Who is repinning? Predicting a brand's user interactions using social media retrieval

Shantanu Singh
Broad Institute of MIT and Harvard

Yan Wang
Columbia University

Lei Ding
Intently.io

ABSTRACT

Despite the fact that firms spend heavily in marketing their brands across social media platforms, very little is understood about what media content, in a predictive manner, can generate high interaction rates among their prospects and customers. However, such understanding can significantly help brand marketers generate desired engagements with their target audience in marketing campaigns. In this paper, we study the problem of predicting a brand's user interactions on social media using the example of Pinterest, an emerging platform that has provided a large volume of brand as well as user data in the form of images. Specifically, we treat the prediction of a brand's user interactions, captured through "repinnings" on Pinterest, as the retrieval of relevant user-pinned images given a brand image. The prototype system that we build incorporates this basic principle, and is tested on a large-scale Pinterest dataset of more than one million images. We demonstrate that our system achieves significant lifts in recalling ground truth repinners of brand images for a variety of brands across several major industry categories.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; I.5.4 [**Pattern Recognition**]: Applications

General Terms

Algorithms, Experimentation, Measurement, Performance

Keywords

Social media retrieval, content marketing, user modeling

1. INTRODUCTION

Social media have revolutionized the way people communicate and share information. Facebook, the largest social media platform, is reported to have more than one billion active users worldwide [12]. As a result, firms worldwide, big or small, are widely using social media for promoting their products or services, engaging customers and strengthening their loyalty [19].

Researchers have studied the effect of social media such as Facebook, Twitter on the audience at a population level [14, 23]. Most recently, studies have been reported on predicting "following" information among social media users on Tencent Weibo, a Chinese microblogging site [7, 22]. Nevertheless, we are not aware of any existing research that has used a predictive framework to explore what media content can create the desired engagements with the audience. For instance, before a brand marketer posts an image on the firm's Facebook page, she would desire to know which candidate image would generate the maximal level of engagement with the target audience. In the past, without access to media data about users' preference on the image content in association with a brand's marketing content, it has been nearly impossible for a brand marketer to quantitatively make such decisions before a campaign.

Fortunately, in recent years it has become increasingly cheap and easy to share image data, driven in part by new image-based platforms such as Pinterest [2] and Instagram [1], where users either pin or post images that they come across on the Internet, or photos that they take using mobile devices. We believe that global marketers can use such data strategically by analyzing the semantic alignment in content between brand images and user images, in order to maximize their campaigns' effectiveness. As an initial attempt towards this objective, we present a prototype system for predicting which users will potentially interact with brand images on Pinterest, a fast-growing platform that has significantly attracted marketers' interest. Given a brand image, we ask the following question: *who are the users that will interact with the image?* We address it by leveraging social media retrieval, and in particular content-based image retrieval [11] on Pinterest collections. Specifically, we treat the brand image as a query, and user images as the repository to rank against the query. Furthermore, our system ranks the users based on their collections' overall similarity measures to the brand image, and thereby predicts a user's propensity to interact with the brand image. This ranking of propensity is evaluated by comparing with the set of ground truth users that have interacted with the brand image.

We organize the paper as follows. In the rest of this section, we provide a brief review of the Pinterest platform, as well as our prototype system. We overview the related back-

MDMKDD'13, August 11, 2013, Chicago, Illinois, U.S.A.

ground and literature in Section 2, and describe the system in Section 3. Experimental results are covered in Section 4, and our conclusions are drawn in Section 5.

Overview of Pinterest. Launched in 2010, Pinterest is a pinboard-style photo sharing website that allows users to create and manage theme-based image collections such as events, interests, hobbies, and more. Individual users and brand marketers alike can browse other pinboards for inspiration, "re-pin" images to their own collections or "like" images.

A user creates a pin by adding an image on Pinterest. A pin can be added directly from a website using the "Pin It" bookmarklet or uploaded from a local source. Any pin on Pinterest can be repinned, and all pins link back to their source. Figure 1 explains how image data are organized on Pinterest. Each user or brand has a main Pinterest page, where a number of pinboards are created. Each of the pinboards contains pins. For each pin, a Pinterest user can "repin" it onto their own pinboards, or "like" it. In this paper, we identify each "pin" with the corresponding pinned image. Besides, we focus on "repinning" as the way of user interaction with brands, as it is the most prominent mode of interaction on Pinterest.

System Components. Next we summarize the key steps that our system performs in order to predict a brand's user interactions.

- Data collection: Our system collects images pinned by brands as well as users. We start with a number of brands and obtain the list of users who have interacted with those brands, whose pinned images are then collected by our system.
- Image representation: Appropriate image-based features are extracted from each image. In specific, we use the well-established scale-invariant feature transform (SIFT) representation [18], and adopt the bag-of-visual-words (BoVW) model [24] for image representation. In this model, each pinned image is represented as a high-dimensional feature vector that can be used for retrieval.
- Image retrieval and prediction of user interactions: For each brand image, our system ranks the users based on their collections' overall similarity measures to the brand image, and thereby predicts a user's propensity to interact with the brand image. The predicted ranking is then compared with users recorded in the ground truth that have repinned the brand image in order to evaluate our system.

2. BACKGROUNDS

In this section, we overview related marketing research work on social media. Next, we discuss techniques in content based image retrieval, which is a key component of the proposed system.

2.1 Social Media Marketing

Social media sites allow individuals to interact with one another and build relationships. When a firm joins such a site, users can similarly interact with the firm or its products. A user interacts with a brand in ways that are similar to interacting with other users. The nature of this relationship makes the experience of interacting with the brand much more personal. In particular, social media sites allow individual followers to "retweet", "repost" or "repin" promotion messages or images made by the firm for the product being promoted. By repeating the message or image, all of the users' connections are able to see it, thereby resulting in the message reaching more potential customers. In this manner social media sites greatly amplify the spread of a message through word of mouth. Through this process, a brand's message becomes self-propagating, resulting in more traffic to the firm [25].

