JOINT LEARNING FOR IMAGE-BASED HANDBAG RECOMMENDATION

Yan Wang, Sheng Li, and Alex C. Kot

Rapid-Rich Object Search (ROSE) Lab
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 639798
wang0696@e.ntu.edu.sg, {lisheng,eackot}@ntu.edu.sg

ABSTRACT

Fashion recommendation helps shoppers to find desirable fashion items, which facilitates online interaction and product promotion. In this paper, we propose a method to recommend handbags to each shopper, based on the handbag images the shopper has clicked. This is performed by **J**oint learning of attribute **P**rojection and **O**ne-class SVM classification (JPO) based on the images of the shopper's preferred handbags. More specifically, for the handbag images clicked by each shopper, we project the original image feature space into an attribute space which is more compact. The projection matrix is learned jointly with a one-class SVM to yield a shopper-specific one-class classifier. The results show that the proposed JPO handbag recommendation performs favorably based on initial subject testing.

***Index Terms*—** Recommender system, handbag, one-class classification, attribute, joint learning

1. INTRODUCTION

As the development of people's living condition, fashion recommendation is receiving an increasing attention in social media these years. Handbag, which has become an irreplaceable item in women's wardrobe ever since 1920s, is one of the most popular fashion items nowadays. Purchasing handbags online is convenient and efficient. However, for some e-commerce websites or manufacturers, how to find the shopper's most preferred handbags is an interesting while challenging problem. Consider the scenarios: 1) a shopper clicks the handbags she finds interesting and the on-line shop provides the shopper a list of recommended handbags she might prefer; 2) the handbag manufacturers recommend some new collections to shoppers according to their search history. These systems ease the task of finding preferred items in the collection or catch shoppers' attention when new collections are available. Therefore, it is desirable to develop such a handbag recommendation engine for e-commerce or shops.

Typically, recommender systems produce recommendations via collaborative filtering or content-based recommendation [1]. Some e-commerce websites consider collaborative recommendation, and they work well when enough ratings are given by a large community of shoppers [2]. Content-based recommendation proposes a way to apply across all products and predicts high interests of certain attributes among products without rating done. Such techniques have been applied to diverse areas such as movie [3, 4], music [5], and articles [1]. Many researchers focus on movie recommendation studies, and the work in [4] combines the concepts of knowledge and collaborative filtering in a principal way, which helps shoppers find content according to their preferences. Automatic music recommendation [5] trains a deep convolutional neural networks to predict latent factors from music audio, which is shown to perform better than the bag-of-words feature representation on an industrial-scale dataset. Wang et. al [1] combine the collaborative filtering and topic model for recommending scientific articles to users in an online scientific community.

For image-based handbag recommendation, the challenge is the sheer variety of different styles, as well as customers' tastes and fashion elements. In this paper, we aim to address the handbag recommendation problem by analyzing the content of the preferred handbags, i.e., the feature of handbag image itself rather than the ratings from other buyers. Empirically, customers tend to prefer handbags with similar (or even the same) attributes, such as the shape, pattern, color or style for a certain period of time. However, defining attributes for handbags is somewhat difficult. Besides, it takes tedious human labor to label the attributes. To address this issue, we project features extracted from handbag images into an attribute space automatically by an attribute projection model. Our work denotes "attribute" as shareable properties of handbags which may not have concise semantic names, like the definition given in the work of [6]. Moreover, instead of directly feeding the extracted attribute features into a shopper-specific classifier to make a recommendation, we

This work was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by a grant from the Singapore National Research Foundation and administered by the Interactive & Digital Media Programme Office at the Media Development Authority.

predict whether the handbag is preferred by a shopper via a joint learning of the projection model and one-class classification model of the attributes.

2. HANDBAG RECOMMENDATION

In a recommender system, the problem of predicting whether a product is preferred by the shopper can be modeled as a binary classification task. Shoppers' behaviors such as click or bookmark offer certain positive data [7]. However, those unselected handbags can be either positive data or not. In this paper, we consider certain positive data and the recommendation can be formulated as a one-class classification problem, where one-class SVM [8, 9] is employed.

