Learning to Predict the Perceived Visual Quality of Photos

Ou Wu, Weiming Hu, Jun Gao
NLPR, Institute of Automation, Chinese Academy of Sciences
{wuou, wmhu, jgao}@nlpr.ia.ac.cn

Abstract

Visual quality (VisQ) representation is a fundamental step in the learning of a VisQ prediction model for photos. It not only reflects how we understand VisQ but also determines the label type. Existing studies apply a scalar value (i.e., a categorical label or a score) to represent VisQ. As VisQ is a subjective property, only a scalar value is insufficient to represent human's perceived VisQ of a photo. This study represents VisQ by a distribution on pre-defined ordinal basic ratings in order to capture the subjectivity of VisQ better. When using the new representation, the label type is structural instead of scalar. Conventional learning algorithms cannot be directly applied in model learning. Meanwhile, for many photos, the numbers of users involved in the evaluation are limited, making some labels unreliable. In this study, a new algorithm called support vector distribution regression (SVDR) is presented to deal with the structural output learning. Two independent learning strategies (reliability-sensitive learning and label refinement) are proposed to alleviate the difficulty of insufficient involved users for rating. Combining SVDR with the two learning strategies, two separate structural-output regression algorithms (i.e., reliability-sensitive SVDR and label refinement-based SVDR) are produced. Experimental results demonstrate the effectiveness of our introduced learning strategies and learning algorithms.

1. Introduction

Visual quality (VisQ) evaluation for photos has received increasing attention recently [1, 2, 7, 10, 12]. Automated evaluation VisQ of photos can facilitate the management of ever-increasing large amounts of online photos [1, 3, 9] and next-generation image retrieval [3]. For example, in Web image search, a photo's VisQ can be incorporated into ranking so that most relevant and best looking photos can be returned [7]. A photo management system can select high-quality images to show and eliminate low-quality ones under space constraints [3]. Automated VisQ evaluation requires learning of a VisQ prediction model. Existing studies on the model learning can be summarized into two classes:

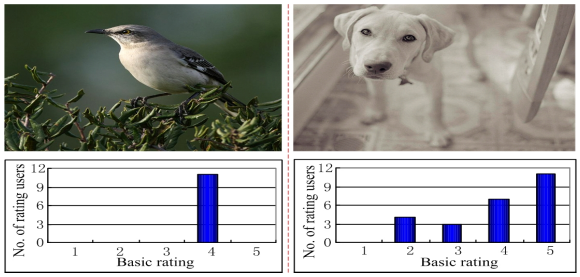


Figure 1. Two photos with their associated user ratings. The histograms show the numbers of users rated for each basic rating.

- *Extraction of more advanced features*. This category focuses on the understanding of the underlying mechanism of how people perceive the aesthetics of photos. Plenty of domain knowledge in photography is brought in to light on the extraction of discriminative features. Most existing studies fall into this category [1, 2, 8, 10, 12].
- *Utilization of more sophisticated learning algorithms*. This category focuses on the application of effective learning algorithms for VisQ classifiers or regression functions over the features of photos. A representative work can be found in [3].

In most of the above studies, training data are collected from online photo sharing communities such as photo.net and dpchallenge.com. In these communities, users rate a photo by choosing one of the predefined ordinal basic ratings, which are consecutive integers with a higher integer indicating better rating. Consequently, a photo is associated with a set of multiple ratings by different users. Figure 1 shows two photos and their associated user ratings over five ordinal basic ratings ("1-5"). In learning a prediction model, each photo's associated user ratings are transformed into a scalar value (e.g., "+1" or "-1"), which is taken as the photo's label. However, VisQ is a subjective property, as different users may perceive inconsistent or even opposite VisQs of the same photo (e.g., the right photo in Fig. 1). A scalar value is insufficient to capture the true nature of the subjectivity of VisQ. Although the two photos in Fig. 1 have equal average rating scores and thus equal

VisQ labels by existing representations, humans do not perceive them equally. Some published studies have noted the limitations of existing VisQ representations. For example, Wu et al.[13] claimed that existing studies ignore the truth that people tend to assign inconsistent labels to the same photo. The statistic in [7] reveals that there is ambiguity in the perceived quality of photos. However, they do not provide corresponding solutions.

1.1. Our work

Unlike previous representation strategies, this study applies a distribution vector on predefined ordinal basic ratings to represent the perceived VisQ of a photo. This new representation is based on a simple viewpoint: given two photos, their perceived VisQs by humans differ in the proportions of human choices on predefined ordinal basic ratings, not of a specific basic rating. For example, the Fig. 1 photos, although they both receive user ratings on "4", also earn different user ratings on other basic ratings. Our new representation is in accord with the truth that most photos have inconsistent user ratings. As will be detailed in Section 2, a distribution representation has several advantages.

According to the new representation, each training and test photo's label type is a distribution vector. Consequently, conventional classification and regression algorithms are unable to be directly used. In addition, as the training data are from online resources, learning about VisQ distribution prediction involves the challenge that the numbers of users that rate some photos are insufficient. Insufficient involved users cause some labels to become unreliable. This study introduces a novel structural regression algorithm (SVDR) to deal with the new label type, and two independent learning strategies (reliability-sensitive learning and label refinement) to deal with unreliable labels.