Existing work on social media in marketing research is concerned with the impact of marketing content at an aggregate level, without any personalization at the user level. For instance, the authors of [5] study why certain pieces of online content are more viral, i.e., have created more user interactions, than others. The article takes a sociopsychological approach to understanding diffusion using a data set of news articles. The effect of emotions on shaping virality is examined and the causal impact of specific emotion on transmission is demonstrated. While these findings shed light on how to design effective viral marketing campaigns, they do not help optimize campaigns in a quantitative fashion, and neither do they predict how a specific user would interact with online content. In [17], the author examines aspects of the seeding strategy by tracking the diffusion of new videos published on YouTube. The results show that the need for using many seed consumers depends on message quality. Besides, the author suggests that one should choose consumers who have strong ties with the advertiser and who also have strong influence on others. Such results, while insightful, are inapplicable for personalization when dealing with large amount of user data, such as in our case of content marketing on Pinterest.

In social network analysis and data mining communities, research on content diffusion has been conducted, but the conclusions do not provide insights on how to optimize content in order to generate interactions or reach. For example, researchers have studied behavioral motifs that are possible to observe at the dyadic level [30]. This study shows significant differences between dyads, or pairs of individuals, that are more versus less engaged in the diffusion process. Dyads that fuel the diffusion process are characterized by stronger relationships, more active and networked receiving party, and higher authority centrality. In another study [16], the authors present an analysis of a person-to-person product recommendation network. While on average recommendations are not very effective at inducing purchases and do not spread very far, they present a model that successfully identifies communities, product and pricing categories for which viral marketing seems to be effective. Recent work has also examined influence maximization for prevalent viral marketing in large-scale social networks [8], where the authors propose an algorithm that can create significant increase in influence spread compared to other heuristics. Their problem setting is to find a small set of seed nodes in a social network that maximizes the spread of influence under certain models. Similarly with existing work in marketing research, these studies are unable to identify the quantitative characteristics in the marketing content that create more interactions from specific users.

2.2 Content-based Image Retrieval

Content-based image retrieval (CBIR) [11, 21] comprises

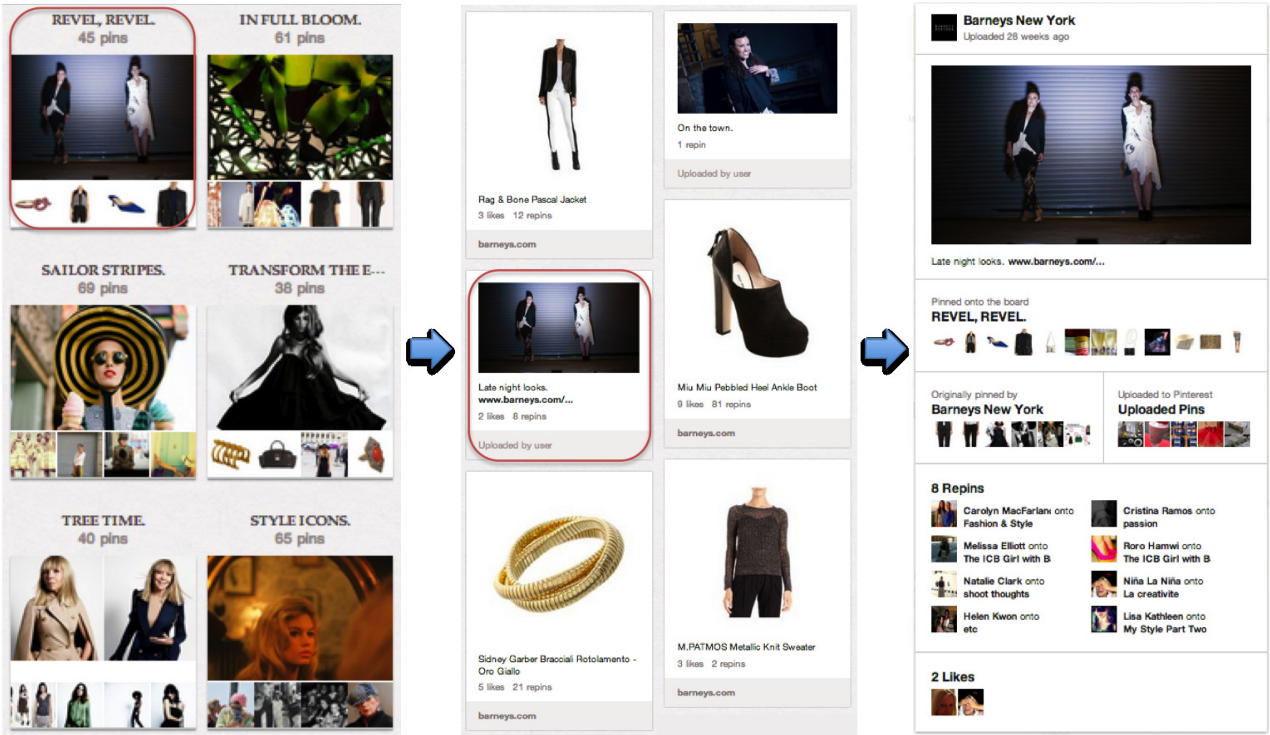


Figure 1: The site structure of Pinterest. Left: a segment of the brand page of Barneys New York, with the pinboard "Revel, Revel" highlighted in a red box. Mid: a pinboard name "Revel, Revel" with a highlighted pin shown on the right. Right: a sample pin from "Revel, Revel" with the list of repinners and likers.

of a set of techniques which use visual contents to search images from large scale image databases according to users' interests. CBIR has been an active and fast-advancing research area since the 1990s. During the past decade, remarkable progress has been made in both theoretical research and system development. The typical image descriptors summarize useful information from images include color, texture and spatial information, among which scale invariant feature transform (SIFT) features have been widely used and achieved high performance in conjunction with bag-of-visual-words (BoVW) image representation [20, 29, 3].

