Image Placement Order Optimization for Mobile Commerce

Yeongnam Chae, Daniel Crane, Björn Stenger and Soh Masuko
Rakuten Institute of Technology, Rakuten, Inc.,
1-14-1 Tamagawa, Setagaya-ku, Tokyo, Japan
{yeongnam.chae, daniel.crane, bjorn.stenger, so.masuko}@rakuten.com

***Abstract*—With the increase in popularity of mobile commerce, the placement order of images on mobile displays is an important factor to attract and retain customers' attention. In this paper, we propose a novel method to optimize the placement of item images based upon the relative attractiveness of each image to the customer. To judge this, the proposed method estimates the leave rate of the customer at each image for each ordering using a model based on unsupervised hierarchical clustering. This allows estimating the expected leave rate for different image placements, solving an optimization problem to obtain the best ordering. The model is evaluated using a dataset collected from Rakuten Ichiba, the largest e-commerce site in Japan.**

I. Introduction

With the transition from PC to mobile devices, the focus of user experience in online shopping is rapidly moving from e-commerce to m-commerce. Due to the major differences between the two platforms, when designing an interface for mobile users these differences must be addressed. One such difference is the relatively restrictive size of mobile displays, due to which the amount of information that can be displayed on the screen at one time is reduced. This especially affects the amount of visual information that can be displayed, with only 1 or 2 images capable of being shown at once (see Fig. 1).

As a result of this restriction, the importance of the order in which images are displayed becomes higher, with products showing less attractive or informative images first potentially retaining the attention of the customer for less time, causing them to leave the page before viewing all of the information.

In this paper, we propose a novel method to optimize the placement order of images based upon the attractiveness to the customer. We propose a model that estimates the attractiveness of an image, based on its content, in terms of visual features, and its particular position within an ordered list of images. Our method applies clustering on visual features and the position within the list. Using this model we estimate the likelihood of a customer staying on an item for various permutations of image placements, and maximize the expected impression by choosing the optimal ordering.

II. Relationship between Customer Attraction, Image Order, and Visual Features

To understand the relationship between attractiveness and image placement, we collected item images from the ladies' fashion genre of Rakuten Ichiba, along with each image's placement order and leave rate. Due to the size constraints

Fig. 1. **Carousel panel on mobile environment.** *Image display interface on mobile environment, along with available actions.*

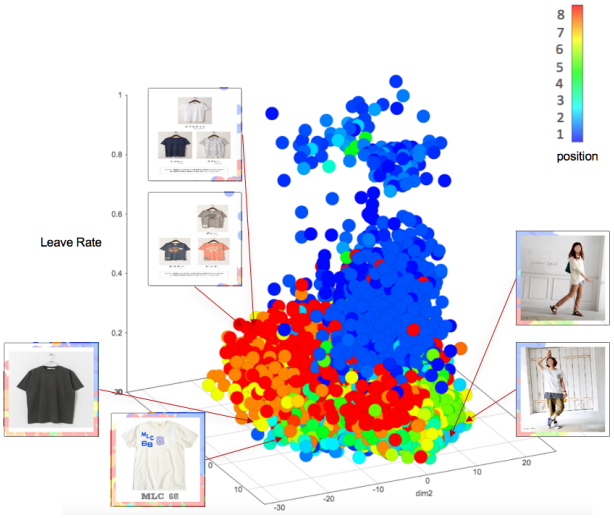

Fig. 2. **Relationship between visual features, placement order, and leave rate.** *The leave rate is affected not only by visual features but also the placement order.*

of mobile displays, the item images are displayed one by one via carousel panel, and the customer can perform one of the three actions shown in Fig. 1: flick to the next image, scroll down to read the item description, or leave the page. We define the attractiveness of the item page by the amount of images viewed by each customer on the item page. Given m images

for an item, all possible permutations of the product's images can be described by the symmetric group S_m [1], [2]. For a given permutation, $s \in S_m$, the number of customers who are viewing the i^{th} image in the permutation, s_i, can be easily estimated using the number of customers viewing the previous image and the leave rate, as follows:

$$n_{s_i} = n_{s_{i-1}} \cdot (1 - l_{s_{i-1}}) \tag{1}$$

where $n_{s_{i-1}}$ describes the remaining number of customers viewing the $i-1^{st}$ image, and $l_{s_{i-1}}$ denotes the leave rate on the $i-1^{st}$ image.

