Learning Visual Elements of Images for Discovery of Brand Posts

FRANCESCO GELLI, National University of Singapore, Singapore
TIBERIO URICCHIO, Università degli Studi di Firenze, Italy
XIANGNAN HE, University of Science and Technology of China, China
ALBERTO DEL BIMBO, Università degli Studi di Firenze, Italy
TAT-SENG CHUA, National University of Singapore, Singapore

Online Social Network Sites have become a primary platform for brands and organizations to engage their audience by sharing image and video posts on their timelines. Different from traditional advertising, these posts are not restricted to the products or logo but include visual elements that express more in general the values and attributes of the brand, called brand associations. Since marketers are increasingly spending time in discovering and re-posting user generated posts that reflect the brand attributes, there is an increasing demand for such discovery systems. The goal of these systems is to assist brand experts in filtering through online collections of new user media to discover actionable posts, which match the brand value and have the potential to engage the consumers. Driven by this real-life application, we define and formulate a new task of *content discovery for brands* and propose a framework that learns to rank posts for brands from their historical timeline. We design a Personalized Content Discovery (PCD) framework to address the three challenges of high inter-brand similarity, sparsity of brand–post interactions, and diversification of timeline. To learn fine-grained brand representation and to generate explanations for the ranking, we automatically learn visual elements of posts from the timeline of brands and from a set of brand attributes in the domain of marketing. To test our framework we use two large-scale Instagram datasets that contain a total of more than 1.5 million image and video posts from the historical timeline of hundreds of brands from multiple verticals such as food and fashion. Extensive experiments indicate that our model can effectively learn fine-grained brand representations and outperform the closest state-of-the-art solutions.

CCS Concepts: • **Information systems → Social networks**; **Recommender systems**; *Users and interactive retrieval*;

Additional Key Words and Phrases: Content discovery, image ranking, computational marketing

ACM Reference format:
Francesco Gelli, Tiberio Uricchio, Xiangnan He, Alberto Del Bimbo, and Tat-Seng Chua. 2020. Learning Visual Elements of Images for Discovery of Brand Posts. *ACM Trans. Multimedia Comput. Commun. Appl.* 16, 2, Article 56 (May 2020), 21 pages.
https://doi.org/10.1145/3385413

NExT++ research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IRC@SG Funding Initiative.
Authors' addresses: F. Gelli and T.-S. Chua, National University of Singapore, Singapore, Singapore; emails: francesco.gelli@u.nus.edu, chuats@comp.nus.edu.sg; T. Uricchio (corresponding author) and A. Del Bimbo, Università degli Studi di Firenze, Florence, Italy; emails: {tiberio.uricchio, alberto.delbimbo}@unifi.it; X. He (corresponding author), University of Science and Technology of China, Hefei, China; email: xiangnanhe@gmail.com.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2020 Association for Computing Machinery.
1551-6857/2020/05-ART56 $15.00
https://doi.org/10.1145/3385413

1 INTRODUCTION

The recent growth of user generated media posted on Social Network Sites (SNS) has greatly impacted media industries such as media and advertising. Most brands and organizations widely use Instagram and other SNS to regularly share posts and engage their audience. However, traditional product-promoting messages are losing their effectiveness due to consumers becoming increasingly resistant [36]. Marketers are therefore exploring not just new platforms to communicate their ads and messages but also new types of content, such as User Generated Content (UGC) [10, 11, 18, 31]. In this scenario, the posts are not restricted to the products or logo only but include certain visual elements that reinforce more in general the values and attributes of the brand, called *brand associations* [5]. For instance, it is common for car brands to post images of landscapes to associate with their organization the brand attributes of freedom and sophistication. There is, in fact, evidence that the consistent use of media relevant to the brand attributes contributes to increase the interactions on the brand posts and improve consumer engagement [43].

However, the discovery of relevant UGC for brands requires a huge amount of manual effort by marketing experts who understand the brand ideas and have the daunting task of browsing through millions of possible posts. As a first step to simplify this process, companies such as Olapic, Chite, and Stackla[1] offer hashtag-based services for the discovery of such actionable UGCs, which may severely limit the amount of relevant content that can be discovered. In fact, brand-related hashtags require the users to be aware of a brand's campaign and deliberately take action to be discovered, missing a large slice of UGC that may be a better match for the brand, because of the lack of related hashtags. To the best of our knowledge, the problem of automatically discovering visual content that matches the unique visual elements and style of a brand is still unexplored. Moreover, existing discovery systems ignore the known associations and attributes of brands, which for example include the brand personality [1] and the Brand Asset Valuator (BAV) framework [32]. There has still been no study on the existence of links between brand attributes and the visual elements in media posts of brands.

Driven by these real-life needs, we explore the task of using historical timelines of brands and their attributes to learn to rank media posts based on their relevance to the brand. We identify three main challenges for this problem. The first is the inter-brand similarity, where the same associations are regularly used by multiple brands, resulting in visual posts having subtle differences from the ones used by the competitors. For example, the visual elements of cars often appear in posts by most automobile companies, and the differences may only be in the logo or visual style. It is thus a fine-grained problem, where it is important to distinguish as much as possible the specific set of unique values and associations of each brand. The second challenge is the sparsity of brand–post interactions. It is extremely rare for different brands to post the same exact post. Marketers tend to post original content, either for reason of copyright infringement or marketing strategy to project brand uniqueness [4]. This is different from the recommendation scenarios, where for each item (corresponding to a post) there are often multiple users (corresponding to brands) interacting with it (e.g., likes and views) [24, 30, 35]. The sparse nature of the brand–post interaction leads to a weak collaborative signal, which needs to be strengthened with content-based approaches. Finally, the third challenge stems from the diversification of the timeline of brands. Figure 1 shows an example of timeline diversification. The figure illustrates a sequence of images and video posts from the Instagram timeline of the retail brand Wallmart. We can easily notice certain recurring themes, as indicated by the four-color marks in the top-right corner of each post: young

[1]Olapic: www.olapic.com; Chute: www.ignitetech.com/chute; Stackla: www.stackla.com.

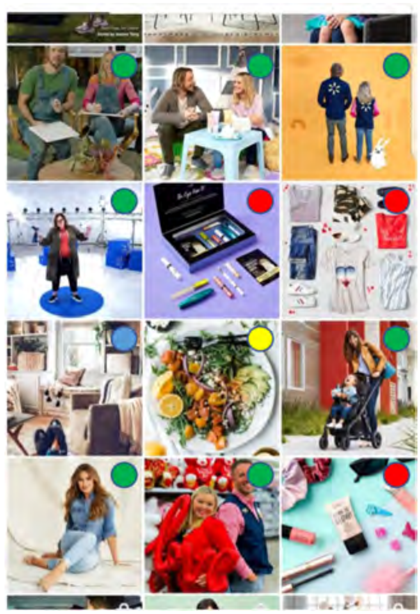


Fig. 1. Example of timeline diversification. Four different colors indicate four different brand associations.

people happily posing in an environment related to retail (green), meticulous product arrangements (red), food (yellow), and interior spaces (blue). In fact, different posts are used to reinforce different brand attributes as well as to engage different groups of users. For this reason each attribute is reflected only by a subset of the posts. In this example, the posts marked in green reflect the brand attributes *Young*, *Trendy*, and *Familiarity*, while the ones in red can easily express *High Quality* and *Chic*. This behaviour makes it difficult to learn which visual elements correspond to a particular brand attribute since the method needs to learn which posts reflect it among the many on the timeline of the brand.

We formulate the new task of *content discovery for brands* and propose the Personalized Content Discovery (PCD) framework to addresses the three above-mentioned challenges. PCD uses the historical timeline to learn the high-level semantics of brand associations, which correspond to recurring visual aspects of the historical timeline of brands. Being purely content-based, PCD can be used to rank a set of unseen posts for a brand by simply measuring their similarity. To improve the performance and generate explanations for the ranking, we introduce a variant of PCD to integrate brand attributes from the domain knowledge of marketing. In fact, a marketer may be interested in discovering a post to communicate a specific brand attribute, as in one of the four cases in Figure 1. The model uses probabilistic optimization to learn which objects in the image correspond to each of the brand attributes, producing a more accurate ranking. By modeling the diversified post timeline as a normal distribution centered on a consensus value, our method is more robust to timeline diversity than other methods that fit all the posts equally to the same brand attribute. We propose a generic solution to integrate any kind of brand attributes, including Aaker's brand personality and BAV dimensions.

