Collaborative Filtering of Color Aesthetics

Peter O'Donovan
University of Toronto

Aseem Agarwala
Adobe

Aaron Hertzmann
Adobe and University of Toronto

Abstract

This paper investigates individual variation in aesthetic preferences, and learns models for predicting the preferences of individual users. Preferences for color aesthetics are learned using a collaborative filtering approach on a large dataset of rated color themes/palettes. To make predictions, matrix factorization is used to estimate latent vectors for users and color themes. We also propose two extensions to the probabilistic matrix factorization framework. We first describe a feature-based model using learned transformations from feature vectors to a latent space, then extend this model to non-linear transformations using a neural network. These extensions allow our model to predict preferences for color themes not present in the training set. We find that our approach for modelling user preferences outperforms an average aesthetic model which ignores personal variation. We also use the model for measuring theme similarity and visualizing the space of color themes.

CR Categories: I.3.0 [Computer Graphics]: General

Keywords: color, design, machine learning, collaborative filtering, aesthetics

1 Introduction

Understanding preferences for visual aesthetics is an important goal for many industries, such as advertising, fashion, and design. Individual preference models could be used for a variety of tasks, such as targeted online advertisements, improved results for visual search engines like Google or Flickr, and better design tools. For example, given a set of rated graphic designs for a user, a design tool could suggest design modifications, such new colours, fonts, or layouts, based on learned preferences. Unfortunately, there is currently little understanding of individual preferences; most methods for evaluating aesthetics currently predict a single value for all users. By contrast, online recommender systems such as Netflix utilize sophisticated techniques to model individual preferences. In this work, we adapt these techniques for visual aesthetics, and demonstrate their use in modelling preferences for color themes.

Color themes are a common shorthand used by designers to describe the palettes of graphic designs, photographs, and fashion outfits, and have been used for several research projects in these domains [Lin et al. 2013; Wang et al. 2010; Lin and Hanrahan 2013; Yu et al. 2012]. O'Donovan et al. [2011] used large-scale datasets of color themes to evaluate and learn models of color compatibility. In that work, a theme rating is the average of all user ratings, and a linear regression model is used to predict ratings. This approach of 'averaging aesthetics' is standard for predicting aesthetic ratings, and has been used for photographs [Datta et al. 2006; Marchesotti et al. 2011], paintings [Li and Chen 2009], and videos [Moorthy et al. 2010]. This approach is reasonable for several reasons. Firstly, it models the overall trend in aesthetics, which can be useful. Second, the learned models are often simple and interpretable. Lastly, subjective preferences are extremely noisy, so averaging reduces noise which makes learning easier. However, this it ignores any subjective variation in ratings due to personal preference. For color aesthetics in particular, subjective preferences are common. The wide variety of color palettes found in clothing and interior design speaks to the extent of individual color preferences. In this work, we use a collaborative filtering (CF) approach to predict per-user ratings for color themes. We show that this approach outperforms an 'average rating' model by a wide margin, indicating the usefulness of modelling individual aesthetic preferences.

One common approach to CF performs a matrix factorization on the rating matrix into latent vectors for users and items. However, this approach is limited in two respects. First, it ignores features of the items, which are often very important for aesthetic items like color themes or photographs. Items with very few ratings also benefit highly from features, as their latent vectors are underconstrained. Second, the model cannot predict ratings for novel items which are not present in the training data. To address these limitations, we use a feature-based approach based on probabilistic matrix factorization [2008b]. We extend this model to learn a latent linear transformation from features, instead of learning a per-theme latent vector. We then extend the model to handle non-linear transformations using a neural network. Feature-based CF methods are not new, but we show that for visual aesthetics, a feature-based approach significantly outperforms the standard approach without features, and can predict aesthetic ratings for themes not seen during training.

We also show that the learned model is useful for other aesthetic tasks. We use the model for measuring the distance between color themes, as nearby themes in the learned latent space have similar styles. For example, two themes with random hues may have a high pixel distance (as measured by comparing individual colours), but have a similar aesthetic style. An analogy with images would be red-eye removal: the before/after images are almost identical in pixel space, but have a high aesthetic distance. We also perform dimensionality reduction using t-SNE [van der Maaten and Hinton 2008] to visualize the space of color themes. We show that the learned latent space improves the embedding, with similarly rated themes clustered together.

More broadly, our work is among the first to train models of individual preference in visual aesthetics. While we examine color aesthetics in particular, our approach could be used for making personalized recommendations for images, videos, or graphic designs.

2 Related Work

2.1 Predicting Aesthetics

While color preferences are important to many industries, aesthetic preferences for colours and color combinations remain poorly understood. In recent decades, psychologists have begun controlled studies of color compatibility and preferences [Granger 1952; Ou and Luo 2006; Szabó et al. 2010; Schloss and Palmer 2010]. Preferences of demographic groups are often investigated, though some researchers investigate individual preferences [Palmer and Griscom 2013]. However, data for this work comes from tightly-controlled laboratory experiments, which forces a small number of participants (usually less than 100), a small range of colors (usually less than 100), and a small number of combinations (usually 1-3). Recently, O'Donovan et al. [2011] explored color compatibility and color preferences using large sets of themes from Kuler and ColorLovers, and MTurk experiments. The aesthetic model was a simple linear regression of features to predict the mean rating of a color theme, i.e., the average rating over all users. O'Donovan et al. also cluster similar users and learn regressors for the clusters, though this approach does not learn an individual model of preference.

