ID is one of the most important types of data in many scenarios, especially in E-commerce.

By jointly using interactive sequences and the structural connections among IDs, we propose an embedding-based framework to learn and transfer low-dimensional representations for all IDs, including user ID, item ID, store ID, brand ID category ID etc.

These low-dimensional representations are deployed and evaluated in several real-world scenarios of Hema App and the results validate the effectiveness.

Thanks