We train our models end-to-end using two large-scale Instagram datasets, which we collected by crawling the historical timeline of two sets of brands from different verticals such as food, fashion, and auto. We benchmark the models against the state-of-the-art method, using five different quantitative metrics and performed ablation studies and qualitative analysis of case studies.

We summarize the contributions of this work as follows: (1) We cast the emerging problem of content discovery for brands as a content-based learning-to-rank problem and highlight the three main challenges. (2) We design and benchmark Personalized Content Discovery (PCD) as a novel content discovery framework that learns the high-level semantics of brand associations. The method is able to outperform the closest state-of-the-art works in content-based recommendations. (3) We extend PCD to integrate brand attributes from the domain knowledge of marketing using probabilistic optimization. Experiment and visualization indicate that our method is able to learn visual elements that correspond to specific attributes as well as generate explanations. (4) We collected two large-scale Instagram datasets containing more than 1.1 million images and video posts with brand attributes. The datasets are released to the research communities for further studies on content discovery, popularity predictions and computational marketing.

2 RELATED WORK

We separate the related work into two parts, with the former describing the background scenario of Social Media Marketing and the latter introducing the most relevant works on image ranking systems.

2.1 Social Media Marketing and Computational Marketing

Social Media Marketing (SMM) is a real-life discipline that involves conducting marketing campaigns on online social networks. Nonetheless, it has constantly solicited a vast amount of research across different disciplines, not only in the marketing community but also in psychology [3] and computer science. For example, Gensler et al. [17] outlined a list of research questions related to SMM, such as investigating what kind of brand content will stimulate consumers and spread in social networks. Such research question inspired scientists to use computational approaches for explaining marketing dynamics in social media, which fall under the sub-field of Computational Marketing. One popular case of these works is the study by De Vries et al. [9], who employed regression models to investigate which features of social brand posts engage users. The authors outlined a set of post features that have a positive impact on popularity indicators such as the number of likes and comments, including image vividness and post interactivity. In the same line, several other works investigated similar indicators with multiple social features on Facebook [23, 38], Twitter [2], and Instagram [34, 37]. A second line of research tackled the problem of identifying the brand associations from social media posts [26, 27]. These papers propose clustering-based methods to automatically detect visual brand associations from social networks post as image clusters. However, they treat each brand independently from the others, while we argue that the same or similar associations are shared among different brands. We thus propose a model that jointly learns brand associations among a large set of brands.

2.2 Image Ranking Systems

Since we formulate the problem of content discovery for brands as a learning-to-rank framework, we here introduce some of the most relevant works on image ranking systems, which we divide into recommendation and retrieval systems. For the former, which solves the problem of recommending images to social users, we focus on the works that combine user-image interaction with image content.

A first group of works is based on Factorization Machines (FM) and their neural extension, which are designed to model the interaction between each combination of entities, including users, items and any additional numerical value information available [6, 21, 39]. Factorization Machines have been extended to content-based scenarios, adopting context information such as hashtags, text in the image [7], or integrating dense image features extracted with a pre-trained CNN [13]. Other frameworks make use of denoising auto-encoders [44, 47], attention models [6] or recurrent neural networks [22] to integrate content features in recommendation architectures. All these works rely on the hypothesis that users interact with multiple items, which does not hold in the scenario of content discovery for brands, where the case of multiple brands using the exact same image post is unrealistic.

The pairwise optimization framework Bayesian Personalized Ranking (BPR) by Rendle et al. [40] inspired another line of recommendation works. Among these, Visual Bayesian Personalized Ranking (VBPR) enriched the item latent vectors with image features [19], while Yu et al. [45] applied BPR to the fashion domain. In a more general setting, Neural Personalized Ranking (NPR) by Niu et al. [35] complements existing matrix factorization and VBPR with a new more flexible neural network architecture. The model supports image features as input, which can be used together with other context clues such as topic or location.

However, these models are hybrid recommender systems, which integrate content-features with the collaborative signal. Purely content-based solutions rely instead on content-features only [16, 24, 30, 33], which have the advantage of preventing the item cold-start problem [42] as well as being more robust to sparsity. The main idea of these works is to map users and images in the same latent space, where recommendations are selected according to proximity. In a more general setting, Lei et al. [30] proposed Comparative Deep learning (CDL) triplet network, where a positive and a negative sample image are simultaneously forwarded through two convolutional sub-networks with shared weights, together with information related to a user who interacted with the positive, but not with the negative photograph. The user features are processed with a separate sub-network and then compared with the two image latent features with element-wise subtraction. However, these approach is unsuitable for our task because of the timeline diversification, since brand attributes are not reflected by all the image posts of a brand.

In the fashion domain, several works propose methods to suggest image products to scenes or users. Kang et al. [24] proposed a method to perform recommendation of product images. They use a siamese network and Bayes Personalized Ranking for learning fashion-aware based image representation. As a follow up work, the same authors proposed the task of "complete the look" where they aim at recommending visually compatible products based on scene images [25]. In [49] the authors considers the aggregation of online posts by influencers as a mean to improve the recommendation with a bidirectional LSTM [49].

Other works investigated the problem of ranking images or videos using attributes of external entities, such as users or brands. Zhang et al. [48] defines nameable and unnameable attributes for content based retrieval, where the former are human-readable labels that are learned from annotated images and the latter are learned with an iterative clustering algorithm. Later, Cui et al. used attributes of users for video recommendations [8]. Based on matrix factorization, their method represents videos with attributes of users and represent users with content attributes of videos. More recently, Gelli at al. [14] proposed a probabilistic framework for ranking images by subjective attributes that use the Kullback-Leibler divergence [29] to learn how brand attributes are reflected in their image posts. However, their framework is designed to rank images by a specific visual attribute, while we aim at ranking image posts for brands instead.

3 PERSONALIZED CONTENT DISCOVERY

We dedicate this section to first presenting the notations, followed by defining the problem of content discovery for brands. We then describe the details of our Personalized Content Discovery (PCD) framework. Finally, we describe the details of the optimization method that we use to train the models.

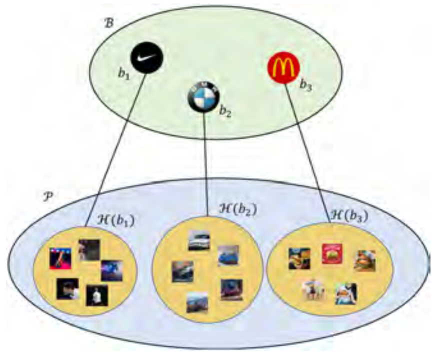


Fig. 2. Illustration of data entities involved in the problem of content discovery for brands.

3.1 Notations and Problem Formulation

In this article, we use bold capital letters and bold lowercase letters to represent matrices and vectors, respectively. In the case of matrices, we use the subscript notation to denote a specific row. For example, $\mathbf{A}_i$ indicates the ith row of matrix $\mathbf{A}$. We use capital cursive letters for sets and regular lower case for the set elements, such as $a \in \mathcal{A}$.

The data entities involved in the problem are illustrated in Figure 2. We indicate the set of brands as $\mathcal{B} = \{b_1, \ldots, b_N\}$, where $b_i \in \mathcal{B}$ is a brand with an active social media account. We denote the total set of image posts with the notation by $\mathcal{P} = \{p_1, \ldots, p_M\}$. We refer to the historical timeline of brand b as $\mathcal{H}(b)$, where $\mathcal{H}(b) \subseteq \mathcal{P}$. With this notation, we indicate that post p is posted by brand b if $p \in \mathcal{H}(b)$. Because of the sparsity of brand–post interactions, timelines of different brands have little or no overlap: $\mathcal{H}(b_i) \cap \mathcal{H}(b_j) \sim 0$. Finally, $\mathbf{a_b}$ denotes the vector of attributes of brand b and, for a post p, the image feature vector and the distribution vector over a set of pre-defined concepts are $\mathbf{x_p}$ and $\mathbf{c_p}$, respectively. For a simpler notation, the brand and post indexes b and p are often omitted in the following.

Given these notations for the input data, the goal of content discovery for brands is to learn a function $f : \mathcal{B}x\mathcal{P} \mapsto \mathbb{R}$ such that for each post $p_i \in \mathcal{H}(b)$ we have:

$$f(b, p_i) > f(b, p_j), \tag{1}$$

where p_j is a post of any brand different from b: $\hat{b} \neq b$. In other words, given a set of brands and related posting histories, we aim to learn a relevance function f that can be used to rank a set of new image posts, such that those that are relevant and likely to be adopted by the brand will be ranked higher.

