Distribution-oriented Aesthetics Assessment for Image Search

Chaoran Cui
School of Computer Science and Technology
Shandong University of Finance and Economics
crcui@sdufe.edu.cn

Huidi Fang
School of Computer Science and Technology
Shandong University
huidif@163.com

Xiang Deng
School of Computer Science and Technology
Shandong University
dxcvai@gmail.com

Xiushan Nie
School of Computer Science and Technology
Shandong University of Finance and Economics
niexsh@sdufe.edu.cn

Hongshuai Dai
School of Statistics
Shandong University of Finance and Economics
math_dsh@163.com

Yilong Yin*
School of Computer Science and Technology
Shandong University
ylyin@sdu.edu.cn

ABSTRACT

Aesthetics has become increasingly prominent for image search to enhance user satisfaction. Therefore, image aesthetics assessment is emerging as a promising research topic in recent years. In this paper, distinguished from existing studies relying on a single label, we propose to quantify the image aesthetics by a distribution over quality levels. The distribution representation can effectively characterize the disagreement among the aesthetic perceptions of users regarding the same image. Our framework is developed on the foundation of label distribution learning, in which the reliability of training examples and the correlations between quality levels are fully taken into account. Extensive experiments on two benchmark datasets well verified the potential of our approach for aesthetics assessment. The role of aesthetics in image search was also rigorously investigated.

KEYWORDS

aesthetics assessment, label distribution learning, image search

ACM Reference format:
Chaoran Cui, Huidi Fang, Xiang Deng, Xiushan Nie, Hongshuai Dai, and Yilong Yin. 2017. Distribution-oriented Aesthetics Assessment for Image Search. In *Proceedings of SIGIR'17, August 7–11, 2017, Shinjuku, Tokyo, Japan, ,* 4 pages.
DOI: http://dx.doi.org/10.1145/3077136.3080704

*This author is the corresponding author.

SIGIR'17, August 7–11, 2017, Shinjuku, Tokyo, Japan

DOI: http://dx.doi.org/10.1145/3077136.3080704

1 INTRODUCTION

Relevance is generally of paramount concern in image search systems, which seek to make the topic of returned images match that of the textual query. However, with rapid advances in information retrieval technologies, modern image search engines are able to achieve the goal in most circumstances, and the relevance may not always be the primary factor in determining user search satisfaction [3]. Meanwhile, users place increasing demands on the aesthetics of returned images, especially for those queries with plenty of relevant results. Therefore, it is desirable for search engines to rank images not only by topical relevance but also by aesthetic quality.

Aesthetics assessment aims to measure whether an image looks beautiful in human's perception, so that aesthetically pleasing images can be singled out. In most existing studies [6–8], aesthetics assessment is transformed to a classification or regression problem, where each image is assigned a single label (i.e., category or score) indicating its aesthetic quality level. However, as the saying goes, "*beauty is in the eye of the beholder*"; aesthetics is essentially a subjective perception, and different people may have different ideas about the beauty of the same image. In light of this, a single label is insufficient to characterize the disagreement among the aesthetic perceptions of users.

To further illustrate this point, we performed a preliminary experiment on the AVA dataset [9], which is currently the largest publicly available dataset for aesthetic visual analysis. The AVA dataset consists of 255,530 images with the counts of aesthetic ratings on the scale of 1 to 10 contributed by different users. For each image, we computed the entropy value of the distribution over its aesthetic ratings. Figure 1 plots the histogram of the entropy values for all images. As can be seen, the distribution over user ratings yields a relatively high entropy value for most images, reflecting the fact that the aesthetic perceptions are inconsistent across users. This underpins our belief that the aesthetic quality of images cannot be measured with merely a single label.

