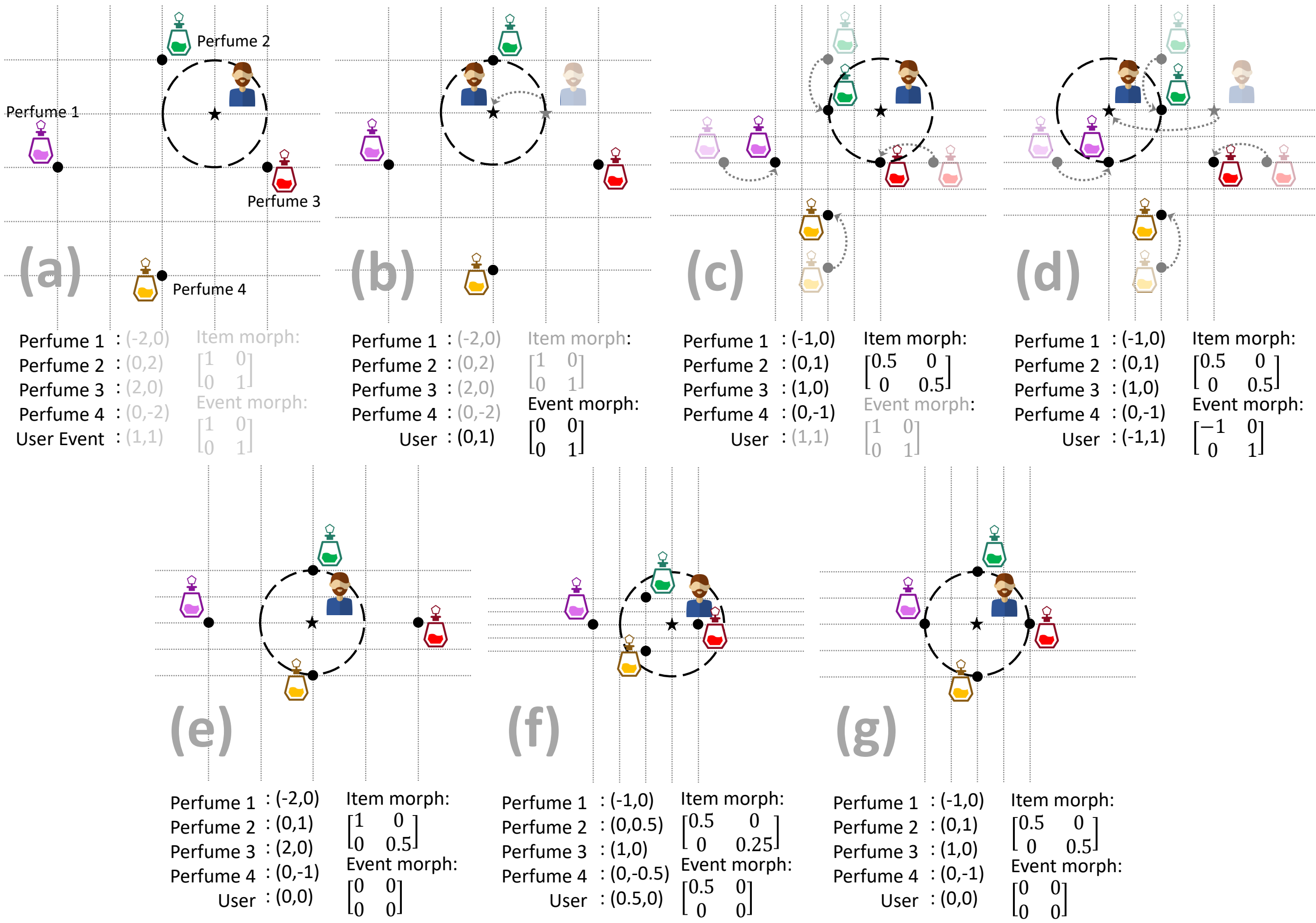

Figure 5: A continuation of the toy example illustrated in Figure 2. In this toy example, it can be shown that no matter what subset of items is desired by the user, there exist item and event morph operators P_u, Q_u that cause exactly that subset to be retrieved.

(1) Figure 5(a) shows that using identity morphs for both item and event embeddings results in the null set getting retrieved.