Besides the theoretical value of this study, VisQ distribution prediction has also practical capabilities. An VisQ distribution prediction algorithm can facilitate professional photographers to obtain detailed information about how the public evaluate their photos. The prediction of VisQ distribution offers several new ways of image ranking for search engines. For instance, images can be ordered according to subjectivity (distribution's variance), the median instead of mean of user ratings[1], or other numerical characteristics such as a specific quantile.

Our contributions can be summarized as follows:

1. A new representation strategy is presented based on the analysis of the representation of the perceived VisQs of photos. This new representation can capture the subjectivity nature of the perceived VisQs of photos better.
2. A new structural regression algorithm (i.e., SVDR) is proposed to handle the challenge that the label type is a (distribution) vector.
3. Two independent learning strategies (reliability-sensitive learning and label refinement) are introduced to cope with the difficulty brought by insufficient involved users for some photos. Respectively combining SVDR with reliability-sensitive learning and label refinement, two separate learning algorithms are generated.

The rest of the paper is organized as follows. Section 2 discusses the proposed distribution representation and the challenges in learning. Section 3 describes SVDR. Section 4 introduces two independent learning strategies regarding unreliable labels and then two concrete learning algorithms. Section 5 reports the experimental results. Finally, conclusions are given in Section 6.

2. Distribution Representation and Challenges in Prediction Model Learning

VisQ representation is a fundamental step in VisQ prediction. It determines label type and subsequent model learning. This section defines our representation strategy and analyzes the challenges in model learning.

Basically, given a photo and its associated users ratings, the representation relies on the transformation from user ratings to a label. Assuming that there are Z ordinal basic ratings and the set of basic ratings is denoted as BRS ($= \{BR_1, \cdots, BR_Z\}$). The user ratings for a photo can be described by $S_k = (S_k(1), \cdots, S_k(L_k))$, where $S_k(i) \in BRS$ is given by the i-th user and L_k is the number of users who have rated the photo (rating users). Existing studies usually employ a categorical label [2, 8] or a score [3] to represent the VisQ. Take photos in Fig. 1 as an example. The left photo's user ratings are (0, 0, 0, 11, 0), while the right photo's are (0, 4, 3, 7, 11). Both their average rating scores are "4" (the calculation of the right photo is $(2 \times 4 + 3 \times 3 + 4 \times 7 + 5 \times 11)/(4 + 3 + 7 + 11) = 4$). According to representation strategies in [2, 8], both their categorical labels are "+1" and both their VisQ scores are 4. A binary classification or regression algorithm can be applied to learning a VisQ classification or scoring model.

As presented earlier, a categorical label or a score is insufficient to represent the perceived VisQ of a photo. We make a statistical analysis on photo numbers (from photo.net) according to average ratings ($avgr$) and standard deviation of their user ratings (sdr). The value of sdr indicates the inconsistence of a photo's user ratings, with a higher value meaning higher inconsistence. The histogram is shown in Fig. 2. Num denotes the photo number located in each $avgr$ and sdr interval. Most photos' $avgr$ values are located in [4.5, 5.5]. Photos in this interval are difficult to handle by the categorical representation. Most photos' sdr values are larger than 0.5, denoting that most photos have inconsistent ratings. Hence, a more appropriate VisQ representation strategy is required. We define a new repre-

[1]For skewed distributions, the median is better than mean. We will in the following section that most typical VisQ distributions are skewed.

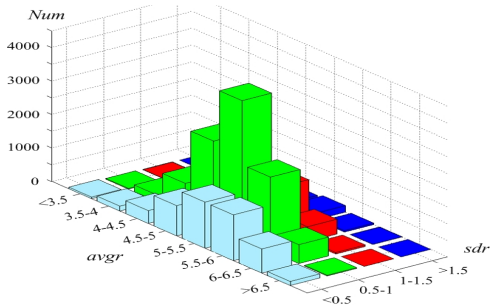


Figure 2. The histogram of numbers of photos located in different intervals.

sentation as follows:

$$y_k = (p_{k1}, \cdots, p_{ki}, \cdots, p_{kZ})^T, \quad (1)$$

where p_{ki} is defined as the proportion that the i-th basic rating is chosen by users. Let $\delta()$ be the indication function. p_{ki} is calculated as:

$$p_{ki} = \sum_{j=1}^{L_k} \delta(S_k(j) = BR_i)/L_k. \quad (2)$$

When $L_k \to +\infty$, y_k in Eq. (1) becomes the stable distribution of the perceived VisQ for a photo. Based on Eqs. (1) and (2), the VisQ labels of the two Fig. 1 photos are $(0, 0, 0, 1, 0)^T$ and $(0, 0.16, 0.12, 0.28, 0.44)^T$, respectively. This representation captures the subjective nature of VisQ better in three folds. (a) It carries more original information about user ratings than a scalar value. (b) It is consistent with the subjectivity of VisQ that a photo could be perceived inconsistently. A distribution's standard indicates how subjective a photo's VisQ is. The larger the standard deviation, the more subjective the VisQ is. (c) A distribution can be transformed to a categorical label and a score in classification or ranking photos according to VisQ.

A clustering analysis based on K-means for the VisQ distributions is performed on the data from photo.net. Figure 3 shows the cluster centers (typical rating distributions) of users' VisQ distributions over basic ratings ("3-7") when the number of clusters is set to 5. In Fig. 3, a curve indicates a VisQ distribution where X-axis denotes the index of basic ratings and Y-axis denotes the proportion of user ratings. It can be observed that most representative distributions are skewed. For skewed distributions, the median value appears to be more appropriate to describe the distributions than the mean value. The proposed distribution presentation can yield the median values, while existing presentations cannot.

