Multi-armed bandits for visual preference prediction

Brennan Young
Electrical Engineering

Jiyun Ivy Xiao
Applied and Computational Math

Tobias Bischoff
Engineering and Applied Science

Abstract

This project attempts to use a multi-armed bandit approach to visual preference prediction. The idea is that different individuals show a different preference for visual features in human faces, i.e., that the preferences can be highly personal. This provides a challenge for machine learning practitioners that are concerned with visual preference prediction, as labeled data is almost always unavailable, impossible to obtain, or both. We tackle this problem by categorizing a set of images of human faces then training a UCB algorithm that treats the different categories as arms of a multi-armed bandit, where each arm has a different reward probability that is different for different individuals. We categorize the images in two different ways: 1) the images are labeled by hand in terms of gender, race, facial expression (happy or not) and 2) we use an unsupervised dimensionality reduction and clustering approach to categorize the images without labeling, thereby providing a pipeline that can process a set of images and learn individual preferences automatically. We build an interface and conduct experiments on human behavior by showing a set of about 1200 images to participants in a trial to learn their visual preferences about the images in the dataset. Lastly, we provide and outlook on the remaining challenges of this problem and present a roadmap for future developments.

1 Introduction

In this project we attempt to use a multi-armed bandit approach for visual preference prediction. The key idea is that different individuals show a different preference for visual features in human faces, i.e., that the preferences can be highly personal. The project presented should be regarded as a first experimental step towards building algorithms that can quickly learn the personal taste in images. A potential application would be dating apps, such as Tinder, that heavily rely on humans rating other humans attractiveness and whose algorithms need to very quickly learn a users personal preferences for facial beauty in order to achieve a high user retention rate. The same approach could also be used for personalized music preference learning.

We approach this problem by categorizing a set of images of human faces and then subsequently training a UCB algorithm that treats the different categories as arms of a multi-armed bandit, where each arm has a different reward probability that is different for different individuals. In order for this to work well, we need to pre-process the images. We do so by categorizing them in two different ways: (1) the images are labeled by hand in terms of gender, race, facial expression (happy or not) and (2) we use an unsupervised dimensionality reduction and clustering approach to categorize the images without labeling, thereby providing a pipeline that can process a set of images and learn individual preferences automatically. In our unsupervised pipeline is highly flexible as the different building blocks (dimensionality reduction, and clustering, and UCB) could be readily switched out for much more advanced algorithms. For the purpose of this project we apply PCA and SparsePCA as dimensionality reduction methods and use K-Means to cluster the reduced set of images into different categories. In the conclusion section, we provide our views on how to upgrade these algorithms with state of the art deep learning tools and contextual bandits.

To study our pipeline, we build an interface and conduct experiments on human behavior by showing a set of about 1200 images to participants in a trial to learn their visual preferences about the images in the dataset. Lastly, we provide and outlook on the remaining challenges of this problem and present a roadmap for future research.

2 Overview of Previous Work

Multi-armed bandit approach has been a area of substantial research and has produced many research-related applications, ranging from multi-armed bandit with similarity information [Slivkins] and multi-armed bandit with changing environment [Besbes] to explore-exploit learning with limited resources [Agarwal] and risk vs. reward trade-off study [Auer]. Among its most widely-used applications are web search and news recommendation [Li], where the multi-armed bandit approach works very well in cases where users have no history information or have fluctuating preferences. In this problem, we would like to extend the multi-armed bandit approach to image recommendation.

Visual preference prediction is a well-studied field with quite a few different methods. The state-of-the-art method is demographic estimation (age and gender classification) [Fu, Ng]. Subjective property estimation improved upon it by studying subjective visual attributes of the image [Marchesotti] and calculating aesthetic quality estimation such as "interestingness" [Dhar]. It can also be achieved using a combination of feature extraction such as SIFT, HoG filters, neural network [Girschick, Krizhevsky], and rating prediction methods such as collaborative filtering [Breese]. In a latest visual preference prediction study, Roth et al. predicted how much a person is attracted to a novel face based on his or her known preferences, using convolutional neural networks and support vector regression for extracting features, and visual regularized collaborative filtering for inferring individual preferences [Rothe]. The method had a 76% accuracy in predicting preference of a large image-set from a dating site as well as images from celebrities, with one image being the sole base of prediction.