While it is well-known that individual differences exist in aesthetic preferences [Martindale et al. 1990], the approach of 'averaging aesthetics' is quite common. Large datasets of visual objects are increasingly used to train aesthetic models, including photographs [Datta et al. 2006; Marchesotti et al. 2011], Impressionist paintings [Li and Chen 2009], and videos [Moorthy et al. 2010]. In all these cases, ratings are averaged over all users to compute an overall prediction of aesthetic quality. Wu et al. [2011] present a structured SVM model which learns a distribution of ratings, though not user preferences. By contrast, we predict theme ratings for individual users using a collaborative filtering approach, and show that this approach substantially improves on averaged predictors.

Reinecke and Gajos [2014] recently examined website aesthetics for demographic groups including gender, age, education, and nationality. They found significant variation between these groups, and learn a model to predict user preferences based on demographic features. In this work, we also use demographics to improve prediction, but learn a per-user latent vector, as prior work has suggested personal preferences outweigh demographic preferences in color aesthetics [O'Donovan et al. 2011].

Researchers have also investigated individual preferences for image enhancement parameters. Kapoor et al. [2013] use a CF approach to predict the desired color correction and tonal adjustments for a user, given a small number of training adjustments. Bychkovsky et al. [2011] similarly predict the tonal adjustment style of expert photographers. Our work differs however, since we model ratings of aesthetic preference, not image enhancement parameters.

2.2 Collaborative Filtering

Over the last decade, recommender systems, such as Netflix, have made significant progress in modelling individual preferences. A common approach is collaborative filtering (CF), where large datasets of preferences for many users are used to aid personalized predictions. We next introduce the relevant research from the CF literature. For a survey, see Su and Khoshgoftaar [2009].

Matrix Factorization. Collaborative filtering often involves a set of ratings for items by users. One common approach uses latent factors, decomposing the rating matrix into the product of two matrices: a matrix U modeling each user, and a matrix V modeling each item. Salakhutdinov and Mnih [Salakhutdinov and Mnih 2008b] presented a simple probabilistic framework, later extended to a full Bayesian model [Salakhutdinov and Mnih 2008a]. The distance between latent vectors can also be used to model relationships between objects. For example, Latent Semantic Analysis [Landauer and Dumais 1997] models the similarity between documents. We use this approach to model similarity between color themes.

Feature-based Collaborative Filtering. One limitation of simple factorization approaches is they ignore valuable features about items. For movies, the date, director, and country are all highly informative. Furthermore, they cannot generalize to unseen items since latent vectors are independent. While using item features has a long history in the collaborative filtering literature [Prem Melville and Nagarajan 2001; Basu et al. 1998], features are less common in matrix factorization approaches. Chen et al. [2011] define a matrix factorization framework which uses item and user features, as well as global features. Our work is similar to this approach, but we present a probabilistic model for features which extends the PMF model of Salakhutdinov and Mnih [Salakhutdinov and Mnih 2008b]. Adams et al. [2010] also extend the PMF framework with Gaussian Process (GP) priors defined over the latent vectors using features. Our work also uses item features, but with a much simpler model: a single-layer neural network that learns a transformation from input features to latent features. Collaborative filtering problems often have tens of thousands of users and items, so GPs are problematic due to their large memory requirements.

3 Color Theme Dataset

We next present a short overview of the MTurk dataset and features from O'Donovan et al. [2011]. In this dataset, 13,343 color themes were randomly from the Adobe Kuler website. Each theme was then rated on a scale of 1-5 stars by 40 participants on Amazon's Mechanical Turk, producing a final dataset includes 528,106 individual ratings. Participants also reported their gender, age, color experience, and nationality.

Following the work of O'Donovan et al., we compute features in several different color spaces: RGB, CIELab, HSV, and CHSV[1]. In each color space, we compute the following features: theme colors, colors sorted by lightness, differences between adjacent colors, sorted color differences, mean, standard deviation, median, max, min, and max minus min across a single channel. We also include plane-fitting features where a 2D plane is fit to the 3D color coordinates using PCA. Lastly, features related to hue entropy (roughly, the spread of hues along the color wheel), and hue probabilities, both unary and for adjacent colors, are included. The final set of 334 features is then normalized to the range 0...1. See O'Donovan et al. [2011] for a full description of these features.

4 Feature-based Matrix Factorization

4.1 Probabilistic Matrix Factorization

We first briefly describe the probabilistic matrix factorization model of Salakhutdinov and Mnih [2008b]. This approach uses a set of M items, N users, and integer rating values from 1 to V. R is a matrix of ratings, usually incomplete, where R_{ij} represents the rating of user i for item j. We first define a latent vector for each user i and item j as U_i and V_j respectively, and model R as the product of the user and item latent vectors. That is, each rating is defined as $R_{ij} = U_i^T \cdot V_j$. The set of all user vectors is given by the matrix U (of dimension $N \times K$), and the item vectors as V (of dimension $M \times K$). The parameter K determines the size of the latent space.

We define the conditional distribution over the observed ratings as:

$$p(R|U,V,\sigma^2) = \frac{1}{2}\prod_{i=1}^{N}\prod_{j=1}^{M}[N(R_{ij}|U_i^T V_j,\sigma^2)]^{I_{ij}} \qquad (1)$$

where is $N(x|\mu,\sigma)$ is a Gaussian distribution with mean μ and variance σ and I_{ij} is the indicator function that is equal to 1 if user

[1]A space where hue θ and saturation s are remapped to Cartesian coordinates: $d_1 = s\cos(\theta)$ and $d_2 = s\sin(\theta)$.