Under the distribution presentation, the label type is a (distribution) vector. The learning for this new label type with online resources encounters two challenges:

i. *Structural output*. Predicting a VisQ distribution label is a type of structural output learning problem. The output label should satisfy: each entry is located in [0, 1] and the sum of all entries is equal to 1. Therefore, conventional classification (e.g., support vector machine) and regression (e.g., neural networks) algorithms are not directly applicable.

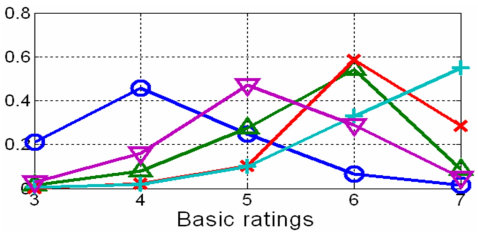


Figure 3. Five cluster centers (VisQ distributions) over basic ratings "3-7". User ratings on "1-2" are rare.

ii. *Unreliable VisQ distribution labels*. When constructing the training set, a reliable VisQ distribution label of a photo is crucial. Naturally, the larger the L_k, the more reliable the distribution by Eq. (1) and Eq. (2) is. However, some photos' rating user numbers (L_k) are limited. For instance, the numbers of rating users for photos in Fig. 1 are 11 and 25, respectively, so their distribution labels obtained by Eq. (1) are not very reliable. As a result, the learned model is likely to be unreliable and biased.

A new support vector distribution regression algorithm (SVDR) is presented toward (i). Two independent learning strategies are proposed toward (ii) and detailed in Section 4.

3. SVDR

Let f represent the pursued prediction function (model) from the feature space X and the distribution label space Y. Let $P(x, y)$ denote the joint distribution and $l(y, f(x))$ denote the prediction loss. Then the learning of f is to minimize the following expected loss

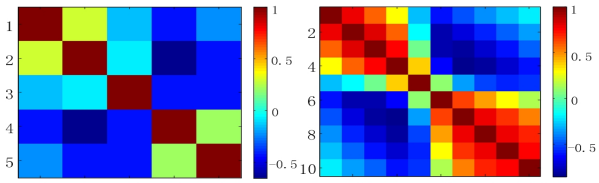
$$R(f) = \int_{X \times Y} l(y, f(x)) dP(x, y) \quad (3)$$

based on training samples $\{(x_1, y_1), \cdots, (x_N, y_N)\} \subset X \times Y$, where x_i is a feature vector and y_i is a distribution vector calculated by Eq. (1). We propose a new algorithm called SVDR to cope with the learning. SVDR is based on a structural support vector machine [11], which provides a natural way to address structural output learning.

Specifically, SVDR aims to learn a discriminate function $\Phi : X \times Y \to R$ over the input feature and output distribution label pairs. With Φ, the prediction model f (or distribution regression function) is

$$f(x) = \arg\max_{y \in Y} \Phi(x, y). \quad (4)$$

$\Phi(x, \cdot)$ can be considered as a matching function. Ideally, the maximum of $\Phi(x, \cdot)$ is found at the desired distribution



(a) (b)

Figure 4. The correlation coefficients of user ratings on different basic ratings.

label for an input x. Φ is assumed to be linear in some combined features (denoted by $(\Psi(x,y))$) of inputs and outputs.

$$\Phi(x,y) =< \mathbf{w}, \Psi(x,y) >, \tag{5}$$

where $\mathbf{w}$ denotes the parameter vector to be learned. Once $\Psi(x,y)$ and the loss function $l(y,f(x))$ are defined, the optimization for SVDR can be summarized as follows.

$$\begin{aligned} &\min_{\mathbf{w},\xi} \tfrac{1}{2}\|\mathbf{w}\|^2 + \tfrac{C}{N}\sum_{i=1}^{N}\xi_i \\ &s.t. \forall i \in [1,N], \xi_i \geq 0 \\ &\forall i \in [1,N], \forall y \in Y/y_i :< \mathbf{w}, \Delta\Psi_i(y) > \ \geq l(y,y_i) - \xi_i \end{aligned} \tag{6}$$

where $\Delta\Psi_i(y) = \Psi(x_i,y_i) - \Psi(x_i,y)$, ξ is a slack variable. and C controls the model complexity. The following parts discuss the definition of $\Psi(x,y)$, the loss function $l(,)$, and the mathematical solution details for Eq. (6).

3.1. The definition of $\Psi(x,y)$

The specific form of Ψ depends on the nature of the problem. In our study, Ψ is divided into two parts. (1) The first part reflects the interactions between input features and output distribution labels. (2) The second part captures the correlations among the entries of output distribution labels.

Let D be the feature dimension. Motivated by multi-class learning [11], the first part is defined as:

$$\Psi_1(x,y) = x \otimes y = (x(1)y(1), ..., x(D)y(Z))^T, \tag{7}$$

where $\otimes$ is the tensor product.

