2009

Image retrieval using noisy query

Jun Zhang
University of Wollongong

Lei Ye
University of Wollongong, lei@uow.edu.au

Image retrieval using noisy query

Abstract

In conventional content based image retrieval (CBIR) employing relevance feedback, one implicit assumption is that both pure positive and negative examples are available. However it is not always true in the practical applications of CBIR. In this paper, we address a new problem of image retrieval using several unclean positive examples, named noisy query, in which some mislabeled images or weak relevant images present. The proposed image retrieval scheme measures the image similarity by combining multiple feature distances. Incorporating data cleaning and noise tolerant classifier, a two-step strategy is proposed to handle noisy positive examples. Experiments carried out on a subset of corel image collection show that the proposed scheme outperforms the competing image retrieval schemes.

Disciplines

Physical Sciences and Mathematics

Publication Details

Zhang, J. & Ye, L. (2009). Image retrieval using noisy query. IEEE International Conference on Multimedia and Expo, 2009. ICME 2009 (pp. 866-869). Piscataway, USA: IEEE.

IMAGE RETRIEVAL USING NOISY QUERY

Jun Zhang and Lei Ye

School of Computer Science and Software Engineering,
University of Wollongong,
Wollongong, NSW, 2522 Australia

ABSTRACT

In conventional content based image retrieval (CBIR) employing relevance feedback, one implicit assumption is that both pure positive and negative examples are available. However it is not always true in the practical applications of CBIR. In this paper, we address a new problem of image retrieval using several unclean positive examples, named noisy query, in which some mislabeled images or weak relevant images present. The proposed image retrieval scheme measures the image similarity by combining multiple feature distances. Incorporating data cleaning and noise tolerant classifier, a two-step strategy is proposed to handle noisy positive examples. Experiments carried out on a subset of Corel image collection show that the proposed scheme outperforms the competing image retrieval schemes.

Index Terms— Content based image retrieval, noisy query, data cleaning, noise tolerant classifier

1. INTRODUCTION

Content based image retrieval (CBIR) is a technique to search for images relevant to the user's query from an image collection, which has got much attention in the last decade. Conventional CBIR schemes employing relevance feedback have achieved certain success. However some disadvantages hinder the practical applications of relevance feedback. First, if not impossible, ordinary users have little patience to persist in the feedback iterations. Especially on the web few people use advanced search interfaces and most would like to complete their search in a single interaction [1]. Second, the manually labeled negative examples may not offer sufficient variety. The previous experiments have shown that a large set of random negative examples is often better than a small set of hand-picked negative examples [2]. Third, most existing retrieval schemes fail to address the problem of noisy examples. Some noisy examples may be present since ordinary users normally have no expertise in constructing a high quality query. In this paper, we address a new problem of image retrieval using several unclean positive examples, named noisy query, in which some mislabeled images or weak relevant images present. Fig.1 gives an example of noisy query

Fig. 1. Noisy query: "Beach"

which is for retrieving "Beach" images. The images labeled by tick are relevant to beach. The image labeled by cross is irrelevant to beach. And the image labeled by circle is weakly relevant to beach. Under this circumstance, user provides several unclean positive example images as a query and CBIR system will return the relevant images from an image collection in a single interaction. The solution of this problem is useful for the applications of CBIR in which relevance feedback is not a suitable choice but several unclean positive example images can be provided.

In this paper, image similarity is obtained by combining multiple feature distances measured in different feature spaces (named feature aggregation). Instead of existing heuristic methods [3], we propose a new way to perform feature aggregation. Given a query image, a new feature similarity space is constructed in which images are represented using the feature distances. Then feature aggregation can be formulated as a classification problem and solved by conventional classification technologies. To handle the noisy positive examples, a new two-step strategy is proposed by incorporating data cleaning and noise tolerant classifier. In step 1, an ensemble of classifiers are constructed in a feature dissimilarity space corresponding to a reliable positive example which are used as consensus filters to identify and eliminate the mislabeled positive examples [4]. In step 2, each retained positive example is associated with a relevance probability to further alleviate the noise influence [5]. Multiple ensembles of classifiers are trained in the feature dissimilarity spaces corresponding to the retained positive examples, which are then combined to get the final scores for ranking images in the collection. Experiments carried out on a sub set of Corel image collection show that the proposed scheme outperforms the competing image retrieval schemes.

2. PROPOSED SCHEME

Let us consider an image collection $\mathcal{I}$ containing N images. The feature representation of an image *I* is a set of m feature vectors, $\{F^k\}_{k=1}^m$. Assume a user provides some positive example images as a query, $\mathcal{P} = \{P_i\}_{i=1}^{n_p}$. In Section 2.1, we propose a new feature aggregation method based on classification technology in which noisy positive examples are not considered. A new two step strategy for handling noisy positive examples is described in Section 2.2 and 2.3, which includes noise identification and elimination, and noise tolerant relevance calculation.