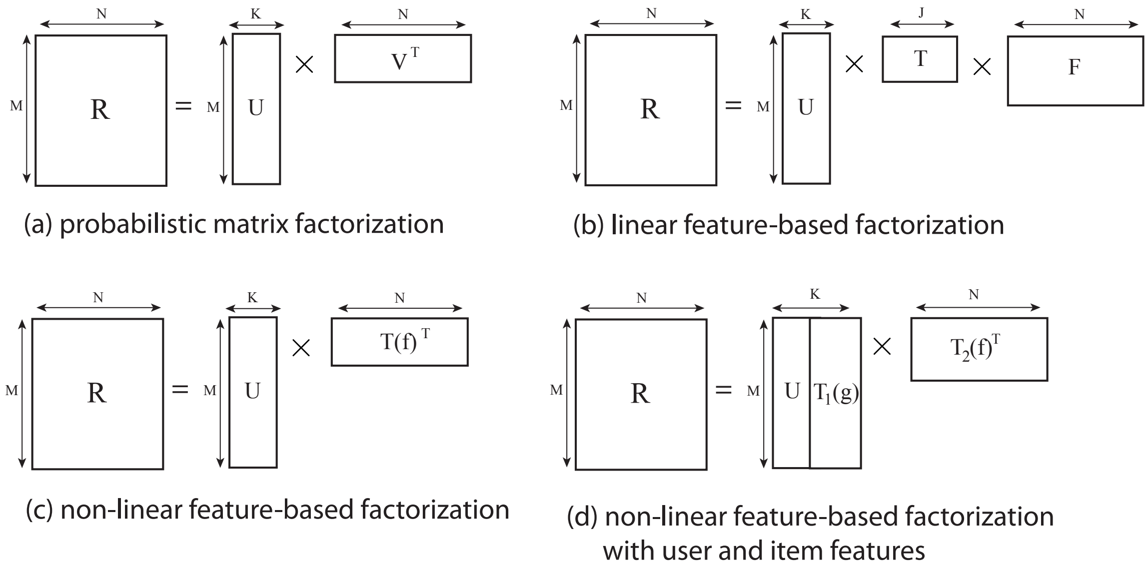

Figure 1: ***Factorization Models***. *(a) standard probabilistic matrix factorization learns latent vectors for each user **U** and each item **V**. (b) feature-based matrix factorization learns a linear transformation **T** from fixed item features **F** to the latent space. (c) features are transformed to the latent space using a neural network. (d) both fixed latent features and non-linear feature transformations are used to model users.*

i rated item j and equal to 0 otherwise. Gaussian priors are also defined for U_i and V_j:

$$p(U|\sigma_U^2) = \prod_{i=1}^{N} N(U_i|0, \sigma_U^2 \mathbf{I}) \quad p(V|\sigma_V^2) = \prod_{i=1}^{N} N(V_i|0, \sigma_V^2 \mathbf{I}) \tag{2}$$

MAP estimation is then used to learn the latent vectors for items and users. The log posterior of Eqn 1 and 2 is used to define a sum-of-squared-errors objective function:

$$\begin{aligned} E(U,V) &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}(R_{ij} - U_i^T V_j)^2 \\ &\quad + \alpha_U \sum_{i=1}^{N} ||U_i||_{Fro}^2 + \alpha_V \sum_{j=1} ||V_j||_{Fro}^2 \end{aligned} \tag{3}$$

The gradients with respect to U_i and V_j are simple to compute, and training done by gradient descent; please see Salakhutdinov and Mnih [2008b] for details. In our experiments, we set the dimensionality of the latent space to $K = 5$ based on a validation set of ratings, described in the next section.

4.2 Linear Feature-based Matrix Factorization

A major disadvantage of the previous approach is it ignores item information which could help rating estimation. For color themes, or other visual stimuli, this information is important for prediction. Another disadvantage is poor generalization; ratings for new items not present in the training data cannot be estimated.

Given a feature vector F_j for each item j, we can learn a mapping from feature space to latent space. We first present a linear transformation T of the feature vector:

$$\begin{aligned} E(U,T) &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}(R_{ij} - U_i^T \cdot (TF_j))^2 \\ &\quad + \alpha_U \sum_{i=1}^{N} ||U_i||_{Fro}^2 + \alpha_T \sum_{k=1} ||T_k||_{Fro}^2 \end{aligned} \tag{4}$$

The gradients with respect to U and T are again straightforward, and training is done with gradient descent. The matrix T is of size $Q \times K$, where $Q = 334$ and $K = 15$.

4.3 Non-linear Feature-based Matrix Factorization

We can also define a non-linear transformation function $T(F_j; W)$ with parameters W:

$$\begin{aligned} E(U,W) &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}(R_{ij} - U_i^T \cdot T(F_j; W))^2 \\ &\quad + \alpha_U \sum_{i=1}^{N} ||U_i||_{Fro}^2 \end{aligned} \tag{5}$$

The non-linear transformation is a neural network trained using back-propagation. When learning the parameters, the user vectors U and the network parameters W are updated alternately; the parameters are fixed for one set and the gradient calculated for the other. The users' latent vectors are updated at each iteration as before. The users' latent vectors act as a final layer of linear weights on the neural network, with the errors are back-propagated through the network to W.

There is also no prior on the network weights. While such regularization is trivial to add, we expect our simple model shared over our hundreds of thousands of datapoints will be robust to overfitting. Therefore, sparsefying or penalizing higher weights may rule out good transformations. In practice, small weights are learned with little overfitting (Fig. 2). Initial experiments with regularization revealed no improvement. The neural network included 200 logistic units, and the dimensionality of the latent space was $K = 15$, set using validation ratings. We found the performance of the FPMF models were fairly robust to parameters changes; results did not change significantly.

