Personalizing Software and Web Services by Integrating Unstructured Application Usage Traces

Longqi Yang†, Chen Fang‡, Hailin Jin‡, Matthew D. Hoffman∗, Deborah Estrin†
†Computer Science, Cornell Tech, Cornell University; ‡Adobe Research; ∗Google
†ylongqi@cs.cornell.edu, destrin@cornell.edu
‡{cfang, hljin}@adobe.com; ∗matt@matthewdhoffman.com

ABSTRACT

Users of software applications generate vast amounts of unstructured log-trace data. These traces contain clues to the intentions and interests of those users, but service providers may find it difficult to uncover and exploit those clues. In this paper, we propose a framework for personalizing software and web services by leveraging such unstructured traces. We use 6 months of Photoshop usage history and 7 years of interaction records from 67K Behance users to design, develop, and validate a user-modeling technique that discovers highly discriminative representations of Photoshop users; we refer to the model as utilization-to-vector, util2vec. We demonstrate the promise of this approach for three sample applications: (1) a practical user-tagging system that automatically predicts areas of focus for millions of Photoshop users; (2) a two-phase recommendation model that enables cold-start personalized recommendations for many new Behance users who have Photoshop usage data, improving recommendation quality (Recall@100) by 21.2% over a popularity-based recommender; and (3) a novel inspiration engine that provides real-time personalized inspirations to artists. We believe that this work demonstrates the potential impact of unstructured usage-log data for personalization.

Keywords
User modeling; application usage; recommendation

1. INTRODUCTION

Modern software services typically record user actions for the purpose of collecting application usage statistics and reproducing program errors. Relative to structured data traces, such as text, image, and search queries, application usage records are noisy and unstructured, and service providers may find it more difficult to extract value from them. Just as social interaction traces have enabled great success in personalizing online communities, we explore how integrating application usage traces can further empower novel, effective and personalized services, as shown in Fig. 1.

In this paper, we explore this largely untapped space using data from a large number of creative professionals who use Photoshop for work and actively socialize on Behance, a popular large-scale online community where millions of professional photographers, designers and artists share their artwork. We demonstrate that by leveraging the data traces from shared users, Photoshop and Behance can provide significantly improved personalized services and create new user experiences. Our contributions in this work are summarized below.

• We develop and evaluate an approach, util2vec, based on distributed representation learning, that produces high-quality representations of Photoshop users. This model encodes the sequence patterns of the actions each user has performed, and significantly outperforms the bag-of-actions representation by 31.72% Mean Reciprocal Rank (MRR) on the user fingerprinting task (Section 4).

• Based on this model, we present three sample applications:

  1. We develop and evaluate a practical tagging system for Photoshop users. The system, for the first time, is able to accurately predict areas of focus for millions of Photoshop users, who may or may not be active on Behance. Our model significantly outperforms popularity-based tagging by 31% (Re-

1https://www.behance.net/
call@1), and is able to accurately predict long-tailed tags that are important but unpopular among the broader population (Section 5.1).

2. We propose a two-phase recommendation method that generates more accurate recommendations for cold-start users on Behance by leveraging their previous Photoshop usage traces. The performance improvements over the popularity baseline are significant on all tested metrics including Area Under Curve (AUC) (6.8%) and Recall@K (21.2%) when K = 100. Ultimately, our model enables personalized recommendations for a massive number of new users with Photoshop usage histories (Section 5.2).

3. We design a novel application, named the inspiration engine, for Photoshop users by leveraging the co-occurrences of application usage traces and uploaded art projects on Behance. The qualitative results demonstrate how integrating these data sources enable new user experiences (Section 5.3).

Although the data used in this paper comes from creative professionals, the models and frameworks studied may be applied to personalize services in a broader range of scenarios. Given the evolution of app ecosystems, user activities across stand-alone software applications and social platforms are more easily associated via a small number of identity systems, e.g., Gmail, Facebook, Creative Cloud, Apple, and Github. This opens up a new and fruitful space of future user-modeling research for service providers as well as open source communities.