Instead of exact matching, content-based image retrieval calculates visual similarities between a query image and images in a database. Accordingly, the retrieval result is not a single image but a list of images ranked by their similarities with the query image. In recent years, many similarity measures have been developed for image retrieval based on estimates of the distribution of features [31, 4]. It is shown that different similarity or distance measures can significantly affect retrieval performances of an image retrieval system [13].

Another important issue in content-based image retrieval is effective indexing and fast searching of images based on visual features. Since the feature vectors of images tend to have high dimensionality, dimension reduction or probabilistic modeling are commonly used by researchers to obtain the semantically useful information for image retrieval. For example, in [27, 6], variants of principal component analysis (PCA), namely generalized PCA and kernel PCA are used for image compression and retrieval. Instead of using the PCA-family of methods, researchers have also used nonlinear manifold embedding is for image retrieval [26]. Additionally, probabilistic methods such as Gaussian mixture learning have been considered for efficient image indexing and retrieval [28, 10].

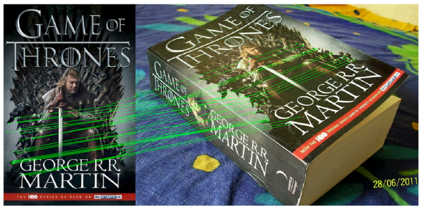
3. SYSTEM DESCRIPTION

In this section, we describe the methodology by which we build the prototype system for predicting user interactions on Pinterest.

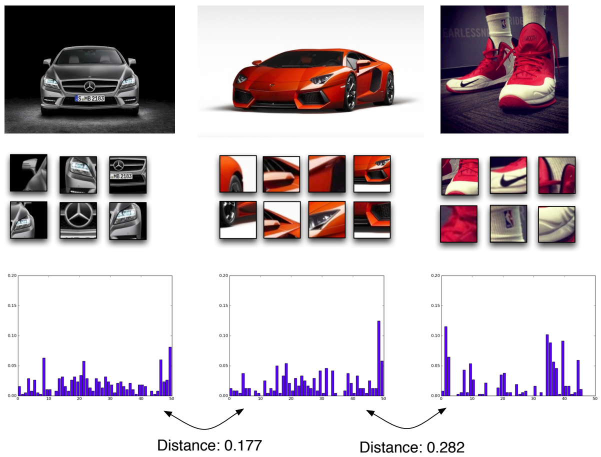
3.1 Collecting image data

In order to collect brands' data from Pinterest website, we start by choosing 20 brands from five broad categories (auto, fashion, food, hotel and retail), which are listed in Table 3. For convenience, we call this set of brands $B = \{b_1, b_2, \cdots, b_{20}\}$. Statistics of the number of pinboards and pins are listed in Table 1. As we can see from this table, the number of pins varies significantly from brand to brand. Thus, in order to construct pin set $\mathcal{P}_{b_i}$ for brand b_i, we choose to randomly sample N_p pins from all the pinboards of brand b_i. By doing so, our results would not be dominated by any one brand. We call the overall set of brand pins $\mathcal{P}_B = \mathcal{P}_{b_1} \cup \mathcal{P}_{b_2} \cup \cdots \cup \mathcal{P}_{b_{20}}$.

Next, we collect users' data. We are interested in users who have repinned our selected pins from the 20 brands. Specifically, if the user has repinned any of the brand image in set $\mathcal{P}_B$, we collect her images. The set of users such generated is $U = \{u_1, u_2, \cdots, u_{|U|}\}$. Similar to what we perform on brand pins, we choose N_p randomly sampled pins for each user u_i in order to construct her pin set $\mathcal{P}_{u_i}$.



(a) SIFT matching



(b) Bag-of-visual-words illustration

Figure 2: SIFT-BoVW image representation model. (a) An example showing the invariance properties of SIFT features. The interest points are extracted from a DVD cover image (left), and an image of the cover of the corresponding novel (right), and then matched using the algorithm described in [18]. (b) An example illustrating how the BoVW model helps recognize object categories. The first row shows the input images, with sample patches detected by SIFT on the second row, followed by normalized BoVW histograms on the third row. Note that while SIFT detects patches of different scales, they have been resized to the same size above for the purpose of visualization.

Collectively, we build the set of user pins $\mathcal{P}_U = \mathcal{P}_{u_1} \cup \mathcal{P}_{u_2} \cup \cdots \cup \mathcal{P}_{u_{|U|}}$.

Problem formulation. Our objective is to generate, for each pin $p \in \mathcal{P}_B$, a ranked list $\mathcal{K}_U = (u_{k_1}, u_{k_2}, \cdots, u_{k_{|U|}})$ of users in U according to semantic similarity between the brand pin p and user pins in $\mathcal{P}_{u_1}, \mathcal{P}_{u_2}, \cdots, \mathcal{P}_{u_{|U|}}$. The higher a user gets ranked in $\mathcal{K}_U$, the higher propensity she is predicted to have in repinning the brand pin. In other words, we address the problem of *predicting the repinners' list of each brand pin*.

3.2 Representing image data

In our approach, we represent each pinned image p with a vector v_p, which will then be fed into the retrieval components for further processing. In order to obtain good performance in retrieval, such a vector representation should be both descriptive enough, i.e., able to distinguish one object

Table 1: List of brands with the corresponding number of pinboards and pins.

Brand	Brds	Pins	Brand	Brds	Pins
Lamborghini	14	91	Hyundaiusa	31	311
Toyotausa	9	154	MercedesBenz	38	804
Anntaylorstyle	24	1237	Giltkids	47	1865
Barneysny	54	2750	Maybelline	19	2153
Chobani	25	4080	Panerabread	14	692
Cabotcheese	30	2210	DunkinDonuts	19	481
Fourseasons	34	1726	Trumpcollection	27	572
Hiltonhotels	19	347	Ritzcarlton	95	653
Homedepot	38	1744	Michaelsstores	79	3769
HobbyLobby	47	2067	Lowes	46	1699

category from another, and robust to the changes of photo shooting conditions, such as camera zooming, rotation, or illumination changes. With these requirements, we adopt a state-of-the-art approach in computer vision and image retrieval, the bag-of-visual-words (BoVW) model [24], to provide such representation for images.