3.2 Personalized Content Discovery

The PCD model is based on the principle of mapping the users and items into the same latent space [41]. We chose to adopt a similar model to learn a common latent space for both brands and image posts since these frameworks can easily be extended to leverage the predictive power of deep neural networks.

The first component of PCD is dedicated to learning the brand representation, while the second learns the post representation. When a post has a similar representation as the brand representation, it is considered as a good match for the brand. Specifically, let $\mathbf{b} \in \mathbb{R}^k$ and $\mathbf{p} \in \mathbb{R}^k$ denote the latent representation of brand b and post p, respectively, the similarity between b and p is defined as:

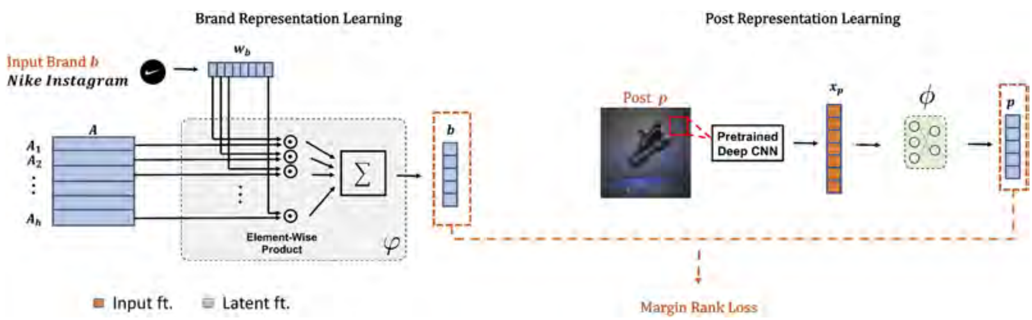


Fig. 3. PCD: Model architecture.

$$f(b,p) = \frac{\mathbf{b}^T\mathbf{p}}{||\mathbf{b}||\ ||\mathbf{p}||}.$$

The whole PCD framework is represented in Figure 3, where the two components of brand and post representation learning are illustrated on the left and right parts, respectively.

3.2.1 Brand Representation Learning. Because of high inter-brand similarity, simply adopting one-hot brand ID may fail to learn the discriminative representation that is capable of making fine-grained distinctions among competitor brands [24, 35]. Our solution is derived from the intuition that each brand is uniquely characterized by multiple themes or brand associations. Different from other works that pre-allocate aspects with fixed handcrafted features [20], PCD learns the importance of every single association in an end-to-end manner for each brand. We automatically learn a fixed-sized latent vector with high-level semantics for each brand association. We add a new set of parameters representing the associations for each brand, which are learned at the training stage. We name them association vectors and arrange them as rows of a matrix $\mathbf{A} \in \mathbb{R}^{hk}$, where h is an arbitrary number of associations and k is the number of dimensions of the latent space. The final brand representation is then learned from both the one-hot brand ID and the association vectors.

The brand representation is computed as:

$$\mathbf{b} = \varphi(\mathbf{A}, \mathbf{w_b}) = \sum_{i=1}^{h} \mathbf{A}_i \circ \mathbf{w_b}, \tag{2}$$

where $\mathbf{w_b} \in \mathbb{R}^h$ are the importance weights for brand b and $\circ$ indicate the element-wise multiplication. Since brands are free to assume any weighted combination of the h association vectors $\mathbf{A}_i$, this method allows a richer learning of fine-grained brand representations compared with the one-hot brand ID.

3.2.2 Post Representation Learning. Because of the sparsity of brand–post interactions, the one-hot post ID as input feature does not provide sufficient information. For this reason, PCD uses uniquely image content to learn image post representation, similarly to the solution by Lei et al. [30]. This approach also prevents the item cold-start problem [42], which is critical for ranking a set of new posts that were not used to train the model.

We design a two-layer neural network ϕ for this task, whose sole input is the image content. To achieve faster training, we adopt the same strategy as in VBPR [19] and utilize input features extracted with a pre-trained CNN. We finally compute the post vector as $\mathbf{p} = \phi(\mathbf{x_p})$, where ϕ follows the following form:

$$\phi(\mathbf{x}) = \mathbf{W}_2(\xi(\mathbf{W}_1\mathbf{x} + \boldsymbol{\gamma}_1)) + \boldsymbol{\gamma}_2, \tag{3}$$

where W_1 and W_2 are learned matrices, γ_1 and γ_1 are bias terms, and $\xi(x)$ is a Leaky Relu with slope 0.01:

$$\xi(x) = \begin{cases} x & \text{if } x > 0 \\ 0.01x & \text{otherwise} \end{cases}. \tag{4}$$

3.2.3 Optimization. Our training data T consist of a set of brand–post pairs (b, p). Similarly to VBPR [19], we adopt pairwise learning, which involves sampling a certain number of negative posts for each pair. With pairwise learning, a single training data point has the following form: (b, p_{pos}, p_{neg}), where b is a brand, $p_{pos} \in \mathcal{H}(p)$ and $p_{neg} \notin \mathcal{H}(p)$. The basic intuition is that for a brand, the positive sample posts should be closer than the negative ones in the learned latent space.

We train our PCD model $f : \mathcal{B}x\mathcal{P} \mapsto \mathbb{R}$ using the following ranking loss:

$$\mathcal{L}(b, p_{pos}, p_{neg}) = max(0, (f(b, p_{neg}) - f(b, p_{pos})) + \eta), \tag{5}$$

where η is the minimum distance desired between the negative and the positive samples. The final ranking loss is computed as:

$$\mathcal{L}_{PCD}(b, p_{pos}, p_{neg}) = \mathcal{L}(b, p_{pos}, p_{neg}) + \alpha * \sum_i |\mathbf{w_b}| + \beta * ||\theta||_2, \tag{6}$$

where θ are the set of all the weights of the model and α and β control the importance of the regularization terms. The two regularization terms are added to increase the interpretability of the aspects associated with a brand and to reduce overfitting. For the first term, we recall that each brand has weights for each association latent vector $\mathbf{w_b}$ (Section 3.2.1). Our desired effect is that these weights operate as "selectors," positively or negatively affecting only a small set of associations for each brand. As a result, we adopt a L1 regularization on $\mathbf{w_b}$ to encourage a sparse representation [28]. For the second term, we adopt L2 regularization on every weight of the model.

Parameters $\mathbf{A}$, $\mathbf{w_b}$ and all other parameters can be learned from the model using any optimizer algorithm. We train our model for 20 epochs using the Adadelta algorithm, which adapts the step interval dynamically over time without the need to set any learning rate [46]. We train using mini batches of 256 brand–post pairs, which we shuffle at the beginning of each epoch before batching. To improve generalization performance, we employ dropout with 50% dropping rate and a loss margin of 0.3.

One of the major difficulties of pairwise learning is negative sampling. Computational requirements prevent training using all the possible brand–post pairs, and hence negative data must be selected using a sampling strategy. For this work we adopt uniform negative sampling: Given each brand–post pair (b, p), $p \in \mathcal{H}(b)$, we randomly sample 10 negative sample posts such that $p_i \notin \mathcal{H}(b)\ \forall i$.

3.3 Personalized Content Discovery with Brand Attributes

The most straightforward solution to integrate brand attributes, such as BAV or brand personality factors, would be to consider such information as an additional brand feature a, affecting all the brand–post pairs indiscriminately. Because of timeline diversification, this approach will introduce a strong noise component, since it would uniformly assign all posts of a brand to the same attribute scores. To address this limitation, we propose a variant of PCD named PCD*. The method is inspired by the probabilistic optimization of Reference [15], which is designed to learn the presence of subjective attributes in images.