(2) Figure 5(b) shows how any single perfume can be retrieved by changing the event morph but keeping an identity item morph.

(3) Figures 5(c,d) show how the perfume pairs (1,2), (2,3), (3,4) and (4,1) can be obtained by using two-sided personalization with both item and event morphs possibly being non-identity.

(4) Figure 5(e) shows how to construct morphs to retrieve the perfume pairs (1,3) and (2,4).

(5) Figure 5(f) shows how to construct morphs to retrieve any triplet of perfumes (2,3,4), (1,3,4), (1,2,4), (1,2,3).

(6) Figure 5(g) shows how to construct morphs to retrieve all 4 perfumes i.e. the set (1,2,3,4).

A HARD NEGATIVE MINING

Hard-negative is the term used to describe items that seem deceptively likely to get clicked/viewed as the next event but were not a part of the ground truth. Negative sampling is necessary [3, 5, 8, 32] with large output spaces since evaluating $\mathcal{L}$ with respect to *all* negative items in $\mathcal{A}$ would take $\Omega(|\mathcal{U}| \cdot |\mathcal{A}|)$ time. XPERT considered two types of negative mining techniques:

(1) *In-batch negative mining* [3, 5, 8]: a mini-batch $\mathcal{S}$ of B users (B was tuned as a hyperparameter) was created and for each user $u \in \mathcal{S}$, the negative set $\mathcal{N}_u$ was created by choosing n random items out of $\left\{a_v^* : v \in \mathcal{S}, v \neq u\right\}$ i.e. the set of the ground truth positive items for the $B-1$ other users in the mini-batch.

(2) *Global negative mining* [32]: for a user $u \in \mathcal{U}$, the generic embedding of the clicked ad $\mathcal{E}(a_u^*)$ was used to query the MIPS structure to retrieve n of the most similar items from $\mathcal{A}$ to create $\mathcal{N}_u$. This offered more informative negatives but could suffer if false negatives (also called missing positives) abound which is indeed the case in P-PER applications.

To avoid potential false negatives, entries in $\mathcal{N}_u$ whose cosine similarity to $\mathcal{E}(a_u^*)$ was higher than 0.5 were excluded. In the above techniques, it was made sure that the hard negative set $\mathcal{N}_u$ did not accidentally contain the ground truth item a_u^* for user u. Appendix D discusses hyperparameter tuning and other implementation details such as embedding model $\mathcal{E}$ training.

B SEED EVENT SELECTION

Two seed event selection policies were explored:

Recency-based: This policy simply took the s recent-most events of user u as the seed event set $\mathcal{H}_u$. Although simple to implement and deploy, this could suffer from lack of diversity if recent events were a result of repetitive activities by the user.

Channel-based: The generic embeddings of events in user histories i.e. $\{\mathcal{E}(e) : e \in \mathcal{H}_u, u \in \mathcal{U}\}$ were clustered into K clusters using the standard k-means algorithm with centroids initialized randomly. Recall that every event in a user's history could be thought of as an item as well e.g. the product title for an ad click event or the webpage title for a webpage visit event. Channels were created one per cluster. Subsequently, the user history $\mathcal{H}_u$ for each user $u \in \mathcal{U}$ was categorized into channels based on the cluster to which they belonged. A single representative was chosen from each channel to act as a seed event for that user. This channel representative could either be the *recent-most* event by the user in that channel or else the *mean* or *medoid* of all events by the user in that channel. Algorithm 1 describes the procedure used to ensure that a user has at most s channels activated at any point of time. An importance decay rate of γ was used. The channel importance vector $\mathbf{f_u} \in \mathbb{R}_+^K$ was initialized to the all-zero vector for every user $u \in \mathcal{U}$. γ and K were tuned as hyperparameters. Note that this policy allows different users to have different representatives for the same channel. In experiments, channel-based seed selection was found to benefit not just XPERT but also competitor algorithms such as DPSR [33]. The centroid and medoid variants were cumbersome to maintain in the face of high frequency user activity. The recent-most variant on the other hand was most convenient to deploy and gave only marginally lower performance than the other variants and was thus chosen as the default option for XPERT.