Roth et al.'s study assumed that the preference was learned beforehand based on users' rating history. However, in many cases, the users' preference can be refined continually as they get more images. Their approach also poses difficulty when users have no previous rating history. In our problem, we will learn the user's preference real-time while recommending images. No known literature has been found on image recommendation based on real-time visual preference prediction. With human psychology that gets first impression within 0.1 second upon seeing faces [Willis], our work will very likely prove applicable in many dating apps.

3 Research Questions

Why are we doing this? There are two main issues with visual preference prediction. The first one revolves around the word "personal". Visual preferences can be highly personal and one of the questions we are interested in answering in this simple experiment is "Can the most basic UCB algorithm be use to learn and predict visual preference categories for different users efficiently?". Here, efficiently means quickly after showing a few pictures. The second question relates to the generation of visual categories, which ideally should be learned in an unsupervised fashion. Here, we compare a simple unsupervised clustering algorithm for images with hand-crafted categories for images. Lastly, we can use this to compare our findings for human test subjects to our current understanding of visual biases. For example in an experimental setup that shows uses a set of images of people from different racial and gender categories, we can use our algorithm to reveal biases in a person's perception of certain character qualities, such as trustworthiness, attractiveness, likability, etc. Under ideal circumstances this approach could be used to learn such biases quickly and efficiently. The mathematical theorems and proofs about regret bounds and other aspects of the UCB algorithm can then be used to quantify potential biases.

4 Approach

Our approach relies on the UCB algorithm and a pipeline for dimensionality reduction, outlined below. We chose UCB1 as algorithm that balances exploration and exploitation because of the strong mathematical framework that comes with it. For unsupervised dimensionality reduction, we use principle component analysis (PCA) and sparse PCA (a variant of PCA that uses lasso regularization to find a sparse set of features) in combination with K-Means clustering. For the experiment, we design an interface that shows users an image recommended by our algorithm and returns their preferences to the algorithm. We then test it on 50 test subjects and calculate the accuracy rate.

4.1 Dataset

Figure 1: Example images from the dataset.

The dataset used to study the performance of the UCB algorithm and for our experiments was obtained from the Chicago Face Database (CFD). It was developed at the University of Chicago by Debbie S. Ma, Joshua Correll, and Bernd Wittenbrink [Ma] and is intended for use in scientific research. It provides high-resolution, standardized photographs of male and female faces of varying ethnicity between the ages of 17-65. These data include both physical attributes (e.g., face size) as well as subjective ratings by independent judges (e.g., attractiveness), which we didn't use in our study, but which could be used in future work. The database contains images of 597 different individuals but for many individual it includes several images with different facial expressions resulting in 1235 different images. The gender and race label for the individual in the database were self-identified.

Figure 2: User interface.

4.2 Interface and experimental design

The user interface for our experiments contains an image box and two buttons that says "yes" or "no". The image box shows an image recommended by our algorithm. Test subjects are instructed to click "yes" if they find this person in the image "likable", and click "no" if otherwise. The selection is then returned to the algorithm to update their preferences. We then test it on a few test subjects, 200 trials per test subject, and record the algorithm's recommendations over time. The rate of correct suggestions should rise significantly after the first few trials for all test subjects. This is shown in Figure 5.2 and discussed further below.

4.3 Clustering by hand

The face dataset we used can easily be clustered by hand. It contains images of adult males and females from four racial categories (Asian, African-American, Caucasian, and Hispanic) and contains faces with neutral face expressions and with a smiles. For example, we could divide the dataset into 4 categories based on race, or 8 categories based on race $\times$ gender or into 16 categories based on race $\times$ gender $\times$ facial expression.

As our approach involves two parts; namely, the clustering of data and the multi-armed bandit problem, it can be useful to modularize these into two parts. Hand-clustered data allows us to consider the multi-armed bandit problem separately, and it also allows us to compare the performance of the entire pipeline with unsupervised clustering versus the pipeline with it.