Using this, the attractiveness to the customer can be estimated by finding the image placement order that maximizes the overall number of staying customers, as follows:

$$\underset{s \in S_m}{\operatorname{argmax}} \sum_{i=1}^{m} n_{s_i} \tag{2}$$

In order to estimate the n_s, we model the leave rate of each image using various placement orders. From Fig. 2 we can see that the leave rate of each image is affected by both its visual features and the placement order. In this figure, the x and y axis represent the visual features that were calculated by embedding the images into the feature space through a deep convolutional neural network. For visualization purposes, we project the high-dimensional visual feature vector to a 2-dimensional space using principal component analysis (PCA) [3], [4].

In reality, however, more than two feature dimensions are required to adequately model the leave rate. The challenge of creating such a model lies in determining which features have an effect on the leave rate, and also determining the optimal number of clusters to represent the different image types.

III. Unsupervised Hierarchical Clustering

In order to predict the expected leave rate of each image from the image set in each possible position we built an unsupervised hierarchical clustering model that classifies the same types of image into the same cluster. To build the model we collected 11,400 images of 1,720 ladies' fashion items from Rakuten Ichiba, the most popular e-commerce service in Japan. Firstly we embed each image into the feature space, we then select the feature dimensions that have an effect on the leave rate, and finally we perform unsupervised clustering using the relevant features and image placement order hierarchically. By using the expected leave rate of each cluster, the proposed method estimates the amount of remaining customers for each image using a given image ordering.

A. Image Embedding

When embedding the images into the feature space, we must ensure we do it in such a way that visually similar images are separated by small distances in the feature space. To construct such an embedding we made use of the GoogleNet [5] deep convolutional neural network architecture, trained on the ImageNet [6] data set. Given an input image, GoogleNet can be used in order to obtain a 1,024 length feature vector of real numbers, with each element roughly corresponding to the likelihood of the given feature being present in the image. Using this embedding, it is possible to quantify the similarity of two images using the Euclidean distance between their feature vectors.

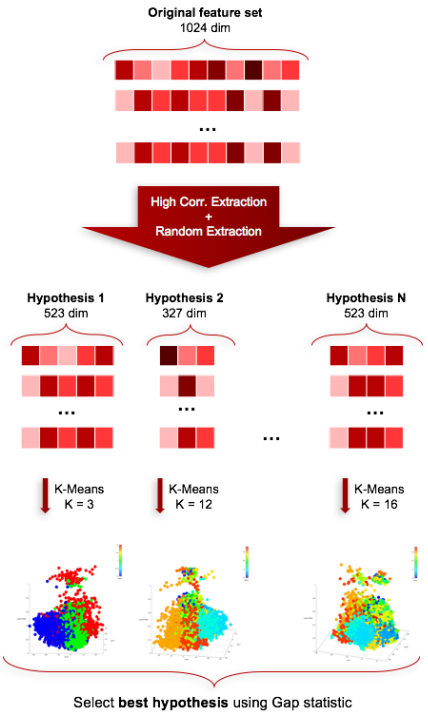

Fig. 3. **Unsupervised clustering of visual features.** *The proposed method generates several hypotheses, performing clustering using each, and evaluates the resultant clustering through the gap statistic.*

B. Feature Selection & Clustering

As the embedding model is trained for general purpose object classification, and we are only interested in ladies' fashion, some of the feature dimensions may not have an effect on the leave rate of each image. Before performing unsupervised clustering, we first rank each feature dimension in terms of mutual information [7] with the leave rate. We then generate 2,000 hypotheses of the number of feature dimensions and cluster numbers, K, required for the clustering. In order to select the number of feature dimensions for each hypothesis, we randomly select the top m highly-correlated feature dimensions from the ranked list, and then select an additional l features randomly from the remaining dimensions; the additional random features are included for the purpose of adding variance to each hypothesis. The number of clusters, K, is selected randomly between 1 and 20. By evaluating each hypothesis with the gap statistic [8], which scores the clustering by comparing the sum of intra-cluster distances with that of the null reference distribution (in our case chosen to be the uniform distribution), we select the hypothesis that gives the best clustering on the visual feature space.