The brand representation block of PCD* has two blocks, the first for learning the latent brand associations and a second to learn visual elements of brand attributes (Figure 4). While a generic brand representation vector $\mathbf{b_x}$ is computed in the same way as b in the standard PCD (Equation (2)), we compute an attributes-specific brand vector $\mathbf{b_a}$ to capture information about the

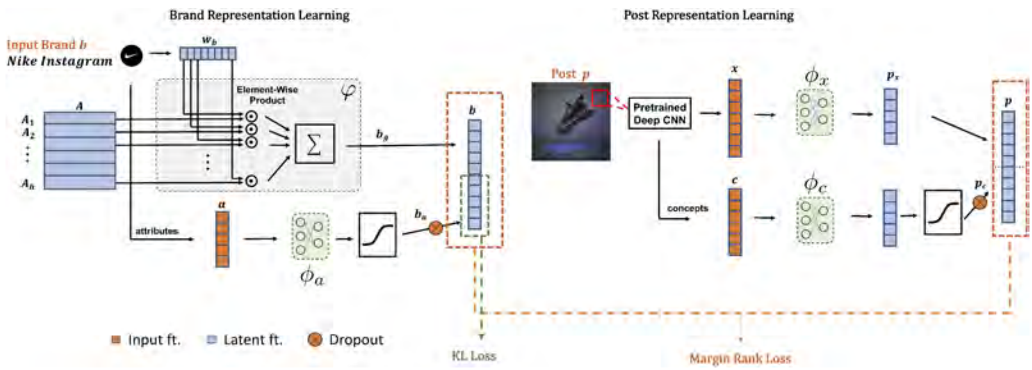


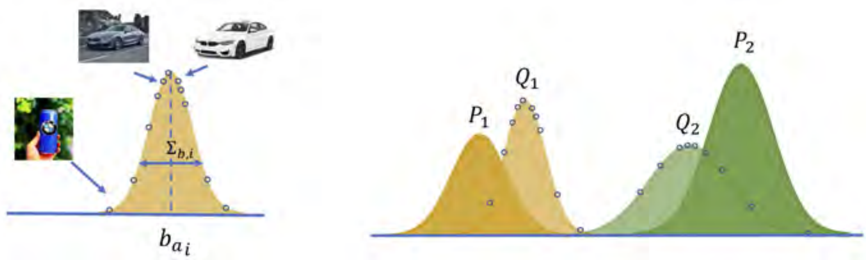
Fig. 4. PCD*: Model architecture.

brand attributes. We compute this vector as $\mathbf{b_a} = \sigma(\phi_a(\mathbf{a}))$, where the function σ is the sigmoid function and ϕ_a is a two-layer perceptron like in Equation (3). The function ϕ_a maps the vector of brand attributes in a new vector of latent attributes with the same dimension. For a brand b, the two intermediate vectors are concatenated in the final brand representation vector $\mathbf{b} = BrandRL(b) = cat(\mathbf{b_x}, \mathbf{b_a})$.

In analogy to the brand representation learning block, for a post p we compute the post representation vector as $\mathbf{p} = PostRL(p) = cat(\mathbf{p_x}, \mathbf{p_a})$. The two components are, respectively, a generic post representation vector $\mathbf{p_x} = \phi_x(\mathbf{x})$ similar to PCD and an attributes-specific vector $\mathbf{p_c} = \sigma(\phi_c(\mathbf{c}))$, where ϕ_c has the same structure of Equation (3) and $\mathbf{c}$ are the probabilities of the 1000 ImageNet concepts derived from the input image using a pre-trained concept classification neural network. We choose to learn $\mathbf{p_c}$ from the concept distribution under the assumption that marketers use associative concepts to build brand attributes [5]. To reduce the correlations between the latent features of the different brand attributes, we set a dropout rate of 0.75 for the elements of the brand and post attributes-specific vectors $\mathbf{b_a}$ and $\mathbf{p_c}$.

Same as in standard PCD, $\mathbf{p_x}$ is a generic image feature expressed as a mixture of latent brand associations. For the attributes-specific component $\mathbf{p_c}$, we want each ith element to indicate how much the image reflects the ith latent brand attribute. We enforce this behaviour by using a probabilistic optimization learning algorithm. The algorithm introduces a new loss term that encourages the vectors $\mathbf{p_c}$ of posts $p, \forall p \in \mathcal{H}(b)$ to be closer to vector $\mathbf{a}$ of brand b. Simply minimising the mean square error between $\mathbf{p_c}$ and $\mathbf{a}$ would be sensitive to outliers because of the challenge of the timeline diversification. For this reason, our algorithm constrains all image vectors $\mathbf{p_c}$ of a brand to follow a desired distribution. We illustrate this concept with an example in Figure 5(a): If a brand has a high score for the attribute *Upper Class*, then we assume that the average *Upper Class* score of all its images is near that score, even if there will be outlier images with much lower *Upper Class* score. In the figure, photos of luxury cars are being used to enforce the attribute of *Upper Class*, while the post of a can with the company logo is certainly an outlier. More formally, we make the assumption that given attribute i for brand b, with score $\mathbf{a}_i$, all of its timeline posts are such that their score $\mathbf{p_c}_i, \forall p \in \mathcal{H}(b)$ follow a one-dimensional normal distribution with mean $\mathbf{c}_i$, with standard deviation $\Sigma_{b,i}$ unknown. In other words, we expect the consensus of the attribute-specific scores $\mathbf{p_c}_i$ being close to the score of brand attribute: $\mathbf{a}_i$.

During each epoch, our training procedure iterates over all brands b using mini-batches of size B, as is shown in Algorithm 1. For each brand b in the mini-batch, we randomly sample K image posts p and compute the brand representations with the neural network *BrandRL* (line 10). We then iterate on the K posts in the image bucket of brand b, computing post representations with *PostRL*



(a) Probabilistic assumption (b) Probabilistic optimization

Fig. 5. (a) Illustration of our probabilistic assumption: Even if not all posts by BMW reflect the high *Upper Class* score, we assume that the mean of the distribution being centered on the attribute score. (b) The probabilistic optimization of two brands (in green and yellow): Q_1 and Q_2 are pushed by the KL divergence loss to match P_1 and P_2, respectively.

ALGORITHM 1: Optimization Algorithm

1: **procedure** TRAIN BATCH(a)
2: $B \leftarrow$ batch size
3: $K \leftarrow$ bucket size
4: $M \leftarrow$ number of brand attributes
5: $loss \leftarrow 0$
6: **for** b=1,…,B **do**
7: $\mathbf{X} \leftarrow$zero matrix$\in \mathbb{R}^{[K,M]}$
8: sample K posts $p_1, p_2, \ldots, p_K \in \mathcal{H}(b)$
9: sample K points $p'_1, p'_2, \ldots, p'_K \notin \mathcal{H}(b)$
10: $\mathbf{b} \leftarrow BrandRL(b)$
11: **for** k=1,…,K **do**
12: $\mathbf{p} \leftarrow PostRL(p_k)$
13: $\mathbf{p}' \leftarrow PostRL(p'_k)$
14: $loss \leftarrow$loss+$\mathcal{L}_{PCD}(\mathbf{b}, \mathbf{p}, \mathbf{p}')$
15: $\mathbf{p_a} \leftarrow$ select attributes-specific vector (last M elements of $\mathbf{p}$)
16: $\mathbf{X}_{k,:} \leftarrow$ add $\mathbf{p_a}$ as kth row
17: **for** i=1,…,M **do**
18: $P_{bi} \leftarrow$ reference normal distribution with mean $\mathbf{a}_i$ and std $\Sigma_{b,i}$
19: $Q_{bi} \leftarrow$ fit normal distribution from $\mathbf{X}_{:,i}$
20: $loss \leftarrow$loss+$\mathcal{L}_{KL}(P_{bi} \mid\mid Q_{bi})$
21: $nn, \Sigma \leftarrow$ update parameters with back-propagation

(line 12). We repeat the same computation for the negative samples (line 13) and compute the margin rank loss $\mathcal{L}_{PCD}$ as in Section 3.2 (line 14). Since vector $\mathbf{p}$ is the concatenation of generic and attributes-specific post representation vectors, we extract the latter by selecting the last m elements of $\mathbf{p}$ and add it as the ith row of matrix $\mathbf{X}$. The goal of the probabilistic loss is to ensure that the elements of each column i of $\mathbf{X}$, (corresponding to a single brand attribute) are generated by a distribution centered on $\mathbf{a}_i$, like in Figure 5(b). To encourage this, we fit a normal distribution Q_{bi} from $\mathbf{X}_{:,i}$ (line 19) and then use the Kullback-Leibler (KL) divergence to compute how well Q_{bi} approximates the ground-truth distributions P_{bi}, which is centered on the brand attribute $\mathbf{a}_i$:

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q(i)}. \tag{7}$$

Since we have no knowledge of the real value of the standard deviation, we propose to learn the standard deviations during training, together with the models' parameters: $P_q \sim \mathcal{N}(\mathbf{a}_i, \Sigma_{b,i})$. There is no supervision on the standard deviation and the model is free to learn any value as long as the distribution is centered to the ground-truth label. During back-propagation, the parameters of *BrandRL*, *PostRL* and the matrix of standard deviations Σ are updated to optimally fit the supervision labels.