The technical content of this paper is structured as follows. We introduce our dataset in Section 3 followed by the util2vec model in Section 4. Then we present three models and applications leveraging usage traces in Section 5.

2. RELATED WORK

Our work benefits from and has implications to multiple threads of user modeling research, and the util2vec model is inspired by previous work on distributed representation learning.

2.1 Distributed representation learning

Distributed representation learning was first introduced in the area of natural language processing [23]. The goal is to learn a vector space for all words so that they can be used as inputs to natural language understanding algorithms [21]. Recently, such an approach has been extended and successfully applied to paragraphs [17], medicine [4] and online purchases [8]. For instance, Grbovic et al. [4] proposed a framework to learn a vector representation for each product and user given the historical purchasing records, and Choi et al. [4] demonstrated that a similar approach can be applied to learn hierarchical representations for medical concepts. Our util2vec framework is inspired by the previous research efforts mentioned above, and to the best of our knowledge, this is the first work to design a distributed representation learning algorithm in the domain of software user modeling.

2.2 Intra- and cross-platform user modeling in online social platforms

For online social platforms, personalization and user modeling are important tasks, since appropriately matching customers and products is a key to satisfactory user experiences [16]. Often, the goal of such modeling is to derive a real-valued vector for each user that summarizes his/her preferences, habits, and traits in online social platforms. Previous work constructs user vectors by leveraging intra-platform interactions [23], e.g., ratings [16], purchases [13], content consumption [12, 25], articles, social networks [10], or cross-platform interactions, e.g., personal data streams across email, Twitter, and Facebook [14], and follower-followee connections across YouTube and Twitter [28].

2.3 Software user and command modeling

Modeling software users’ proficiency based on the actions they perform has been previously studied in the context of command-recommendation systems [20, 18, 6]. The goal of such a recommender is to help users learn commands in a complex software application. However, the user modeling under such a circumstance is limited to a specific application because of the narrow scope that the modeling system is exposed to. In this work, we show that by integrating application usage traces with online social interactions, the potential applications that such data traces can empower are much broader and diverse. Specifically, we demonstrate that the Photoshop service provider can conduct better user tagging and create new user experiences.

Another line of related work around application usage records attempts to understand the semantic meanings of software actions [1, 7]. By training a word2vec model [21] on online documents, previous work [1] discovered correspondences and relationships between natural language and software actions, which was used to fuel tutorial-recommender systems. Although our work is not directly optimized for this task, we can still extract semantic meanings of actions and their relationships to users’ social interactions, because the actions, along with the users, are embedded in the same high-dimensional feature space (Section 4).

Although previous research on user modeling has achieved great success, most models only consider data from within the online social platforms. In our work, we demonstrate that by leveraging users’ digital traces from application usage records, online social platforms can better understand users and provide more effective recommendations.

3. DATASET

We associate action histories from Photoshop with social interactions on Behance through Creative Cloud accounts as people use them to log into both services. The reasons why we choose these two platforms are three-fold. (1) Photoshop is one of the most popular computer software applications used by creative professionals, and it is an indispensable daily component for people across many creative occupations including graphical designer, photographer, and architect. Therefore, it is an ideal context in which to study and impact users’ working behavior at a large scale. (2) Behance possesses an abundant user base as millions of creative professionals share their work and socialize with each other on the platform. Also, it is one of the major websites for creative talent search. (3) As Photoshop and Behance both serve creative professionals, there are many shared users for us to investigate.

In Photoshop, all of the actions performed in the application, e.g., buttons clicked and features applied, are collected from the users who enabled application usage reporting. An example of the action sequence is shown in Fig. 2. We target a group of users from the U.S. and their action histories from January 2015 to June 2015. We selected 22 billion actions from 3 million unique Photoshop users. From the Behance platform, we collected users’ social interactions in three categories: (1) self-disclosed areas of focus, e.g., Cartooring, Interaction Design and Fashion; (2) user-uploaded projects; and (3) users’ view and appreciate history on these projects. An example of the collected information from Be-
4. SOFTWARE USER REPRESENTATION

In this section, we propose an accurate and robust user modeling framework to model the action histories of software users. We start by introducing the model, followed by implementation details and performance evaluations.