2.1. Classification based feature aggregation

Given a query image $P \in \mathcal{P}$, we use it as a prototype and construct a feature dissimilarity space by modifying the method proposed by Duin and Pekalska [6]. For each collection image I_i, we have

$$S_i = (s_{i1}, s_{i2}, \ldots, s_{im}), \qquad (1)$$

where s_{ij} represents the dissimilarity between I_i and P on the jth feature, and S_i is a vector in a m-dimensional space $\mathcal{S}$, called feature dissimilarity space. In this paper the dissimilarity is defined by a feature distance. We denote $D_j(\cdot,\cdot)$ as a specified distance metric for the j-th visual feature, then $s_{ij} = D_j(F_{ij}, F_{Pj})$. Therefore, similarities between all images in $\mathcal{I}$ to P are represented by matrix with size $N \times m$.

The key difference between the feature dissimilarity space and conventional dissimilarity space [6] is that the feature dissimilarity space is introduced to address the feature aggregation problem which has only one prototype, while the conventional dissimilarity space has multiple prototypes selected by the system designer. Compared with original feature space, the feature dissimilarity space inherits the advantages of the dissimilarity space. Sometimes it is difficult to construct a combined feature space with a unified distance metric for multiple features, but we always can construct the feature dissimilarity space [6].

In feature dissimilarity space, feature aggregation can be formulated as a binary classification problem. The positive class consists of relevant images to the query and the negative class consists of all irrelevant images. Such that feature aggregation can be solved in a principle way based on conventional classification technologies. In this paper we choose support vector machine (SVM) algorithm [7] to design the specific binary classifier. The SVM based retrieval scheme has shown promising results as SVM has good generalization and noisy tolerant ability which is well suitable to our retrieval scheme. In this way, the decision value produced by SVM, $f(x)$, is used as the result of feature aggregation, i.e., combined similarity between a collection image and a query image. If a linear kernel is used in SVM, it will lead to a linear feature aggregation method. If a non-linear kernel is used in SVM, it will lead to a non-linear feature aggregation method.

Table 1. Consensue filtering

Input: query image set $\mathcal{P}$, SVM classifier, integer T (the number of bagging classifiers), and the image collection $\mathcal{I}$
1. P_o = select prototype from $\mathcal{P}$
2. $\mathcal{S}$ = construct feature dissimilarity space for P_o
3. For i=1 to T {
4. $\mathcal{N}_i$ = random sampling from $\mathcal{I}$, with $\lvert\mathcal{N}_i\rvert = n_p$
5. C_i = SVM($\mathcal{P}, \mathcal{N}_i$)
6. }
7. $\bar{\mathcal{P}}$ = consensus filters $\{C_i(\mathcal{P})\}$
Output: $\bar{\mathcal{P}}$

2.2. Noise identification and elimination

Before computing the relevance of an image to the query, we try to eliminate some mislabeled positive images by constructing an ensemble of SVM classifiers as consensus filters [4] in feature dissimilarity space. Enlightened by the k-medoids approach [8], we choose an actual positive image as the prototype which is more robust to the noisy examples than random selection and average point of positive examples through our studies. The strategy can be represented as

$$P_o = \arg\min_{P \in \mathcal{P}} \sum_{i=1}^{n_p} \sum_{j=1}^{m} s_{ij}^{(p)}, \qquad (2)$$

where $s_{ij}^{(p)}$ is the distance of ith positive examples P_i to a candidate center P on the jth feature. In this strategy the positive example with the minimum sum of distances between it and all positive examples will be chosen as the prototype.

In the feature dissimilarity space, the positive examples and an equal number of negative examples randomly selected from image collection are used to train SVM. Since some noisy examples are present, the SVM classifier will be unstable. We use different negative example sets to train SVM and get multiple classifiers. A similar strategy, named asymmetric bagging, has been used in [9], which can effectively handle the unstable and unbalanced classifiers. After that, we apply all classifiers to classify all positive examples. The examples labeled by all SVM classifiers as negative will be identified as mislabeled examples and eliminated. The consensus filtering algorithm is summarized in Table 1.