To explore the correlations among entries of distribution labels, we calculate the correlation of user ratings on different basic ratings using data from photo.net and dpchallenge.net[2]. Their correlation efficiency maps are shown in Figs. 4 (a) and (b), respectively. User ratings on most pairs of basic ratings are correlated, especially the adjacent basic ratings on the data from dpchallenge.net. To capture the dependency between basic ratings, the second part of $\Psi(x,y)$ is defined as follows:

$$\begin{aligned} \Psi_2(x,y) = (&y(1)y(2), \cdots, y(1)y(Z), y(2)y(3), \\ &\cdots, y(2)y(Z), y(3)y(4), \cdots, y(Z-1)y(Z))^T \end{aligned} \tag{8}$$

Hence, the definition of $\Psi(x,y)$ is

$$\Psi(x,y) = [\Psi_1(x,y)^T, \Psi_2(x,y)^T]^T. \tag{9}$$

[2]The number of basic ratings in photo.net is seven. However, the first two basic ratings receive few users' ratings. Only the remaining five basic ratings are considered and still named "1-5" in this study. The number of basic ratings in dpchallenge.net is ten.

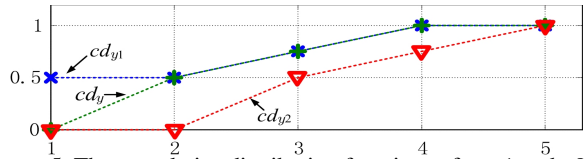


Figure 5. The cumulative distribution functions of y, $y1$ and $y2$.

3.2. The definition of loss function ($l(,)$)

A common loss function is:

$$l^{(1)}(y,\widehat{y}) = \left\|y - \widehat{y}\right\|_2^2. \tag{10}$$

This loss function ignores that the pre-defined basic ratings are ordinal. Therefore, it is inappropriate and beyond our intuition when calculating losses. Let us consider the following exemplar distribution labels (five ordinal basic ratings): $y = (1,0,0,0,0)^T$, $y' = (0,1,0,0,0)^T$, $y'' = (0,0,0,0,1)^T$. According to Eq. (10), we have $l^{(1)}(y,y') = l^{(1)}(y,y'') = 2$. These results are inconsistent with our intuition that y' is closer to y than y'' because the basic ratings are ordinal.

Assume that two distribution labels over BRS are denoted as y_a and y_b, respectively. At first, the cumulative distribution functions (cd_a and cd_b) of these two labels (distributions) are calculated by

$$cd_a(i) = \sum_{j=1}^{i} y_a(j) \quad \text{and} \quad cd_b(i) = \sum_{j=1}^{i} y_b(j). \tag{11}$$

where $i = 1, \cdots, Z$. Then a new loss is defined as:

$$l^{(2)}(y_a,y_b) = \sum_{i=1}^{Z-1} [cd_a(i) - cd_b(i)]^2 \tag{12}$$

With this new loss, the losses between the three exemplar distributions are: $l^{(2)}(y,y') = 1 < l^{(2)}(y,y'') = 4$.

Another example is: $y = (0.5,0,0.25,0.25,0)^T$, $y1 = (0,0.5,0.25,0.25,0)^T$, and $y2 = (0,0,0.5,0.25,0.25)^T$. $y1$ is closer to y than $y2$. If Eq. (10) is used, $l^{(1)}(y,y1) = 0.5 > l^{(1)}(y,y2) = 0.375$; if our proposed loss Eq. (12) is used, $l^{(2)}(y,y1) = 0.25 < l^{(2)}(y,y2) = 0.75$. The results indicate the reasonableness of the proposed loss function. Their cumulative distribution functions (cd_y, cd_{y1}, and cd_{y2}) are shown in Fig. 5.

3.3. The solution of Eq. (6)

As y is continuous, the second class of constraints for each y_i in Eq. (6) is infinite. To handle this problem, for each y_i, a small working set storing most active constraints is constructed to replace the complete infinite constraints. Following the method proposed by [11], the construction of the working set for a training sample (x_i,y_i) is given in Algorithm 1. The maximum optimization problem in Algo-

Algorithm 1 Update the working set (WS) for (x_i, y_i) in the t-th iteration

Input: (x_i, y_i), ε, $\mathbf{w_{t-1}}$, working set $WS_{t-1}(i)$.
Output: working set $WS_t(i)$.
Steps:

1. Compute $\bar{y} = \arg\max_{y \in Y} Q(y)$, where $Q(y) = l(y_i, y) - < \mathbf{w}_{t-1}, \Delta\Psi_i(y) >$.
2. Compute $\eta_i = \max\{0, \max_{y \in WS_{t-1}(i)} Q(y)\}$.
3. If $Q(\bar{y}) > \eta_i + \varepsilon$, then $WS_t = WS_{t-1} \cup \bar{y}$, else, $WS_t = WS_{t-1}$.

rithm 1 is denoted as follows:

$$\begin{aligned} &\max_{y} \sum_{d=1}^{D} \sum_{\varsigma=1}^{Z} \mathbf{w}_{t-1}(Z \cdot (d-1) + \varsigma) x_i(d) y(\varsigma) + \\ &\sum_{\varsigma=1}^{Z} \sum_{\eta=\varsigma+1}^{Z} \mathbf{w}_{t-1}(Z(D+\varsigma-1) - \tfrac{\varsigma(\varsigma+1)}{2} + \eta) y(\varsigma) y(\eta) + l(y_i, y) \\ &s.t. \sum_{\varsigma=1}^{Z} y(\varsigma) = 1, \quad y(\varsigma) \geq 0, \quad \varsigma = 1, \cdots, Z \end{aligned} \tag{13}$$

The above problem can be solved via conventional optimization techniques such as quadratic programming. Once all the working sets are obtained, the second class of constraints of Eq. (6) becomes:

$$\forall i \in [1, N], \forall y \in WS_{t-1}(i) : < \mathbf{w}, \Delta\Psi_i(y) > \geq l(y_i, y) - \xi_i. \tag{14}$$

Then $\mathbf{w}$ is updated by solving the dual form of Eq. (6) using the cutting-plane algorithm [6]. The algorithm stops if each sample's working set remains unchanged.