4.1 util2vec framework

Given the action history $H_u = (a_1^u, a_2^u, \ldots, a_t^u)$ from a software user $u$, our goal is to learn a fixed-length real-valued vector $v_u$ that represents his/her software usage pattern. We propose a framework named util2vec to learn the user representation. In our framework, each user or action is mapped to an $M$-dimensional vector, and the vectors are trained to maximize the log probability, as defined in eqn. 1 across all users.

$$\frac{1}{T - 2K} \sum_{u} \sum_{t=1}^{T-K} \log p(a_t^u | a_{t-K}^u, \ldots, a_{t+K}^u \setminus a_t^u)$$

where $T$ is the total number of actions from a given user, and $K$ is the farthest action before/after the prediction target that is used as the context. In other words, the size of the sliding window is $2K + 1$. Intuitively, the model optimized for the objective defined in eqn. 1 will be able to predict any action given the context of the user and the surrounding actions.

For the prediction, we use the softmax function to model the conditional probability $p(a_t^u | a_{t-K}^u, \ldots, a_{t+K}^u \setminus a_t^u)$ as follows (We omit the superscripts of $a_t^u$ where they are clear from context).

$$p(a_t^u | a_{t-K}^u, \ldots, a_{t+K}^u \setminus a_t^u) = \frac{e^{v_{y_t}}} {\sum_{y_t} e^{v_{y_t}}}$$  

where the vector $y = b + WH(a_{t-K}^u, \ldots, a_{t+K}^u \setminus a_t^u; V, X)$; the bias vector $b$ and weight matrix $W$ are parameters of the model, and the columns of the matrices $V$ and $X$ store the user and action representations respectively, i.e., $v_u = V[:, u]$ and $x_u = X[:, u]$ in numpy-style notation. The parameters $b, W, V, X$ are learned during training. In the util2vec framework, we use a transfer function $h$ that averages or concatenates a user representation with representations from $2K$ context actions, as Fig. 3 shows.

4.2 Implementation details

Along with util2vec, we use negative sampling and additional action preprocessing steps to speed-up the training and reduce the noise, which will be discussed next.

4.2.1 Negative sampling

It is expensive to compute the softmax function in eqn. 2 since the denominator involves a sum over a large number of unique actions. To avoid this cost, we replace the softmax loss with a negative-sampling loss. This strategy has been successfully applied in the word2vec model [21]. Specifically, for each instance, we randomly sample $S$ actions that are different from the target action $a_t$ and approximate the log probability $\log p(a_t | a_{t-K}, \ldots, a_{t+K} \setminus a_t)$ as follows:

$$\log(\sigma(y_{a_t})) + \sum_{s \in S} \log(\sigma(-y_{a_t}))$$

where $S$ is a set of randomly sampled actions such that $a_t \notin S$, and $\sigma$ is the sigmoid function $\sigma(x) \equiv \frac{1}{1+e^{-x}}$.

4.2.2 Preprocessing and parameter settings

**Preprocessing:** For each action, we keep it in the vocabulary only if it is used by at least 100 unique users, and the final size of the vocabulary is 1990. During the preprocessing, we also add a special separation token $[E]$ between two sessions to indicate the boundary of action sequences.

**Parameter settings:** The hyper-parameters of our model are set as follows: (1) the dimensionality of the representations, $M$, is set to 500. (2) sampling window size, $2K + 1$, is set to 11, i.e., $K = 5$. During training, we use 0.025 as the initial learning rate and subtract it by 0.005 for each subsequent epoch (5 epochs in total). For inference, we use 0.1 as the initial learning rate and subtract it by 0.02 for each subsequent epoch (5 epochs in total). Our parameter settings are consistent with the previous work on word2vec [21], although further tuning might yield better performance.
Table 1: User fingerprinting (hold-out session retrieval) performance regarding Mean Reciprocal Rank (MRR). The improvement is respect to the bag-of-actions+tf-idf.