Similar to the bag-of-words model in text mining, the bag-of-visual-words model also uses a dictionary to turn each document into a vector representation by counting the words, while ignoring the order of the words. In the bag-of-words model, a dictionary is used to transform a document into a vector representation by counting the frequency of words. While this general concept can be naturally extended from documents to images, the problem of defining words and the dictionary for images is not straightforward. The BoVW model addresses this problem by taking as input a set of image-based features, and using them to generate a set of visual words through clustering.

Local feature of images. We first turn the images from raw pixels to intermediate representations robust to geometric transforms such as translation, rotation and scaling using the scale invariant feature transform (SIFT) representation, a well-established technique in computer vision and image retrieval [18]. This is achieved by first identifying corner points, i.e., areas with large local contrast in at least two orientations of the image. The corner points are identified with the difference-of-Gaussian detector performed at multiple scales. Next, at each detected corner point, the gradient histogram is calculated to describe the neighborhood information. A subset of these points are further selected as interest points based on certain metrics of robustness. By explicitly considering multiple geometric transforms, SIFT provides a 128-dimensional descriptor for each interest point in each image. This ensures that similar descriptors for an interest point are obtained even when it is translated, rotated, or scaled. SIFT can also provide a certain degree of invariance on illumination and 3D viewpoint changes, which also benefits our application in image retrieval.

In panel (a) of Figure 2, we illustrate how SIFT can achieve robust matching against the transforms mentioned above with the matching algorithm proposed in [18]. Specifically, the interest points are extracted from a photo of the cover of a novel, and from a similar photo from the cover of the accompanying DVD. Then, for each interest point in the left image, we identify the corresponding point in the right image by selecting the nearest neighbor in the SIFT feature space, where the neighbors are points from the right image. After the initial matching, we also perform spatial verification to select the matches with strong geometric consistency [24]. It can be seen that SIFT features enable robust matching between interest points in these two images even after multiple transformations and illumination changes. While we do not use strict matching for the Pinterest images, the example above illustrates that SIFT is indeed able to capture the interest points across images, which can later be leveraged to build robust visual representations.

Dictionary construction and image representation. Next we construct a dictionary of SIFT features and represent the image with respect to the constructed dictionary. First, k-means clustering is performed on the collected SIFT descriptors, resulting in c cluster centers, which are treated as the visual words. Collectively, the set of all visual words are called the visual dictionary. Given the dictionary, for each image, we map every SIFT descriptor to the nearest visual word in the SIFT feature space, and use a c-dimensional normalized histogram of the mapped visual words as the final image representation. By this process, we obtain model robustness from the invariance against geometric transformations and illumination changes provided by the SIFT feature and descriptiveness provided by the BoVW representation. Extensive experiments in computer vision [24, 20, 29, 3] have shown the BoVW model has reliable performance in tasks such as image retrieval and classification.

In panel (b) of Figure 2, we show how BoVW model can measure similarity among images. We use three images from Pinterest—a user-pinned Mercedes-Benz car, an image from Lamborghini's collections, and a user-pinned Nike sneaker, as shown on the first row of the figure. We also show some examples of detected patches by SIFT on the second row. The normalized histograms on the third row illustrate the BoVW representation for each image, with the horizontal axis as the visual words, and the vertical axis as visual word frequency. While our system is built with $c = 500$ as the vocabulary size, the dictionary in this example is built with $c = 50$ for illustration. We can see here that the two car images are close in terms of Euclidean distance in the BoVW representation, although they are merchandise of different brands, while the sneaker's representation is obviously different from that of the cars. Consequently, using SIFT-BoVW enables us to make predictions about images' categories and thereby infer users' propensity to interact with an image.

3.3 Retrieving image data

Once all brand pins in $\mathcal{P}_B$ and user pins in $\mathcal{P}_U$ are represented using SIFT-BoVW, we leverage image retrieval to compute the propensity that a user is to repin a brand pin. Our assumption is that, everything else being equal, the chance that a user would repin a brand image is higher if she has collected similar image content onto her own Pinterest collection. Specifically, we compute the similarity measure $s(p, \mathcal{P}_{u_i})$ for a brand pin $p \in \mathcal{P}_B$, and a user's pin set $\mathcal{P}_{u_i}$. Next we will detail two approaches commonly used in image retrieval to accomplish this task.

Distance-based approach. This is a model-free approach. First, we select a distance measure d (e.g. Euclidean distance), and compute the $d(v_p, v_q)$ for $q \in \mathcal{P}_{u_i}$ using their BoVW feature vectors v_p and v_q. Next, we compute the similarity measure s as following

$$s(p, \mathcal{P}_{u_i}) = f\big(\Gamma_{q\in\mathcal{P}_{u_i}}\{d(v_p, v_q)\}\big), \tag{1}$$

where $f(\cdot)$ is a monotonically decreasing function as we require a similarity measure. Note that the particular choice of the function $f(\cdot)$ is not relevant here, since we are only interested in ranking the users. The choice of summary function Γ is significant, as evidenced by the results discussed in Section 4.2.

Model-based approach. To begin with, we train a Gaussian mixture model on the collection of all BoVW feature vectors using the standard expectation-maximization (EM) procedure [9] that maximizes the overall log-likelihood of data, which is $\sum_p \log Pr(v_p)$, where v_p is a BoVW feature vector corresponding to a pin. After we discard the mixing weights, we obtain the learned global BoVW dictionary $\Omega = \{\mathcal{N}(\mu_i, \Sigma_i)\}_{i=1}^K$, where K is the total number of components. The global dictionary tells us about the important BoVW feature vectors and their variances and covariances.