4. Learning Regarding Unreliable Labels

SVDR does not consider that some training samples' labels (distributions) are relatively unreliable. Two independent strategies are proposed to alleviate the problem of unreliable labels.

At first, a reliability factor (r_k) is introduced to measure the reliability of a distribution label (y_k) which is calculated by Eq. (1) and Eq. (2) from the user ratings (S_k) of a photo. The reliability factor depends on the number of rating users L_k ($= |S_k|$). The larger the L_k, the larger the reliability factor r[3]. The relationship between L_k and r_k is:

$$r_k = \mu(L_k). \tag{15}$$

where μ is required to be a non-decreasing function. As r_k should be set to 0 if L_k is 0 and to 1 if L_k is sufficiently large, and each additional label should have marginal utility, the function used in this study is:

$$\mu(L_k) = \begin{cases} \frac{\ln(L_k+1)}{\ln(L_k+1)+1} & \text{if} \quad L_k <= \tau \\ 1 & \text{otherwise} \end{cases}, \tag{16}$$

where τ is a threshold. With the defined function above, two independent learning strategies are introduced as follows.

[3] L_k reflects a photo's popularity, importance, or interestingness. In a statistical viewpoint, if each rating is taken as a random event, L_k reflects the sample size. The larger the sample size, the larger the confidence level.

4.1. Reliability-sensitive learning (RSL)

Intuitively, incorrect prediction losses of reliable photos should be higher than those of less reliable ones. To this end, the expected loss defined in Eq. (3) is modified to

$$R(f) = \int_{X \times Y} r(y) l(y, f(x)) dP(x, y), \tag{17}$$

where $r(y)$ is defined by Eq. (15). Eq. (17) places higher punishments upon more reliable samples with inaccurate predictions. When all labels are reliable, Eq. (17) is equal to Eq. (3). In this study, learning under the risk of Eq. (17) is called reliability-sensitive learning (RSL). RSL is similar to the cost-sensitive learning [5]. A simple learning strategy toward Eq. (17) is to define a new loss:

$$l'(y, f(x)) = r(y) l(y, f(x)). \tag{18}$$

With the above loss, RSL can be solved directly using the learning algorithms designed for Eq. (3).

4.2. Label refinement (LR)

Unlike RSL, this strategy aims at refining unreliable distribution labels. Motivated by the label propagation in semi-supervised learning [14], the lower-reliable labels are iterated refined via the propagation of information regarding more reliable labels to less reliable labels.

Given N training data $(x_i, y_i, r_i), i = 1, \cdots, N$, which are divided into two subsets: V, containing the data with $r = 1$, and U, containing the data with $r < 1$. First, a transformation matrix $\mathbf{M}$ is calculated and represented by:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{VV} & \mathbf{M}_{VU} \\ \mathbf{M}_{UV} & \mathbf{M}_{UU} \end{bmatrix}, \tag{19}$$

where each entry is the normalized similarity of two corresponding training data for the sum of each row equaling 1; $\mathbf{M}_{VV}$ is the submatrix that describes the similarities between samples in V; likewise, $\mathbf{M}_{VU}$, $\mathbf{M}_{UV}$, $\mathbf{M}_{UU}$ also describe similarities of samples in corresponding subsets V and U. Let $Y_V = \left[y_1, \ldots, y_{|\mathrm{V}|}\right]^{\mathrm{T}}$ and $Y_U = \left[y_{|\mathrm{V}|+1}, \ldots, y_{|\mathrm{V}|+|\mathrm{U}|}\right]^{\mathrm{T}}$. The refinement rule is defined as:

$$\left[\ {Y_V}^T, {Y_U}^T\ \right] \leftarrow \mathbf{r} \left[\ Y_V(0)^T, Y_U(0)^T\ \right] + (1 - \mathbf{r}) \mathbf{M} \left[\ {Y_V}^T, {Y_U}^T\ \right], \tag{20}$$

where $Y_V(0)$ and $Y_U(0)$ are the samples' initial distribution labels, and

$$\mathbf{r} = \begin{bmatrix} r_1 & 0 & 0 \\ & \ddots & \\ 0 & 0 & r_N \end{bmatrix} = \begin{bmatrix} \mathbf{r}_V & \mathbf{0} \\ \mathbf{0} & \mathbf{r}_U \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{r}_U \end{bmatrix}. \tag{21}$$

Then the following theorem is obtained.