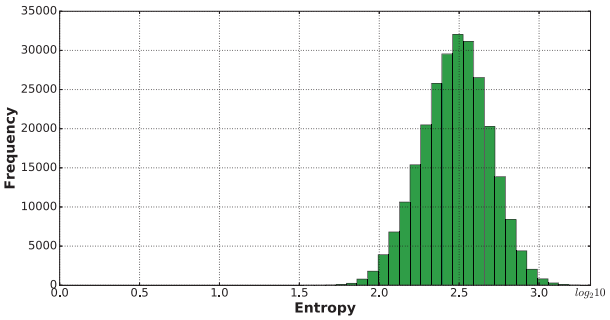

Figure 1: Histogram of the entropy values of the distribution over user ratings. Note that the upper limit of the entropy value is $log_2 10$, since the ratings are discretized into 10 levels.

Motivated by the above discussions, in this paper, we propose to apply a distribution to depict the aesthetic quality of an image. Each component of the distribution indicates the probability of users assigning a specific quality level to the image. The form of distribution offers advantages in two aspects: 1) Distinguished from a single label, a distribution quantifies the uncertainty in the process of aesthetic evaluation, so that the disagreement about users' aesthetic perceptions can be effectively captured; 2) Depending on practical needs, a distribution can be easily converted to a category or score with its numerical characteristic, such as the expectation or median. This ensures that the aesthetic classification or ranking task can be smoothly performed with the distribution representation as well.

Our framework is developed on the foundation of Label Distribution Learning (LDL) [4], which seeks to learn the mapping from an instance to its distribution over multiple labels in a supervised manner. Specifically, we adopt the Multivariate Support Vector Regression (M-SVR) [10] as the backbone of our learning algorithm. To reduce the effects of the shortage of rating users for training examples, the reliability of training examples is explicitly defined and incorporated into our approach. Besides, we take account of the correlations between quality levels to further enhance the robustness of our approach.

Extensive experiments on two benchmark datasets verify the promise of our approach in both scenarios of distribution prediction and label prediction for aesthetics assessments. In addition, we also demonstrate the efficacy of our approach in the task of aesthetics-based image reranking.

2 FRAMEWORK

2.1 Aesthetic Distribution

In this paper, we target at measuring the aesthetic quality of images by an aesthetic distribution, so that the disagreement among aesthetic perceptions of users can be effectively characterized. Formally, let $\mathcal{X}$ be the image feature space, and $\mathcal{L} = \{l_1, l_2, \ldots, l_c\}$ denotes the set of c predefined quality levels. Our goal is to learn a hypothesis $f : \mathcal{X} \to \mathbb{R}^c$. Given an image $\mathbf{x} \in \mathcal{X}$, $f(\mathbf{x})$ is a c-dimensional vector with the i-th element $f_i(\mathbf{x})$ indicating the likelihood of users perceiving the image $\mathbf{x}$ as being at the quality level of l_i. We subsequently normalize $f(\mathbf{x})$ to make $f_i(\mathbf{x}) \in [0, 1]$ and $\sum_{i=1}^{c} f_i(\mathbf{x}) = 1$ to constitute the aesthetic distribution for $\mathbf{x}$.

Previous work [11] viewed the distribution as a structure and solved the problem via structured learning. However, there is no intuitive way to define the auxiliary compatibility function measuring how well a possible aesthetic distribution fits for an image. In our study, we consider LDL as a more rational solution to this issue. Specifically, we apply the M-SVR algorithm to model the aesthetic distribution for the image $\mathbf{x}$ by

$$\mathbf{z} = \mathbf{W}\varphi(\mathbf{x}) + \mathbf{b} \ , \tag{1}$$

where $\varphi(\mathbf{x})$ is a nonlinear transformation of $\mathbf{x}$ to a higher dimensional feature space $\mathbb{R}^{\mathcal{H}}$. In practice, the kernel tricks can be harnessed to avoid the explicit computations of $\varphi(\mathbf{x})$. $\mathbf{W} \in \mathbb{R}^{c \times \mathcal{H}}$ and $\mathbf{b} \in \mathbb{R}^c$ are the model parameters. Moreover, we constrain each component of the output of our model within the range of a probability, i.e., $[0, 1]$. Therefore, the sigmoid function is applied to each component of the vector $\mathbf{z}$, and the final formulation of our model is represented by

