REFINING IMAGE RETRIEVAL USING ONE-CLASS CLASSIFICATION

Jie Xiao[1], *Yun Fu*[2], *Yijuan Lu*[3], *Qi Tian*[1]

[1]Dept. of Computer Science, University of Texas at San Antonio, TX 78249, USA
[2]BBN Technologies, Cambridge, MA 02138, USA
[3]Dept. of Computer Science, Texas State University, San Marcos, TX 78747, USA

ABSTRACT

Can we take advantage of the huge number of online images to improve image search quality? Motivated by this question, we propose a novel model to re-rank Google image search results by exploring the latent characteristic of massive unrelated images as a clue to filter them in the re-ranking. Inspired by the characteristic of the intrinsic diversity and the unwanted availability of the unrelated images, in our model, we adopt one-class classification to build a hyper-sphere for the target objects, unrelated images, and construct a robust boundary to distinguish them from the related images effectively. Then the Google results can be easily re-ranked by filtering the unrelated images with the built-up model. Extensive experiments demonstrate our approach outperforms Google image search engine's results, even if its baseline is high.

Index Terms— One-class classification, re-ranking, relevance feedback, image retrieval

1. INTRODUCTION

There are billions of online images available to represent the visual world. Can we find images that we really want by exploring a keyword search in such a huge visual dataset? In fact, we are often frustrated by those irrelevant images returned from web search engines based on keyword search. People find visual information conveys much more information than keyword description. However, how to improve image search quality by using their visual information is not an easy issue, because it involves image collection, image representation, image categorization, image ranking based on visual and surrounding text information [1], re-ranking, relevance feedback and many other topics.

Among all of these topics, image representation, image categorization and image re-ranking have drawn much more attention. In image representation, visual words are proposed as analogous to words appearing in documents. [2] [3] treat clustered affine-invariant point descriptors as visual words. Under this model, images are regarded as documents and represented by a histogram of visual words.

In recent years, image re-ranking becomes more and more popular and a lot of work [4, 5] has been devoted to improve the quality of image and video search.

The particular objective of our work is to improve the search quality by filtering unrelated images in re-ranking. In this paper, we focus on exploring the latent characteristic of massive unrelated images as a clue to filter them in re-ranking. One-class classification is an effective solution[6]. Different from two-class or multi-class problem, there is only one target class, the others are outliers. Its goal is to distinguish the target from outliers, which are difficult or expensive to measure. In our case, we treat the massive unrelated images data as target class and build up a hyper-sphere model for it. Then we re-rank the Google images by filtering the unrelated images with the model. Extensive experiments show that our approach improves Google's results.

2. FRAMEWORK

We propose a framework using the unrelated images returned by image search engine to re-rank the images based on one-class classification (OCC) model. It also supports relevance feedback, a good way to improve the precision.

2.1. One-class classification

Against two-class or multi-class classification, one-class classification just studies the pattern for the target class. Even if the unrelated data are diverse and massive, there are still some clues to reject them. In our work, we treat the unrelated images as target data for a given keyword search. And we try to distinguish them from the related data, the outliers in our one-class classification model.

For one-class classification, the goal is to detect the small percent of data (outliers) against the large number of target data. Several approaches including novelty detection [7], outlier detection [8] or concept learning [9] are proposed to solve the detecting problem for the imbalanced data. Common solutions resort to density evaluation, fitting the data into a statistical distribution to the target data, but it depends on whether the sample size is sufficient and whether we choose an appropriate distribution to the data. To avoid dependency in the case of limited amount of data, a good solution is to use support vector data description (SVDD) [6,10,11], which fixes the boundary by minimizing the volume of hypersphere rather than evaluating the density.

2.2. Kernel whitening

The performance of one-class classifier depends on the scaling of data and is harmed by data distribution in (nonlinear) subspace. To obtain good data representation, Kernel PCA [12] is used to get the non-linear principle components, which outperform the same number of linear principle components, and then they are rescaled by the corresponding eigenvalues. After transforming, data is represented with equal variance [13] in each feature direction with eigenvalue larger than zero.

Assume a set of data $\mathbf{X}$ is mapped to the kernel space F by some mapping Φ: $\mathbf{x}_k \rightarrow \mathbf{y}_k \in$ F, and the transformed data has zero mean. $\sum \Phi(\mathbf{x}_k) = 0$ for k=1, …., M. We use Gaussian kernel to form the kernel matrix $\mathbf{K}_{ij} = (K(x_i,x_j))_{ij}$. Solve the eigenvalue problem with

$$\lambda\, \boldsymbol{\alpha} = \mathbf{K}\, \boldsymbol{\alpha} \qquad (1)$$

by diagonalizing $\mathbf{K}$ and normalize the eigenvector expansion coefficients α^n by requiring $\lambda_n^2\ (\alpha^n \cdot \alpha^n) = 1$. For normal kernel PCA the eigenvectors should be normalized to unit length. It is to rescale the data in the kernel space with unit variance. A new object $\mathbf{z}$ can be mapped onto eigenvector by

$$(\mathbf{z})_n = \sum \alpha_i^n K(\mathbf{x}_i \cdot \mathbf{z}),\ i = 1,\ldots,M \qquad (2)$$

where $(z)_n$ refers the n-th component of vector $\mathbf{z}$.