4.4 Unsupervised clustering

Figure 3: First 8 cluster-mean images from the dataset using unsupervised clustering retaining 50% of the variance after PCA and K-Means with 8 clusters. We can see that the unsupervised clustering algorithm has difficulties finding facial expressions that are different from neutral. This is because only a small subset of the people in the dataset have images with more than one facial expression (neutral). It is intriguing that our clustering algorithm seemingly is able to identify different racial and gender features, but also hairstyles and face shapes.

In step one, we use a face dataset (the details of the dataset are given below) and reduce the resolution of each image from 1700 $\times$ 1700 to 128 $\times$ 128 in an attempt to keep the computation cost for the subsequent parts of the clustering pipeline low. We then perform an eigen-decomposition on the whole image dataset using the typical eigenface approach, where we first remove the global average face from each face in the dataset, then construct a data matrix X and digagonalize the covariance matrix as

$$X^T X = U^T D U, \tag{1}$$

where U is an orthogonal matrix. We then project the data onto a reduced set of eigenfaces corresponding to the largest eigenvalues in the diagonal matrix D. We pick the number of eigenfaces here such that 50% of the variance of the dataset are retained, which for the dataset used in this project amounts to the first three eigenfaces. After this is done, we have effectively reduced the dimensionality of the dataset from 1700x1700 to 3.

It is this three dimensional space in which we perform K-Means clustering, which groups the reduced dataset in three dimensions into K clusters by finding the datapoints closest to each cluster center. This type of algorithm in its simplest form has trouble with high-dimensional datasets, hence the reduction to three dimensions from originally 1700×1700. The combinations of PCA and K-Means tested in this projects are shown in Table 4.4.

Dimensionality reduction	Clustering
PCA to 3 dimensions	K-Means to 4 clusters
PCA to 3 dimensions	K-Means to 8 clusters
PCA to 3 dimensions	K-Means to 16 clusters
SparsePCA to 3 dimensions	K-Means to 4 clusters
SparsePCA to 3 dimensions	K-Means to 8 clusters
SparsePCA to 3 dimensions	K-Means to 16 clusters

Table 1: These are the different combination of clustering algorithms used for unsupervised clustering of the face dataset. We tested clustering consistency between PCA and SparsePCA, which yielded no significant different for the clustering behavior. The unsupervised clustering procedure used for our preference prediction experiments is PCA to 3 dimensions with K-Means to 8 clusters because of the relatively small number of images in the dataset used (1235 images in the dataset).

Figure 4.4 shows the first 8 cluster-mean images from the face dataset using unsupervised clustering retaining 50% of the variance after PCA and K-Means with 8 clusters. We can see that the unsupervised clustering algorithm has difficulties finding facial expressions that are different from neutral. This is because only a small subset of the people in the dataset have images with more than one facial expression (neutral). It is intriguing that our clustering algorithm seemingly is able to identify different racial and gender features, but also hairstyles and face shapes. The results from the unsupervised clustering procedure are encouraging because at least visual inspection shows that the cluster means are very different from each other, implying that we are actually able to meaningfully separate the data for the UCB algorithm. We find that cluster number beyond 8 do not give different or improved results and given the size of the dataset (1235 images) we stick to 8 clusters.

We also compared two different PCA variants, one of which uses L1-regularization (SparsePCA) to induce sparsity. We don't find that it improves the cluster selection or that the clusters selected differ from vanilla PCA.

The implementation of this part of the project was accomplished using Python 2.7 with Scikit-Learn and the Python Image Library (PIL).

4.5 Multi-armed bandits

We modeled each cluster as an arm in a stochastic bandits setting, where there a finite number of arms each with a stationary and unknown reward distribution. The choice of a stochastic bandits setting makes two rather idealistic simplifications to the task: firstly, we do not account for the possibility that different users may have different preferences, and secondly, we assume that humans can be grouped into a finite number of clusters. However, reducing to setting to a stochastic bandits setting makes our task simpler, and we accept this simplification because we consider the exact multi-armed bandits model to be only one part of our overall pipeline.