Theorem 1 *Y_V and Y_U in Eq. (20) are converged to*

$$\begin{aligned} Y_V &= Y_V(0) \\ Y_U &= (\mathbf{1} - (\mathbf{1} - \mathbf{r}_U)\mathbf{M}_{UU})^{-1} \mathbf{r}_U Y_U(0) \\ &\quad + (\mathbf{1} - (\mathbf{1} - \mathbf{r}_U)\mathbf{M}_{UU})^{-1} (\mathbf{1} - \mathbf{r}_U) \mathbf{M}_{UV} Y_V(0) \end{aligned}. \tag{22}$$

Algorithm 2 Label refinement

Input: $(x_1, S_1), \cdots, (x_N, S_N)$, ε, $t = 1$.
Output: New distribution labels (Y).
Steps:

1. Compute the distribution label (y_i) and reliability factor (r_i) of each sample.
2. Calculate $\mathbf{M}$ and $\mathbf{r}$.
3. Calculate new distribution labels using Eq. (20).
4. If $\|Y_U(t) - Y_U(t-1)\|_1 < \varepsilon$, return $[Y_V(0)^T, Y_U(t)^T]^T$; otherwise $t = t + 1$ and goto Step 3.

The proof is omitted and is similar to the proof in [14].

In Eq. (22), Y_U consists of two components: values from their initial labels $Y_U(0)$, and values from reliable labels $Y_V(0)$. The higher the $\mathbf{r}_U$, the larger determinant of $Y_U(0)$. $\mathbf{r}_U = 0$, Eq. (22) becomes a common form of label propagation; when $\mathbf{r}_U = 1$, $Y_U = Y_U(0)$. When the number of samples is large, an iterative LR is shown in Algorithm 2.

4.3. SVDR within the two learning strategies

4.3.1 Reliability-sensitive SVDR (R-SVDR)

Section 4.1 suggests that RSL learning can be solved by conventional learning algorithms after reshaping the loss function (see Eq. (18)). Hence, the loss function in Eq. (12) is re-formulated as: $l^{(3)}(y, f(x)) = r(y) \cdot l^{(2)}(y, f(x))$. Then a reliability-sensitive SVDR (R-SVDR) can be obtained by replacing the loss functions in Eq. (6), Eq. (13) and Eq. (14) with $l^{(3)}$. Algorithm 3 shows the steps.

Algorithm 3 R-SVDR

Input: samples $(x_1, S_1), \cdots, (x_N, S_N)$, ε, $\mathbf{w}_0 = null$, working set $WS_0(i) = null, i = 1, \cdots, N$, $Maxnum$, $t = 1$.
Output: $\mathbf{w}$.
Steps:

1. Compute the distribution (y_i) and reliability factor (r_i) of each training sample.
2. Update the working set for each sample using Algorithm 1 based on $l^{(3)}$; if all working sets remain unchanged, return current $\mathbf{w}$ and exit.
3. Replace the second class of constraints of Eq. (6) using Eq. (14) based on the updated working sets.
4. Solve the dual form of Eq. (6) with the replaced constraints by Eq. (14) and $l^{(3)}$ to obtain a new $\mathbf{w}$.
5. $t = t + 1$, if $t > Maxnum$, return $\mathbf{w}$; otherwise goto Step 2.

4.3.2 Label refinement-based SVDR (L-SVDR)

Label refinement (LR) is first used to update unreliable distribution labels. Then SVDR is performed on refined labels. The integrated algorithm is called L-SVDR, as shown in Algorithm 4.

Algorithm 4 L-SVDR

Input: samples $(x_1, S_1), \cdots, (x_N, S_N)$, ε, $\mathbf{w_0} = \mathbf{null}$, working set $WS_0(i) = null, i = 1, \cdots, N$, $Maxnum$, $t = 1$.
Output: $\mathbf{w}$.
Steps:

1. Compute the distribution (y_i) and reliability factor (r_i) of each training sample.
2. Refine distribution labels using Algorithm 2.
3. Update the working set for each sample using Algorithm 1 based on $l^{(2)}$; If all working sets remain unchanged, return $\mathbf{w}$ and exit.

4-5 are as the same as Steps 4-5 in Algorithm 3, where the loss function is replaced by $l^{(2)}$.

In the experiments, to test the performances of L-SVDR, the LR algorithm is performed on the training set and test set independently.

5. Evaluations

This section verifies the effectiveness of the proposed learning algorithms for the prediction of perceived VisQ distributions of photos. The proposed algorithms will be compared with methods that are slightly modified from classical algorithms such as support vector machine (SVM). In addition, the performance of the proposed algorithms using for classification and scoring is also evaluated.

5.1. Experimental data and setup

Two photo sets were constructed as the experimental data. One called DS1 is collected from photo.net. The other, called DS2, is collected from dpchallenge.net and based on the data compiled by Datta [4]. DS1 contains 1224 images and each image has 56 dimensional features described by [2]. DS2 contains 9000 images, and the features are also those described by [2]. The average of the number of rating users per photo on DS1 is approximately 28, while the average is about 189 on DS2. Both sets are randomly divided into two equal parts: one for training and the other for testing. This division is repeated five times and the average results are recorded. In calculating reliability factors, τ is set to 200. When running R-SVDR and L-SVDR, the radial basis kernel is chosen. The parameters C and g are searched via 5-cross validation in $\{0.1, 1, 10, 50, 100\}$ and $\{0.01, 0.1, 1, 10\}$, respectively.

Baseline methods: support vector regression (SVR) and back propagation neural networks (BP) are used as competing methods. To make the presentation clear, SVR is still named SVM in the following part. When applying SVR, each dimension of distribution outputs is separately learned and predicted. The parameters of SVM are searched similar to SVDR. For BP, the hidden neurons' number is determined from $\{50, 100, 150, 200, 250\}$.

Three different losses are calculated: $l^{(2)}$, $l^{(3)}$, and the

Table 1. Competing algorithms on each loss.