A further extension is to add feature-based latent vectors for users as well. Each user in our dataset self-reported gender, experience, country, and age. We can therefore use these binary features with a second transformation:

$$E(U, W_1, W_2) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\left(R_{ij} - [U_i T_1(G_i; W_1)]^T \cdot T_2(F_j; W_2)\right)^2 + \alpha_U \sum_{i=1}^{N} ||U_i||^2_{Fro}$$

Where the vector $[U_i T_1(G_i; W_1)]$ is the concatenation of the user's latent vector U_i with the output the neural network T_1 given the user features G_i and parameters W_1. We use 200 logistic units for the item-feature network, and 50 logistic units for the user-feature network. U_i has dimension 15, and T_1 has dimension 5.

5 Experimental Results

The first baseline we compare our CF approach against is an average aesthetic model. This model is trained on the average of all training ratings for a theme. Testing is however done for each individual rating, not the average. This baseline indicates how much individual user preferences affect the rating. We use the approach of O'Donovan et al. [2011]: linear regression with an L1-norm [Tibshirani 1996]. A training set of 300,000 ratings was used, with a testing set of 128,106 ratings. A separate validation set of 100,000 ratings was used to select model parameters.

We also compare our feature-based models with regular PMF to evaluate how important features are for modeling visual aesthetics. As mentioned earlier, one important advantage of feature-based models is the ability to handle test themes not seen in training. We therefore test on a dataset ('Novel') where all test ratings are for new themes. Feature creation can be time-consuming and often requires expert knowledge. We therefore also explore the models' performance with a reduced feature set. With a smaller set of features, we expect non-linear FPMF should perform better than linear FPMF, as the non-linear transformation should compensate for less hand-crafted features. We therefore test a feature set ('Reduced') with only the 15 CIELab colors, and compare to the full 334-dimensional feature vector.

Table 1 shows our main result: the error for the average predictor is substantially higher than those which model individual user preferences. We also show the value of using features when modeling visual aesthetics, as the feature-based FPMF model performs much better than PMF at predicting theme ratings. Non-linear FPMF also out-performs linear FPMF, with better relative performance with fewer features. However, adding demographic user features only gives a very small improvement. Fig. 2 plots the error of the validation set.

We next investigate the effects of user modelling and demographic features. As a baseline, we trained the non-linear FPMF model with a constant U_i for all users and no demographic features, using the 'Novel' dataset. This model gave a RMSE of 1.079, as compared to 0.828 for the non-linear FPMF with user modelling, again demonstrating its value. Note that this approach closely matched the RMSE of 1.081 for the averaged predictor of O'Donovan et al. [2011], which also ignores user modelling.

Method	Seen	Reduced	Novel
Averaged	1.082	1.107	1.081
PMF	0.964	0.964	-
Linear FPMF	0.842	0.969	0.841
NL FPMF (V)	0.831	0.945	0.829
NL FPMF (U+V)	0.829	0.944	0.828

Table 1: ***Model Testing.*** *We evaluate various models using the RMSE of test theme ratings. 'Averaged' is a linear regressor trained on mean theme ratings ([O'Donovan et al. 2011]) Non-linear FPMF (V) uses a neural network with theme features. Non-linear FPMF (U+V) uses a neural network with user and theme features. 'Seen' and 'Reduced' include previously seen users and themes. The 'Novel' set has no themes used in training. 'Reduced' uses only 15 features (the theme's CIELab color values); 'Seen' and 'Novel' sets use the full 334-dimensional feature vector.*

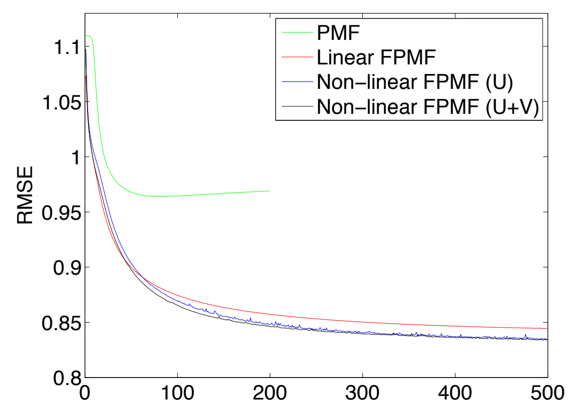

Figure 2: ***RSME of validation set during training epochs***

We then tested using only demographic features to model the user. Specifically, we removed the latent vector U_i from Eqn. 6, and modelled users only by the neural network $T_1(G_i)$, where G_i are the demographic features of user i. This model produced a RMSE of 1.066, suggesting that while demographic features are informative, they are far less important than modeling individual preferences for color themes. The marginally better performance with demographic features is slightly surprising. Previous research on webpage aesthetics, including colourfulness, found significant differences between demographic groups [Reinecke and Gajos 2014]. One reason may be that Reinecke's dataset included a broad sampling of countries, whereas the vast majority of the MTurk color dataset are from USA or India. It is also likely that color theme preferences have more variation within groups than across them, particularly compared to webpage aesthetics.