In stochastic bandits, we look to find an algorithm which minimizes the pseudo-regret (or expected regret). The UCB1 algorithm, is able to achieve logarithmic pseudo-regret, and so we employ it here.

For a set of arms $\{1, \dots, M\}$, we define $\bar{x}_j$ to be the average reward currently observed at arm j. We define n_j to be the number times time arm j has been played, and n to be $\sum_{j=1}^{M} n_j$. In the context of our specific problem, the reward probability distributions are Bernoulli random variables with unknown biases, and the number of arms is equal to the number of clusters that we have chosen to segment our data into.

The UCB1 algorithm is:

1. Play each arm once
2. While $n < T$, play the arm which maximizes the quantity $\bar{x}_j + \sqrt{\frac{2\ln n}{n_j}}$

Intuitively speaking, the UCB1 balances exploration of arms which may be good with exploitation of arms which are known to be good. The term $\bar{x}_j$ is the exploitation term: arms with higher running averages are believed to be better arms, encouraging their exploitation. The $\sqrt{\frac{2\ln n}{n_j}}$ term is the exploration term: as we play a specific arm more, we increase the size of n_j relative to n, decreasing the value of this exploration term and encouraging the playing of different arms.

It can be shown that the pseudo-regret bound of the UCB1 algorithm is

$$\mathcal{O}(\frac{K}{\epsilon}\ln T) \tag{2}$$

where K is the number of arms (in our case number of clusters), ϵ is the difference between the largest mean and the second-largest mean of the probability distributions of the arms, and T is the time horizon; the number of times that the multi-armed bandit game has been played.

5 Experiments

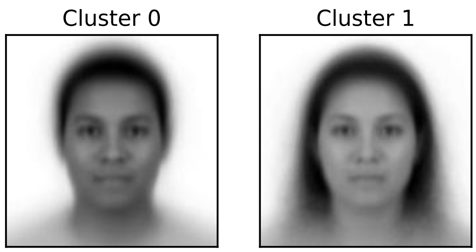
5.1 Testing

Testing procedures for our pipeline (PCA–K-Means-UCB) is difficult for personal visual preference prediction because we cannot compare our algorithms performance with the ground truth, nor do we have sufficient time and resources to design large-scale experiments that can unambiguously show convergence to reasonable results. Because of this shortcoming, we decided to test our pipeline by performing a simple two-arm selection experiment. First we perform unsupervised clustering with two clusters on the whole dataset. The results of this clustering procedure are shown in Figure 5.1 (left panel). We can easily see that our clustering algorithm, on average, divides the dataset into a male and a female cluster. In order to test the performance of the UCB algorithm, we simply accept images that show females and reject images that show males. The results of this testing procedure are shown in Figure 5.1 (right panel). Because we are dealing with only two very well separated clusters, the UCB algorithm converges very quickly, as shown by the histograms in Figure 5.1 (right panel).

5.2 Visual preference prediction: Single-subject results

For this experiment, we asked three male (Yay Caltech!) test subjects to test our algorithm. None of them was informed about how the algorithm works to keep potential biases low. We had our test subjects perform 200 UCB trials using the 8 unsupervised clusters shown in Figure 4.4 with the instruction to select the more attractive image. There was no time limit on the selection task. After two introduction rounds with 10 images each, we had the test subject run the selection experiment for the full 200 trials. Figure 5.2 shows the cluster-mean corresponding to the cluster selected by our test subject JEREMY. After the experiment was completed, we asked the test subjects to rate the cluster-mean images corresponding to the 8 clusters used for this experiment. Every single test subject rated the cluster-mean image highest for the cluster selected by our pipeline (PCA–K-Means–UCB). Figure 5.2 shows the performance of the UCB component of our algorithm again for the test subject JEREMY. We can see that the

Figure 6: The cluster-mean image corresponding to the cluster selected by our example test subject.



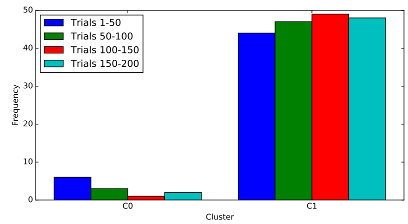


Figure 4: Mean images for our unsupervised clustering algorithm for the case of only two clusters. Clearly, our algorithm picks out the individuals' sex as feature. Amazingly, the algorithm keeps this property even for larger cluster numbers.