<table>
<thead>
<tr>
<th>Modeling framework</th>
<th>MRR (± standard error of mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>util2vec</td>
<td>0.8238± 0.0029</td>
</tr>
<tr>
<td>bag-of-actions + tf-idf (baseline)</td>
<td>0.6037± 0.0037</td>
</tr>
<tr>
<td>bag-of-actions</td>
<td>0.5944± 0.0037</td>
</tr>
<tr>
<td>% of improvement</td>
<td>31.72%</td>
</tr>
</tbody>
</table>

4.3 User profiling performance

We evaluate the profiling performance of the util2vec model with a user fingerprinting task. We start by holding out the 200 most recent sessions from Photoshop users who have at least 400 sessions in the first 6 months of 2015 (In total, 15,369 unique users are selected). We then train the util2vec model over the rest of the action sequences from 3 million Photoshop users. For each of the 15,369 users, her action history $H_i$ has been divided into training sub-sequence, i.e., $H_i^{\text{train}} = H_i[1:200]$ and validation sub-sequence, i.e., $H_i^{\text{val}} = H_i[200:]$, and an ideal model should be able to link $H_i^{\text{val}}$ with $H_i^{\text{train}}$ based on generated profiles. We infer the user’s representation based on the two subsequences respectively, i.e., infer $\phi_i^{\text{train}}$ from $H_i^{\text{train}}$ and $\phi_i^{\text{val}}$ from $H_i^{\text{val}}$. For each user $i$ and her profile $\phi_i^{\text{train}}$, we predict which validation subsequence belongs to her using cosine similarities. More specifically, we sort all the validation subsequence $H_i^{\text{val}}$ by the similarities between $\phi_i^{\text{val}}$ and $\phi_i^{\text{train}}$ in a descending order, and the ranking of the user’s real validation subsequence $H_i^{\text{val}}$ is denoted as rank$_i$. Finally, Mean Reciprocal Rank (MRR), as defined in eqn. 4, is used to evaluate the overall fingerprinting accuracy across $N$ users.

$$\text{MRR} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\text{rank}_i}$$

(4)

We compare util2vec to the bag-of-actions model, which counts the frequency with which each action occurred. As shown in Table 1 our framework outperforms the baselines by 31.72% even when tf-idf is leveraged to down-weight the frequent actions. The experimental results demonstrate that our model is able to produce user vectors that are more representative and have stronger discriminative power. Generally speaking, for a given user, our representation is able to discriminate against 31.72% more distractors, and in practice, software service providers can use such representations to better fingerprint each user.

Along with the user representations, util2vec also learns an action embedding $X$ that encodes semantic similarities between actions. For example, we present the nearest neighbors of five Photoshop actions in Table 2 (the neighbors are ranked by the cosine similarities between action embeddings in descending order), and the retrieval results show that the actions are grouped by their functionalities and usage affinities. The action embeddings may also be useful for the service improvements as it tells the common software usage patterns among the population. Given the scope of this paper, we leave further investigation as future work.

5. APPLICATIONS

In this section, we build and present three applications that can benefit from the integration of such usage traces: software user tagging, cold-start art project recommendation and inspiration engine.
average recall rate, Recall@K, as defined below, to quantitatively compare the tagging performance.

\[
\text{Recall@K} = \frac{\text{number of correct tags in top } K \text{ predictions}}{\text{total number of tags in the ground truth set}}
\]  

The results in Table 4 show that the models leveraging software usage history significantly outperform the baseline that is agnostic to such information, and the improvements are particularly remarkable for top-ranked tags—the system achieves 31.0% and 35.0% improvements in terms of Recall@1 and Recall@2 respectively. This justifies that, practically, our system can not only predict tags that are popular, but the ones that are long-tailed. Ultimately, our tagging system can make accurate predictions for millions of Photoshop users, who may or may not be active on Behance, and it is valuable to enable customized business service for the provider.