Original image features may not capture the shareable properties of a shopper's preference due to the diversity of handbags. It is not appropriate to feed original features into a one-class SVM directly. For online shopping, the most direct way to recommend handbags is to return the ones with the same (or similar) attributes as those the shopper clicks. Feed the attributes extracted from positive images into one-class SVM is direct and simple.

However, it is difficult to standardize the attributes labeling and it will require a great effort to label attributes for each handbag. We think that for a certain period of time, there exists some common attributes among the preferred handbags clicked by each shopper. We propose to project features from the original feature space into an attribute space for extracting the common properties of positive data. We jointly learn the projection and one-class SVM classification to build the binary classifier for handbag recommendation. Next, we briefly review the one-class SVM, followed by the introduction of our joint learning algorithm.

2.1. One-class SVM

One-class SVM is a one-class classifier which samples from positive class and aims at finding minimal circumscribing hyperball in a high-dimensional space. If newly encountered data is too far from the learned hypeball, it will be labeled as out-of-class, otherwise it is regarded as in-class. This approach is similar with density estimation [10] because it captures regions where the probability density of the data lives. Given a set of positive data $\{\boldsymbol{x}_i\}_{i=1}^N$, the quadratic program minimization function [8] is:

$$\min_{\boldsymbol{\omega},\xi_i,\rho} \frac{1}{2}||\boldsymbol{\omega}||^2 + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i - \rho, \qquad (1)$$

subject to:

$$(\boldsymbol{\omega}\cdot\boldsymbol{x_i}) \geq \rho - \xi_i, \xi_i \geq 0.$$

where $\boldsymbol{\omega}$ is a weight vector, ρ is an offset parameterizing the hyperball, ν decides the tradeoff between maximizing the margin and the training errors, and slack variables ξ_i allow some data points to lie within the margin.

2.2. Joint learning of attribute projection and one-class SVM classification

In this section, we propose a **J**oint learning of handbag attribute **P**rojection and **O**ne-class SVM classification, which is termed as JPO. We aim to learn a projection that can automatically project the original features onto an attribute space, and meanwhile learn a hyperball to separate the learned attribute features from the origin (negative class). The learned attribute features may not necessarily be semantically meaningful, but they represent some projections which capture the common properties of all features in the positive class. Let $\{\boldsymbol{x}_i\}_{i=1}^N$ be the set of original features extracted from the shopper-clicked images, the objective function is defined as

$$\begin{aligned} C = &\frac{1}{2N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{A}\boldsymbol{x}_i - \boldsymbol{A}\boldsymbol{x}_j||^2 + \\ &\lambda\left(\frac{1}{2}||\boldsymbol{\omega}||^2 + \frac{1}{\nu N}\sum_{i=1}^{N}(\rho - \boldsymbol{\omega}\cdot\boldsymbol{A}\cdot\boldsymbol{x}_i)^2 - \rho\right) + \\ &\frac{\gamma}{2}||\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{I}||^2, \end{aligned} \qquad (2)$$

where the first term,

$$\frac{1}{2N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{A}\boldsymbol{x}_i - \boldsymbol{A}\boldsymbol{x}_j||^2,$$

is the attribute-based projection term. $\boldsymbol{A} \in \mathbb{R}^{M\times H}$ is a learned projection matrix, which projects H-dimensional original features onto M-dimensional attribute features, where $M \ll H$. The second term,

$$\lambda\left(\frac{1}{2}||\boldsymbol{\omega}||^2 + \frac{1}{\nu N}\sum_{i=1}^{N}(\rho - \boldsymbol{\omega}\cdot\boldsymbol{A}\cdot\boldsymbol{x}_i)^2 - \rho\right),$$

is the attribute-based one-class SVM classification. In this term, λ is a scalar that trades off between this term with others. $\{\boldsymbol{\omega},\rho\}$ are weight vector and learned hyperball for classification based on attribute features respectively. ν is a smoothness parameter, and it trades off model losses on the positive data with other components. This term separates the positive data from the origin. We borrow the idea from one-class SVM, but use square loss rather than hinge loss as in (1). The least square model has an exact closed form solution instead of a quadratic programming and it learns the hyperball closely to target values. The component $-\rho$ is to ensure a hyperball far from the origin, which sets an upper bound on the fraction of outliers. And the third term

$$\frac{\gamma}{2}||\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{I}||^2,$$

is the regularization term, which is used to force the projection matrix to be orthogonal. γ is a trade off parameter.