Figure 5: Evolution of the cluster histogram with trial number for the male-female experiment using our UCB algorithm. The separation between the two clusters is so strong that the UCB algorithm converges immediately.

performance of the UCB algorithm is not as good as in our simple male-female test, but after 150 trials, the histogram converges to JEREMY's preferred cluster. The reason for this reduced performance lies in the facts that JEREMY prefers individuals from more than one cluster, which reduces the convergence speed of our algorithm, see Figure 5.2 (upper panel).

The results for the two other test subject of this experiment are qualitatively identical to JEREMY, see Figure 5.2 (lower panel), except that they prefer different clusters. Given the results of this experiment, we are confident that a pipeline of the form reduction–clustering–UCB can be successful in principle in determining the visual preference of different individuals. The main shortcoming of our approach as it stands right now is that when different clusters are acceptable to the test subject, the convergence speed of the UCB component of our pipeline is dramatically reduced. We think that using an entirely different framework based on contextual bandits may alleviate this problem and also eliminate the need for clustering altogether.

5.3 Additional Testing

An alternative metric which can measure the performance of our algorithm is to measure the average number of positive responses that a user makes as a function of time. For instance, in a test of 200 trials, if a user rates the first 100 images less favorably on average than the second 100, we have an indication that our clustering method and our UCB algorithm are effective.

This metric is operationally useful; a new user to Tinder, for instance, may be presented a series of random images at first, but as they use the application and provide information about their preferences, the images shown should ideally be images that the application believes the user will respond positively to.

In a separate series of tests from the previous section, we asked three male test subjects, who we name ARAGORN, BOROMIR, CIRDAN, to perform 200 trials each when shown images drawn from 4 unsupervised clusters, and additionally asked two male test subjects, who we name DENETHOR and ELROND, to perform 200 trials each when shown images drawn from 16 unsupervised clusters. We additionally asked a single male subject, named FEANOR, to perform 200 trials with both hand-clustered images and unsupervised clustered, with 8 cluster in each case.

We plot moving-averages of the user's response, where a "yes" corresponds to a response with value 1 and a "no" corresponds to a response of value 0. We choose the filter to have length $n = 25$.

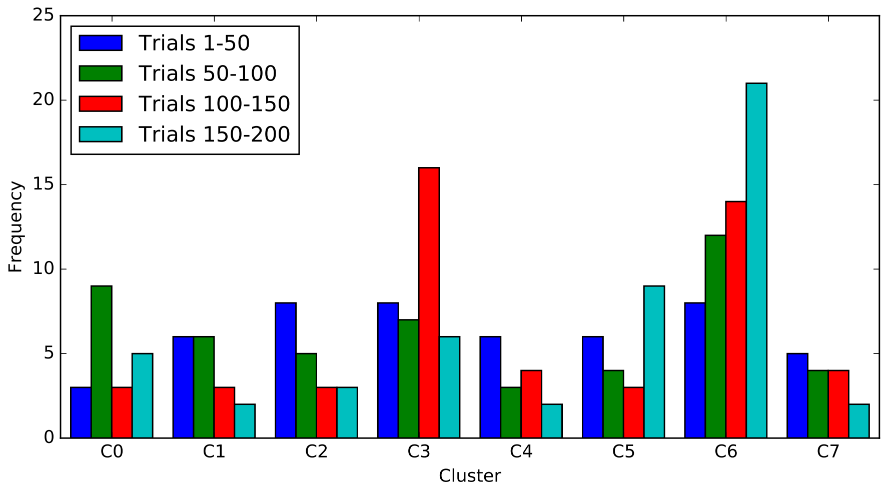


Figure 7: This figure shows a set of cluster histograms for the visual preference experiment performed by the test subject JEREMY. In this experiment, we used the clusters that were obtained from our unsupervised clustering algorithm. The UCB algorithms performed 200 trials on the test subject, showing one image per trial. We can see that the cluster histogram evolves as the number of trials increases, eventually showing one preferred cluster for this test subject. After the experiment had ended, we showed the mean cluster faces from Figure 4.4 to our test subject to confirm that the average face corresponding to cluster C6 is indeed preferred over all other mean faces. This was the case.