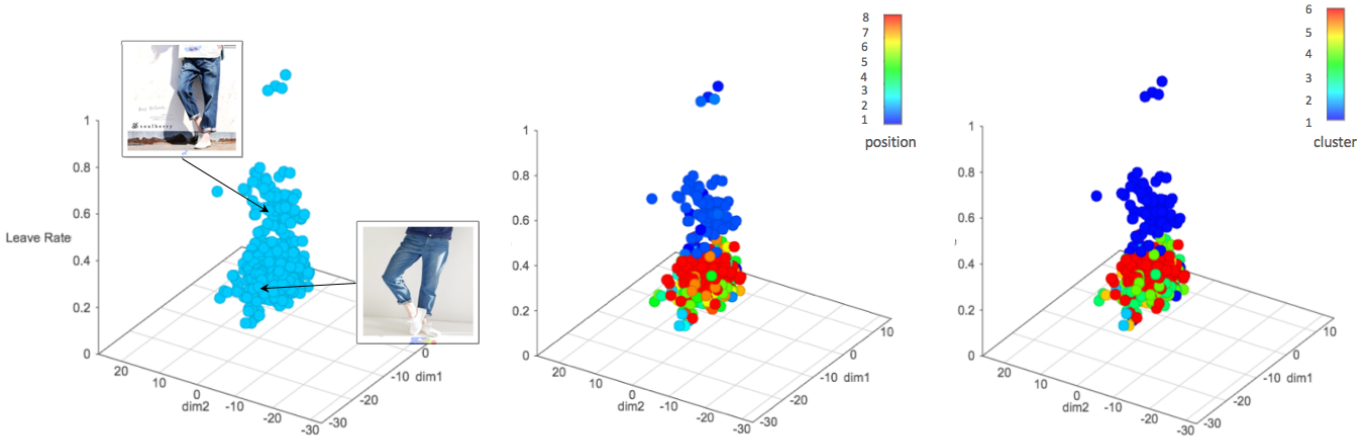

Fig. 4. **Visualization of a single feature-based cluster.** *(left) Points and leave rates within the cluster, and two typical examples of images within the cluster. (middle) Leave rates based on position, from real data. (right) Results of applying order-based clustering.*

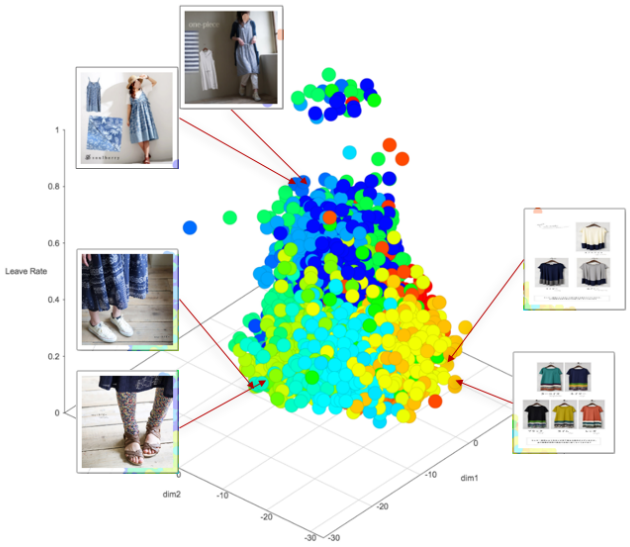

Fig. 5. **Visualization of order-based clustering results.** *Colored dots represent the clusters to which the image belongs, dim1 and dim2 are chosen as the first two principal components of the feature-space.*

C. Order-based Clustering

Unfortunately, clustering based upon visual features alone does not allow us to accurately predict the leave rate for each cluster in many cases. As seen in Fig. 4 (*left*), which provides a visualization of one of the clusters of our model, the intra-cluster leave rate variability is extremely high despite the cluster containing similar images, and so making any meaningful assumption about the leave rate for an image in this cluster would be impossible.