Algorithm 1: JPO: Joint learning of attribute projection model and one-class classification model

Input:
$\boldsymbol{X}$: Positive training data
λ, γ: Trade off parameters for controlling weight of attribute-based projection term, attribute-based one-class SVM classification term and regularization term
ν: Trade off parameters for controlling losses on the positive data
Output:
$\boldsymbol{A}$: Projection matrix
$(\boldsymbol{\omega}, \rho)$: weight vector and parameter for characterizing the hyperball
Initialize $\boldsymbol{A}$ based on (7);
while $\boldsymbol{A}$ *and* $(\boldsymbol{\omega}, \rho)$ *not converge* **do**
    Fix $\boldsymbol{A}$ and solve (3)-(4) by updating $(\boldsymbol{\omega}, \rho)$;
    Fix $(\boldsymbol{\omega}, \rho)$ and solve (6) by updating $\boldsymbol{A}$ via gradient descend: $\boldsymbol{A} \leftarrow \boldsymbol{A} - \beta \frac{\partial C}{\partial \boldsymbol{A}}$;
return $\boldsymbol{A}$, $(\boldsymbol{\omega}, \rho)$;

JPO minimizes two types of errors, one is feature-to-attribute error and the other is attribute-to-preference error. We jointly learn the projection matrix and classifier to make the distances among learned attribute features close, such that the common properties of positive data are captured. On the other hand, the classifier leads to a more proper projection matrix. Meanwhile, the learned projection matrix in turn helps the one-class classifier learning.

Fig. 1. Examples of handbag images in our dataset.

2.3. Optimization

Our objective function is not convex. However, when $\boldsymbol{A}$ is fixed, it is convex regarding $\boldsymbol{\omega}$ and ρ, and when $\boldsymbol{\omega}$ and ρ are fixed, a suboptimal $\boldsymbol{A}$ can be obtained by a gradient descend technique. Hence we employ an alternating optimization algorithm to iteratively update $\boldsymbol{A}$ and $(\boldsymbol{\omega}, \rho)$.

First, we consider the $(\boldsymbol{\omega}, \rho)$-subproblem of (2) with $\boldsymbol{A}$ fixed. This problem has an optimum closed form solution by taking the derivative with respect to $\boldsymbol{\omega}$ and ρ:

$$\frac{\partial C}{\partial \boldsymbol{\omega}} = \lambda \boldsymbol{\omega} + \frac{2\lambda}{\nu N}\left(\boldsymbol{\omega}\boldsymbol{A}\boldsymbol{X}\boldsymbol{I}^T\boldsymbol{X}^T\boldsymbol{A}^T - \rho \boldsymbol{d}\boldsymbol{X}^T\boldsymbol{A}^T\right), \quad (3)$$

and

$$\frac{\partial C}{\partial \rho} = \frac{\lambda}{\nu N}(2N\rho - 2\boldsymbol{\omega}\boldsymbol{A}\boldsymbol{X}\boldsymbol{d}^T) - \lambda. \quad (4)$$

where $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_N) \in \mathbb{R}^{H\times N}$ is a matrix composed of positive training data. $\boldsymbol{I} \in \mathbb{R}^{N\times N}$ is an identity matrix. $\boldsymbol{d} = (1, ..., 1) \in \mathbb{R}^{1\times N}$ is a row vector.