Algorithm 1 XPERT: User-wise Channel Maintenance

Require: User channel importance vector $\mathbf{f}_u \in \mathbb{R}_+^K$ for user $u \in \mathcal{U}$, importance decay rate $\gamma \in (0, 1)$
Ensure: An active set of at most s out of K channels for this user

```
Receive new event e for user u
f_u = γ · f_u        //Reduce imp. of all channels as time has passed
Use MIPS search to find the cluster C(e) ∈ [K] of event e
if f_u[C(e)] > 0 then        //Is this channel already active for u?
    f_u[C(e)]+ = 1        //Bump up its importance as it got touched
else
    f_u[C(e)] = 1        //Assign initial importance
end if
Update the representative of the channel C(e) for u
if ||f_u||_0 > s then        //Are there more than s active channels?
    Set importance of all but the top s channels to 0 for u
end if
```

C OFFLINE METRICS

Results are reported on Recall@k, nDCG@k, MRR@k and AUC@k ($k \in \{10, 50, 100\}$). For a predicted score vector $\hat{\mathbf{y}} \in \mathbb{R}^{|\mathcal{A}|}$ and ground truth vector $\mathbf{y} \in \{0, 1\}^{|\mathcal{A}|}$ where $\mathcal{A}$ is the candidate item set, the various metrics are computed as follows:

$$Recall@k = \frac{1}{\|\mathbf{y}\|_0} \sum_{l \in rank_k(\hat{\mathbf{y}})} y_l$$

$$DCG@k = \frac{1}{k} \sum_{l \in rank_k(\hat{\mathbf{y}})} \frac{y_l}{log(l+1)}$$

$$nDCG@k = \frac{DCG@k}{\sum_{l=1}^{min(k, \|\mathbf{y}\|_0)} \frac{1}{log(l+1)}}$$

The MRR metric is defined as the mean reciprocal of the first rank at which a relevant item was observed in the ranked list of retrieved items. AUC is defined as the average number of inversions in the ranked list of retrieved items i.e. the fraction of (relevant,irrelevant) item pairs where the irrelevant item was given a higher score than the relevant item.

D IMPLEMENTATION DETAILS AND HYPERPARAMETERS TUNING

20% training users were used as a held-out validation set. The final values used in the XPERT model are described below. Training was found to converge within 20 epochs and was robust to tuning of these hyper-parameters due to which default values were used for most. Table 8 gives further details on the range of values over which hyperparameters were tuned for the U2A datasets. These tuning exercises resulted in only a small boost in performance indicating that XPERT is robust to most of its hyperparameters. For example, increasing the # of heads in Segment 2 resulted in a 1.5% increase in recall@100 while tuning the # of transformer layers and attention heads in Segment 1 resulted in a 0.5% increase in recall@100.

Hyper-parameters for baselines were set as suggested by their authors wherever applicable or finely-tuned otherwise.

(i) S1: A 2-layer transformer with 8 attention heads were used with internal dimensionality 512 for U2A datasets and 1024 for

Table 8: Various hyperparameter values on which XPERT was tuned

Name	Range	Tuned value
# seed events s	1, 5, 10, 20, 50	10
Batch size B	50, 100, 200	200
Learning rate	1e-4, 5e-4, 1e-3	5e-4
# negatives n	5, 10, 20	10
Loss λ_+	0.8, 0.9, 1.0	1.0
Loss λ_-	0.1, 0.2, 0.3, 0.5	0.3
Seg1 # layers	1, 2, 4, 8, 16	2
Seg1 # heads	1, 2, 4, 8, 16	8
Seg2 # heads	1, 2, 5, 10, 50	10
# clusters K	500K, 1M, 2M	1M
Imp decay γ	$e^{0.01}, e^{0.02}, \ldots, e^{0.09}, e^{0.1}, e^{0.2}$	$e^{0.08}$