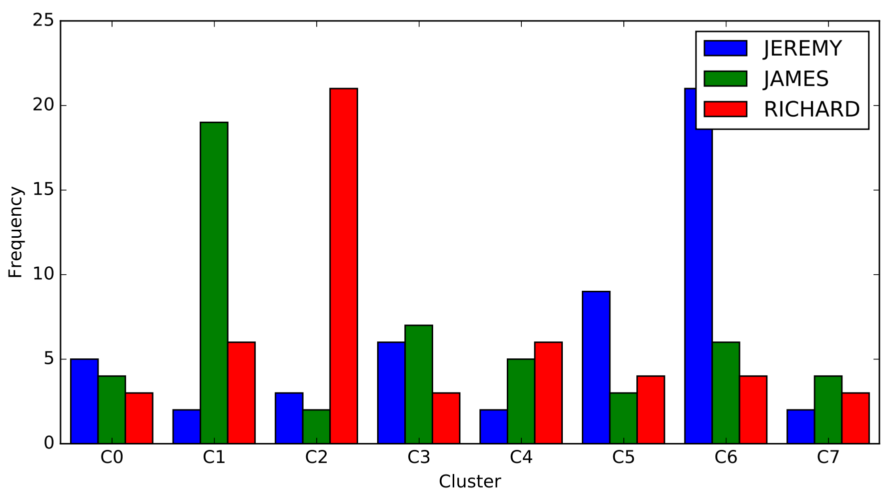


Figure 8: This figure shows the histograms for the final 50 trials for each of the three test subjects JEREMY, JAMES, and RICHARD. All of which prefer different clusters and hence have different personal visual preferences. The PCA–K-Means–UCB pipeline is able to determine each person's preference rather well, as determined by a post-experiment interview.

Perhaps in competition with the results above, it is difficult to draw conclusions from the moving-average plots of the users' responses. With 4 clusters, it is not clear if the UCB algorithm shows any improvement with time. On 16 clusters, it is more believable, especially in the case of ELROND, that the UCB algorithm converged to a favorable set of images.

For the single subject testing both with hand-clustered images and unsupervised clusters, the implications are slightly more positive. The results show that unsupervised clustering is able to perform roughly as well as hand clustering. In this case, the tie favors the unsupervised setting, as hand-labeling clusters can be a laborious process.

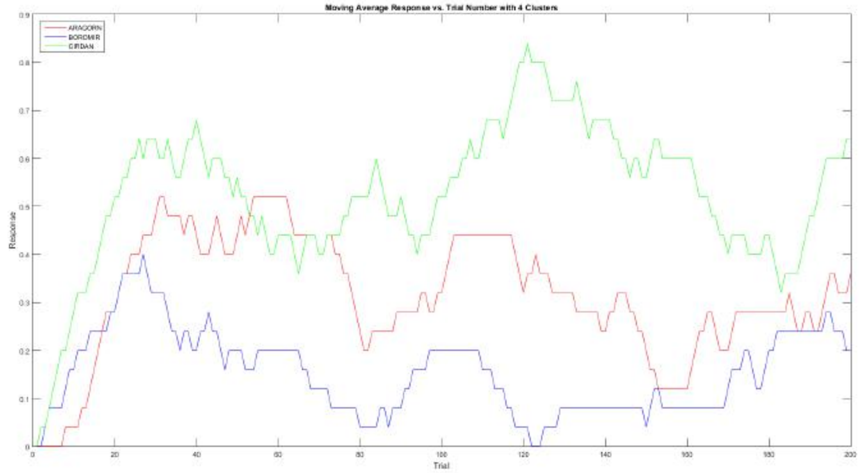


Figure 9: This figure plots the responses of users ARAGORN, BOROMIR, and CIRDAN over time, with a moving average filter with memory $n = 25$ applied. It is difficult to draw any conclusion from this set of data, as the different users each have a different baseline average response (the average taken over all of their responses). There is no upward trend apparent, which indicates that for these users the clustering and UCB algorithm could not effectively determine their preferences.