$$f(\mathbf{x}) = s(\mathbf{z}) = \frac{1}{1 + \exp(-\mathbf{z})} \ . \tag{2}$$

2.2 Distribution-oriented Learning

In our study, we seek to learn the model parameters $\mathbf{W}$ and $\mathbf{b}$ in a supervised manner. Typically, a training set is available in the form of $\mathcal{T} = \{(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), (\mathbf{x}^{(2)}, \mathbf{y}^{(2)}), \ldots, (\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}$, where $\mathbf{x}^{(i)}$ is an image instance, and $\mathbf{y}^{(i)} = [y_1^{(i)}, y_2^{(i)}, \ldots, y_c^{(i)}]^T$ is the vector whose element $y_j^{(i)}$ denotes the number of users choosing l_j for $\mathbf{x}^{(i)}$. Assume $\mathbf{d}^{(i)} = [d_1^{(i)}, d_2^{(i)}, \ldots, d_c^{(i)}]^T$ is the true aesthetic distribution associated with $\mathbf{x}^{(i)}$, which can be approximately estimated by

$$d_j^{(i)} = y_j^{(i)} \Big/ \sum\nolimits_{k=1}^{c} y_k^{(i)} \ . \tag{3}$$

With the training set, the model parameters $\mathbf{W}$ and $\mathbf{b}$ can be determined by minimizing the objective

$$\frac{\lambda}{2}\|\mathbf{W}\|_F^2 + \sum\nolimits_{i=1}^{n} L(f(\mathbf{x}^{(i)}) - \mathbf{d}^{(i)}) \ , \tag{4}$$

where $\|\mathbf{W}\|_F^2$ is the regularization term with the Frobenius norm, and $L(f(\mathbf{x}^{(i)}) - \mathbf{d}^{(i)})$ is the loss function for the i-th training example. λ is a hyperparameter controlling the tradeoff between the model complexity and the corresponding loss. In M-SVR, the loss function is defined based on the hinge loss, i.e.,

$$L(\mathbf{v}) = \begin{cases} 0 & \|\mathbf{v}\|_2 < \varepsilon \ , \\ \left(\|\mathbf{v}\|_2 - \varepsilon\right)^2 & \|\mathbf{v}\|_2 \geq \varepsilon \ . \end{cases} \tag{5}$$

However, as pointed in [5], a potential problem with the objective in Eq. (4) lies in that it is non-convex and cannot be solved via the kernel tricks, due to the involvement of the sigmoid function. To address the problem, instead of comparing $f(\mathbf{x}^{(i)})$ against $\mathbf{d}^{(i)}$, we measure the loss by computing the distance between $\mathbf{z}^{(i)}$ and $\widehat{\mathbf{z}}^{(i)}$. $\mathbf{z}^{(i)}$ is the intermediate transformation of $\mathbf{x}^{(i)}$ as shown in Eq. (1), and

$\widehat{\mathbf{z}}^{(i)}$ is attained by solving the equation $\mathbf{d}^{(i)} = s(\widehat{\mathbf{z}}^{(i)})$. It has been demonstrated that $||\mathbf{z}^{(i)} - \widehat{\mathbf{z}}^{(i)}||_2 \geq 4||f(\mathbf{x}^{(i)}) - \mathbf{d}^{(i)}||_2$ always holds [5]. Therefore, we replace $(f(\mathbf{x}^{(i)}) - \mathbf{d}^{(i)})$ with $\frac{1}{4}(\mathbf{z}^{(i)} - \widehat{\mathbf{z}}^{(i)})$ and derive a new objective, i.e.,

$$\Gamma = \frac{\lambda}{2}\|\mathbf{W}\|_F^2 + \sum\nolimits_{i=1}^{n} L(\frac{1}{4}(\mathbf{z}^{(i)} - \widehat{\mathbf{z}}^{(i)})) \ . \quad (6)$$

Γ is a upper bound on the original objective in Eq. (4). By minimizing Γ, we can effectively reduce the value of the original objective. Finally, we resort to an iterative quasi-Newton method called Iterative Re-Weighted Least Square (IRWLS) [5] for the minimization problem.