2.3. Noise tolerant relevance calculation

The consensus filters are suitable to a paucity of data since they are conservative at throwing away good data at the expense of retaining bad data [4]. Motivated by the idea of noise tolerant classifier [5], we associate with each retained positive example a relevance probability to alleviate the influence of retained noisy positive examples. To estimate the relevance probability of an retained positive example using

Table 2. Relevance Calculation

Input: filtered query image set $\bar{\mathcal{P}}$, SVM classifier,
integer T (the number of bagging classifiers),
and the image collection $\mathcal{I}$
1. For i=1 to $|\bar{\mathcal{P}}|$ {
2. $\quad \mathcal{S}^i$ = create feature dissimilarity space for $\bar{P}_i \in \bar{\mathcal{P}}$
3. $\quad$ For j=1 to T {
4. $\quad\quad \mathcal{N}_{ij}$ = random sample from $\mathcal{I}$, with $|\mathcal{N}_{ij}| = |\bar{\mathcal{P}}|$
5. $\quad\quad C_{ij} = \text{SVM}(\bar{\mathcal{P}}, \mathcal{N}_{ij})$
6. $\quad$ }
7. }
6. $C^*(\mathcal{I}) =$ classifier combination $\{C_{ij}(\mathcal{I})\}$

Output: C^*

a small number of samples, we propose a probability estimation algorithm based on ensemble method. This algorithm also can be seemed as a by-product of consensus filtering presented in last Section. For a retained positive example E, the sigmoid function combined with the output of SVM can be used to estimate the class-conditional probability [10] by

$$P\left(L_k|C_i, E\right) = \frac{1}{1 + \exp\left(-\left|f_i(E)\right|\right)}, \tag{3}$$

where $f_i(E)$ is the decision value produced by the ith SVM classifier C_i and L_k is a class label. L_0 and L_1 denote positive and negative class, respectively. After that we apply these SVM classifiers to classify all retained positive examples. All classifiers are then combined to get the conditional probabilities based on Bayes Sum Rule (BSR),

$$P\left(R|E\right) = \frac{1}{T}\sum_{i=1}^{T} P\left(L_0|C_i, E\right), \tag{4}$$

where $P\left(R|E\right)$ is the relevance probability.

In the proposed scheme, the similarity between a collection image and a specific positive example is measured by an ensemble of SVM classifiers. Each filtered positive example is associated with a probability that it is relevant to the query. For ranking images, we combine multiple ensembles of SVM classifiers to get the relevance of a collection image to the user's query. The relevance calculation algorithm is summarized in Table 2. Three classifier combination models [11] are evaluated in this paper.

SVM-Weighted-MVR: For a given image, we first use the weighted Majority Vote Rule (MVR) to recognize it as query relevant or irrelevant, which can be represented as follows:

$$C^*(I) = \text{sgn}\left[\sum_{i,j} w_i \cdot C_{ij}(I) - \frac{T\sum_i w_i}{2}\right] \tag{5}$$

where I is a collection image, $C_{ij}(\cdot)$ is the jth classifier trained in the feature dissimilarity space corresponding to the ith retained positive example, and w_i is the weighting assigned to this positive example. In this paper, $w_i = P\left(R|E\right)$ represents the relevance of an positive example. Then, we measure the relevance between the image and the query as the output of the individual SVM classifier, which gives the same label as the weighted MVR and produces the highest weighted confidence value (the weighted absolute value of the decision function of the SVM classifier).

SVM-Weighted-BSR: For a given image, we first use the weighted BSR to recognize it as query relevant or irrelevant. The weighted BSR can be represented as follows:

$$C^*(I) = \arg\max_k \left[\sum_{i,j} w_i \cdot P\left(L_k|C_{ij}, I\right)\right] \tag{6}$$

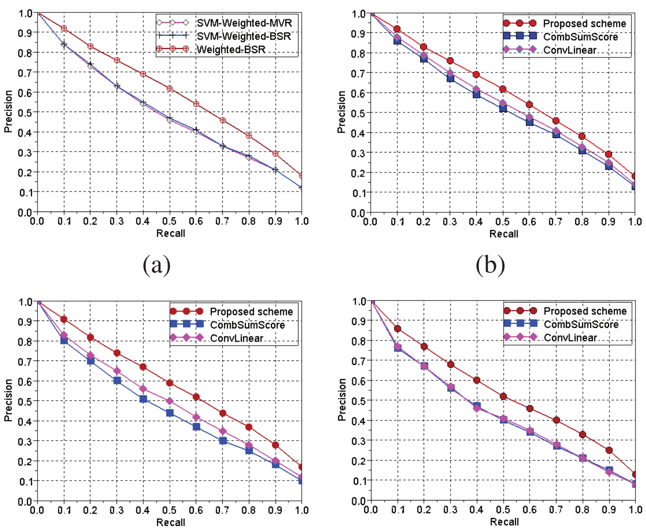
where $P\left(L_k|C_{ij}, I\right)$ represents the class-conditional probability which can be computed by Eq.(3). Then, we measure the relevance between the image and the query using the individual SVM classifier, which gives the same label as the weighted BSR and has the highest weighted confidence value.

