Active learning based on maximizing information gain for content-based image retrieval

Xu Jie Shi Pengfei

(School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: This paper describes a new method for active learning in content-based image retrieval. The proposed method firstly uses support vector machine (SVM) classifiers to learn an initial query concept. Then the proposed active learning scheme employs similarity measure to check the current version space and selects images with maximum expected information gain to solicit user's label. Finally, the learned query is refined based on the user's further feedback. With the combination of SVM classifier and similarity measure, the proposed method can alleviate model bias existing in each of them. Our experiments on several query concepts show that the proposed method can learn the user's query concept quickly and effectively only with several iterations.

Key words: active learning; content-based image retrieval; relevance feedback; support vector machines; similarity measure

Content-based image retrieval (CBIR) has become one of the most active research areas in the past few years. At the early stage of CBIR, research primarily focused on exploring visual feature representations, evaluating distance metrics and efficient searching schemes. While these research efforts establish the basis of CBIR, the major obstacle facing the proposed approaches is the gap between high-level query concepts and low-level features. A promising approach to this problem is online learning technique.

Initiated in the document retrieval field, relevance feedback as an on-line learning mechanism is adopted in the image retrieval system[1]. In such an interactive system, a user gives the system feedbacks on which of the images returned by the system are relevant to the current query. Then a learning algorithm automatically adjusts the query using the user's feedback in each iteration such that the adjusted query is a better approximation of the user's query.

However, since most users may not be patient enough to provide endless feedback, the labeled training sample set from the user's query and feedback may be very small relative to the dimension of the feature space, and pure supervised on-line learning from such a small training data set will have poor generalization performance. This makes it an obvious target for active learning, where the learning program asks for labels only on the items that will most help the learning, resulting in fewer examples being used as compared to supervised learning.

The main issue in active learning is to find a way to choose the most useful examples to ask a user to label; some related studies have been conducted mostly in the context of machine learning. Query by committee (QBC)[2-4] is a general approach to active learning first proposed by Seung, et al. The method queries an example based upon the degree of disagreement between the committee of classifiers. Dagan and Engelson[5] proposed a similar method, termed committee-based sampling. While Cohn[6] proposed selective sampling method to choose for labeling the instance that the current classifier is most uncertain about. Lewis and Gale[7] also developed a similar method, called uncertainty sampling for text categorization. Of late, uncertain-sampling-like methods based on support vector machines (SVMs) have been proposed[8-10].

However, active learning technique is still an open issue for future research. In this paper, a new active learning for CBIR is proposed. The method uses the expected information gain to signal the need for requesting the actual value of each example's label from the user. To alleviate the model bias existing in the learners, the method attempts to integrate two very different learning models into the active learning framework, SVM classifier and similarity measure. The experimental results show that the proposed method can converge to the current user's query concept acutely and quickly only after several iterations.

Received 2003-11-26.

Foundation item: The National Natural Science Foundation of China (No. 69931010).

Biographies: Xu Jie (1975—), female, graduate; Shi Pengfei (corresponding author), male, professor, pfshi@sjtu.edu.cn.

1 SVMs for CBIR

CBIR system uses the visual content of the images, such as color, texture, and shape features, as the image index. Each image is transformed into a point in the feature space and the learning algorithm is applied in this space. Traditional learning methods in relevance feedback considered image retrieval as a problem of similarity comparison between images, which can be expressed as a weighted linear combination of similarities in features. While some recent learning systems assume that query concepts can be learned through binary class. However, such learning work suffers from model bias. When a query concept does not fit the model assumption, these systems perform poorly. We propose to alleviate model bias through the combination of these two different kinds of models into an active learning framework, where the images with maximum expected information gain are selected to query the user for the true class labels.

Firstly, SVM classifiers are used to learn an initial target concept by separating the relevant images from the irrelevant ones with a hyperplane in a projected space H. SVMs[11] are powerful tools for data classification. They are designed to minimize structural risk so that they are less vulnerable to overfitting problems and more suitable for learning from small training sets than techniques based on minimization of empirical risk.