2.3 Model Reinforcement

In the earlier discussions, we treat all training examples uniformly, and estimate their true aesthetic distributions with Eq. (3). However, it is clear that the more users have rated an image, the more reliable is the aesthetic distribution computed by Eq. (3). In view of this, we assign each training example a weight to reflect its reliability. Specifically, the weight of $\mathbf{x}^{(i)}$ is defined by

$$w^{(i)} = s\left(\frac{\gamma \cdot \left(\sum_{k=1}^{c} y_k^{(i)} - \mu\right)}{\sigma}\right) \ , \quad (7)$$

where μ and σ are the mean and standard deviation of the number of rating users for an image in the training set, respectively. $s(\cdot)$ is the sigmoid function, and γ is a smoothing hyperparameter. By introducing the weights of training examples into Eq. (6), a reliability-sensitive objective is achieved, i.e.,

$$\Gamma_{rel} = \frac{\lambda}{2}\|\mathbf{W}\|_F^2 + \sum\nolimits_{i=1}^{n} w^{(i)} L(\frac{1}{4}(\mathbf{z}^{(i)} - \widehat{\mathbf{z}}^{(i)})) \ . \quad (8)$$

Minimizing Γ_{rel} results in the best model that gives priority to ensuring the correct predictions for more reliable training examples.

It is also accepted that the probabilities of different quality levels assigned to an image are correlated. For instance, the pairs of adjacent quality levels generally have highly positive correlations, while negative correlations are allocated to those apart ones. To capture such dependencies between quality levels, the loss function in Eq. (5) is modified to

$$\widehat{L}(\mathbf{v}) = \begin{cases} 0 & \|\mathbf{v}\|_{\mathbf{C}} < \varepsilon \ , \\ \left(\|\mathbf{v}\|_{\mathbf{C}} - \varepsilon\right)^2 & \|\mathbf{v}\|_{\mathbf{C}} \geq \varepsilon \ . \end{cases} \quad (9)$$

Here, $\|\cdot\|_{\mathbf{C}}$ denotes the ellipsoidal norm. We define $\|\mathbf{v}\|_{\mathbf{C}} = \sqrt{\mathbf{v}^T\mathbf{C}\mathbf{v}}$, where $\mathbf{C} \in \mathbb{R}^{c\times c}$ is the matrix with the (i,j)-th entry indicating the correlation coefficient between l_i and l_j. On the basis of the new loss function $\widehat{L}$, a correlation-embedded objective can be represented by

$$\Gamma_{col} = \frac{\lambda}{2}\|\mathbf{W}\|_F^2 + \sum\nolimits_{i=1}^{n} \widehat{L}(\frac{1}{4}(\mathbf{z}^{(i)} - \widehat{\mathbf{z}}^{(i)})) \ . \quad (10)$$

3 EXPERIMENTS

3.1 Experimental Configuration

We adopted two benchmark datasets for aesthetics assessment, i.e., AVA [9] and Photo.net (PN) [2]. On both datasets, we took half of the total images for training and the rest for testing. Each image was represented with the same features as described in [8]. The proposed three approaches for aesthetic distribution learning, namely, ADL, ADL_{rel}, and ADL_{col}, were obtained by minimizing the objectives of Γ, Γ_{rel}, and Γ_{col}, respectively. Hyperparameters of our approaches were tuned via 5-fold cross-validation.

Table 1: Distribution prediction results on AVA.