Weighted-BSR: The output of the weighted BSR, $\sum_{i,j} w_i \cdot P\left(L_k|C_{ij}, I\right)$, can be directly used as a relevance measure between a given image and the query.

3. EXPERIMENTS & RESULTS

The experiments are carried out on a subset of Corel image collection, which consists of 20 image class and each image class includes 100 Corel images. The images in one class have the same perceptual meaning, so the ground truth is based on the image class. Two state-of-the-art feature aggregation based retrieval schemes are implemented for comparison. One is the CombSumScore scheme which is the best one in all schemes using fixed aggregation function evaluated by Donald *et.al.* [3]. The other is an enhanced version of CombSumScore, named ConvLinear, which uses the linear weighting method proposed by Rui [12] to combine multiple feature distances. Five standardized MPEG-7 visual descriptors and the recommended distance metrics [13] are used in the experiments. The retrieval performance in terms of average precision and recall on 400 random queries are reported. $T = 5$ for both of Tables 1 and 2 is based on experimental results. Considering normally ordinary users can provide only a small number of examples, each query includes 5 images.

Fig.2(a) shows the retrieval performance of the proposed scheme using different classifier combination models. In the experiments, no mislabeled positive examples are introduced and negative examples are obtained from the image collection randomly. We can see that Weighted-BSR outperforms both SVM-Weighted-BSR and SVM-Weighted-MVR significantly. The reason may be that Weighted-BSR can combine the outputs of all weak SVM to get a more confidently decision score for relevance measurement, while no best individual SVM can be used to measure the image relevance. Fig.2(b)-(d) show the retrieval performance using unclean queries. The mislabeled positive examples are manually introduced to highlight the noise influence. In the proposed scheme we choose Weighted-BSR combination model based on previous experimental results. Fig.2(b) is for query without mislabeled positive examples. Fig.2(c) and (d) are for query with 1 or 2 mislabeled positive examples,respectively. The results show that the proposed scheme outperforms ConvLinear and CombSumScore. The reason is the proposed scheme can handle noisy positive examples while other schemes can't.



(a) (b) (c) (d)

Fig. 2. Image retrieval performance

4. CONCLUSIONS

In this paper, we addressed a new problem of image retrieval using noisy query. In the proposed scheme, feature aggregation is addressed in a new way using classification technologies. Incorporating data cleaning and noise tolerant classifier, a new two-step strategy is proposed to handle the noisy positive examples. The preliminary experimental results show: (1) classifier combination model can affect the retrieval performance significantly. Weighted-BSR is better than SVM-Weight-BSR and SVM-Weight-MVR when a small number of examples are available. (2) the proposed image retrieval scheme can handle noisy positive examples effectively, which outperforms CombSumScore and ConvLinear schemes significantly under noisy circumstance.

5. REFERENCES

[1] C. D. Manning, P. Raghavan, and H. Schtze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[2] K. Tieu and P. Viola, "Boosting image retrieval," *International Journal of Computer Vision*, vol. 56, no. 1/2, pp. 17–36, 2004.

[3] K. M. Donald and A. F. Smeaton, "A comparison of score, rank and probability-based fusion methods for video shot retrieval," in *Proc. of IEEE Int. Conf. on Image And Video Retrieval*, Singapore, Jul. 2005, pp. 61–70.

[4] C. Brodley and M. Freidl, "Identifying mislabeled training data," *J. Artif. Intell. Res.*, vol. 11, pp. 131–167, 1999.

[5] N. D. Lawrence and B. Schlkopf, "Estimating a kernel fisher discriminant in the presence of label noise," in *Proc. of the 18th Int. Conf. on Machine Learning*, Morgan Kaufmann, San Francisco, USA, 2001, pp. 306–313.

[6] E. Pezkalska and R. P. Duin, "Dissimilarity representations allow for building good classifiers," *Pattern Recognition Letters*, vol. 23, no. 8, pp. 943–956, Jun. 2002.

[7] J. C. Burges, "A tutorial on support vector machine for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[8] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, pp. 85–126, 2004.

[9] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088–1099, Jul. 2006.

[10] J. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods," in *Proc. Advances in Large Margin Classifiers*, 2000, pp. 61–74.

[11] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons, July 2004.

[12] Y. Rui and T. S. Huang, "Optimizing learning in image retrieval," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1, Hilton Head Island, South Carolina, Jun. 2000, pp. 1236–1243.

[13] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.