Following that, for a user's pin set $\mathcal{P}_{u_i}$, we process it into a single vector of mixing coefficients in a Gaussian mixture model with the previously learned means and covariances. Since each user has a relatively small number of pins, we choose not to re-learn the Gaussian components each time. Specifically, we model BoVW feature vector v_q for $q \in \mathcal{P}_{u_i}$ as a finite mixture of Gaussian components:

$$v_q \sim \mathcal{M}_{u_i} = \sum_{i=1}^{K} \alpha_i \mathcal{N}(\mu_i, \Sigma_i), \tag{2}$$

where α_i's are the mixing weights.

We next present the EM procedure for each user's pin set $\mathcal{P}_{u_i}$. Let $v_{p_1}, \ldots, v_{p_m}$ be the BoVW feature vectors corresponding to pins in $\mathcal{P}_{u_i}$, z_i's be latent variables on which Gaussian component is used, and

$$\theta_t = \{Pr(z = 1), \ldots, Pr(z = K)\}$$

is the set of unknown model parameters at iteration t.

E-step:

$$Pr(z_j = i|v_{p_j}, \theta_t) = \frac{p(v_{p_j}|z_j = i)Pr(z_j = i|\theta_t)}{\sum_{l=1}^K Pr(v_{p_j}|z_j = l)Pr(z_j = l|\theta_t)}, \tag{3}$$

M-step:

$$Pr(z = i|\theta_{t+1}) = \frac{1}{|\mathcal{P}_{u_i}|} \sum_{j=1}^{|\mathcal{P}_{u_i}|} Pr(z_j = i|v_{p_j}, \theta_t). \tag{4}$$

Therefore, at the end of the iterative EM procedure, we learn a model $\mathcal{M}_{u_i}$ for user u_i. We then define the similarity measure $s(p, \mathcal{P}_{u_i}) = \mathcal{M}_{u_i}(v_p)$. That is, we evaluate the brand pin's feature vector v_p against the user model $\mathcal{M}_{u_i}$.

4. EXPERIMENTAL RESULTS

In this section, we discuss our experimental setup, present our results and discuss them.

4.1 Experimental Setup

For each user u_i, we have the set of pins $\mathcal{P}_{u_i}$. We would like to predict, given a pin $p \in \mathcal{P}_{b_j}$ from a brand b_j, as to how likely the user is to repin it. In order to remove any bias in performance, we first construct a new pin set $\mathcal{P}'_{u_i} = \mathcal{P}_{u_i} - \mathcal{P}_{b_j}$. That is, we remove all brand pins belonging

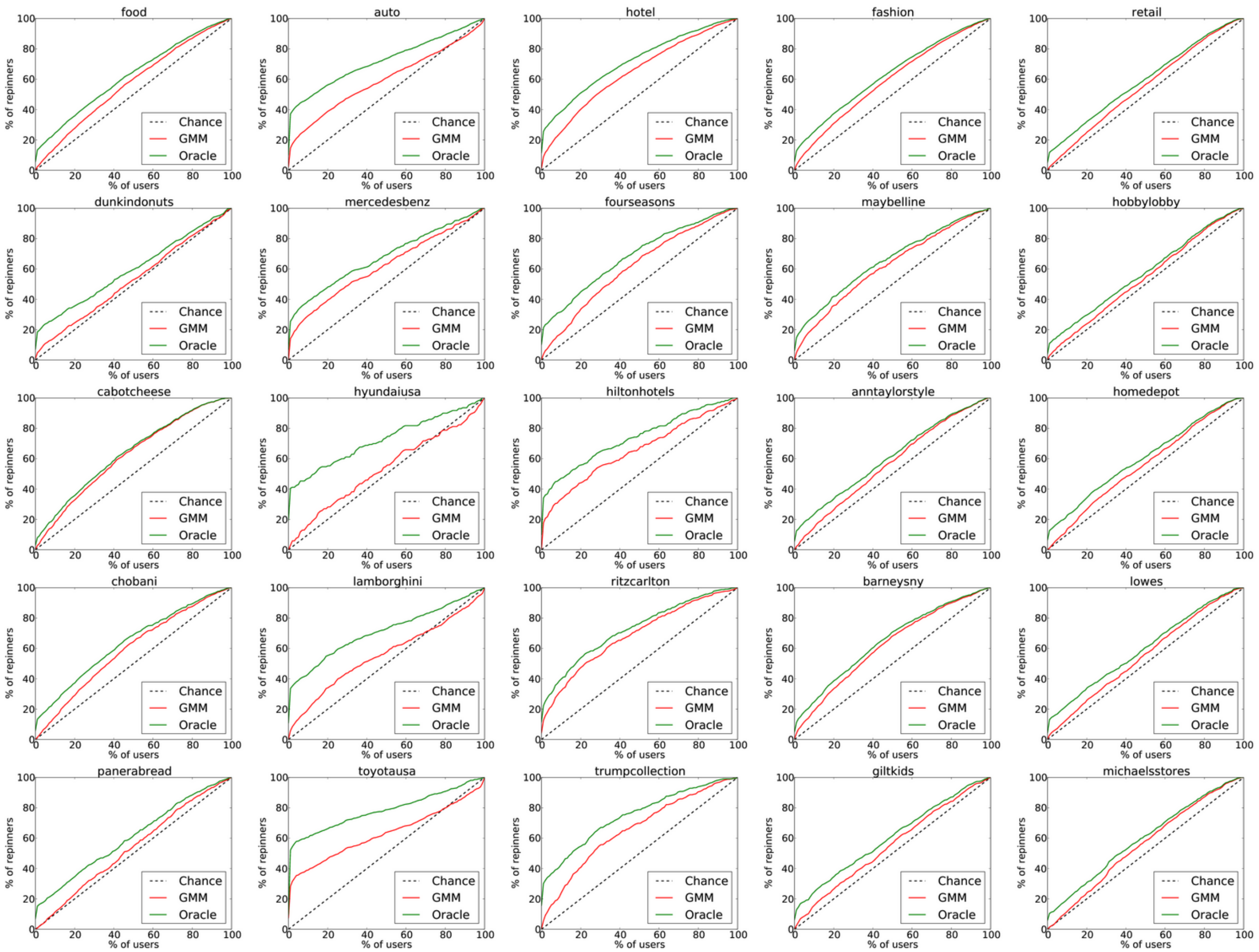


Figure 3: Gains charts by the proposed GMM scheme, by chance and by the oracle. Each column shows the gains chart for all brands from a category, with the top-most row showing aggregate performance across all brands for the category, and the rows below showing that for individual brands. It is observed that significant lifts are created for all brands using the proposed GMM scheme.

to b_j when considering the signature for the user. In the presented results $N_p = 100$ has been considered. In Table 2, we show the number of users who have repinned each brand's selected pins, as well as the total number of pins of these repinners. Collectively, this results in a dataset of more than 1 million pins.