The expected behavior of our training procedure is illustrated in Figure 5(b). In this example with two brands, the KL loss is pushing Q_1 and Q_2 closer to P_1 and P_2, respectively, while at the same time adjusting the broadness of the latter. In traditional neural networks, images that are far away from the consensus (in the figure those small circles on Q_1 and Q_2 that are far from their mean) would act as noisy inputs and would try to fit them close to the label. However, by modeling distributions and not the points themselves, our learning method is not affected by such cases and will only fit those images close to the consensus.

The final expression of the loss can be written as:

$$\mathcal{L}_{PCD^*} = \mathcal{L}_{PCD} + \alpha \sum_{b=1,\ldots,B} \sum_{i=1,\ldots,M} D_{\mathrm{KL}}(P_{bi} \parallel Q_{bi}) + \beta ||\theta||_2, \tag{8}$$

where θ is the set of parameters and α and β control the importance of the regularization terms.

Because of the different batching mechanisms, we train our model for 10,000 epochs. We use mini-batches of 64 brands and we sample $K = 200$ posts per brand. As initial value for Adadelta, we set the learning rate to 1.0. Since different brand attributes are defined in different ranges, we scale them in the interval $[0, 1]$. All the other settings are the same as those for the standard PCD.

4 EXPERIMENTS

In this section, we first describe the collection of datasets and the experimental setup. We then report several experiments and studies to demonstrate the effectiveness of our method and visualization of the results. Experiments include a comparison with baselines, the performance of the individual blocks of PCD, the impact of modeling brand attributes, and the effectiveness of PCD*. Finally, we report several qualitative experiments where we visualize the learned brand associations and perform case studies on a selection of brands.

4.1 Datasets

Given the large amount of data required by neural networks, we collected two large-scale image datasets containing the historical timeline of social media postings of brands from different verticals. Beside posts of brands in SN websites, we also collected brand information on brand attributes to test the effectiveness of PCD*.

4.1.1 Existing Datasets. Among the public datasets of brand image posts in social networks, Gao et al. released a popular large-scale collection of over 1.2 million microblogs [12]. The authors searched image microblogs by text keywords and selected those cases where a brand logo was found. However, the dataset is not suitable for our task, because it only contains images with a brand logo, while one of the aims of our work is to find images exhibiting a more complex set of brand attributes beyond the product. Other works that study the popularity of brands [2, 37, 38] use datasets where the task of discovering the content for brands may be applied. However, they are private, made ad-hoc for the popularity task, or publicly unavailable. We hence decided to build our own datasets.

Table 1. Number or Brands for Each Vertical in Dataset1

Alcohol 69	Airlines 57	Auto 83	Fashion 98	Food 85	Furnishing 49	Electronics 79
Non-profit 71	Jewellery 71	Finance 37	Services 69	Entertainment 88	Energy 4	Beverages 67

4.1.2 Two New Datasets for Content Discovery. We chose Instagram as our source of data, since it has a more vibrant brand community due to its higher engagement rate.[2]

For our first dataset (Dataset1), we selected fourteen verticals on the marketing Website Iconosquare and collected the list of Instagram accounts from the Website. We filtered out brands with less than 100 posts to avoid sample insufficiency, retaining a set of 927 brands. For each of these, we crawled at most 2,000 recent posts from their historical timeline, for a total of 1,158,474 posts (approximately 1,250 posts per brand on average). For each post, we collected the image or video together with all the metadata available such as the posting time, number of likes and comments. Table 1 shows the distribution of brands in each of the 14 verticals. The vertical with the lowest number of brands is *Energy* with only 4 brands, We believe that energy brands, such as OCTG oil & gas, target businesses rather than end customers and hence do not have a wide presence on social media. We split the dataset into training and testing sets, where the test set contains the 10 most recent posts for each brand, and all the remaining data were used for training. This gives rise to a total of 1,149,204 posts for training and 9,270 for testing. We denote the training posts for a brand b as $H_{train}(b)$ and the testing posts as $H_{test}(b)$.

We collected a second dataset (Dataset2) to include brand attributes from the domain of marketing. We use the same dataset as in Reference [15], which consists of 698,230 posts (693,230 for training and 4,939 for testing) from 462 brands. For each brand the dataset has hundreds of brand attributes, from which we sampled 89 variables, including BAV metrics (*differentiation*, *relevance*, *esteem* and *knowledge*), brand personality attributes (*cutting edge*, *classic*, *superior*, *chic*, *customer centric*, *outgoing*, *no nonsense*, *distant*), and other attributes from surveys (*daring*, *trendy*, *excitement*, emphcool, emphyoung, emphleader, etc.). More details can be found in Reference [32].

Figure 6 shows an example of the positive correlations of concepts from ImageNet with brand attributes, computed from Dataset2.

4.1.3 Metrics. We adopted the metrics described in Table 2. To measure the probability of choosing the most relevant examples, we adopt *AUC* as in Reference [40]:

$$AUC = \frac{1}{|B|} \sum_b \frac{1}{|E(b)|} \sum_{(p_i,p_j)\in E(b)} \delta(f(b,p_i) > f(b,p_j)), \tag{9}$$

where δ is the indicator function and the evaluation pairs per brand b are

$$E(b) = \{(p_i,p_j)|p_x \in H_{test}(b) \wedge p_y \in H_{test}(c), c \neq b\}. \tag{10}$$

To assess the ability of our model to learn fine-grained brand representation for discriminating between subtle differences among competitor brands, we introduce a novel metric called Competitors AUC (cAUC). The metric is computed exactly as the regular AUC but restricts the evaluation pairs to only those involving competitor brands,

$$E(b) = \{(p_i,p_j)|p_x \in H_{test}(b) \wedge p_y \in H_{test}(c), V(c) = V(b)\}, \tag{11}$$

where $V(b)$ is the vertical of brand b.

[2] https://locowise.com/blog/instagram-engagement-rate-is-higher-than-facebook.

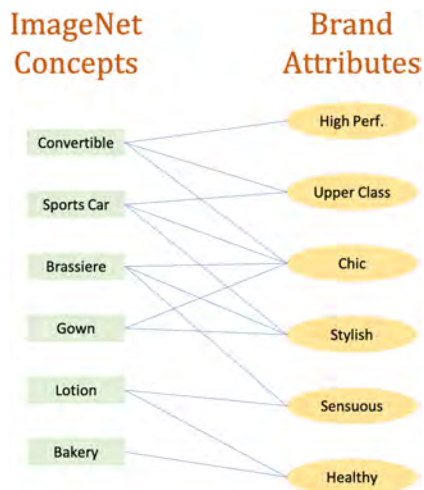


Fig. 6. Positive correlations between six ImageNet concepts and six attributes in dataset 2. All correlations are greater than 0.09.

Table 2. Metrics for Comparison of PCD with Baselines

Metric	Range	Description
AUC	[0-1]	Probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative one.
cAUC	[0-1]	Probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative sample from a competitor.
NDCG$_\mathbf{x}$	[0-1]	Measures the quality of a ranking list based on the post position in the sorted result list. Truncated at x. The higher the better.
MedR	[0-inf]	The median position of the first relevant document. The lower the better.

We also employ Normalized Discounted Cumulative Gain (NDCG) to evaluate the performance of post discovery by taking into account the ranking of relevant posts. We compute the relevance score of position i as follows:

$$r_b(i) = \begin{cases} 1 & \text{if } p_i \in H_{test}(b) \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

where b is the target brand, p_i is the post ranked at position i, and $H_{test}(b)$ are the testing set images posted by brand b. Intuitively, given a brand, a high-performance discovery model will rank as high as possible the test images posted by that brand. Thus, in addition to NDCG, we also introduce the metric MedR, which is the median position of the first relevant post retrieved. A low MedR indicates that the first relevant post is ranked as the most relevant results most of the times.

4.1.4 Baselines. Since there are no existing methods specifically designed for content discovery for brands, we compare our method against a set of baselines inspired by the pairwise models in image recommendations, which are the closest to PCD.
Random: We generate a random ranking for testing posts.