Qualitatively, we show the outputs of two tagging approaches for 6 representative Photoshop users in Fig. 4. For each user, we present the ground truth tags, the tags predicted by our system and the popular tags. In addition, we include user’s Behance portfolio (uploaded projects) side-by-side for the illustration purpose. But this information is not available to the tagging algorithm under any circumstance. From Fig. 4, we find that our tagging model is especially advantageous in the following aspects.

- **Tag diversity.** We can accurately predict a diverse array of areas of focus based on the Photoshop usage traces, e.g., Photography (U1, U4), Fine Arts (U1), Web Design (U5), Typography (U2), Cartooning (U3), Animation (U4) and Painting (U6) etc. The tags can be popular on the platform, e.g., Graphic Design and Photography, or long-tailed (infrequent), e.g., Motion Graphics, Cartooning and Painting etc. The prediction results justify the robustness of our system when it is applied to diverse application usage patterns.

- **Generalization power.** Although there is a high correlation between user tags and appearances of uploaded art projects, as shown in Fig. 4, some users didn’t exhaustively select all of the tags that are related. This limitation is partially addressed by the generalization power of our linear classifier. For example, based on U1’s portfolio (images with same content but different coloring), there is a high chance that she is focusing on retouching, which is not selected by herself. Nevertheless, our system can still make reasonable predictions that include retouching in the top tags. This characteristic is further verified in the U6 example (tag Drawing).

Overall, we have shown that by modeling application usage traces, we are able to build an accurate and practical user tagging system for Photoshop with minimal human effort.

### 5.2 Cold-start art project recommendation

Cold-start is a well-known hard problem in the design of modern recommender systems. Specifically, user-cold-start [14] refers to the scenario where the recommendations are targeting new users, and item-cold-start [12] describes the case when a new item needs to be included in the recommendation pool. In user-cold-start, since we lack the information of her activities within a platform, a typical solution is to either recommend the most popular items, which is not personalized, or leverage side information, such as gender, age [22] and personal data traces [14]. However, in many cases, when a new user shows up in online social platforms, their application usage records are already available. If we could leverage these data traces and properly use them to inform the recommender, there is a great potential for the social platforms to improve their cold-start recommendations. For example, Behance might be able to generate better recommendations for 3 million Photoshop users, which is almost 4 times the current number of Behance users. In this section, we propose a two-phase recommendation framework that leverages Photoshop usage data in recommending artistic projects on Behance.

#### 5.2.1 Two-phase recommendation framework

Our recommendation framework is inspired by previous research on content-based music recommendation [25], that incorporates au-
**Note that users’ portfolios are included only for illustration purpose. All of the tag predictions are solely based on users’ Photoshop usage traces.**