In Fig. 3, we show a concrete example for two users with different aesthetic styles. We show highly and poorly rated themes for the two users, along with predicted ratings for new themes using the non-linear FPMF model, demonstrating that our CF model accurately captures the users' aesthetic preferences. Our method can also predict ratings distributions, by predicting ratings for all users

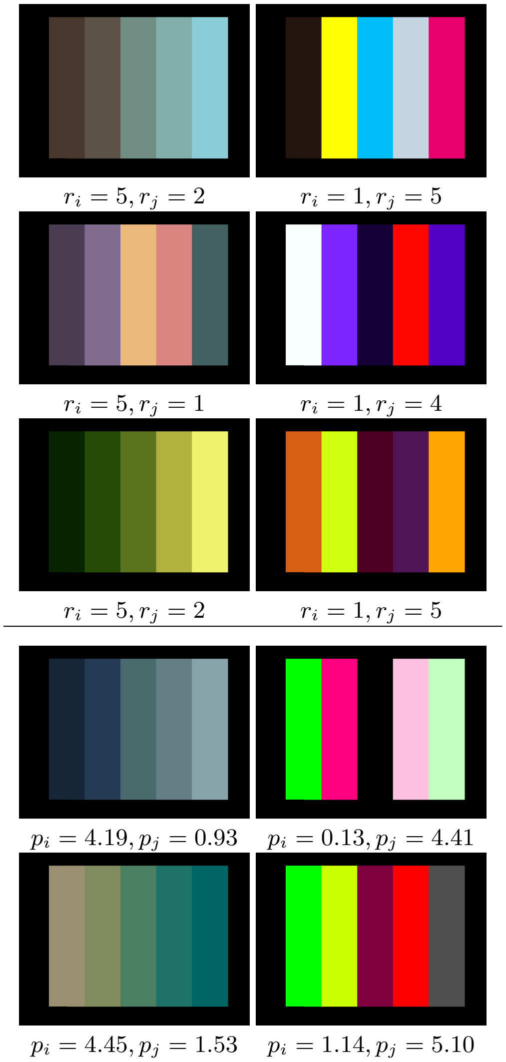

Figure 3: ***Collaborative Filtering Example.*** *The top three rows show highly and poorly rated themes from two users (i and j) with different aesthetic preferences. The ratings for the two users are denoted as r_i and r_j. The bottom two rows shows our predicted ratings p_i and p_j for new themes using non-linear FPMF.*

in the training set. In Fig. 4, we show the distribution for two novel testing themes not seen in the training data. This figure shows that there can be large differences between distributions; the variance of the top theme is much higher, indicating more disagreement in ratings than the bottom theme.

6 Applications

Navigating the space of color themes is a difficult problem with little previous work. User-specified tags are often used for searching similar themes (e.g., 'pastel', 'venice' , 'stone', 'rose') but this approach is limited. The main problem is the lack of a distance metric for themes. We wish to find 'similar' themes, but similarity is poorly understood for color combinations. One simple solution is to take the sum of color differences in a perpetually uniform color space like CIELab. However, this naive approach does not model the relationships between colours, or the overall style of the theme. For example, a color theme which lies along a gradient (for ex, dark to light) should be closer to the flipped theme (i.e., light to dark) than to a random permutation of the colors which does not preserve the gradient, though it may have a lower CIELab distance. An analogy for images would be before and after red-eye removal. While both images are extremely similar in pixel distance, they have a large aesthetic disparity.

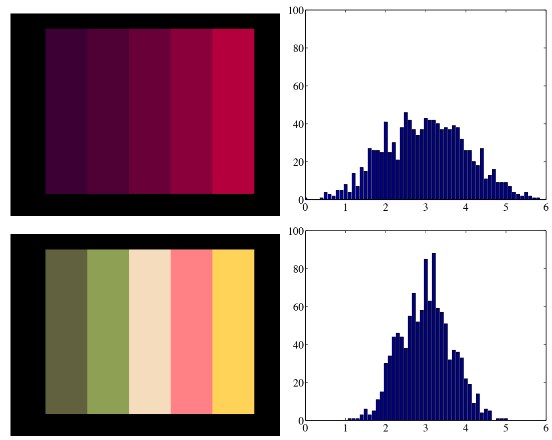

Figure 4: ***Predicting Rating Distributions.*** *Given a novel theme (i.e., one not present in the training set), we predict the ratings for all users in the training set, and plot the distribution of their ratings. In this example, both themes have a mean rating of 3.00, but the top theme has greater disagreement (std. dev. of 1.03 vs 0.64).*

Instead of color differences, we propose a similarity metric for color themes which measures differences in aesthetic style. Unfortunately, specifying such a distance is not intuitive. However, theme ratings can be used as a proxy for measuring aesthetic distances; themes which are aesthetically similar will tend to have similar ratings. We would like a transformation for themes such that, in this new space, a small distance results in a small rating difference. FPMF produces such a transformation, incorporating aesthetic differences and grouping similarly rated themes. Similar latent factor approaches have been used to detect synonyms [Landauer and Dumais 1997], and to visualize similar movies [Koren et al. 2009].

In Fig. 5 we next show several themes with a large CIELab distance but a small distance in the FPMF latent space, and vice-versa. Since the scales of the spaces are different, we also report the distance sort order in each space. That is, for each theme, we first calculate the distances to every other theme. These distances are then sorted, and the order number reported. A value of 0 indicates the second theme is the closest theme to the first in this space. A value of 1 is the most distant. This metric gives a relative sense of the distances.

In the top three examples, we show themes which are visually quite similar have a large CIELab distance. By contrast, the latent distance is much smaller. In the top example, the two hues are switched; in the second, the gradient is reversed; in the third, both themes are poorly ordered with bright primary colors. In the bottom three examples, themes with a small CIELab distance are visually quite distinct, which is reflected by a larger latent distance. A naive CIELab distance does not account for contrast between colors. If two themes have similar lightness and saturation, they will have a fairly low CIELab distance. However, a single modified hue can greatly decrease the perceived similarity and aesthetic rating.