Table 3. Comparison of MedDiscR with the Baselines

	AUC	cAUC	$NDCG_{10}$	$NDCG_{50}$	MedianRank
Random	0.503	0.503	0.001	0.003	568
BrandAVG	0.796	0.687	0.068	0.105	29
DVBPR	0.862	0.734	0.059	0.102	20
CDL	0.807	0.703	0.079	0.119	19
NPR	0.838	0.716	0.040	0.076	33
PCD	**0.880**	**0.785**	**0.151**	**0.213**	**5**

We use cut-offs of 10 and 50 for NDCG.

BrandAVG: We perform nearest neighbor retrieval with respect to brand representation, which is the mean feature vector among the image features of all images appearing in the brand's posting timeline.
DVBPR [24]: Visually-aware Deep Bayesian Personalized Ranking is a pairwise model inspired by VBPR [19] that excludes non-visual latent factors. We adopt the variant with pre-trained image features as described in the article.
CDL [30]: Comparative Deep Learning is a pure content-based pairwise architecture. We use pre-trained image features and one-hot brand ID as user information.
NPR [35]: Neural Personalized Ranking is one of the most recent pairwise content-based architecture. Since our formulation is a pure content-based scenario, we use image features as the sole item input, using pre-trained image features after PCA as described in the article.

4.2 Experiment 1: PCD vs. Others

We perform this experiment to evaluate quantitatively the performance of PCD versus the baselines. The results of PCD against the baselines on our two datasets are reported in Table 3. From the table, it can be seen that our method has the best performance according to all metrics, except for the metric in the second dataset. The reason is that DVBPR directly maximises the AUC, while PCD is based on a margin rank loss. However, in terms of cAUC, PCD outperforms the baselines for both datasets, confirming that our method learns a more fine-grained brand representation compared to content-based recommender systems. Moreover, we notice that the cAUC values are consistently lower than AUC, confirming that, for the challenge of inter-brand similarity, it is hard to discriminate between the subtle differences of competitors.

By inspection of the NDCG and Medr metrics, we notice that the performance of NPR is inferior when compared to the other baselines. We believe the reason is that NPR is the only non content-based method, which is designed for a less sparse collaborative filtering scenario.

Finally, PCD has the highest NDCG values and the lower MedR, indicating that the learned brand and post embedding have a higher capability of discovering a small number of relevant posts in the large test set.

4.3 Experiment 2: Latent Brand Associations

This experiment is an ablation study we design to investigate if explicitly modeling brand associations yield better rankings than directly learning brand representation from the one-hot brand ID. For this purpose, we define PCD1H, which is a variant of PCD without the brand representation learning component, learning a brand embedding from the one-hot brand ID instead.

In Figure 7, we compare the NDCG of PCD and PCD1H for increasing cut-off values on dataset 1. We notice that PCD values are consistently higher than that of PCD1H at all cut-offs points. This

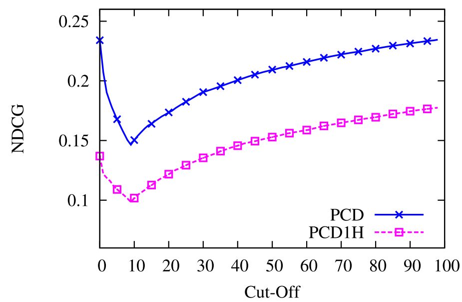


Fig. 7. PCD vs. PCD1H: The improved performance of learning brand representation from associations is shown by a higher NDCG curve in the case of PCD for all the cut-off points.

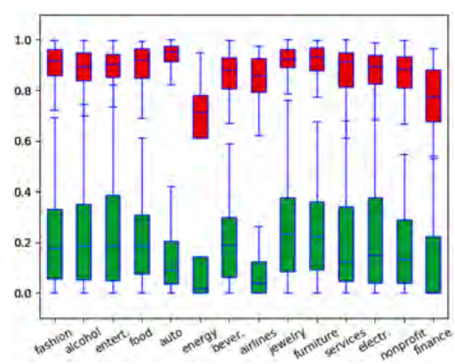


Fig. 8. Box plot of performances for brand verticals. AUC and NDCG@50 are represented in red and green boxes, respectively.

result confirms the effectiveness of our brand representation learning and the importance of explicitly modeling brand associations.

We observe that PCD has a more marked v-shape than PCD1H, particularly on the left side. For example, considering the cut-off of 1, PCD retrieves a relevant result in the top position for 202 of 927 brands, while in the case of PCD1H, this only happens in 127 cases. This indicates that our brand representation learning is particularly attractive in retrieving the relevant posts at the top 10 positions of the ranking. The reason for the curves to invert their trends is due to the discount effect of NDCG.

4.4 Experiment 3: Brand Verticals

Different from previous works on brand associations [27], PCD processes all the brands of our datasets with an end-to-end framework. For this reason, the evaluation metrics we use in the previous studies are computed as average among the 927 brands (except Median Rank). We perform an additional study to investigate the performance of the model one step deeper in terms of brand verticals.

We compute AUC and NDCG@50 for each brand in dataset 1 and plot these results respectively using a red and green box plot, organized by verticals. Each box in Figure 8 represents the distribution of performance scores for all the brands belonging to a certain vertical, such as food or alcohol. The boxes indicate the median value, the upper and lower quartile and extremes. We omit the outliers for a clearer representation. Similar performance is observed among brands in fashion, alcohol, entertainment, food, beverage, jewelry, and furniture, with a median NDCG and AUC of

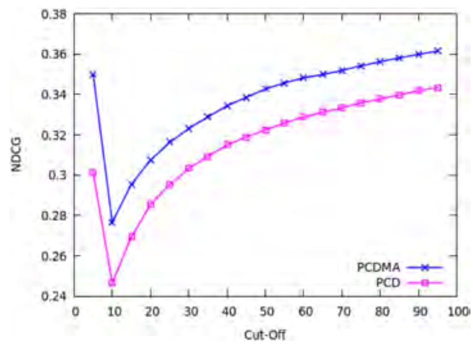


Fig. 9. PCD* vs. PCD: The increase in performance after integrating brand attributes is shown by a higher NDCG curve in the case of PCD* for all the cut-off points.

approximately 0.19 and 0.86, respectively. In terms of NDCG, we learn that images for brands in automobiles, airlines, energy, and finance are the hardest to discover. However, the former two verticals still achieve high AUC, while for the latter ones, this metric is also poor. One possible explanation for energy and finance having lower performance is that they lack clear recurring visual elements. However, images from automobile and airline verticals commonly share similar visual elements, and thus we believe that these are the hardest brands in terms of fine-grained distinctions between competitors.

4.5 Experiment 4: PCD*

Finally, we evaluate the performance of PCD* over PCD. The goal of this experiment is to assess whether there is a link between image posts and the brand attributes, as well as to evaluate whether PCD* is able to leverage them to generate a better ranking. To carry out this experiment, we use dataset 2, since dataset 1 does not have information on brand attributes.

We follow the first setting as in experiment 4.3 and compare the NDCG for different cut-off values. We show the NDCG curves in Figure 9. From the curves, we observe that PCD* has an NDCG@5 score of approximately 0.35, while that for the standard PCD is 0.3. This indicates that brand attributes are reflected in social media posts, and integrating them with PCD* yields better results.

4.6 Case Study 1: Ranking Visualization

To achieve a deeper understanding of what kind of image posts our model discovers, we offer a qualitative analysis of eight case studies. We select eight among the most popular brands in our dataset and we use PCD to rank the posts of our testing dataset for each of them. We aim to show for each of these brands what kind of posts were correctly discovered, what relevant posts were missed, and what are the posts from other brands that erroneously obtained high positions in the ranking. Figure 10 tabulates results for the eight selected cases: beer brand *Carlsberg*, *Qatar Airways*, computer manufacturer *Lenovo*, *Ford* motor company, *Coca-Cola*, Italian fashion brand *Gucci*, and video-game companies *Nintendo* and *Ubisoft*. For each brand, the first column shows a relevant post that PCD ranked in one of the top-10 positions of the ranking (true positive), while the central example is another relevant post for which the model failed to attribute a high ranking score (false negative). Finally, the third column shows an example image of a post from another brand (either competitors or otherwise) that erroneously achieve the top-10 positions in the ranking (false positives), together with the name of the brand the image belongs to.