Next we compute the similarity measure between user pins and the brand pin $s_{p,i} = s(p, \mathcal{P}'_{u_i})$ using an appropriate function for $s(\cdot,\cdot)$ depending on whether a model-based or model-free approach is being used. For each pin, we also create ground truth based on the available list of repinners collected from the website, denoted by $t_{p,i}$, which is defined as following:

$$t_{p,i} = \begin{cases} 1 & \text{if the user } u_i \text{ has repinned } p \\ 0 & \text{if otherwise} \end{cases} . \tag{5}$$

Gains charts. We generate a gains chart to demonstrate the performance of our system. A gains chart provides a visual summary of the usefulness of the information provided by one or more models for predicting a binomial outcome variable, and is commonly used in predictive modeling for marketing [15]. Specifically, in our setup, we evaluate the performance of the model prediction given by $s_{p,i}$ with respect to the outcome variable $t_{p,i}$. The gains chart is computed as follows. We sort $s_{p,i}$ where $u_i \in U$ in descending order, producing a new sequence s_{p,k_l} with a corresponding mapping of indices k_l. In other words, $s_{p,k_1} \geq s_{p,k_2} \geq s_{p,k_2} \cdots \geq s_{p,k_{|U|}}$.

A gains chart is then generated by computing the percentage of ground truth repinners recalled in a sorted shortlist of users, versus the percentage of users in that shortlist. Specifically, for each percentage level $r\%$ (x-axis), we compute the y-axis in gains chart to be:

$$gains(r\%) = \frac{\sum_{i=1}^{|U| \times r\%} t_{p,k_i}}{\sum_{i=1}^{|U|} t_{p,i}}. \tag{6}$$

Intuitively, a gains chart conveys the predictive power of our system, as it shows, for shortlists consisting of highest-

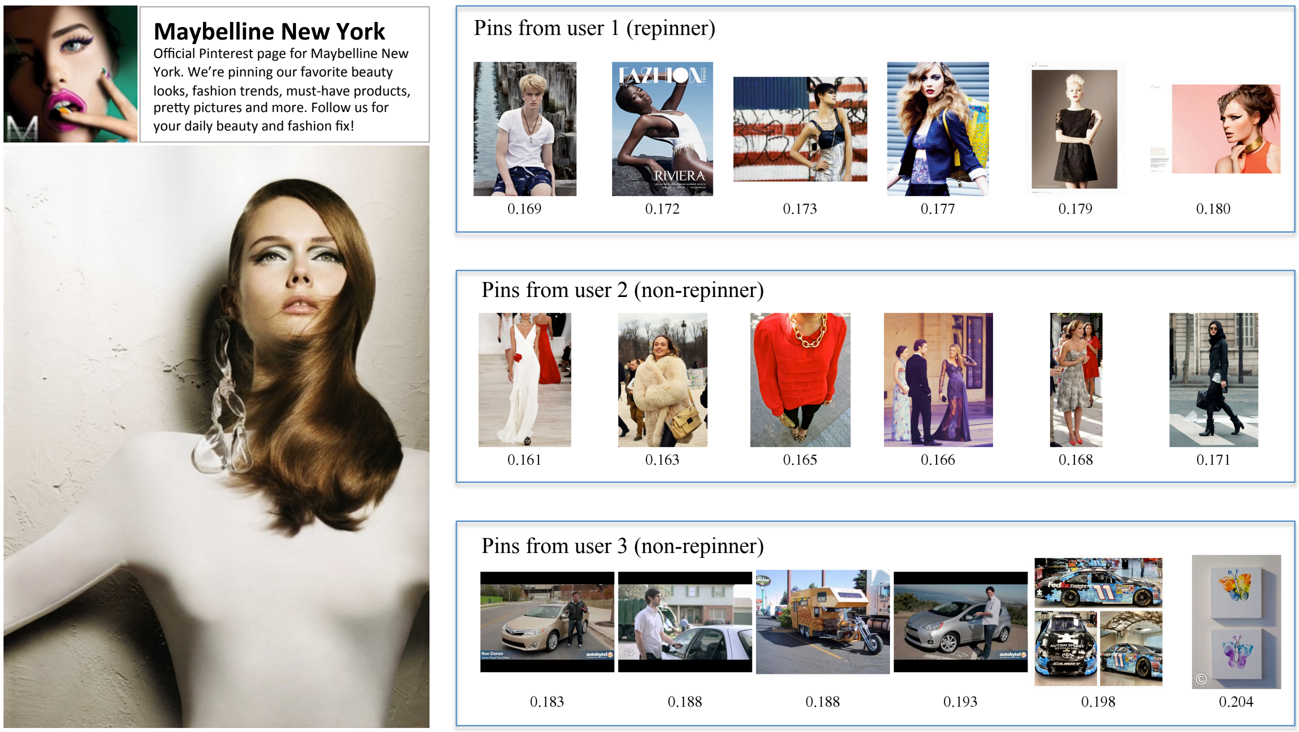


Figure 4: A brand pin by Maybelline (left bottom) and retrieved best-matching pins from three users (right). User 1 is a repinner with high similarity measure to the Maybelline pin, User 2 is a non-repinner with high similarity measure to the brand pin, and User 3 is a non-repinner with low similarity measure to the brand pin. It can visually verified that the best-matching pins by User 1 have high visual similarity to the brand pin, so are the best-matching pins by User 2, but the same is not true for those by User 3.