To visualize the space of color themes, we use t-SNE [van der Maaten and Hinton 2008] to create a 2-D embedding. In Fig. 6, we compare an embedding using the CIELab distance with one

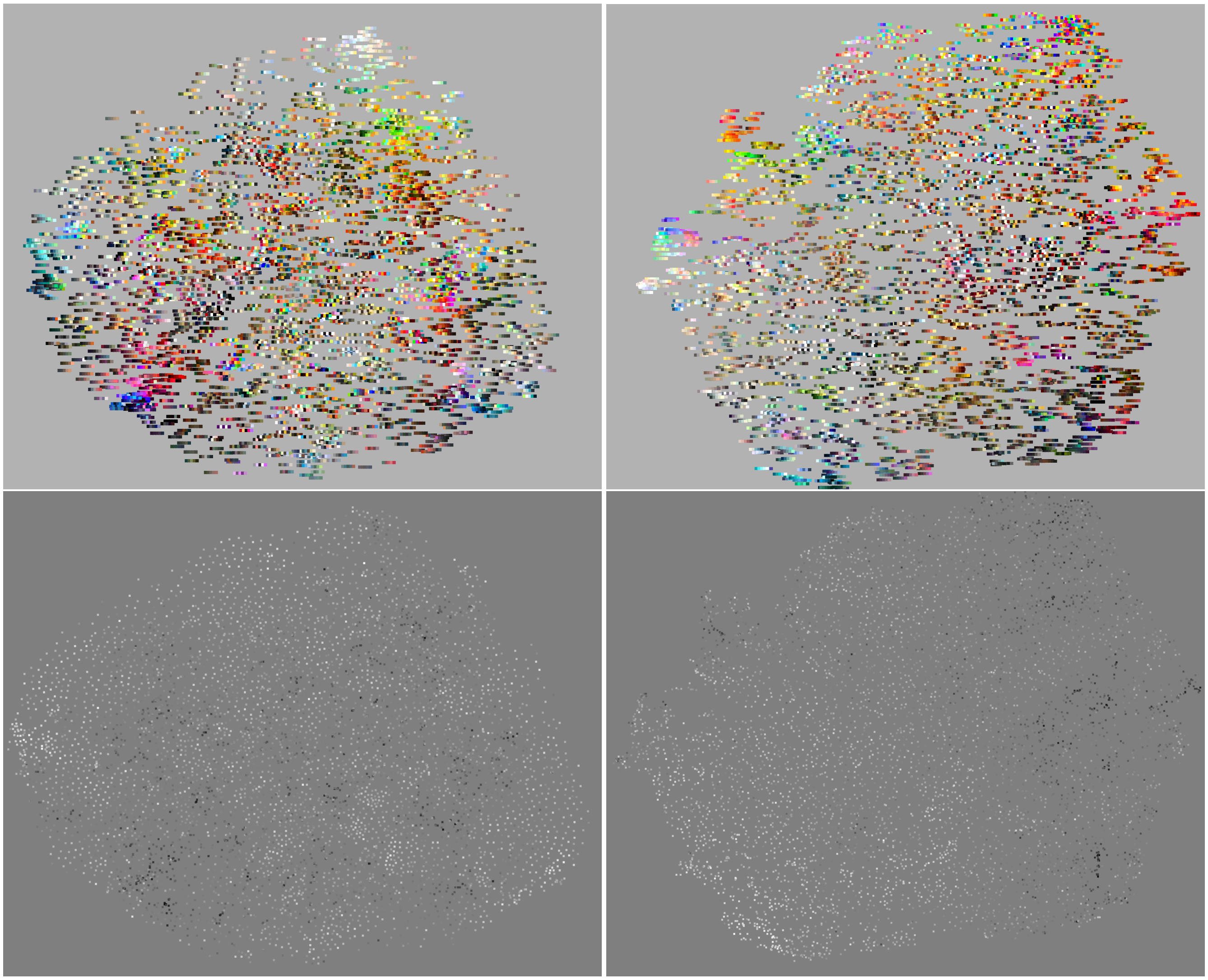

Figure 6: ***t-SNE embedding of 2000 color themes. Top left:*** *Embedding of CIELab color values.* ***Top right:*** *Embedding of FPMF latent features.* ***Bottom left:*** *Mean user ratings for each theme (CIELab embedding).* ***Bottom right:*** *User ratings of FPMF embedding.* ***Please zoom in for detail.*** *The FPMF embedding clusters similarly rated themes better than the CIElab embedding.*

using the latent vectors (please zoom in for more detail). While both embeddings lack clear clusters, the results are improved with FPMF in several ways. First, there is an overall diagonal light to dark trend with FPMF not present in the CIELab embedding. Second, bright themes with significant color variation (e.g., the third theme of Fig. 5) are clustered in the top right whereas they are spread out in the CIELab embedding. Particular hues are also better grouped (e.g., the blue theme of Fig. 5). We also plot the embeddings with the mean user ratings in Fig. 6 (bottom). This shows that similarly rated themes are being placed closer together using the latent vectors; with CIELab distances, poorly rated themes are spread throughout.

We can also use t-SNE to visualize users. Fig. 7 show 2D embeddings of users' latent vectors, coloured by demographic features. The figure does show some degree of clustering, indicating users with similar preferences. Some clusters are predominately of one country, though the map fails to show a clear separation between users of different countries. There is also little separation of users based on their gender. We also tried labelling the users by their age, but there was similar degree of inter-group variation. These findings reinforce the claim that differences in color preferences between demographic groups are lower than differences within the groups.