Fig. 10. Case study for eight brands. For each brand, the three columns show each one example of true positive, false negative, and false positive. Please note that all examples were manually selected, and hence this picture has no indication of performance.

While the examples in the first column evidently match their brand expected style and content, the ones in the second column are much harder to discover. For example, in the case of *Carlsberg*, the method is able to easily retrieve an image of a beer glass on a green background but fails to retrieve another image posted by the brand, featuring two astronauts visiting the beer company on the first anniversary of their space mission. Also in the other cases, we notice that the false negatives are to a certain extent sporadic and occasional for the brands, which partially explains why they were missed.

Finally, the false positives in the third columns are those image posts that were mistakenly included in the top 10. In the case of Carlsberg, PCD selects another picture of a beer glass on a grass background by German beer brand *Astra*, most likely because it learned to associate the green color with the brand. We observe that not all the false positives are from the competitor brands. For example, a post from the financial services company *Allianz* is retrieved for the brand *Ford*, featuring a truck in an outdoor environment. This confirms that in these cases PCD captured fine-grained stylistic features rather than stopping at high-level category level representation.

4.7 Case Study 2: Brand Attributes

We design this case study to assess the ability of PCD* to learn brand attributes in images as well as generate explanations.

We first assess whether the model correctly learns to rank images by marketing attributes. To conduct this study, we use the network ω_a to rank the images in the testing set by six specific attributes. To generate a ranking, we forward the distribution of image concepts $\mathbf{c_p}$ to ω_a, computing the attributes-specific post vector $\mathbf{p_a}$ for each testing image. The ith element of $\mathbf{p_a}$ indicates the strength of the ith brand attribute in the post. We then select the six brand attributes of *Upper Class*, *Fun*, *Stylish*, *Cool*, *Corporate*, and *Prestigious* and rank the testing images by the corresponding element of $\mathbf{p_a}$ We sample five posts from the top 20 and show them in Figure 11, with the only exception for the attribute *Prestigious*, for which we sample among the last 20 results. From the figure, we learn that images of cocktails are often used by brands that are considered being *Upper Class* and *Cool* by their consumers. Posts related to cartoons are instead related to attribute *Fun*, since the brands in our dataset with the highest score for these attributes are in the entertainment vertical. Unsurprisingly, the top *Corporate* posts feature skyscrapers and men in business attire, while fashion posts occupy the top positions in the *Stylish* rank. Finally, images about drinks in plastic bottles are ranked last for the attribute *Prestigious*.

Fig. 11. Top results for five brand attributes and last results for the attribute *Prestigious*.

Fig. 12. Brand attributes as explanation to Content Discovery for Brands.

Beside ranking by the attributes themselves, another important application of brand attributes is to generate explanations for the original task of ranking images for a specific brand. Once the testing images are ranked, we can look at the attributes-specific component $\mathbf{p_a}$ of the image representation vector $\mathbf{p}$ to produce a list of top brand attributes of the ranked images. For this study, we rank images for the brands: *BMW*, *Burger King*, *Gucci*, and *Nike*. We sample four posts among the top 20 results for each brand and list the three top attributes as explanations.

Figure 12 shows how different image recommendations can be explained with different brand attributes. For example, a black SUV on a sunny landscape can be used to express the ideas of a young, cool, and spirited brand; while a white vehicle on a snow-covered background is more suitable to communicate reliability and trustworthiness. Such results can be used by marketers as an additional help to pick the most appropriate post according to their campaign goals.

4.8 Case Study 3: Latent Brand Associations

To visualize which latent associations are automatically learned by the model, we use our trained model to project all training images into the latent space and select the nearest neighbor posts to the association latent vector using cosine similarity. Figure 13 shows a selection of qualitative examples of brand associations. The example shows some high-level semantics aspects that are captured by latent vectors of brand associations. For example, dedicated brand associations are

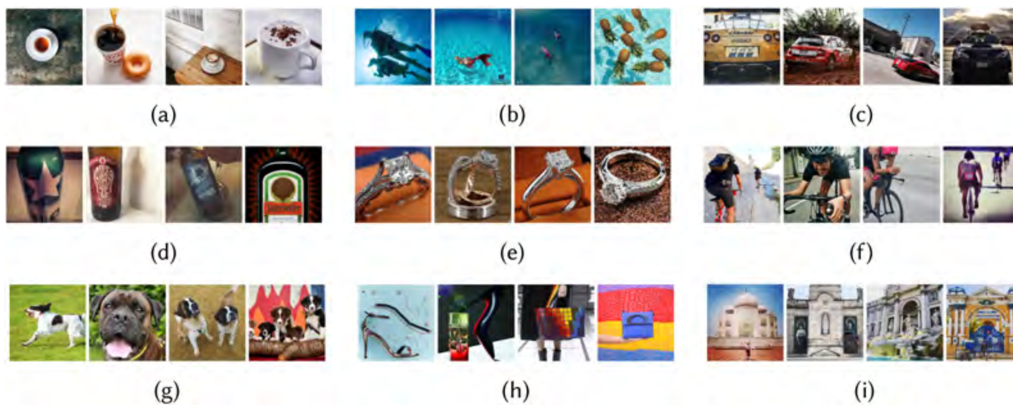


Fig. 13. Nine examples of brand associations. For each association, four example images are selected for display from the training set using nearest neighbors.

automatically learned to represent the visual aspects of coffee cups, seawater, cars, alcohol bottles, rings, cyclists, dogs, fashion items, and classical buildings.

By looking into a brand's weight vector $\mathbf{w_b}$, we can understand which brands adopt a particular brand association the most. For the associations in Figure 13, we retrieve the 50 brands that were positively affected the most. Among these, association (a) affects brands such as *Costa Coffee*, *Starbucks*, and *Salt Spring Coffee*, while association (d) is adopted by alcohol brands such as *Dom Pérignon* and *Moët & Chandon*. Finally, association (c) is adopted by most of the car manufacturer brands in our dataset, such as *Rolls-Royce*, *Tesla*, *Cadillac*, and *Volvo*.

5 CONCLUSIONS

Inspired by a real-life marketing problem, we introduced the new problem of content discovery for brands to automatically filter images that match the brand associations, namely the set of values of the brand. We proposed a content-based method called PCD to generate a personalized ranking by learning the fine-grained brand representation from their historical timeline. We integrated brand characteristics from surveys to improve the performance and to generate explanations with the top results of the ranking. We tested our methods on two Instagram datasets in terms of performance comparisons, ablation studies, and qualitative visualizations. Our findings first indicate that, thanks to modelling brand associations and brand attributes, our method can outperform state-of-the-art solutions. Second, our model can be used to visualize the visual elements of brand attributes, like in the case of skyscrapers and suits that are learned for attribute *Corporate*.

A promising future direction is to integrate a knowledge graph built with domain knowledge in marketing to better understand the similarities between brands in terms of featured products and other brand attributes.

REFERENCES

[1] Jennifer L. Aaker. 1997. Dimensions of brand personality. *J. Market. Res.* 34, 3 (1997), 347–356.
[2] Hassan Alboqami, Wafi Al-Karaghouli, Yasser Baeshen, Ismail Erkan, Chris Evans, and Ahmad Ghoneim. 2015. Electronic word of mouth in social media: The common characteristics of retweeted and favourited marketer-generated content posted on Twitter. *Int. J. Internet Market. Advert.* 9, 4 (2015), 338–358.
[3] Christy Ashley and Tracy Tuten. 2015. Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement. *Psychol. Market.* 32, 1 (2015), 15–27.
[4] Janelle Barlow and Paul Stewart. 2004. *Branded Customer Service: The New Competitive Edge*. Berrett-Koehler.

[5] Chia Yu Chang. 2014. *Visualizing brand personality and personal branding: case analysis on Starbucks and Nike's brand value co-creation on Instagram.* MA (Master of Arts) thesis, University of Iowa. http://ir.uiowa.edu/etd/1304.

[6] Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, and Tat-Seng Chua. 2017. Attentive collaborative filtering: Multimedia recommendation with item- and component-level attention. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'17).* 335–344.

[7] Tao Chen, Xiangnan He, and Min-Yen Kan. 2016. Context-aware image tweet modelling and recommendation. In *Proceedings of the 2016 ACM on Multimedia Conference (MM'16).* 1018–1027.

[8] Peng Cui, Zhiyu Wang, and Zhou Su. 2014. What videos are similar with you?: Learning a common attributed representation for video recommendation. In *Proceedings of the 22nd ACM International Conference on Multimedia.* 597–606.