Second, we fix $(\boldsymbol{\omega}, \rho)$ to update $\boldsymbol{A}$, the cost can be rewritten in the matrix form as:

$$C = \frac{1}{N-1}\left(Tr(\boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{X}) - \frac{1}{N}\boldsymbol{d}\boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{X}\boldsymbol{d}^T\right) + \frac{\lambda}{\nu N}\left(N\times\rho^2 + \boldsymbol{\omega}\boldsymbol{A}\boldsymbol{X}\boldsymbol{I}\boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{\omega}^T - 2\rho\boldsymbol{\omega}\boldsymbol{A}\boldsymbol{X}\boldsymbol{d}^T\right) - \lambda\rho + \frac{\gamma}{2}||\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{I}||^2. \quad (5)$$

where $Tr(\cdot)$ denotes the trace of the matrix, and the gradient is written as:

$$\frac{\partial C}{\partial \boldsymbol{A}} = \frac{1}{N(N-1)}\boldsymbol{A}\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^T + \frac{\lambda}{\nu N}2\boldsymbol{\omega}^T\boldsymbol{\omega}\boldsymbol{A}\boldsymbol{X}\boldsymbol{I}\boldsymbol{X}^T - \frac{2\rho\lambda}{\nu N}\boldsymbol{w}^T\boldsymbol{d}\boldsymbol{X}^T + 2\gamma\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{I}\right). \quad (6)$$

where $\boldsymbol{G} = 2N\boldsymbol{I} - 2\boldsymbol{Q}$ and $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ is an all one matrix.

Here, we adopt L-BFGS to optimize our cost function. The elements of the projection matrix $\boldsymbol{A}$ is first initialized with all zeros, except that for those corresponding to the feature dimension with small variance, we set it to one. More specifically, we rewrite the training data matrix according to different feature dimensions: $\boldsymbol{X} = \left(\boldsymbol{z}_1^T, ..., \boldsymbol{z}_H^T\right)^T$. Then we rank $\{\boldsymbol{z}_j\}_{j=1}^H$ according to the ascending order of their variances. We denote the ranking as $R = (r(1), ..., r(H))$ where $r(j) \in \{1, ..., H\}$ indicates the rank of the variance of feature $\boldsymbol{z}_{r(j)}$ is j. After the initialization, the elements in $\boldsymbol{A}$ are zero except that:

$$\boldsymbol{A}(i, r(i)) = 1 \quad \text{for } i = 1, ..., M. \quad (7)$$

At each iteration, we alternatively update $(\boldsymbol{\omega}, \rho)$ and $\boldsymbol{A}$ until convergence. Experimentally, it shows that the algorithm converges within dozens of iterations. The learning algorithm procedure is shown in Algorithm 1.

Fig. 2. Sampled selected handbags from five subjects, where one row indicates handbags clicked by one subject.

For any testing image, we first project its original feature into the attribute feature, then the learned attribute-based one-class SVM is applied to obtain the prediction score.

3. EXPERIMENTS AND ANALYSIS

3.1. Dataset construction

To the best of our knowledge, no existing handbag dataset for recommender purpose is available. In order to compare our JPO algorithm with other approaches, we create a handbag dataset which consists of 835 handbag images from *Amazon.com* with examples given in Fig. 1. The handbags in our dataset vary a lot, due to diverse colors, patterns or styles.

As the preference of each person is different, we asked 30 subjects (with different nationalities, and age ranges from 19-35) to choose handbags that they like from the dataset. To imitate the online shopping scenario, we reshuffle the handbags for each subject and choose a display of 10 images to show each timestep. The subject can leave the system at any timestep. To obtain sufficient images for training and testing, we retain the subjects' data containing at least 15 handbag images and end up with a set of 26 subjects' data.

3.2. Experimental settings

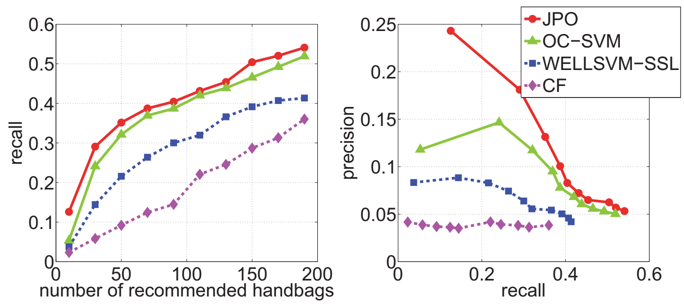
We build a model per user, and among the handbag images clicked from each subject, we randomly sample 10 images as positive data for training. The rest of the dataset is for testing. Fig. 2 shows some clicked handbag images from different subjects. It can be observed that although the varieties of handbags are selected by a subject, they roughly share some common properties, such as the denim hobo bags, plaid pattern, plain pattern, horizontal pattern, and lovely style.