Loss type	Algorithms
$l^{(2)}$	SVM, BP, SVDR
$l^{(3)}$	SVM, BP, R-SVDR
$l^{(4)}$	SVM, BP, L-SVDR

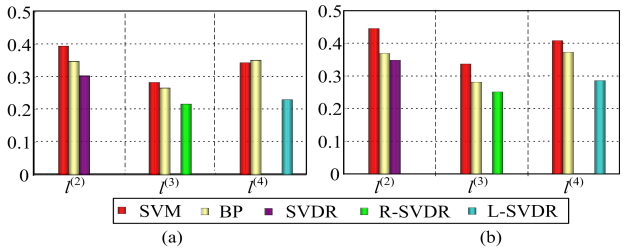


Figure 6. Results on DS1.

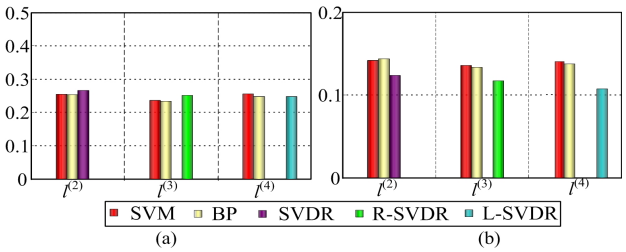


Figure 7. Results on DS2.

$l^{(2)}$ on the refined distributions by LR are also calculated, which is called $l^{(4)}$. The reason for using three different losses is that there is no reliable ground-truth (note that some distribution labels are unreliable), so only a single measuring criterion appears to be unfair. Yet the foci of each loss are different. The main goal of using $l^{(2)}$ is to compare SVDR with classical algorithms. The goal of using $l^{(3)}$ is to evaluate the performances of R-SVDR. The goal of using $l^{(4)}$ is to evaluate the performances of L-SVDR. Consequently, when different losses are measured, the results of different algorithms are reported. To clarify the comparisons, Table 1 lists each loss's corresponding competing algorithms. As introduced in Section 4.3.2, LR is independently run on the test set.

5.2. Results on the DS1 set

On DS1, most samples' reliability factors range from [0.65, 0.85]. Figure 6 (a) shows the loss values of the compared methods. Overall, the SVDR-series (SVDR, R-SVDR, L-SVDR) achieves better results against the rest algorithms. SVM and BP's results are similar and inferior to others.

What a user usually cares about is the proportion of rating distributions on "high", "mediocre", and "low" for a photo's VisQ. We repeated all of the above experimental steps after transforming the BRS $\{1, 2, 3, 4, 5, 6, 7\}$ to a new BRS $\{1, 2, 3\}$ by mapping the ratings "1-4" to "1" (There is no user ratings on "1-2"), mapping "5" to "2", and mapping "6-7" to "3". The results are shown in Fig. 6 (b). SVDR-series obtains the lowest loss values.

Table 2. Results on different features for original BRS.

	DS1		DS2	
	All	Subset	All	Subset
SVDR ($l^{(2)}$)	0.3017	0.2798	0.3500	0.3401
R-SVDR ($l^{(3)}$)	0.2113	0.2056	0.2511	0.2524
L-SVDR ($l^{(4)}$)	0.2509	0.2430	0.2871	0.2620

Table 3. Results on different features for transformed BRS.

	DS1		DS2	
	All	Subset	All	Subset
SVDR ($l^{(2)}$)	0.2687	0.2590	0.1248	0.1203
R-SVDR ($l^{(3)}$)	0.2550	0.2325	0.1207	0.1076
L-SVDR ($l^{(4)}$)	0.2506	0.2470	0.1098	0.1061

5.3. Results on the DS2 set

On DS2, most samples' reliability factors range in [0.8, 1]. Figure 7 (a) shows the overall loss values of the competing methods. It can be observed that SVDR-series achieves similar results with BP and SVM. The underlying reason is that: the number of ordinal basic ratings is ten, so the dimension of $\Psi(x, y)$ is high (56*10+9*10/2). Hence the optimizations in SVDR-series are quite challenging and the learned model may be underfitting. A larger number of training samples and more complicated optimization algorithms may be helpful to alleviate the underfitting.

We also repeated all of the above experimental steps after transforming the BRS $\{1, 2, \cdots, 10\}$ to a new BRS $\{1, 2, 3\}$ by mapping the ratings "1-5" (User ratings on "1-2" are few.) to "1", and "6-7" to "2", and "8-10" to "3". The results are shown in Fig. 7 (b). Similar observations to Fig. 6 are obtained.

5.4. Results on different features

Now we investigate how the performances of the proposed algorithms vary with respect to feature selection. In [2], fifteen image features (e.g., lightness, texture, etc) are revealed to be more discriminative. Hence, all the competing algorithms are run with the fifteen features on DS1 and DS2. Tables 2 and 3 show the results on the original BRS and the transformed BRS as used in Sections 5.2 and 5.3, respectively ('All' denotes all the original features, while 'Subset' denotes the fifteen ones). The comparison indicates that only using partial more discriminative features can lead to better results.

5.5. Results on VisQ scoring

As presented earlier, a VisQ distribution can also be applied to classifying and scoring a photo based on the distribution's mean. We take the VisQ scoring as an example to compare the proposed algorithms with existing regression methods. In VisQ scoring or regression, the residual sum-of-squares error (RSSE) is reported. On DS1 (with

Table 4. Results of SVDR-series on reliable samples.