[9] Lisette De Vries, Sonja Gensler, and Peter S. H. Leeflang. 2012. Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *J. Interact. Market.* 26, 2 (2012), 83–91.

[10] Yi Ding, Chee Wei Phang, Xianghua Lu, Chuan-Hoo Tan, and Juliana Sutanto. 2014. The role of marketer- and user-generated content in sustaining the growth of a social media brand community. In *Proceedings of the 2014 47th Hawaii International Conference on System Sciences.* 1785–1792.

[11] Aleksandr Farseev, Kirill Lepikhin, Kenny Powar, Eu Khoon Ang, and Hendrik Schwartz. 2018. SoMin.ai: Social multimedia influencer discovery marketplace. In *Proceedings of the 26th ACM International Conference on Multimedia (MM'18).*

[12] Yue Gao, Fanglin Wang, Huanbo Luan, and Tat-Seng Chua. 2014. Brand data gathering from live social media streams. In *Proceedings of the International Conference on Multimedia Retrieval (ICMR'14).* 169–176.

[13] Francesco Gelli, Xiangnan He, Tao Chen, and Tat-Seng Chua. 2017. How personality affects our likes: Towards a better understanding of actionable images. In *Proceedings of the 2017 ACM on Multimedia Conference (MM'17).* 1828–1837.

[14] Francesco Gelli, Tiberio Uricchio, Xiangnan He, Alberto Del Bimbo, and Tat-Seng Chua. 2018. Beyond the product: Discovering image posts for brands in social media. In *Proceedings of the 2018 ACM Multimedia Conference on Multimedia Conference (MM'18).* 465–473.

[15] Francesco Gelli, Tiberio Uricchio, Xiangnan He, Alberto Del Bimbo, and Tat-Seng Chua. 2019. Learning subjective attributes of images from auxiliary sources. In *Proceedings of the 27th ACM International Conference on Multimedia (MM'19).* 2263–2271. DOI: https://doi.org/10.1145/3343031.3350574

[16] Xue Geng, Hanwang Zhang, Jingwen Bian, and Tat-Seng Chua. 2015. Learning image and user features for recommendation in social networks. In *Proceedings of the International Conference on Computer Vision (ICCV'15).* 4274–4282.

[17] Sonja Gensler, Franziska Völckner, Yuping Liu-Thompkins, and Wiertz Caroline. 2013. Managing brands in the social media environment. *J. Interact. Market.* 27, 4 (2013), 242–256.

[18] Khim-Yong Goh, Cheng-Suang Heng, and Zhijie Lin. 2013. Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content. *Inf. Syst. Res.* 24, 1 (2013), 88–107.

[19] Ruining He and Julian McAuley. 2016. VBPR: Visual Bayesian personalized ranking from implicit feedback. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence.* 144–150.

[20] Xiangnan He, Tao Chen, Min-Yen Kan, and Xiao Chen. 2015. TriRank: Review-aware explainable recommendation by modeling aspects. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM'15).* 1661–1670.

[21] Xiangnan He and Tat-Seng Chua. 2017. Neural factorization machines for sparse predictive analytics. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'17).* 355–364.

[22] Balázs Hidasi, Massimo Quadrana, Alexandros Karatzoglou, and Domonkos Tikk. 2016. Parallel recurrent neural network architectures for feature-rich session-based recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys'16).* 241–248.

[23] Sudarsan Jayasingh and Venkatesh Rajagopalan. 2015. Customer engagement factors in facebook brand pages. *Asian Soc. Sci.* 11, 26 (2015). DOI: 10.5539/ass.v11n26p19

[24] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian McAuley. 2017. Visually-aware fashion recommendation and design with generative image models. In *Proceedings of the International Conference on Data Mining (ICDM'17).* 207–216.

[25] Wang-Cheng Kang, Eric Kim, Jure Leskovec, Charles Rosenberg, and Julian McAuley. 2019. Complete the look: Scene-based complementary product recommendation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 10532–10541.

[26] Gunhee Kim and Eric P. Xing. 2013. Discovering pictorial brand associations from large-scale online image data. In *Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops (ICDMW'15).* 404–411.

[27] Gunhee Kim and Eric P. Xing. 2014. Visualizing brand associations from web community photos. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM'14).* 623–632.

[28] Jingu Kim, Renato D. C. Monteiro, and Haesun Park. 2012. Group sparsity in nonnegative matrix factorization. In *Proceedings of the 2012 SIAM International Conference on Data Mining (SDM'12)*. 851–862.
[29] Solomon Kullback. 1997. *Information Theory and Statistics*. Courier Corporation.
[30] Chenyi Lei, Dong Liu, Weiping Li, Zheng-Jun Zha, and Houqiang Li. 2016. Comparative deep learning of hybrid representations for image recommendations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16)*. 2545–2553.
[31] Tereza Litsa. 2016. How the rise of user-generated content changes marketing. Retrieved from https://www.clickz.com/how-the-rise-of-user-generated-content-changes-marketing.
[32] Mitchell Lovett, Renana Peres, and Ron Shachar. 2014. A data set of brands and their characteristics. *Market. Sci.* 33, 4 (2014), 609–617.
[33] Shuang Ma and Chang Wen Chen. 2018. D-Sempre: Learning deep semantic-preserving embeddings for user interests-social contents modeling. *CoRR* abs/1802.06451 (2018).
[34] Masoud Mazloom, Robert Rietveld, Stevan Rudinac, Marcel Worring, and Willemijn van Dolen. 2016. Multimodal popularity prediction of brand-related social media posts. In *Proceedings of the 2016 ACM on Multimedia Conference (MM'16)*. 197–201.
[35] Wei Niu, James Caverlee, and Haokai Lu. 2018. Neural personalized ranking for image recommendation. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM'18)*. 423–431.
[36] Steve Olenski. 2016. Can Your Content Sell a Lifestyle and Not Just Products? Retrieved from https://www.forbes.com/sites/steveolenski/2016/07/02/can-your-content-sell-a-lifestyle-and-not-just-products.
[37] Gijs Overgoor, Masoud Mazloom, Marcel Worring, Robert Rietveld, and Willemijn van Dolen. 2017. A spatio-temporal category representation for brand popularity prediction. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval (ICMR'17)*. 233–241.
[38] Irena Pletikosa Cvijikj and Florian Michahelles. 2013. Online engagement factors on Facebook brand pages. *Soc. Netw. Anal. Min.* 3, 4 (2013), 843–861.
[39] Steffen Rendle. 2010. Factorization machines. In *Proceedings of the IEEE 10th International Conference on Data Mining (ICDM'10)*. 995–1000.
[40] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI'09)*. 452–461.
[41] Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic matrix factorization. In *Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS'07)*. 1257–1264.
[42] Martin Saveski and Amin Mantrach. 2014. Item cold-start recommendations: Learning local collective embeddings. In *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys'14)*. 89–96.
[43] Shaojung Sharon Wang, Yu-Ching Lin, and Ting-Peng Liang. 2018. Posts that attract millions of fans: The effect of brand-post congruence. *Electr. Commerce Res. Appl.* 28 (2018), 73–85. DOI: 10.1016/j.elerap.2017.12.010
[44] Yao Wu, Christopher DuBois, Alice X. Zheng, and Martin Ester. 2016. Collaborative denoising auto-encoders for Top-N recommender systems. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM'16)*. 153–162.
[45] Wenhui Yu, Huidi Zhang, Xiangnan He, Xu Chen, Li Xiong, and Zheng Qin. 2018. Aesthetic-based clothing recommendation. In *Proceedings of the 2018 World Wide Web Conference (WWW'18)*. 649–658.
[46] Matthew D. Zeiler. 2012. ADADELTA: An adaptive learning rate method. *CoRR* abs/1212.5701 (2012).
[47] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*. 353–362.
[48] Hanwang Zhang, Zheng-Jun Zha, Yang Yang, Shuicheng Yan, Yue Gao, and Tat-Seng Chua. 2013. Attribute-augmented semantic hierarchy: Towards bridging semantic gap and intention gap in image retrieval. In *Proceedings of the 21st ACM International Conference on Multimedia (MM'13)*. 33–42.
[49] Yin Zhang and James Caverlee. 2019. Instagrammers, fashionistas, and me: Recurrent fashion recommendation with implicit visual influence. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1583–1592.

Received January 2020; revised February 2020; accepted February 2020