In our experiment, instead of applying feature extraction methods from the whole handbag image, we first adopt a saliency detection technique [11], and then extract the bag-of-words [12] (1st level pyramid only, codebook size 256) of SIFT feature [13] from the bounding box covering top 10 saliency windows. We apply saliency detection because some of the handbags in our dataset are not well-segmented. This 256-dimensional feature vector is regarded as the original feature and fed into all content-based comparison methods.

Among the set of 26 subjects' data, in order to set the parameters, 10 are randomly chosen to form a validation set, and the rest 16 are used as testing data. We use the grid search to find that $\lambda = 4$, $\gamma = 0.1$ and $\nu = 0.1$ give good performance for JPO on the validation dataset, which are applied in our implementation. We also find that by setting $M \ll H$ (such as $M = 2$ compared with $H = 256$), we can obtain good results for the handbag recommendation.

3.3. Evaluation Protocol

We present each subject (one of the 16 subjects for testing) with K handbags in our dataset except for the 10 training handbags, sorted by the subject's predicted scores. As suggested in [1], we evaluate *recall@K* based on which of these handbags were selected by the subject. The definition of *recall@K* [1] is



(a) (b)

Fig. 3. Comparisons of (a) recall curves and (b) precision-recall curves by varying the number of returned handbags over all 16 subjects for testing.

(a)

(b)

Fig. 4. Handbags recommended by (a) proposed JPO and (b) OC-SVM in the top ranks for subject $\#1$.

$$\text{recall}@K = \frac{\text{number of handbags the subject likes in top } K}{\text{total number of handbags the subject likes}} \tag{8}$$

The recall values are averaged for all 16 subjects and plotted for each top K recommended handbags. For each subject, we list the subject's mean recall value for comparison. We also evaluate our algorithm by precision-recall curves as they are widely used in information retrieval and they give an informative picture of the algorithms [14].

3.4. Comparisons

We test and compare different recommender algorithms on the remaining 16 subjects' data, including the content-based methods: proposed JPO, one-class SVM (OC-SVM) and a public semi-supervised learning method [15] (WELLSVM-SSL), since semi-supervised learning is also a good choice for the recommender systems [3]. We also compare with a conventional recommender relying solely on user-item matrix, which is the item-based collaborative filtering (CF) [16]. For OC-SVM, it inputs the original feature from 10 positive images and trains a one-class classifier. In WELLSVM-SSL, we train a semi-supervised model for each subject, where the labeled data is the same as OC-SVM and unlabeled data are original features from all the other $835 - 10 = 825$ images. We apply the cosine similarity and weighted sum to obtain the predictions in CF.

Fig. 3 (a) shows the recall curves for handbag recommendation over all 16 subjects, where we vary the number of recommended handbags $K = 10, 30, ..., 200$. Table 1 lists the average recall for each of the subjects and summarizes the overall results. It can be seen that our method performs the best for most subjects. We show some recommended handbags given by JPO and OC-SVM for subject $\#1$ as an example (see Fig. 4). Compared with the handbags recommended by OC-SVM, those recommended by the proposed JPO are closer with the handbags selected by the subject (first row of Fig. 2). The reason is that OC-SVM may not well capture the common properties of the preferred handbags from merely original features, as the distances among original features are large. While the proposed JPO jointly learn the feature and the classifier, where the distances among newly learned features (potential attributes) are closer and the new classifier is more reliable to differentiate the positive data with others. As expected, CF does not work well due to the small data.

Table 1. Comparisons of mean recall for each subject. Number of recommended handbag is $K = 10, 30, ..., 200$.