2.3. Support vector data description

Support vector data description (SVDD) is a method to fit the closed boundary for a class by minimizing the volume of the target data without requiring knowledge of density evaluation. The objects inside the boundary are classified as target objects while others are treated as the outliers. We enclose the data by a hyper-sphere with minimum volume. Denote the center by α and radius by R. Instead of retricting the distance from kernel feature space object $\mathbf{y}_i$ to the center $\boldsymbol{\alpha}$ strictly smaller than R^2, we penalize the larger distance. An error function is built by allowing the training set probably containing a few outliers to make it more robust. The error contains two parts: the structural error and the empirical error. After optimization, we have

$$\mathbf{L} = \sum_i \beta_i\, (\mathbf{y}_i\, \mathbf{y}_i) - \sum_{ij} \beta_i \beta_j\ (\mathbf{y}_i\, \mathbf{y}_j\,). \qquad (3)$$

where β_i is associated with each object $\mathbf{y}_i$. For more details please refer to [10]. An object with $\beta_i > 0$ can be on the boundary or outside the boundary. Since the center $\boldsymbol{\alpha}$ can be represented as the linear combination of objects with weights β_i, only the objects with $\beta_i > 0$ are used to describe the hyper-sphere center and the radius R, and they are called support objects or support vectors. A new object z is classified as target object if

$$f(z) = \left\| \mathbf{z} - \boldsymbol{\alpha} \right\|^2 = (\mathbf{z} \cdot \mathbf{z}) - 2\sum_{k,j=1}^{M} \beta_k \beta_j\ (\mathbf{y}_k \cdot \mathbf{y}_j\,) \le R^2 \qquad (4)$$

otherwise, it is classified as outlier.

2.4. Relevance feedback

Our framework supports the relevance feedback from users. The feedback images are added into our training dataset to learn the model. Since our model is learned to distinguish the related data by using the diverse unrelated data, the more diverse the unrelated data, the more confident prediction the model will produces. So the feedback from users will make our model more robust.

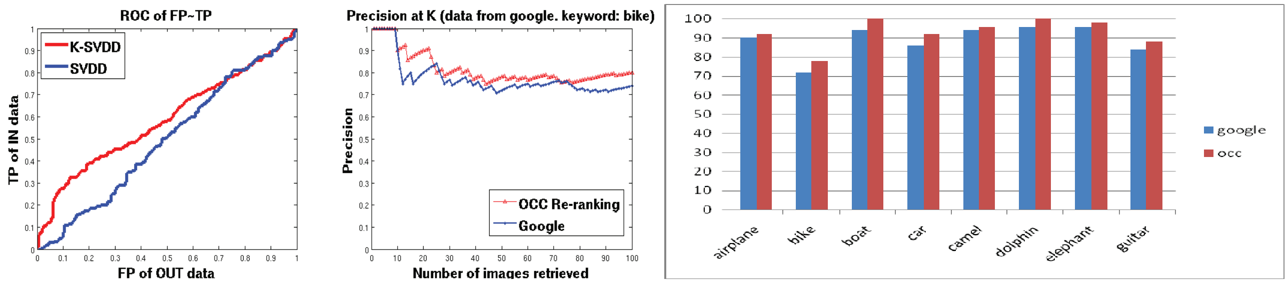
2.5. Our algorithm

The algorithm framework is: (1) Crawl all available images from public image search engines, such as Google and Yahoo and represent them as histogram over codebook. (2) Label the images as completely related, partially related and unrelated. (3) Form training data by randomly sampling unrelated images. (4) Do Kernel whitening on training data and fix the boundary with SVDD. (5) Test with a small number of related and unrelated images. (6) Apply the model to filter the unrelated images from the returned images.

3. EXPERIMENTS

3.1 Data preparation

We crawl top 1000 images from Google and Yahoo! image search engines for a group of keywords and filter the abstract images [1], such as drawings, non realistic paintings, comics etc. Images are rescaled to the same width. Interest points are detected by hessian-affine method [14] and presented by 128 dimension SIFT [15] feature. We generate a hierarchy vocabulary tree [16] using k-means for over 297k features sampled from [1]'s dataset. And then quantize the feature vectors as a histogram over the codebook for each image. Images are labeled as completely related, partially related and unrelated. The completely related images contain the whole object without major occlusion; the partially related images contain parts of the object, or the object only takes up a small percentage of space in the whole image; the unrelated images don't contain the object for the given topic.