![User2vec predictions](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>R1 tag</th>
<th>R2 tag</th>
<th>R3 tag</th>
<th>R4 tag</th>
<th>R5 tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>popular tags</td>
<td>Graphic Design</td>
<td>Illustration</td>
<td>Photography</td>
<td>Branding</td>
<td>Art Direction</td>
</tr>
<tr>
<td>disclosed tags (Ground truth)</td>
<td>Photography</td>
<td>Digital Photography</td>
<td>Fine Arts</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![User2vec predictions](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>R1 tag</th>
<th>R2 tag</th>
<th>R3 tag</th>
<th>R4 tag</th>
<th>R5 tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>popular tags</td>
<td>Graphic Design</td>
<td>Illustration</td>
<td>Photography</td>
<td>Branding</td>
<td>Art Direction</td>
</tr>
<tr>
<td>disclosed tags (Ground truth)</td>
<td>Graphic Design</td>
<td>Print Design</td>
<td>Typography</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![User2vec predictions](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>R1 tag</th>
<th>R2 tag</th>
<th>R3 tag</th>
<th>R4 tag</th>
<th>R5 tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>popular tags</td>
<td>Graphic Design</td>
<td>Illustration</td>
<td>Photography</td>
<td>Branding</td>
<td>Art Direction</td>
</tr>
<tr>
<td>disclosed tags (Ground truth)</td>
<td>Motion Graphics</td>
<td>Photography</td>
<td>Animation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![User2vec predictions](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>R1 tag</th>
<th>R2 tag</th>
<th>R3 tag</th>
<th>R4 tag</th>
<th>R5 tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>popular tags</td>
<td>Graphic Design</td>
<td>Illustration</td>
<td>Photography</td>
<td>Branding</td>
<td>Art Direction</td>
</tr>
<tr>
<td>disclosed tags (Ground truth)</td>
<td>Motion Graphics</td>
<td>Typography</td>
<td>Digital Art</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![User2vec predictions](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>R1 tag</th>
<th>R2 tag</th>
<th>R3 tag</th>
<th>R4 tag</th>
<th>R5 tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>popular tags</td>
<td>Graphic Design</td>
<td>Illustration</td>
<td>Photography</td>
<td>Branding</td>
<td>Art Direction</td>
</tr>
<tr>
<td>disclosed tags (Ground truth)</td>
<td>Illustration</td>
<td>Digital Art</td>
<td>Painting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![User2vec predictions](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>R1 tag</th>
<th>R2 tag</th>
<th>R3 tag</th>
<th>R4 tag</th>
<th>R5 tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>popular tags</td>
<td>Graphic Design</td>
<td>Illustration</td>
<td>Photography</td>
<td>Branding</td>
<td>Art Direction</td>
</tr>
<tr>
<td>disclosed tags (Ground truth)</td>
<td>Illustration</td>
<td>Digital Art</td>
<td>Painting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![User2vec predictions](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>R1 tag</th>
<th>R2 tag</th>
<th>R3 tag</th>
<th>R4 tag</th>
<th>R5 tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>popular tags</td>
<td>Graphic Design</td>
<td>Illustration</td>
<td>Photography</td>
<td>Branding</td>
<td>Art Direction</td>
</tr>
<tr>
<td>disclosed tags (Ground truth)</td>
<td>Illustration</td>
<td>Digital Art</td>
<td>Painting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![User2vec predictions](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>R1 tag</th>
<th>R2 tag</th>
<th>R3 tag</th>
<th>R4 tag</th>
<th>R5 tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>popular tags</td>
<td>Graphic Design</td>
<td>Illustration</td>
<td>Photography</td>
<td>Branding</td>
<td>Art Direction</td>
</tr>
<tr>
<td>disclosed tags (Ground truth)</td>
<td>Illustration</td>
<td>Digital Art</td>
<td>Painting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Six user tagging examples with two different approaches. For each user, we show her portfolio, top 5 tag predictions with util2vec feature, top 5 most popular tags and self-disclosed tags. The tags with orange color are the correct predictions, and the ones with green color are the ones that are inferable from the portfolio but not explicitly selected by the user.

![Two-phase recommendation framework](image)

**Figure 5: Two-phase recommendation framework.** In step 1, we derive users’ latent factors and items’ latent factors and bias from their implicit feedback (project views). In step 2, we learn a projection function \( f \) to map software usage features to the corresponding users’ latent factors.

![Item-loading function](image)

**Step 1.** The goal of the first step is to learn each user \( u \)'s latent factors \( l_u \), and each item \( e \)'s latent factors \( l_e \) and bias \( b_e \), such that the value of \( r_{ue} \), which is defined as \( r_{ue} = l_u^T l_e + b_e \), is proportional to user \( u \)'s preference level towards item \( e \). We learn the parameters by leveraging users’ project views on the platform. Considering that such signals are implicit feedback, as suggested by [27], we propose to minimize the following Weighted Approximately
Table 4: Art project recommendation performance for cold-start users in terms of Recall@K and Areas Under Curve (AUC). We use bold font for the best performed approach and feature set. The percentage of improvements are the comparison between the approach with bold font, and baseline method (popular items).