By looking at the real data of this cluster, we can see from Fig. 4 (*middle*) that the leave rate is not only based on visual features, but also on image position. Intuitively, this means that given a particular type of image, users prefer to see this image in a particular position in the display order - for example in ladies' clothing, it would be preferable to show images of the clothing item itself in the first few images, and tables with details of the clothing later on. As such, after performing the clustering based upon visual features, we perform a second level of clustering based upon the leave rate of images in each visual feature-based cluster. In this case, we generate several hypotheses by randomly selecting the number of level-2 clusters K' in the range 1,...,10 for each of the K level-1 clusters (resulting in $10\times K$ hypotheses); as previously, each hypothesis is evaluated using the gap statistic.

The results of combining both feature-based and order-based clustering on the cluster of Fig. 4 can be seen on the right sub-figure; here we can see a close resemblance between the ground-truth (*middle*) and the level-2 clustering (*right*). The most obvious difference between them is that for 8 positions we only have 6 clusters, and so for example positions 1 and 2 are combined into one cluster; as these two positions share similar leave rates, we can safely assume that the amount of information loss through this combination is minimal.

Fig. 5 shows the overall results of the hierarchical clustering. As described previously, the proposed clustering model performs clustering based not only on visual features (vertical separation) but also placement order (horizontal separation)

IV. Selecting Image Order

By classifying the images based upon their type and placement order through clustering, it is possible to estimate the expected leave rate for each image in any given position, and judge the attractiveness of the product based upon these values. In order to find the optimal placement of images of a given product, we should search over permutations of the available placements. However, comparing all permutations may be time consuming when the number of images is large, with 40,320,

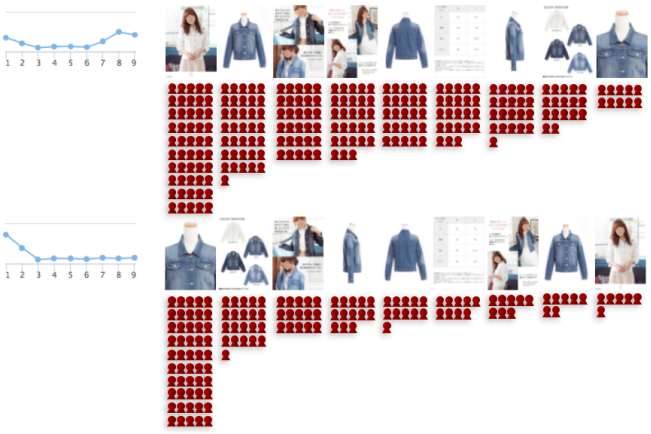

Fig. 6. **Visualization of attractiveness of an image permutation.** *Attractiveness is estimated based on the leave rate of each image (left graph) of the item; two example orderings of the same item are shown.*

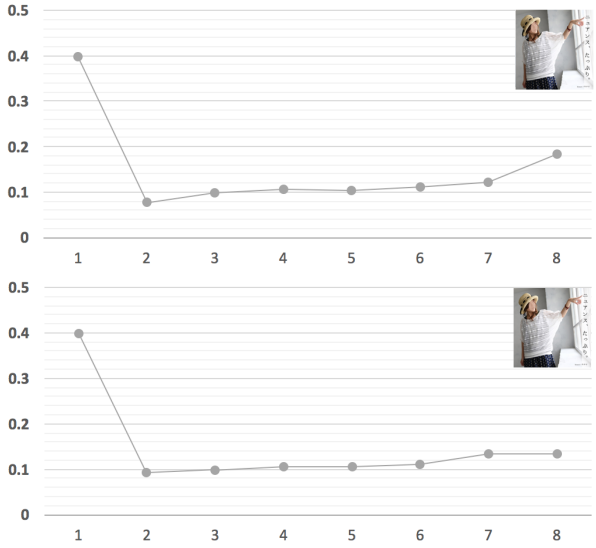

Fig. 7. **Actual leave rates vs estimates.** *Average leave rates for each image position of a chosen level-1 cluster. (top) From real data. (bottom) Estimated by model.*

362,880, and 3,628,800 possible permutations for items with 8, 9, and 10 images respectively.