Table 2: List of brands with the number of repinners (Rp) across all the sampled pins, and the total number of pins across all the corresponding repinners.

Brand	Rp	Pins	Brand	Rp	Pins
Lamborghini	231	21k	Hyundaiusa	127	11k
Toyotausa	173	16k	MercedesBenz	309	29k
Anntaylorstyle	571	57k	Giltkids	393	39k
Barneysny	696	69k	Maybelline	633	63k
Chobani	750	74k	Panerabread	619	61k
Cabotcheese	719	71k	DunkinDonuts	505	49k
Fourseasons	511	51k	Trumpcollection	238	22k
Hiltonhotels	213	20k	Ritzcarlton	264	25k
Homedepot	764	76k	Michaelsstores	821	81k
HobbyLobby	765	76k	Lowes	737	72k

score users of different lengths, how much lift over chance we can get in predicting the ground truth repinners of a brand image.

Gains charts in aggregate. We also consider gains charts in aggregate to demonstrate the performance of our system for each brand and each industry category. Given the set of pins $\mathcal{P}$ (a brand, or an industry category), for each $p \in \mathcal{P}$, we similarly sort $s_{p,i}$ where $u_i \in U$ in descending order, producing a new sequence s_{p,k_l} with a corresponding mapping of indices k_l. In other words, $s_{p,k_1} \geq s_{p,k_2} \geq s_{p,k_2} \cdots \geq s_{p,k_{|U|}}$.

In generating the gains chart at $r\%$, we compute the following at the $r\%$ level:

$$gains(r\%) = \frac{\sum_{p\in\mathcal{P}} \sum_{i=1}^{|U|\times r\%} t_{p,k_i}}{\sum_{p\in\mathcal{P}} \sum_{i=1}^{|U|} t_{p,i}}. \tag{7}$$

4.2 Detailed Results

The performance of the proposed technique on the task of predicting the repinners' list is evaluated using four different schemes of retrieval:

- Min-pooling: Distance-based approach with Γ set to the *min* operator in Eqn. 1.
- Average-pooling: Distance-based approach with Γ set to the *average* operator in Eqn. 1.
- GMM (proposed): Model-based approach with similarity given by Eqn. 2.
- Oracle: The min-pooling scheme with $\mathcal{P}'_{u_i} = \mathcal{P}_{u_i}$. That is, we retain brand pins when considering the signature for the user. Note that the oracle's performance is in general not achievable, but is introduced for benchmarking purpose.

Table 3 summarizes the results across these schemes using the gains ratio, defined as the ratio of the area under the

Table 3: **Performance evaluation using gains ratio for brands and categories across three different schemes for computing similarity.**

Brand/Category	GMM	Min	Avg
MercedesBenz	0.91	0.90	**0.97**
Hyundaiusa	**0.74**	0.71	**0.74**
Lamborghini	0.79	0.74	**0.91**
Toyotausa	**0.79**	**0.79**	0.72
Auto	0.82	0.79	**0.85**
Maybelline	**0.94**	0.91	0.80
Anntaylorstyle	**0.93**	0.92	0.83
Barneysny	**0.95**	0.91	0.84
Giltkids	**0.92**	0.87	0.83
Fashion	**0.94**	0.89	0.83
DunkinDonuts	**0.88**	**0.88**	0.68
Cabotcheese	0.97	0.92	**0.99**
Chobani	**0.94**	0.89	0.78
Panerabread	**0.88**	0.87	0.82
Food	**0.92**	0.89	0.82
Fourseasons	**0.90**	0.82	0.87
Hiltonhotels	**0.89**	0.86	0.77
Ritzcarlton	**0.94**	0.88	0.83
Trumpcollection	**0.88**	0.78	0.86
Hotel	**0.90**	0.83	0.85
HobbyLobby	**0.95**	**0.95**	0.81
Homedepot	**0.92**	0.90	0.82
Lowes	**0.93**	0.90	0.83
Michaelsstores	**0.93**	**0.93**	0.81
Retail	**0.93**	0.92	0.81

curve for a given scheme divided by that for the oracle. It is observed that the GMM scheme is top-performing across 4 out of the 5 categories, and is also the top performer for most of the individual brands.

In order to get a graphical summary of the prediction performance, the gains charts for the GMM scheme and the oracle are plotted in Figure 3. Each column shows the gains chart for all brands from a category, with the topmost row showing aggregate performance across all brands for the category, and the rows below showing that for individual brands. It is observed that significant lifts are created for all brands using the proposed GMM scheme. For instance, for the auto category, nearly 15% of the repinners are captured in the top 1% of the predicted list.

In order to better understand the working of proposed framework, we consider an example of a pin from Maybelline's collections, and show examples of how the users are ranked on their propensity to repin that pin, based on images that have been previously pinned by them. In Figure 4, we show on the left side a sample pin by Maybelline which is a model's photo, and on the right side a number of best-matching user pins with the highest similarity measures by three users. Collectively, User 1's similarity measure to the Maybelline pin is at the 99^{th} percentile among its repinners, User 2's similarity measure to the brand pin is at the 99^{th} percentile among its non-repinners, and User 3's similarity measure to the brand pin is at the 1^{th} percentile among its non-repinners. It can visually verified that User 1's best-matching pins have high visual similarity to the brand pin, so are User 2's best-matching pins, but the same is not true for User 3's. It is also interesting to note that, the semantic content of the Maybelline pin is captured well by the best-matching pins of both User 1 and User 2, which validates the quantitative results achieved by our system.