Consider SVMs in the binary classification setting for CBIR. Given training examples $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ in image feature space $F \subseteq \mathbf{R}^d$, and their labels $\{y_1, y_2, \cdots, y_n\}$, where $y_i \in \{-1, 1\}$ stands for whether the image is a positive or negative instance to the query concept. SVMs map the original training data in space F to a higher dimensional space H via a Mercer kernel operator K where it may become linearly separable. Then the form of the set of the classifiers is like this:

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \boldsymbol{\alpha}_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) \tag{1}$$

For convenience, it is assumed that there is no bias weight. When K satisfies Mercer's condition, it can be written as: $K(u, v) = \boldsymbol{\Phi}(u) \cdot \boldsymbol{\Phi}(v)$ where $\boldsymbol{\Phi}: F \to H$ and " · " denotes an inner product. Then f is rewritten as

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \boldsymbol{\alpha}_i y_i \boldsymbol{\Phi}(\boldsymbol{x}_i)\boldsymbol{\Phi}(\boldsymbol{x}) = y_i \boldsymbol{w}\ \boldsymbol{\Phi}(\boldsymbol{x}) \tag{2}$$

where $\boldsymbol{w} = \sum_{i=1}^{n} \boldsymbol{\alpha}_i \boldsymbol{\Phi}(\boldsymbol{x}_i)$.

Thus, by using K the training data are implicitly projected into a higher dimensional space H. The SVM then computes $\boldsymbol{\alpha}_i$ that corresponds to the maximal margin hyperplane separating the training data into two classes labeled as -1 and 1 in H.

$$\begin{aligned} &\max_{\boldsymbol{w} \in F} \min\{y_i(\boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{x}_i))\} \\ &\text{subject to } \|\boldsymbol{w}\| = 1 \\ &\qquad y_i(\boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{x}_i)) > 0 \quad i = 1, 2, \cdots, n \end{aligned} \tag{3}$$

The training instances that lie closest to the hyperplane are called support vectors. Algorithmically, $\boldsymbol{\alpha}_i$ parameters that specify the SVM can be found in polynomial time by solving a convex optimization problem as follows[11]:

$$\begin{aligned} &\max \sum_i \boldsymbol{\alpha}_i - \frac{1}{2}\sum_{i,j} \boldsymbol{\alpha}_i \boldsymbol{\alpha}_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}) \\ &\text{subject to } \boldsymbol{\alpha}_i > 0 \qquad i = 1, 2, \cdots, n \end{aligned} \tag{4}$$

There are several commonly used kernel functions for nonlinear mapping in SVMs. We choose the Gaussian radial basis function (GRBF) in our experiments, which has the form,

$$k(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{2\delta^2}\right) \tag{5}$$

where parameter δ is the width of the Gaussian function.

2 Active Learning Based on Maximizing Information Gain

An SVM has learned a query concept by dividing the relevant images from the irrelevant ones with a hyperplane. The images on the query side of the hyperplane are considered relevant to the query concept and the rest, irrelevant. The farther the image away from the hyperplane, the more relevant or irrelevant to the query target.

The existing SVM active methods attempt to justify that selecting the images nearest the hyperplane, whose labels are most uncertain according to the current classifier, can approximately halve the version space each time and so can reduce the expected size of it as fast as possible. It has essentially placed a uniform distribution over the current space of consistent hypotheses and relies on the assumption that the version space is fairly symmetric and that $\boldsymbol{w}_i$ is centrally placed. However, both in theory and in practice, these assumptions can fail significantly. Furthermore, to work well, this method also requires seeding a query with good examples.

We propose to select the most informative instances to request for user's labels. The expected information gain from knowing the real label of an instance is computed based on the degree of hypothesis difference between SVM classifier and similarity measure. More important, with the combination of the two different learning models, we wish to mitigate the model bias existing in either of them.

For all the unlabeled images that have been classified as relevant or irrelevant by the current SVM hyperplane, we compute their class probabilities according to their distances to the hyperplane as

$$P^{+}_{\mathrm{svm}}(\boldsymbol{x}) = |d_{\boldsymbol{x}} - d^{+}_{\min}| / |d^{+}_{\max} - d^{+}_{\min}| \tag{6a}$$

$$P^{-}_{\mathrm{svm}}(\boldsymbol{x}) = |d_{\boldsymbol{x}} - d^{-}_{\min}| / |d^{-}_{\max} - d^{-}_{\min}| \tag{6b}$$

where P^{+}_{svm} and P^{-}_{svm} are the relevance (irrelevance) probabilities of the image that is on the query side (the other side) of current SVM hyperplane; $d^{+,-}_{\max}$ and $d^{+,-}_{\min}$ are the maximum and minimum distances for the relevant or irrelevant images, respectively; $d_{\boldsymbol{x}}$ is the distance of image $\boldsymbol{x}$ to the learned hyperplane, which is defined as

$$d(\boldsymbol{x},\ \theta) = \sum_{i=1}^{N_{\mathrm{s}}} \boldsymbol{\alpha}_i k(\boldsymbol{s}_i,\ \boldsymbol{x}) \tag{7}$$

where $\boldsymbol{\alpha}_i$ and $\boldsymbol{s}_i$ are parameters of the learned SVM hyperplane, N_{s} is the number of support vectors. The class probability is in direct proportion to the absolute value of the distance, and the image with maximum distance corresponds to class probability of 1.