(a) ROC Curve for OCC (b) precision of OCC (c) comparison of eight topics on precision at 50

Fig. 1 TP-FP ROC for OOC model and precision in re-ranking. Comparison between OCC and Google image search engine.

Table 1 Precision for top 50 results of eight topics: airplane, bike, boat, car, camel, dolphin, elephant and guitar.

topic	airplane (%)		bike (%)		boat (%)		car (%)		camel (%)		dolphin (%)		elephant (%)		guitar (%)	
top	google	occ	google	occ	google	occ	google	occ	google	occ	google	occ	google	occ	google	occ
5	80	100	100	100	100	100	100	100	100	100	100	100	80	100	100	100
10	90	100	90	90	100	100	90	90	100	100	100	100	90	100	100	100
20	95	100	80	90	100	100	90	90	95	95	90	100	95	95	95	100
30	90	93.3	76.7	80	96.7	100	90	90	96.7	96.7	93.3	100	93.3	96.7	90	93.3
40	92.5	95	75	77.5	97.5	100	87.5	90	92.5	95	95	100	95	97.5	87.5	95
50	90	92	72	78	94	100	86	92	94	96	96	100	96	98	84	88
100	68	68	74	80	92	94	84	83	91	91	93	92	93	94	74	79

3.2 One-class classification

Given a search word, we build our one-class classification model using 50 unrelated images from the search engine as target data. Although 50 images are not a big dataset to represent the diverse nature of visual world, the experiment results show that with this limited training dataset, the model still exhibits a good performance in separating the target objects from the outliers. The model has good potential in a real-world web-based image search, where massive and more diverse unrelated data is sufficient to build it.

In our experiments, we compare the performance of two models with different data representation: Kernel whitening + SVDD (K-SVDD) and SVDD only. The ROC curve (Figure 1(a)) shows that the model trained with K-SVDD outperforms the other one. With kernel-whitening, data has unit length in all variance direction, which makes it more likely to fix a hypersphere boundary. For the latter one, the fixed boundary is apt to be harmed by the data distribution.

3.3 Re-ranking

Applying the model built in section 3.2, we predict the target objects by our one-class classifier for the top 500 images on eight different topics from Google image search engine (shown in Table 1). We focus on improving the precision for the very top results, like top 50, which receives more attention in reality. We re-rank the Google's result by filtering the predicted unrelated images. There is no doubt that Google has high precision for the top results. For example, its precision on top 50 images is 89% on average for eight topics listed in Table 1, which leaves us a limited

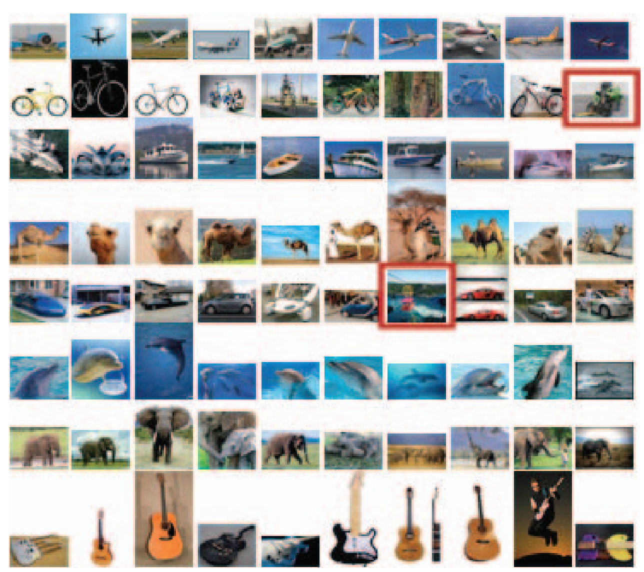


Fig. 2 Top 10 ranked images of airplane, bike, boat, camel, car, dolphin, elephant and guitar based on our OCC re-ranking model. The red boxes indicate the false positives.

improvement space. Even if in this case, our model still can make improvement for the top 50 results. For example, on topic bike, our model improves the precision by 10%, 6% and 6% on top 20, 50 and 100 images respectively as shown in Figure 1(b). And on topic boat and dolphin, our model even achieves 100% while Google has 94% and 96% precision respectively (Figure 1(c)). Due to the small number of images used in building the model, it is reasonable to see Google sometimes has a better precision for the top 100 results.

Note that, we label images into three classes: completely related, partially related and unrelated. And we use the unrelated data to distinguish them from the others. By definition, bike refers to bicycle not motorbike, which makes the 10th image as a false-positive, while the 7th image is regard as partially related because in that image there are parts of bike appearing in trees (Figure 2, 2nd row).