4.3 Discussions

Per-category performance. We observe that certain categories perform better than other—for instance, as seen in Figure 3, hotel and auto are the top performing categories, whereas the performance of retail is significantly worse. This is arguably because images in the auto and hotel category tend to be relatively more homogeneous whereas those from retail tend to be diverse.

Interesting, we also observe from Table 3 that that more heterogeneous the collection of brand or category images, the more the improvement of the GMM scheme with respect to the average pooling scheme. For example, we observe that in hotel and auto, the difference between the GMM and average pooling schemes is the least across categories. On the contrary, for retail–the category with the most perceived diversity in the collection of pins–the improvement gained by GMM is the most.

This phenomenon is explained by the fact that heterogeneity in the pin collection is well modeled by the multimodality of mixture models, whereas a single summary statistic such as the average is incapable of capturing this diversity in the underlying data. Further, GMMs also provide an improvement over min-pooling, because as with average-pooling, considering only the closest user pin to the given brand pin does not capture the diversity of the user's interest, and hence their interaction patterns.

Potential applications. First, given the findings that we made in this paper, it is plausible that brand marketers can optimize each of their campaign images based on the similarity to the image collections of the target audience. In particular, when a brand marketer has a set of candidate images, she can run them through a system such as our prototype, given a sample set of brand target audience. By this process, the candidate with the best average score with respect to the sample target audience can be chosen for the marketing campaign.

Second, brands can strategically leverage our system's predicted users with high propensity to interact with their pins. Since these users' pins align semantically with the brand's pins, they are potentially high-quality brand prospects. However, many of these users may not have formerly interacted with the brand pins, either by repinning or by being "followers" of the brand. Given the knowledge of these users, a brand marketer can proactively "like" these users' pins and consequently draw their attention to her brand.

5. CONCLUSIONS

We have presented a prototype system in this paper that is able to predict who interacted with a brand image on Pinterest. We approached this problem using social media retrieval techniques; that is, given a brand image, we retrieved similar user images and based on the computed similarity measures, we ranked users and inferred their propensity to interact with the brand image. We note that our system is purely based on image features, which makes it complementary to the existing body of work leveraging social and textual content, among other sources of information. Our system has been systematically tested on a large-scale Pinterest dataset of more than 1 million images and produced promising results.

6. REFERENCES

[1] Instagram website. http://www.instagram.com.

[2] Pinterest website. http://www.pinterest.com.

[3] R. Arandjelovic and A. Zisserman. Three things everyone should know to improve object retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[4] M. Arevalillo-Herraez, J. Domingo, and F. J. Ferri. Combining similarity measures in content-based image retrieval. *Pattern Recognition Letters*, 29(16), 2008.

[5] J. Berger and K. L. Milkman. What makes online content viral. *Journal of Marketing Research*, 49(2):192–205, 2012.

[6] G.-H. Cha. Kernel principal component analysis for content based image retrieval. In *Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining*, 2005.

[7] T. Chen, L. Tang, Q. Liu, D. Yang, et al. Combining factorization model and additive forest for collaborative followee recommendation. In *KDD Cup*, 2012.

[8] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010.

[9] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[10] L. Ding and M. Belkin. Component based shape retrieval using differential profiles. In *ACM International Conference on Multimedia Information Retrieval*, 2008.

[11] D. Feng, W. C. Siu, and H. J. Zhang. *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*. Springer, Berlin Heidelberg, Germany, 2003.

[12] G. A. Fowler. Facebook: One billion and counting. *The Wall Street Journal*, Oct 2012.

[13] T. Hertz, A. Bar-Hillel, and D. Weinshall. Learning distance functions for image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[14] A. M. Kaplan and M. Haenlein. Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53:59–68, 2010.

[15] P. Leeflang, D. Wittink, M. Wedel, and P. Naert. *Building Models for Marketing Decisions*. Springer, 2000.

[16] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral marketing. In *ACM Conference on Electronic Commerce*, 2006.

[17] Y. Liu-Thompkins. Seeding viral content: The role of message and network factors. *Journal of Advertising Research*, 59, 2012.

[18] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 2004.

[19] S. Neti. Social media and its role in marketing. *International Journal of Enterprise Computing and Business Systems*, 1(2), 2011.

[20] K. A. Peker. Binary SIFT: Fast image retrieval using binary quantized sift features. In *International Workshop on Content-Based Multimedia Indexing*, 2011.

[21] N. Ramzan, R. van Zwol, J.-S. Lee, K. Clüver, and X.-S. Hua. *Social Media Retrieval*. Springer, 2013.

[22] S. Rendle. Social network and click-through prediction with factorization machines. In *KDD Cup*, 2012.

[23] C. Savio and J. Raroque. Social media's growing influence among high net worth investors. Technical report, Cogent Research and Linkedin, May 2012.

[24] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *IEEE International Conference on Computer Vision*, 2003.

[25] M. Trusov, R. E. Bucklin, and K. Pauwels. Effects of word-of-mouth versus traditional marketing: Findings from an Internet social networking site. *Journal of Marketing*, 73(5), 2009.

[26] C. Wang, J. Zhao, X. He, C. Chen, and J. Bu. Image retrieval using nonlinear manifold embedding. *Neurocomputing*, 72(16–18):3922–3929, 2009.

[27] J. Ye, R. Janardan, and Q. Li. GPCA: An efficient dimension reduction scheme for image compression and retrieval. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004.

[28] H. Yuan. Texture image retrieval based on a gaussian mixture model and similarity measure using a kullback divergence. In *IEEE International Conference on Multimedia and Expo*, 2004.

[29] X. Yuan, J. Yu, Z. Qin, and T. Wan. A SIFT-LBP image retrieval model based on bag-of-features. In *IEEE International Conference on Image Processing*, 2011.

[30] A. Zbieg, B. Zak, J. Jankowski, R. Michalski, and S. Ciuberek. Studying diffusion of viral content at dyadic level. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2012.

[31] Z.-H. Zhou and H.-B. Dai. Query-sensitive similarity measure for content-based image retrieval. In *IEEE International Conference on Data Mining*, 2006.