<table>
<thead>
<tr>
<th>cold-start recommendation with software usage data</th>
<th>Recall@K</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>util2vec features (500 dim)</td>
<td>0.0143</td>
<td>0.0213</td>
<td>0.0261</td>
<td>0.0313</td>
<td>0.0356</td>
<td>0.8202</td>
<td></td>
</tr>
<tr>
<td>bag-of-actions+tf-idf features (1990 dim)</td>
<td>0.0138</td>
<td>0.0209</td>
<td>0.0269</td>
<td>0.0309</td>
<td>0.0350</td>
<td>0.8166</td>
<td></td>
</tr>
<tr>
<td>cold-start recommendation without software usage data</td>
<td>Recall@K</td>
<td>0.0118</td>
<td>0.0188</td>
<td>0.0218</td>
<td>0.0281</td>
<td>0.0297</td>
<td>0.7683</td>
</tr>
<tr>
<td>popular items (baseline)</td>
<td>% of improvements</td>
<td>21.2%</td>
<td>13.3%</td>
<td>23.4%</td>
<td>11.4%</td>
<td>19.9%</td>
<td>6.8%</td>
</tr>
</tbody>
</table>

Figure 6: Four image retrieval results of the inspiration engine using single action. The retrieval results reflect the context where each action is often used. For example, with fade_smart_blur, returned images have blurred background and fading effects, and with rotate_canvas, images tend to have repetitive patterns.

Figure 7: The algorithm framework for the inspiration engine. We learn a function $g$ to project image features to the util2vec embedding space such that true actions-image pairs are close to each other and false pairs are farther away.

Step 2. In the second step, we learn a projection function $f$ that takes a user’s software usage feature $\psi_u$ as input and produces output $f(\psi_u)$ to approximate her latent factors $I_u$. We propose to minimize the l2 loss $\sum_{u,e}(f(\psi_u) - I_u)^2$ for regression.

In this paper, we use a linear function $f$, i.e., $f(\psi_u) = b + W\psi_u$. However, any non-linear function should be directly applicable here, which we leave to future work. During training, for step 1, the parameters are learned with mini-batch Adagrad [5], the dimensionality of the latent factors and learning rate are set to 50 and 0.05 respectively. For step 2, we use l-bfgs to find the optimal solution since the optimization target is convex.

In practice, for any cold-start user $u$ and her Photoshop usage feature $\psi_u$, the items’ recommendation rankings are based on the value of $r_{ue} = f(\psi_u)^T I_e + b_e$ where the item with higher value of $r_{ue}$ will be recommended earlier.

5.2.2 Evaluation and analysis

We evaluate the performance of our cold-start recommendation system by holding out a validation set from the view histories of 67,805 users. We randomly sample 10,000 users among the people who have viewed at least one project after July 1st, 2015 and regard them as the cold-start users. The most recent viewed project $e_{\text{rec}}$ from each cold-start user $u$ is then held for validation, and the rest 57,805 users’ complete view histories are used for training. All the items appear in the training set are included in the items pool, which yields 5.8 million candidates for recommendation. The time...
restraint is used to guarantee the causality of recommendation as the Photoshop usage data is collected from the first 6 months of 2015. During the validation, for each cold-start user, we only use her software usage data to make the preference prediction, without relying on any previous views. Therefore, our evaluation results can properly reflect the system performance when serving cold-start users in practice.

We compare our recommender to the baseline algorithm that ranks the items based on their popularity (total number of views received). This is shown to be a very strong baseline for the cold-start recommendations [14]. Similar to user tagging, we use Recall@K defined in eqn.[8] and Area Under ROC Curve (AUC) defined in eqn.[9] to evaluate the recommendation performance (N=10,000).

\[
\text{Recall@K} = \frac{1}{U} \sum_{u=1}^{U} \delta(e_{pu} \text{ in the top K items for u})
\]

\[
\text{AUC} = \frac{1}{U} \sum_{u=1}^{U} \frac{\delta(f(v_{u})^T I_{e_{pu}} + b_{e_{pu}} > f(v_{u})^T I_{c_{pu}} + b_{e_{pu}})}{\text{size of the items pool}}
\]

The experimental results shown in Table. 4 demonstrate that all of the recommenders that leverage the Photoshop usage traces and two-phase recommendation framework perform significantly better than the baseline in terms of Recall@K and AUC. For top-ranked items (Recall@100), in particular, our recommender outperforms the popularity-based recommendation by 21.2%, which means that the users will potentially appreciate 21.2% more items among which we recommend. Also, the performance improvement suggests that we are able to personalize item recommendations to creative professionals who are new to the Behance platform.