Data set	R-SVDR	L-SVDR	SVDR
DS2 (10 basic ratings)	**0.1459**	0.1496	0.1685
DS2 (3 basic ratings)	**0.0556**	0.0561	0.0582

original BRS), the RSSEs of BP, SVR (support vector regression), SVDR, R-SVDR, L-SVDR are 0.7123, 0.6909, 0.6812, 0.6917, 0.6687, respectively. On DS2 (with original basic ratings), the RSSEs of BP, SVR, SVDR, R-SVDR, L-SVDR are 0.7527, 0.7312, 0.7456, 0.7301, 0.7263, respectively. The results show that our algorithms can achieve slightly better performances than methods used in previous work. Results for transformed BRS also indicate the similar conclusion and are omitted due to lack of space.

5.6. Discussions

We have considered three losses ($l^{(2)}$, $l^{(3)}$, and $l^{(4)}$) to measure performances. In terms of $l^{(2)}$, SVDR is superior to other methods. In terms of $l^{(3)}$, R-SVDR achieves the best results. In terms of $l^{(4)}$, L-SVDR achieves the best results. The comparison reveals that the SVDR-series captures the nature of the learning problem better, especially the relationships between input features (X) and output distribution labels (Y). Meanwhile, the experiments reveal that only a subset of more discriminative features can improve performances, and our algorithms can also be used to score photos' VisQs.

As there is no reliable ground truth, it is hard to compare R-SVDR with L-SVDR. They are based on different loss measurements ($l^{(3)}$ and $l^{(4)}$). We note that the losses of reliable samples under different loss functions are identical. Hence, only the reliable samples on the test sets are selected to compare R-SVDR with L-SVDR. As the number of reliable samples on the DS1 set is limited, only reliable samples from DS2 are tested. The results are shown in Table 4 (SVDR is also compared). R-SVDR is slightly better than L-SVDR. This holds partly because R-SVDR directly imposes high punishments to reliable samples in learning. In addition, R-SVDR has lower computational complexity than L-SVDR. Consequently, R-SVDR appears to be superior to L-SVDR. We note that the loss values on reliable samples are lower than the average loss values on all the test samples (see Figs. 6 and 7). SVDR is inferior to both R-SVDR and L-SVDR on the reliable samples. These observations indicate that the proposed learning strategies are effective.

6. Conclusions

Aesthetics is usually considered as subjective. In our point of view, subjectivity should not be ignored in VisQ research. This paper has reviewed VisQ representations in previous literature. To capture the subjective nature of VisQ better, a new representation that leverages a distribution vector over predefined ordinal basic ratings has been provided. To cope with the difficulties in learning of a VisQ distribution prediction model, a structural regression algorithm (SVDR) and two separate learning strategies (RSL and LR) are proposed. Their combinations yield two concrete algorithms: R-SVDR and L-SVDR. The experimental results demonstrate the better performances of the introduced learning algorithms over several classical methods (SVM and BP). Additionally, our algorithms achieve comparable results to existing work on scoring. In terms of the performances on reliable samples, R-SVDR is better than L-SVDR.

Future work will adapt the proposed distribution representation and learning algorithms to other kinds of emotion distribution prediction as discussed in [4].

7. Acknowledgement

We would like to thank Dr. Bing Li and Wei Li for their useful suggestions. This work is partly supported by NSFC (Grant No. 60825204, 60672040, 60723005, 61003115).

References

[1] S. Bhattacharya and M. S. R. Sukthankar. A framework for photo-quality assessment and enhancement based on visual aesthetics. *ACM MM*, 2010. 1

[2] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. *ECCV, LNCS*, pages 288–301, 2006. 1, 2, 6, 7

[3] R. Datta, J. Li, and J. Z. Wang. Learning the consensus on visual quality for next-generation image management. *ACM MM*, pages 533–536, 2007. 1, 2

[4] R. Datta, J. Li, and J. Z. Wang. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. *ICIP*, 2008. 6, 8

[5] C. Elkan. The foundations of cost-sensitive learning. *IJCAI*, pages 973–978, 2001. 5

[6] V. Franc and S. Sonnenburg. Optimized cutting plane algorithm for support vector machines. *ICML*, pages 320–327, 2008. 5

[7] Y. Ke and et al. The design of high-level features for photo quality assessment. *CVPR*, pages 419–426, 2006. 1, 2

[8] I. Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. *ECCV*, pages 386–399, 2008. 1, 2

[9] P. Obrador, X. Anguera, R. Oliveira, and N. Oliver. The role of tags and image aesthetics in social image search. *WSM*, pages 65–72, 2009. 1

[10] X. Sun and et al. Photo assessment based on computational visual attention model. *ACM MM*, pages 541–544, 2009. 1

[11] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun. Support vector machine learning for interdependent and structured output spaces. *ICML*, pages 104–112, 2004. 3, 4

[12] L.-K. Wong and K.-L. Low. Saliency-enhanced image aesthetics class prediction. *ICIP*, pages 997–1000, 2009. 1

[13] Y. Wu and et al. The good, the bad, and the ugly: Predicting aesthetic image labels. *ICPR*, pages 1586–1589, 2010. 2

[14] X. Zhu and Z. Ghahramani. Learning from labeled and unlabeled data with label propagation, 2002. Technical Report CMU-CALD-02-107, Carnegie Mellon University. 5, 6