Metric	AAkNN	IISLLD	BFGS	ADL	ADL_{rel}	ADL_{col}
Cheb	0.152	0.166	0.156	0.146	0.147	**0.143**
Euc	0.379	0.382	0.348	0.310	0.308	**0.303**
KL	0.446	0.393	0.393	0.332	0.328	**0.326**

Table 2: Distribution prediction results on PN.

Metric	AAkNN	IISLLD	BFGS	ADL	ADL_{rel}	ADL_{col}
Cheb	0.305	0.323	0.335	0.299	0.291	**0.290**
Euc	0.477	0.552	0.522	0.461	**0.448**	0.450
KL	0.531	0.599	0.690	0.517	**0.505**	0.507

3.2 Aesthetic Distribution Prediction

We first set up experiments in the scenario of aesthetic distribution prediction for images. Our approaches were compared against three existing LDL algorithms [4], including AAkNN, IISLLD, and BFGS. We evaluated each method by measuring the average distance between the predicted and true aesthetic distributions of images. Three metrics were adopted, i.e., Chebyshev distance (Cheb), Euclidean distance (Euc), and Kullback-Leibler divergence (KL).

Table 1 and Table 2 display the comparison results on AVA and PN, respectively. As can be seen, all the proposed approaches outperform the other competitors in different metrics. This verifies the promise of our approaches for aesthetic distribution learning. Besides, both ADL_{rel} and ADL_{col} achieve better performance than ADL in most cases, which highlights the importance of taking account of the reliability of training examples as well as the correlations between quality levels in our framework.

3.3 Aesthetic Label Prediction

As aforementioned, the aesthetic distribution can be converted to a single label with its numerical characteristics. Therefore, we further evaluate our approaches in the task of image aesthetic classification. On both datasets, each image was labeled as "high quality" or "low quality" according to the average of its ground-truth ratings. Given a test image, the expectation of the predicted aesthetic distribution was used as an aesthetic score, indicating the confidence of its belonging to the high quality category. We introduced several existing aesthetic classification algorithms as the competitors, i.e., the methods of Marchesotti [8], Ke [6], and Lo [7].

Figure 2 presents the ROC curves for the classification performance of different methods. More precisely, the AUC value achieved by ADL_{col} is 0.685 on AVA, and substantially

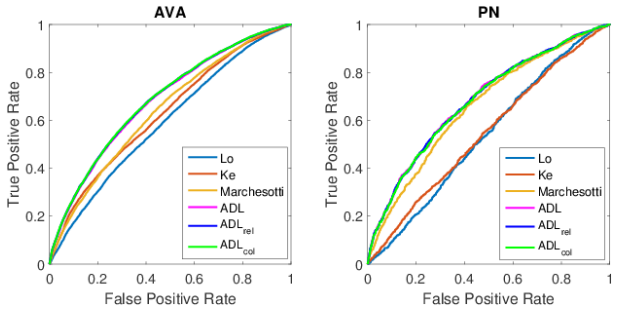

Figure 2: ROC curves for the classification performance of different methods.

higher than the values of 0.640, 0.629, and 0.592 for Marchesotti, Ke, and Lo, respectively. On PN, our approaches also enjoy up to 26.7% relative improvement over the existing algorithms. Such results suggest that our approaches, while specialized for aesthetic distribution learning, still emerge as highly effective tools for image aesthetic classification. Moreover, it should be noted that there seems no significant difference in the AUC values of ADL, ADL_{rel}, and ADL_{col}.

3.4 Aesthetics-based Reranking

In order to investigate the role of aesthetics in image search, we manually selected 10 queries, and collected the top 50 images returned by a well-known search engine for each query. The aesthetic quality of a returned image was evaluated based on its aesthetic score predicted by ADL_{col}, which was used as the representative of our approaches. Following the Borda count method [1], we developed a new reranking strategy by fusing the initial ranking and the ranking in order of aesthetic quality for each query. We term the resulting ranking as the aesthetic-based ranking.