The relevant images with high class-probabilities, which are the farthest from the hyperplane on the query concept side, are supposed to capture the query concept. But they are chosen only based on their distances to the hyperplane. There is no evidence that this kind of distance can be used as a perceptual similarity or dissimilarity measure. And due to the model bias and other causes, it cannot be ensured that the current classifier captures all the relevant images and all images considered as relevant by the current SVM classifier are sure to be relevant to the query. Some relevant images may not be enclosed by the learned query hypotheses, while some dissimilar images may be enclosed. Therefore we switch to the similarity measure to check the probabilities, and select the most informative images for soliciting user’s feedback based on maximizing difference of the prediction.

Similarity is one of the central theoretical constructs in CBIR. To distinguish images that are similar to a query image from others, kinds of similarity measures have been proposed. Euclidean distance is the most popular one. We employ Euclidean distance to measure the similarity between the query $\boldsymbol{q}$ and the image $\boldsymbol{x}$ in database, which is computed as

$$S(\boldsymbol{x},\ \boldsymbol{q}) = \| \boldsymbol{x} - \boldsymbol{q} \|_2 \tag{8}$$

The smaller the distance is, the more similar the image to the query is. Then we also assign class probability to image $\boldsymbol{x}$ according to the calculated Euclidean distance, which is defined as

$$P_{\mathrm{s}}(\boldsymbol{x}) = \frac{S_{\boldsymbol{x}} - S_{\min}}{S_{\max} - S_{\min}} \tag{9}$$

where $S_{\max}$ and $S_{\min}$ are the maximum and minimum Euclidean distances of the unlabeled image in database to the query, $P_{\mathrm{s}}(\boldsymbol{x})$ is ranging between 0 to 1. The value is closer to 0, the corresponding image is more relevant to the query, otherwise, it is more irrelevant to the query.

If the prediction about the label of the image according to similarity metric were almost consistent with the already learned hypothesis by the SVM classifier, the expected information gain from this image would be zero. On the contrary, if the prediction of similarity metric about this image were different from the already learned hypothesis by the SVM classifier, the information gain from knowing its real label would be high. We compute the degree of the difference as

$$\mathrm{dif}(\boldsymbol{x}) = |P_{\mathrm{svm}}(\boldsymbol{x}) - P_{\mathrm{s}}(\boldsymbol{x})| \tag{10}$$

A higher value of $\mathrm{dif}(\boldsymbol{x})$ for a relevant image or a lower value of $\mathrm{dif}(\boldsymbol{x})$ for an irrelevant image means more prediction difference between the SVM classifier and similarity metric. We can always get more information through soliciting user’s label on image $\boldsymbol{x}$ with more difference. Then based on maximizing difference, we define the expected information gain from query on the label of image $\boldsymbol{x}$ to be:

$$\left.\begin{aligned} M^{-} &= \max_{\boldsymbol{x}} \left(\ln \frac{1 + \mathrm{dif}(\boldsymbol{x})}{1 - \mathrm{dif}(\boldsymbol{x})} \right) \\ M^{+} &= \min_{\boldsymbol{x}} \left(\ln \frac{\mathrm{dif}(\boldsymbol{x})}{2 - \mathrm{dif}(\boldsymbol{x})} \right) \end{aligned}\right\} \tag{11}$$

The images with maximum expected information gain are presented to the user. Once the user’s feedback on the selected images is seen, the learner updates dynamically according to the errors in previous learning and the hypotheses are updated with learning too. Thus we increase the probability of finding relevant instances in the next feedback iteration. This active selection strategy ensures fast convergence to the query concept in a small number of feedback rounds. It also can sustain the learning process without good initial examples and works quite well with moderate model bias or noisy feedback. Once the classifier is trained, SVM active learning returns the top-k most relevant images, which are the farthest to the learned SVM hyperplane on the query side and whose value of dif($\boldsymbol{x}$) is no more than a certain threshold.