Subject #	WELLSVM-SSL	OC-SVM	CF	JPO
1	74.71	87.65	46.47	**88.82**
2	28.33	32.50	0.00	**34.17**
3	8.57	13.21	**19.64**	16.07
4	43.33	45.76	34.85	**50.00**
5	49.09	70.00	25.45	**74.55**
6	31.43	27.86	15.71	**45.71**
7	37.86	22.86	**42.86**	40.71
8	18.57	19.29	15.00	19.29
9	9.09	15.45	2.73	**17.27**
10	55.38	59.62	17.31	**61.54**
11	27.14	**53.57**	13.57	52.14
12	16.00	32.00	18.00	**36.00**
13	20.00	21.82	13.33	**28.18**
14	0.00	3.00	**9.00**	7.00
15	25.00	**55.00**	20.00	52.50
16	12.99	13.64	11.82	**17.66**
Average	28.59	37.08	17.86	**40.10**

In Table 1, the reason for a high recall of subject $\#1$ is that the subject clicks similar types of handbags (Stonewashed Denim Hobo bags), which ends up with a denser distribution of original features and easier-to-learned attributes than other subjects. For other subjects, however, the selected handbags have a certain degree of diversity. We also investigate the learned attribute features for positive training data (i.e., $\boldsymbol{AX}$) of different subjects, and find that usually a higher value in $\boldsymbol{AX}$ yields a higher performance. A high $\boldsymbol{AX}$ indicates high consistency among the appearances of the chosen handbags.

We then vary the number of returned handbags $K = 10, 30, ..., 200$ and draw an average precision-recall curve for all 16 subjects. As shown in Fig. 3 (b), the proposed JPO can always perform better than OC-SVM, WELLSVM-SSL and CF methods, especially when recall is small. The experiment shows an overall low precision, which may be caused by the uncertainty of the unlabeled handbags.

4. CONCLUSION

In this paper, we address the image-based handbag recommendation problem. We treat each shopper's clicked handbags as positive data and analyze the content of those handbag images. As shoppers tend to prefer handbags with similar or the same attributes (e.g., color, pattern or style) for a certain period of time, we propose a joint learning to learn the attribute projection and one-class SVM classification (JPO) together. This joint learning enables the projection of high-dimensional feature space into lower dimensional attribute space. The experimental results show that our method is promising on handbag recommendation.

5. REFERENCES

[1] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011, KDD '11, pp. 448–456.

[2] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, "Combining content-based and collaborative filters in an online newspaper," in *Proceedings of the ACM SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation*. 1999, ACM.

[3] C. Christakou, L. Lefakis, S. Vrettos, and A. Stafylopatis, "A movie recommender system based on semi-supervised clustering," in *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, Nov 2005, vol. 2, pp. 897–903.

[4] W. Carrer-Neto, M. L. Hernández-Alcaraz, R. Valencia-García, and F. García-Sánchez, "Social knowledge-based recommender system. application to the movies domain," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 10990–11000, Sept. 2012.

[5] A. V. D. Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation.," in *NIPS*, 2013, pp. 2643–2651.

[6] F.X. Yu, Liangliang Cao, R.S. Feris, J.R. Smith, and Shih-Fu Chang, "Designing category-level attributes for discriminative visual recognition," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 771–778.

[7] R. Pan, Y. Zhou, B. Cao, N.N. Liu, R. Lukose, M. Scholz, and Q. Yang, "One-class collaborative filtering," in *Data Mining, 2008. ICDM '08. Eighth IEEE International Conference on*, Dec 2008, pp. 502–511.

[8] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, July 2001.

[9] L. Cui and Y. Shi, "A method based on one-class {SVM} for news recommendation," *Procedia Computer Science*, vol. 31, pp. 281 – 290, 2014.

[10] S. Ben-David and M. Lindenbaum, "Learning distributions by their density levels: A paradigm for learning without a teacher," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 171 – 182, 1997.

[11] J. Feng, Y. Wei, L. Tao, C. Zhang, and J. Sun, "Salient object detection by composition," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1028–1035.

[12] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, 2006, CVPR '06, pp. 2169–2178.

[13] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[14] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd International Conference on Machine Learning*, 2006, ICML '06, pp. 233–240.

[15] Y.-F. Li, I. W. Tsang, J. T. Kwok, and Z.-H. Zhou, "Convex and scalable weakly labeled svms," *Journal of Machine Learning Research*, vol. 14, pp. 2151–2188, 2013.

[16] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th International Conference on World Wide Web*, 2001, WWW '01, pp. 285–295.