AmazonReviews datasets. The output dimensionality was set to $D = 64$ for U2A and $D = 600$ for AmazonReviews.
(ii) S2: 10 heads were used.
(iii) Training: A batch size of $B = 200$ with a contrastive loss with margin parameters $\lambda_+ = 1.0$ and $\lambda_- = 0.3$ (see Section 4) for U2A and $\lambda_+ = 1.2$ and $\lambda_- = 0.1$ for AmazonReviews datasets. The Adam optimizer was used with learning rate set to $5e - 4$.
(iv) Clustering: $K = 1$ million global clusters were used for U2A. It is notable that clustering is applied to the set of unique events in user histories which for this data was around $\approx$ 120M making the choice of K reasonable. We clarify that clusters were not created over the set of candidate items $\mathcal{A}$ (which were far fewer in number $\approx$ 1.02M for this task). $K = 75k$ clusters were chosen for the AmazonReviews-1M dataset. Yet again, it is notable that the total number of unique historical events in the user histories was $\approx$ 300K for this dataset. K-means was executed for 20 iterations on all datasets.

E BASELINE IMPLEMENTATION

Due to the lack of publicly available codebase for all the baselines, implementations were done by closely following the details from their corresponding papers. All hyperparameters were picked as suggested by their respective papers and further fine tuned using a fine-grained grid search strategy. The architectural description of each implemented baseline is given below:

- **SUR**: Both SUR-DNN and SUR-BERT compute the representation for a user $u \in \mathcal{U}$ using an aggregation model and perform retrieval based on this user embedding. While SUR-DNN uses a simple MLP to perform aggregation, SUR-BERT passes the user history through a BERT model to get the final user representation. A BERT-base model was used for the SUR-BERT. For SUR-DNN item embeddings were allowed to be fine-tuned by introducing a small feed-forward network before the MLP-based aggregation model.
- **DPSR**: The DPSR method was adapted to per-event retrieval tasks by considering each seed event as a query for retrieval. User embeddings were calculated as an aggregation of the user history (using a transformer architecture instead of the simpler mean aggregation suggested by [33]) that was concatenated with the seed event embedding and passed through a feed-forward network to get the final personalized embedding for the seed event.
- **PinnerSage**: Ward clustering was used to cluster items in a user's history and the corresponding medoids were taken to be the cluster representatives. The clusters were ranked according to a cluster importance score as defined in [23]. For each user, representatives of the top 3 clusters were chosen as seed events for retrieval.

F EXAMPLES STUDY

Table 9 shows examples of clusters created by Algo 1. It is evident that the clusters capture distinct product categories. Table 10 shows an example to illustrate how XPERT is able to increase the diversity in the retrieval set compared to NP-PER. Table 11 shows an example study to show how XPERT is able to personalize the retrievals.

Table 9: Subjective study of clusters used for seed event selection. 4 clusters out of the total 1M are shown. For each cluster, 5 random members are sampled for brevity

Cluster members
snowblower under $90.00 Best Price Snow Blowers, Homeandgardenideas.com best snow blowers for sale snow blowers lowest price Cheap Snow Blowers, Fastquicksearch.com
silacone kitchen ware Search Results | Hobby Lobby silcon cookware silacone utensils cooks essentials sillicone pmats - Walmart.com SILACONE PANS Search Results - QVC.com
where do i get a new charger for my at&t phone att phone chargers ATT phone charger - Walmart.com does the at&t store have chargers Wall Chargers & Wireless Phone Chargers for Cell Phones - AT&T
Wooden Photo Frames - Baker Ross Custom Walnut Solid Wood Photo Frame with Silver Metal Edge | Etsy Light pine wood photo frames | Etsy Black Wooden Picture Frame Landscape – Abstract House Geometric Carved Wood Photo Frame | Best Price and Reviews | Zulily

Table 10: Subjective comparison of seed events selected by XPERT and NP-PER for the 1st user in Table 11

Method	Events selected for retrieval
NP-PER	Restorative Compression - Blue Dots - Primes Compression Restorative Compression - Blue Dots - Primes Compression Restorative Compression - Aqua Confetti Restorative Collection - Primes Compression Restorative Compression - Aqua Confetti Restorative Collection - Primes Compression Restorative Compression - Aqua Confetti Restorative Collection - Primes Compression Your Shopping Cart - Primes Compression Thank you for your purchase! - Primes Compression
XPERT	My T-Mobile | Usage Treatment for Venous Insufficiency | Stanford Health Care Facts about Edema Treatment - Can we Cure Edema or Swollen Meridian Stretches and Breathing Exercises - Biogetica compression socks for women best compression socks for women Your Shopping Cart - Primes Compression Restorative Compression - Aqua Confetti Restorative Compression - Mermaid Thank you for your purchase! - Primes Compression

Table 11: Subjective comparison of XPERT and NP-PER retrievals on the U2A-4M dataset. For each user, the user history, the ground truth , the event based on which retrievals are made, and 10 retrievals both by XPERT and NP-PER are mentioned.