7 Conclusion

Modeling aesthetic preferences is an exciting new area with many potential applications from music, to image processing, to fashion and design. Large-scale datasets also offer the opportunity for greater understanding of aesthetic preferences. To our knowledge, collaborative filtering approaches have not been explored previously for modeling aesthetic ratings. Previous approaches average over all ratings to measure an overall aesthetic score. This approach is appropriate when no information is available about a new user. However, when previous information is available, modeling individual user preferences can achieve significantly better performance than average aesthetic models.

In our work, we use a feature-based probabilistic matrix factorization (FPMF) model to predict individual user ratings. We introduce

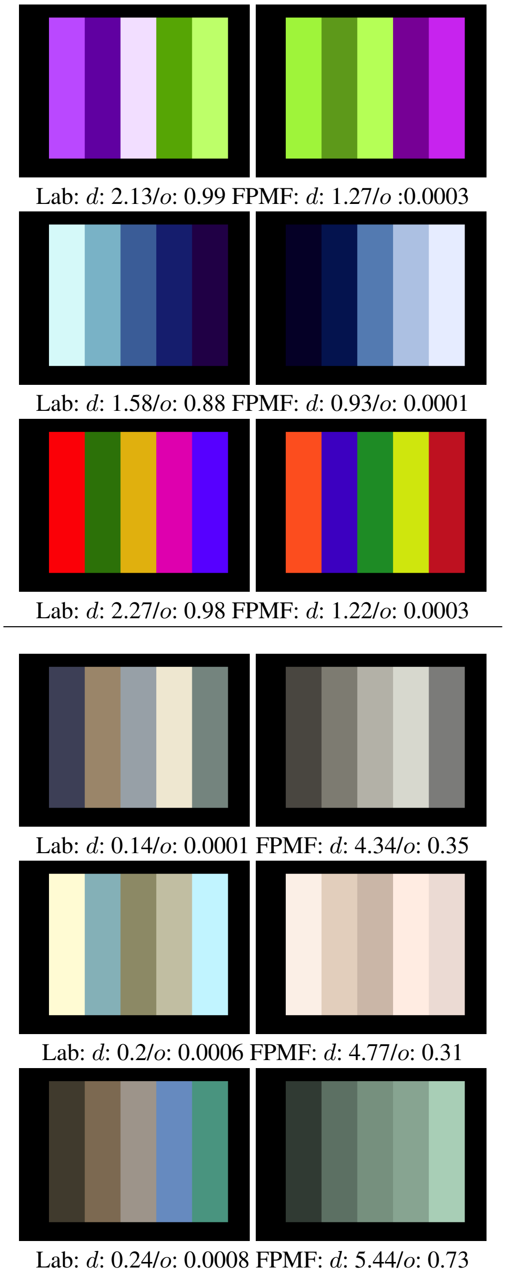

Figure 5: ***Distances Between Themes. Top:*** *themes with a large CIELab distance but small FPMF latent distance.* ***Bottom:*** *themes with a small CIELab distance but large latent distance. The (d)istances are computed in CIELab and FPMF. Note that distances in CIELab and FPMF are not directly comparable as they are different spaces (i.e., a distance of 1 is not equivalent in both spaces). To compare the different spaces, we report the sorted distance (o)rder for all themes. 0 indicates no other theme is closer, 1 indicates no theme is farther away. These results indicate the latent features are better for measuring visual similarity.*

two simple extensions to the original PMF framework. First, instead of solving for a latent vector for each color theme, we solve for a transformation from theme features to the latent space. Second, we propose a non-linear transformation within the factorization using a neural network. We show a feature-based approach significantly

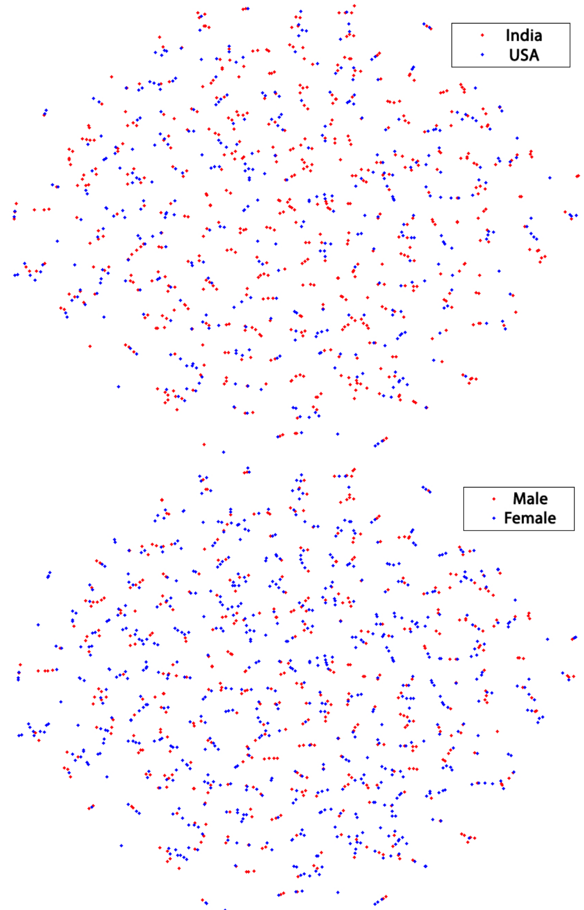

Figure 7: *t-SNE embedding of users with country-of-origin and gender labels. While there is some clustering of user preferences, there is substantial variation within the demographic groups.*

outperforms one which ignore features. We also show the model's usefulness for understanding and visualizing color themes. Latent factor transformations can measure the aesthetic distance between visual stimuli which can be difficult to specify directly. We also use this representation to visualize the space of color themes. Given the vast datasets of color themes and images available online, building interfaces which use aesthetic models to help navigate these spaces is an exciting area of research.