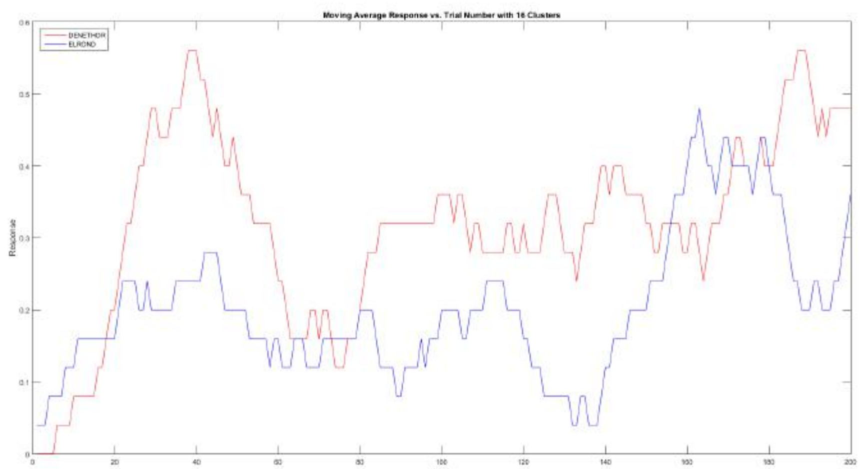


Figure 10: This figure plots the responses of users DENETHOR and ELROND, over time, with a moving average filter with memory $n = 25$ applied. Here, it is more apparent that there is an upward trend in the response, especially with ELROND, indicating that the clustering and UCB algorithm with 16 clusters for these users could determine the users' preferences.

5.4 Summary of Experiments

Due to constraints on time and because individual experiments could take a long time to run with the number of trials needed, we were not able to collect a large amount of data that would conclusively point to the value of our approach in either direction. However, our experiments make for strong proof-of-concept demonstrations of our approach. While a more thorough set of experiments with greater care taken to avoid biased setups and confounding

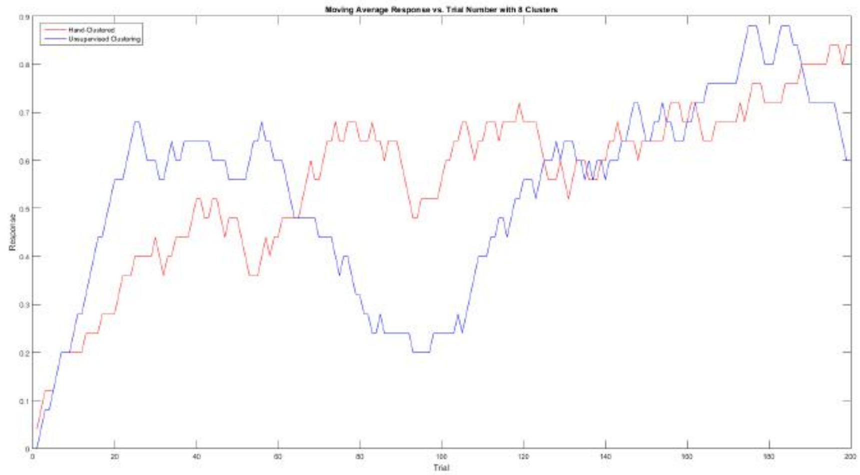


Figure 11: This figure plots the response of the user FEANOR for both hand clustering and unsupervised clustering. The data does not indicate that hand clustering possesses any inherent advantage.

variables would be ideal, we believe that our approach is fundamentally effective and that there is evidence to support this.

6 Future Work

6.1 Better data

A crucial problem with the kind of research presented in this project is the quality of the available data. While we were able to obtain high-resolution images of a diverse group of people, 1235 images are not enough if we want to use more advanced machine learning techniques such as neural networks for the dimensionality reduction task. We estimated that we would need several 10,000 to 100,000 images for training the dimensionality reduction algorithm alone. However, companies and dating apps that are interested in this problem usually have millions of images of user faces.