5.3 Inspiration engine

In this section, through a sample application named inspiration engine, we demonstrate that the data integration can also enable innovative user experiences. The goal of inspiration engine is to provide real-time and personalized inspirations for creative professionals when they are working in Photoshop, and the system is able to show the potential outcomes of the actions that have been or are likely to be performed. Such presentations are inspiring because the artists can explore a wider range of possibilities that are related but different from their current work.

Technically speaking, the core component of such an application is a search engine that returns art projects likely to be produced by a given sequence of Photoshop actions. We build the system by leveraging the weak correspondence between the pairs of users’ Photoshop usage traces and the projects that they uploaded to Behance. With such pairs, we can learn a heterogeneous joint embedding where the true actions-image pairs are close to each other, and the false pairs are further away. As shown in Fig. 4 for each actions-image pair \((a_1, a_2, ..., a_n, e)\), we first extract features for the action sequence and the image respectively, denoted as \(v_i\) and \(z_i\). In our prototype, we extract \(v_i\) from util2vec and \(z_i\) from the pooling layer (2048 dim) of pre-trained ResNet [11], the state-of-the-art image feature extractor. Then we learn a function \(g\) to project image features \(z_i\) to the util2vec embedding space such that the objective \(\sum ||v_i - g(z_i)||^{2}\) is minimized.

To prototype the system, we train a linear projection function \(g\) with 353,205 actions-image pairs from 43,441 users and validate it over 20,000 held-out pairs from 20,000 users, i.e., each user contributes exactly one pair in the validation. There is no user overlap in the training and validation set, and the training is conducted using bfgs algorithm. Quantitatively, we use Recall@K and AUC as defined in Section 5.2 to evaluate the system performance, and the results are shown in Table. 5. The improvements over the random guess baseline justify that there is a close relationship between the Photoshop usage pattern and visual appearance of art project. In addition, in Fig. 6 we show four qualitative retrieval results of the inspiration engine (we only show retrieval with a single action, but our technique is applicable to action sequence as well). The nearest neighbors of each action reflect the scenarios where it is often used. For example, the action \(\text{drag\_path}\) is heavily used in web design, and the \(\text{rotate\_canvas}\) is typically leveraged to create repetitive patterns. We will conduct an end-to-end further user study in the future to evaluate the engine.

Through three applications, we observe that the amount of improvements brought by util2vec, compared to the intuitive \(\text{bag-of-actions+tfidf}\) model, are contingent on the context of end applications. Nevertheless, the performance improvements are significant under most of the metrics except Recall@300 in the cold-start recommendation task, so we can safely conclude that util2vec is beneficial in modeling unstructured application usage traces, and we may get further gains in the future by tuning the model parameters and training methods.

6. CONCLUSIONS AND FUTURE WORK

In this work, we personalize software and web applications for creative professionals by leveraging Photoshop usage traces. These systems can enhance existing services provided by Photoshop (i.e., accurate prediction of users’ areas of focus, Section 5.1), and Behance (i.e., personalized recommendation for cold-start users, Section 5.2), and enable new experiences (i.e., inspiration engine, Section 5.3) for millions of users.

Although we focus on platforms for creative professionals, our results suggest that such an integration may be fruitful for personalization research more generally. For example, personalized applications can be built for programmers based on their Github usage records, and for journalists based on the usage of document editing tools. As people’s work and leisure lives are increasingly accompanied by applications, understanding and integrating digital breadcrumbs that they leave behind can lead to truly user-centric personalization.

7. ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their insightful comments. This work is supported by Adobe Research gift funding and The AOL-funded Connected Experiences Lab. The first author conducted part of this research at Adobe Research as a summer intern and is further supported by the small data lab at Cornell Tech, which receives funding from NSF, NIH, RWJF, MacArthur Foundation, Google, and UnitedHealth Group.
8. REFERENCES