3.4 Relevance feedback

In our framework, we also allow users to pick out some unrelated images they think, and add them to our training dataset to learn the model and apply it to re-rank the Google's image search results.

In Figure 3, the precision is improved by allowing the user to pick 10 unrelated images for further training. The OCC feedback method has the highest precision for the top 50 retrieved images, which makes it possible to work on web, because users always pay much more attention to the very first few pages, and ignore the rest in web search.

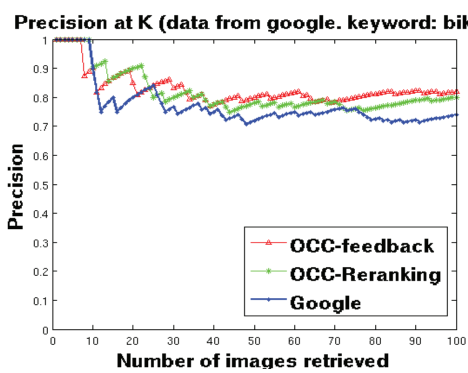


Fig. 3 Comparison on precision for OCC with Feedback, OCC re-ranking and Google image search engine.

4. CONCLUSION AND DISCUSSION

In this paper, we propose a novel model to re-rank the Google image search results by exploring the latent characteristic of the unrelated images as a clue to filter them in re-ranking. Our approach provides a potential practical application to filter the unrelated images in web image search. Even if the data used to train the model is small, which is not diverse enough to describe everything, the one-class classifiers with kernel-whiten and SVDD still show a good performance.

The contribution of this paper can be highlighted as follows: (1) We explore the unrelated data as a clue to distinguish them from the related data. It is helpful when we are lack of clean data and full with contaminated unrelated data. (2) We prove that for one-class classification model, a few new added unrelated data is helpful to fix the boundary to distinguish the outliers in web images search. (3) We apply one-class classification in the web-based image search. Its good performance in image re-ranking and relevance feedback makes it possible to improve the search quality of web image search engine, especially for the search of non-popular or non-typical categories or topics.

In the future, we will consider the pseudo relevance feedback to avoid human interference, which would make it more practical to improve the quality in web image search engine and learn the model automatically.

5. REFERENCES

[1] F. Schroff, A. Criminisi, and A. Zisserman, "Harvesting Image Databases from the Web," *IEEE ICCV*, pp. 1-8, 2007.

[2] C. Dance, J. Willamowski, L.X. Fan, C. Bray, G. Csurka, "Visual categorization with bags of keypoints," *ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1-22, 2004.

[3] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," *IEEE ICCV*, pp. 1470-1477, 2003.

[4] L. S. Kennedy and S. F. Chang, "A reranking approach for context-based concept fusion in video indexing and retrieval," *ACM CIVR*, pp. 333-340, 2007.

[5] W. H. Hsu, L. S. Kennedy, and S.-F. Chang, "Video search reranking through random walk over document-level context graph," *ACM Multimedia*, pp. 971-980, 2007.

[6] Z. Zeng, Y. Fu, Y. Hu, T.S. Huang, et al., "One-Class Classification for Spontaneous Facial Expression Analysis," *IEEE FG*, pp. 281-286, 2006.

[7] G. Ritter, M.Gallegos, "Outliers in statistical pattern recognition and an application to automatic chromosome classification," *PR Letters*, vol. 18, pp. 525-539, 1997.

[8] C. Bishop, "Novelty detection and neural network validation," *IEE Proc. on Vision, Image and Signal Processing. Special Issue on Applications of Neural Networks*, pp. 217-222, 1994.

[9] N. Japkowicz, C. Myers, M. Gluck, "A novelty detection approach to classification," *Proc. of the Fourteenth International Joint Conference on Artificial Intelligence*. pp. 518-523, 1995.

[10] D. Tax, "One-class classification," Ph.D. thesis Delft university of Technology, 2001.

[11] X.D.Yu, D. DeMenthon, D. Doermann, "Support Vector Data Description for Image Categorization from Internet Images," *ICPR*, 2008

[12] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp.1299–1319, 1998.

[13] D. Tax, P. Juszczak, "Kernel whitening for one-class classification," *Int'l Journal of Pattern Recognition and Artificial Intelligence*, vol. 17, no. 3, pp. 333-347, 2003.

[14] http://www.robots.ox.ac.uk/~vgg/research/affine/detectors.html

[15] D. G. Lowe, "Object recognition from local scale-invariant features," *IEEE ICCV*, Corfu, Greece, pp. 1150-1157, 1999.

[16] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree, " *Proc. of CVPR*, vol. 2, pp. 2161-2168, 2006.