One important application of our model is in improved design and photography tools. Color themes are commonly used to describe the color palettes of graphic designs, ranging from websites to posters, as well as photographs. Our approach could be used to make recoloring suggestions for designs or photographs which match user preferences. Finally, while this work examines color aesthetics specifically, our approach could easily be applied to making recommendations beyond colors, to photographs, videos, or graphic designs. These domains have large datasets and rich feature sets, and are therefore applicable to our FPMF model. Personalized recommendations for photographs or graphic designs have the potential to greatly improve numerous industries, such as online advertising and search.

References

ADAMS, R. P., DAHL, G. E., AND MURRAY, I. 2010. Incorporating Side Information Into Probabilistic Matrix Factorization Using Gaussian Processes. In *Proc. UAI*.

BASU, C., HIRSH, H., AND COHEN, W. 1998. Recommendation as Classification. In *Proc. AAAI*.

BYCHKOVSKY, V., PARIS, S., CHAN, E., AND DURAND, F. 2011. Learning Photographic Global Tonal Adjustment with a Database of Input / Output Image Pairs. In *Proc. CVPR*.

CHEN, T., ZHENG, Z., LU, Q., Z. W., AND YU, Y. 2011. Feature-based Matrix Factorization. Tech. rep., Apex Data and Knowledge Management Lab, Shanghai.

DATTA, R., JOSHI, D., LI, J., AND WANG, J. Z. 2006. Studying Aaesthetics in Photographic Images Using a Computational Approach. In *Proc. ECCV*, 7–13.

GRANGER, G. W. 1952. Objectivity of Color Preferences. *Nature 170*, 4332.

KAPOOR, A., CAICEDO, J., LISCHINSKI, D., AND KANG, S. 2013. Collaborative Personalization of Image Enhancement. *IJCV*.

KOREN, Y., BELL, R., AND VOLINSKY, C. 2009. Matrix Factorization Techniques for Recommender Systems. *IEEE Computer*.

LANDAUER, T. K., AND DUMAIS, S. T. 1997. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*.

LI, C., AND CHEN, T. 2009. Aesthetic Visual Quality Assessment of Paintings. *Sel. Topics in Signal Processing 3*, 2.

LIN, S., AND HANRAHAN, P. 2013. Modeling How People Extract Color Themes from Images. In *Proc. CHI*.

LIN, S., RITCHIE, D., FISHER, M., AND HANRAHAN, P. 2013. Probabilistic Color-by-Numbers: Suggesting Pattern Colorizations Using Factor Graphs.

MARCHESOTTI, L., PERRONNIN, F., LARLUS, D., AND CSURKA, G. 2011. Assessing the Aesthetic Quality of Photographs Using Generic Image Descriptors. In *Proc. ICCV*.

MARTINDALE, C., MOORE, K., AND BORKUM, J. 1990. Aesthetic Preference. *The American Journal of Psychology*.

MOORTHY, A. K., OBRADOR, P., AND OLIVER, N. 2010. Towards Computational Models of Visual Aesthetic Appeal of Consumer Videos. In *Proc. ECCV*.

O'DONOVAN, P., AGARWALA, A., AND HERTZMANN, A. 2011. Color Compatibility from Large Datasets. *ACM Trans. Graphics (Proc. SIGGRAPH)*.

OU, L.-C., AND LUO, M. R. 2006. A Color Harmony Model for Two-Color Combinations. *Col. Res. Appl 31*, 5, 191–204.

PALMER, S., AND GRISCOM, W. 2013. Accounting for Taste: Individual Differences in Preference for Harmony. *Psychon Bull Rev.*.

PREM MELVILLE, R. M., AND NAGARAJAN, R. 2001. Content-boosted Collaborative Filtering. In *Proc. SIGIR Workshop on Recommender Systems*.

REINECKE, K., AND GAJOS, K. Z. 2014. Quantifying Visual Preferences Around the World. In *Proc. CHI*.

SALAKHUTDINOV, R., AND MNIH, A. 2008. Bayesian Probabilistic Matrix Factorization Using Markov Chain Monte Carlo. In *Proc. ICML*.

SALAKHUTDINOV, R., AND MNIH, A. 2008. Probabilistic Matrix Factorization. In *Proc. NIPS*.

SCHLOSS, K. B., AND PALMER, S. E. 2010. Aesthetics of Color Combinations. In *Human Vision and Elec. Imaging*, 752719.

SU, X., AND KHOSHGOFTAAR, T. M. 2009. A Survey of Collaborative Filtering Techniques. *Adv. in Artif. Intell.*.

SZABÓ, F., BODROGI, P., AND SCHANDA, J. 2010. Experimental Modeling of Colour Harmony. *Col. Res. Appl 35*, 1, 34–39.

TIBSHIRANI, R. 1996. Regression Shrinkage and Selection Via the Lasso. *Royal. Statist. Soc B 58*, 1, 267–288.

VAN DER MAATEN, L., AND HINTON, G. 2008. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research 9*, 2579–2605.

WANG, B., YU, Y., WONG, T.-T., CHEN, C., AND XU, Y.-Q. 2010. Data-Driven Image Color Theme Enhancement. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*.

WU, O., HU, W., AND GAO, J. 2011. Learning to Predict the Perceived Visual Quality of Photos. In *Proc. ICCV*, 225–232.

YU, L.-F., YEUNG, S. K., TERZOPOULOS, D., AND CHAN, T. F. 2012. DressUp!: Outfit Synthesis Through Automatic Optimization.