To reduce the number of permutations to evaluate, we adopt a genetic optimization approach. We start by generating an initial seed permutation by placing the images of minimum leave rate at each consecutive position. By randomly mutating this seed we generate candidate permutations, and then compare the attractiveness of each. In our case, we chose to generate 10,000 permutations from the original seed. By summating the number of expected remaining customers at each image using the leave rate, we estimate the item's attractiveness for each permutation. Fig. 6 provides a visualization of the expected number of remaining customers for two different permutations of a given product (*right*), and the predicted leave rate at each image using the proposed model (*left*). Through the use of this attractiveness measure, we assign a numerical score to each permutation through the use of (2), and choose the placement order that maximizes this score. Another set of test results can be seen in Fig. 8, where we can see the attractiveness scores of the top 4 image orderings (*top*) and the bottom 4 image orderings (*bottom*). From this example we can gain an intuitive understanding of what constitutes a good image placement order in the genre of ladies' fashion; our result shows strong preference for a model shot as the main image, followed by close-up shots of the item itself, and finally group shots and item specification tables towards the end.

V. Experimental Results

To evaluate the proposed model, we split the collected dataset into a training and test set, with the training set containing 9,180 images and the test dataset 2,220 images, all from the ladies' fashion genre. Using the above feature selection & clustering procedure on the training set, we select a hierarchical clustering model using 471 of the original 1,024 feature dimensions, and containing K=14 level-1 clusters; the number of level-2 clusters vary for each of these 14 clusters, totaling approximately 70.

We validate the trained model by comparing the actual leave rates to those obtained by applying this model to the test dataset. The overall mean-squared error (MSE) of actual vs estimated leave rates is low, at 5.89%. An example of the comparison between actual leave rates vs estimated leave rates are shown in Fig. 7, where we can see the true average leave rate by position for a chosen cluster (*top*) and the average of the estimated leave rates for this cluster (*bottom*).

VI. Conclusion

In this paper, we proposed a novel method to optimize the image placement order based on the attractiveness to the customer for m-commerce. We generated an unsupervised hierarchical clustering model that can classify an image's leave rate at a given position through visual similarity to similar images. By estimating the attractiveness of various image placement orders, the model can find the optimal placement order for the given product. As the model was trained using an unsupervised clustering methodology, it can be equally applied to any genre, and is not limited to ladies' fashion.

The results of this model rely heavily upon the accurate estimation of perceptual difference between images, as this is the foundation upon which the hierarchical clustering is based. As such, by using a more advanced network architecture (i.e. ResNet, Inception) we can expect to see an increase in accuracy of model predictions.

Due to the influence on the leave rate of the current image by the previous images, the image placement problem can be viewed as a time-series problem. As such, future work will focus on investigating the combination of the model proposed in this paper with time-series analysis models such as Recurrent Neural Networks (RNN) [9], including Long-Short Term Memory (LSTM) [10] models.

Fig. 8. **Permutation hypotheses.** *(top) Estimation of most attractive permutations (bottom) least attractive permutations*

Other avenues for future work include studying the applicability of this model to product categories other than fashion, and studying the effects of external factors such as display size.

REFERENCES

[1] B. Sagan, "The symmetric group: representations, combinatorial algorithms, and symmetric functions." Vol. 203. Springer Science & Business Media (2013).

[2] E. W. Weisstein, "Symmetric Group." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/SymmetricGroup.html

[3] A. M. Martínez, A. C. Kak, "PCA versus LDA". IEEE transactions on pattern analysis and machine intelligence 23.2 (2001): 228-233.

[4] H. Hotelling, "Analysis of a complex of statistical variables into principal components." Journal of educational psychology 24.6 (1933): 417.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition (2015); 1–9.

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei- Fei, "Imagenet: A large-scale hierarchical image database." Proceedings of the IEEE computer Vision and Pattern Recognition (2009).

[7] A. M. Fraser, and H. L. Swinney, "Independent coordinates for strange attractors from mutual information." Physical review A 33.2 (1986): 1134.

[8] R. Tibshirani, G. Walther, T. Hastie, "Estimating the number of clusters in a data set via the gap statistic." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63.2 (2001): 411–423.

[9] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, S. Khudanpur, "Recurrent neural network based language model." Interspeech. Vol. 2. (2010).

[10] S. Hochreiter, J. Schmidhuber, "Long short-term memory." Neural computation 9.8 (1997): 1735–1780.