6.2 Improved dimensionality reduction and clustering

When we first set out to study this problem, we hoped to be able to used Stacked Denoising Autoencoders as technique for dimensionality reduction [Hin]. Extended to convolutional autoencoders [Masci], this type of algorithm has been very successful at compression high dimensional data to low dimensional features spaces [Hin]. It is likely that these types of high-powered algorithms are much better at reliable compressing the data in our images and similar techniques have been used in state-of-the-art attractiveness prediction [Rothe]. Unfortunately, as mentioned above, we need a lot more images for training in order to achieve acceptable performance with these kind of algorithms.

Why are these high-powered algorithms necessary? One main obstacle for visual preference prediction is that humans likely judge the attractiveness of faces in images not solely based on their physical experience but also on what "other impressions" are contained in the image, such as outfit and background. It is likely that to properly identify such aspects, deep neural networks will be necessary.

6.3 Contextual bandits

As additional improvement for visual preference predictions, contextual bandits come to mind. Instead of using K-Means clustering to define facial clusters, we envision a contextual bandit algorithm similar to the one used for news article recommendation in [Li], where the feature vector of each image is obtained from encoder component of the aforementioned convolutional autoencoders. As mentioned before, we think that contextual bandits might be helpful in increasing the convergence to an acceptable solution in our pipeline because, at least in principle, we can sidestep the need for a clustering procedure altogether and instead apply the contextual bandit algorithm straight to the reduced dataset.

6.3.1 References

[Ma] Ma, Debbie S., Joshua Correll, and Bernd Wittenbrink. "The Chicago face database: A free stimulus set of faces and norming data." Behavior research methods 47.4 (2015): 1122-1135.

[Hin] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." Science 313.5786 (2006): 504-507.

[Masci[Masci, Jonathan, et al. "Stacked convolutional auto-encoders for hierarchical feature extraction." Artificial Neural Networks and Machine LearningICANN 2011. Springer Berlin Heidelberg, 2011. 52-59. APA

[Rothe] Rothe, Rasmus, Radu Timofte, and Luc Van Gool. "Some like it hot-visual guidance for preference prediction." arXiv preprint arXiv:1510.07867 (2015). APA

[Li] Li, Lihong, et al. "A contextual-bandit approach to personalized news article recommendation." Proceedings of the 19th international conference on World wide web. ACM, 2010. APA

[Slivkins] Slivkins, Aleksandrs. "Contextual bandits with similarity information." The Journal of Machine Learning Research 15.1 (2014): 2533-2568.

[Besbes] Besbes, Omar, Yonatan Gur, and Assaf Zeevi. "Stochastic multi-armed-bandit problem with non-stationary rewards." Advances in Neural Information Processing Systems. 2014.

[Agarwal] Agarwal, Deepak, Bee-Chung Chen, and Pradheep Ela. "Explore/exploit schemes for web content optimization." Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on. IEEE, 2009.

[Auer] Auer, Peter, and Ronald Ortner. "UCB revisited: Improved regret bounds for the stochastic multi-armed bandit problem." Periodica Mathematica Hungarica 61.1-2 (2010): 55-65.

[Fu] Fu, Yun, Guodong Guo, and Thomas S. Huang. "Age synthesis and estimation via faces: A survey." Pattern Analysis and Machine Intelligence, IEEE Transactions on 32.11 (2010): 1955-1976.

[Ng] Ng, Choon Boon, Yong Haur Tay, and Bok-Min Goi. "Recognizing human gender in computer vision: a survey." PRICAI 2012: Trends in Artificial Intelligence. Springer Berlin Heidelberg, 2012. 335-346.

[Dhar] Dhar, Sagnik, Vicente Ordonez, and Tamara L. Berg. "High level describable attributes for predicting aesthetics and interestingness." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.

[Breese] Breese, John S., David Heckerman, and Carl Kadie. "Empirical analysis of predictive algorithms for collaborative filtering." Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 1998.

[Girschick] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[Krizhevsky] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[Willis] Willis, Janine, and Alexander Todorov. "First impressions making up your mind after a 100-ms exposure to a face." Psychological science 17.7 (2006): 592-598.