Given a query, five volunteers were invited to select the 20 most agreeable results, from the set composed of the top 20 images in the initial ranking and those in the aesthetic-based ranking. For either ranking, the precision metric is defined as the proportion of selected images from it:

$$Prec = \sum_{k=1}^{20} \frac{I(k)}{20} \cdot \frac{1}{O(k)+1}, \tag{11}$$

where $I(\cdot)$ and $O(\cdot)$ are two indicator functions. $I(k) = 1$ if the k-th selected image came from the given ranking and zero otherwise. Likewise, $O(k)$ indicates whether the image appears in the initial ranking and aesthetic-based ranking simultaneously. The average value of precision over all volunteers was reported to evaluate the overall performance.

Table 3 lists the performance comparison between the initial ranking and aesthetic-based ranking with respect to different queries. We can see that for most queries, volunteers prefer the images from the aesthetic-based ranking rather than those from the initial ranking. This points clearly to the importance of the aesthetic quality of results in image search. In addition, the aesthetic-based ranking does not exhibit improved performance on some queries about products or people, such as "Phone" and "Trump". We conjecture that users aim to gain related information for such queries, and the aesthetics may not be the major concern in their search process.

Table 3: Performance comparison between the initial ranking and aesthetic-based ranking in terms of precision.

Query	Initial	Aesthetic	Query	Initial	Aesthetic
Lamp	0.455	**0.545**	Sunset	0.481	**0.519**
Car	**0.503**	0.497	Phone	**0.529**	0.471
Penguin	0.491	**0.509**	Rain	0.479	**0.521**
Party	0.493	**0.507**	Cat	0.448	**0.552**
Trump	**0.533**	0.467	Dandelion	0.479	**0.521**

4 CONCLUSIONS

In this paper, we investigate the problem of image aesthetics assessment from a new perspective of learning the distribution over quality levels. Our approach is thereby able to characterize the disagreement among the aesthetic perceptions of users. We build upon the framework of label distribution learning, and integrate disparate sources of information in our model. For future work, we plan to experiment with other label distribution learning algorithms to augment our current scheme.

5 ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (61573219, 61671274), NSFC Joint Fund with Guangdong under Key Project (U1201258), China Postdoctoral Science Foundation (2016M592190), and the Fostering Project of Dominant Discipline and Talent Team of Shandong Province Higher Education Institutions.

REFERENCES

[1] J.A. Aslam and M. Montague. 2001. Models for metasearch. In *SIGIR*. 276–284.
[2] R. Datta, J. Li, and J.-Z. Wang. 2008. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. In *ICIP*. 105–108.
[3] B. Geng, L. Yang, C. Xu, X.-S. Hua, and S. Li. 2011. The role of attractiveness in web image search. In *MM*. 63–72.
[4] X. Geng. 2016. Label distribution learning. *TKDE* 28, 7 (2016), 1734–1748.
[5] X. Geng and P. Hou. 2015. Pre-release prediction of crowd opinion on movies by label distribution learning. In *IJCAI*. 3511–3517.
[6] Y. Ke, X. Tang, and F. Jing. 2006. The design of high-level features for photo quality assessment. In *CVPR*. 419–426.
[7] K. Lo, K. Liu, and C. Chen. 2012. Assessment of photo aesthetics with efficiency. In *ICPR*. 2186–2189.
[8] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. 2011. Assessing the aesthetic quality of photographs using generic image descriptors. In *ICCV*. 1784–1791.
[9] N. Murray, L. Marchesotti, and F. Perronnin. 2012. AVA: A large-scale database for aesthetic visual analysis. In *CVPR*. 2408–2415.
[10] M. Sánchez-Fernández, M. de Prado-Cumplido, J. Arenas-García, and F. Pérez-Cruz. 2004. SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems. *TSP* 52, 8 (2004), 2298–2307.
[11] O. Wu, W. Hu, and J. Gao. 2011. Learning to predict the perceived visual quality of photos. In *ICCV*. 225–232.