User history	Event	Method	Retrieved Ads
My T-Mobile Recently Sold Homes in Pittsburgh PA - 29,866 Transactions Peters Township, PA Real Estate - Peters Township Homes for Sale venous insufficiency symptoms extremity edema treatment for lower extremity edema Facts about Edema Treatment - Biogetica Meridian Stretches and Breathing Exercises - Biogetica venous insufficiency treatment Treatment for Venous Insufficiency	Restorative Compression - Aqua Confetti - Primes Compression	NP-PER	compression shirts walmart.com mens compression tee walmart.com compression shirts for men walmart.com amazon com men s compression shirts sports outdoors mens compression shirt before and after walmart.com compression shirt target men s compression t shirt walmart.com compression shirt for men walmart.com compression t shirt men walmart.com amazon com compression shirts for men
compression socks for women best compression socks for women Restorative Compression - Nude - Primes Compression Restorative Compression - Aqua Confetti - Primes Compression Restorative Compression - Blue Confetti - Primes Compression Restorative Compression - Mermaid - Primes Compression Restorative Compression - 3 Pack Surprise - Primes Compression Ground truth ad click: **best compression socks reviews**		XPERT	affordable compression socks for men and women support stockings 5 affordable compression socks for support women stockings men hose plus size **best compression socks reviews** best compression sock stockings sale for men women best compression socks sale paypal wholesale compression socks findinfoonline.com top rated compression socks walmart.com 5 best compression socks may 2021 bestreviews performance compression socks walmart.com best compression socks you can buy pro compression procompression.com pro compression our 1 pick for best compression socks of 2020
Buy Used Cars, Find Used Vehicles for Sale - Enterprise Car Sales CarMax - Browse used cars and new cars online Used cars $12,878-$17,171 for Sale Used Toyota Highlander $12,878-$17,171 for Sale 10 Best AWD Cars for 2021 Used SUV / Crossover for Sale Right Now - CarGurus Used SUV / Crossover for Sale in Chicago, IL - CarGurus Used Ford Edge for Sale in Chicago, IL - CarGurus 2019 Ford Edge SEL AWD - $21,788 - CarGurus Used 2018 Toyota Highlander in Columbus, Ohio | Carmax	2019 Jeep Grand Cherokee Limited RWD - $27,995 - CarGurus	NP-PER	find your jeep grand cherokee all the latest models and great deals on are on surprisesavings.net jeep grand cherokee suv walmart.com jeep grand cherokee walmart.com new jeep grand cherokee for sale cargurus ca new jeep grand cherokee for sale cargurus amazon com 2018 jeep grand cherokee limited automotive jeep grand cherokee 2018 walmart com 2021 jeep grand cherokee prices reviews trims photos truecar menu 2021 jeep cherokee edmunds grand cherokee walmart com
Used Toyota Highlander for Sale in Chicago, IL - CarGurus 2018 Jeep Grand Cherokee for Sale in Chicago, IL - CarGurus 2019 Jeep Grand Cherokee Limited 4WD - $29,988 - CarGurus 2021 Jeep Grand Cherokee | Edmunds 2021 Jeep Cherokee | Edmunds Used Jeep Grand Cherokee for Sale in Chicago, IL - CarGurus Ground truth ad click: **used jeep for sale**		XPERT	used jeep grand cherokee for sale get low jeep grand cherokee price quotes at carpricesecrets.com used jeep grand cherokee 2019 for sale used jeep grand cherokee 2018 for sale jeep grand cherokee clearance prices autoweb.com get low jeep grand cherokee price quotes at newcars.com used jeep grand cherokee 2016 for sale used jeep grand cherokee 2015 for sale 2021 jeep grand cherokee prices reviews trims photos truecar menu Titused jeep grand cherokee for sale with photos cargurus