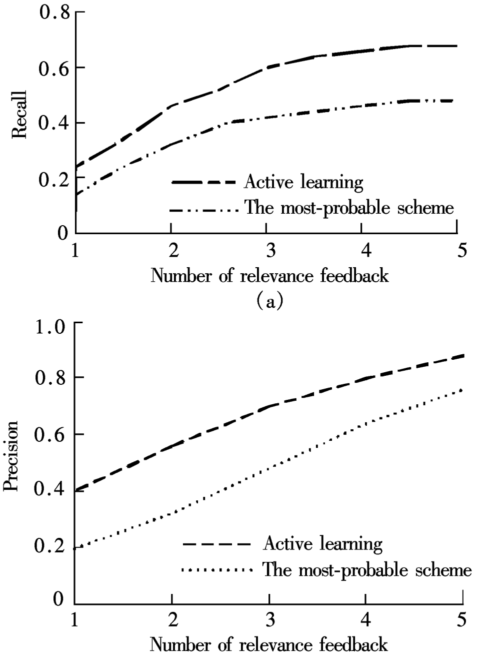
3 Experiments

A series of experiments have been performed on a data set of 1 600 images from the Corel image database, which has eight classes such as airplane, bird, flower, cloud, horse, elephant, bear and tiger. The images in our dataset are preprocessed with the Blobworld system[12], where an image is segmented into blobs and each characterized by color, texture and shape descriptors. However, to overcome the inaccurate segmentation, we use each blob and the corresponding background (image area except for the blob) to represent the image. Both the blob and the background are characterized by color and texture descriptors. The color descriptor includes mean and the transformed histogram. The texture descriptor consists of values of contrast and products of anisotropy and contrast[12].

In initial iteration of the online learning process, the randomly selected images are shown in the screen, and on subsequent rounds of query, active learning with 20 images is invoked. The user clicks on desired blobs of images as positive examples while leaving the unclicked ones as negative examples. All these responses from the user's interaction are taken with the system to refine the learned query concept, and then iterate. The retrieval performance is evaluated by precision and recall. Precision is the ratio of the number of relevant images returned to the total number of images returned. Recall is the ratio of the number of relevant images returned to the total number of relevant images in the database.

To testify the effects of the active selection approach, we have performed a set of experiments, where the performances of the proposed active learning method with the most-probable selection strategy are compared. The most-probable scheme, which is the current popular method in CBIR, chooses the images that possess the highest probability of being the target for the next display.

We have run each experiment through up to five rounds of relevance feedback and ten times with different initial starting samples, and computed the average value of precision and recall to evaluate the performance of the two different methods. The experimental results are shown in Fig.1. The proposed active learning algorithm makes significant improvement over the most-probable scheme, and it enhances the learning process and improves classification significantly.



(a)

(b)

Fig.1 Comparison of the active learning with the most-probable scheme

From the results, we also notice that the computed recall improves with the number of returned images, but the increase slows down when the number of the returned images exceeds 200. This phenomenon implies that the proposed method can rank most of the relevant images' priority with others, and can achieve good retrieval performance with a relative small number of returned images. However, complete retrieval of all the relevant image in database is still not a reality.

4 Discussion

For online learning technique, due to the limit on the number of instances presented to the oracle, the choice of instances becomes important. In this paper, we have presented an active learning algorithm based on maximizing information gain for CBIR. The active learning part of the proposed method computes the information gain from knowing the label of an image based on the class prediction differences between SVM classifiers and similarity metric and selects the most informative ones to query the user. Through the combination of the two different learning models, the method can cope well with model bias and user's error. The experimental results show that our active learning scheme improves the retrieval performance significantly in CBIR when the number of labeled instances is limited.

The proposed approach in this paper also can be used in a variety of widely applicable scenarios, thereby vastly reducing the amount of data that needs to be gathered while, at the same time, increasing the quality of the resulting models, classifiers and conclusions.

References

[1] Rui Y, Huang T S, Ortega M, et al. Relevance feedback: a power tool in interactive content-based image retrieval [J]. *IEEE Trans Circuits and Systems for Video Tech*, **1998, 8**(5): 644 − 655.

[2] Seung H S, Opper M, Sompolinsky H. Query by committee [A]. In: *Proc of the Fifth Workshop on Computational Learning Theory* [C]. San Mateo, CA, 1992. 287 − 294.

[3] Freund Y, Seung H, Shamir E, et al. Selective sampling using the query by committee algorithm [J]. *Machine Learning*, 1997, **28**(2, 3): 133 − 168.

[4] Chang E, Li B. MEGA — the maximizing expected generalization algorithm for learning complex query concepts [J]. *ACM Transactions on Information Systems*, **2003, 21**(4): 347 − 382.

[5] Dagan I, Engelson S. Committee-based sampling for training probabilistic classifiers [A]. In: *Proc of the Twelfth International Conference on Machine Learning* [C]. Tahoe City, California, 1995. 150 − 157.

[6] Cohn D, Atlas L, Ladner R. Improving generalization with active learning [J]. *Machine Learning*, **1994, 15**(2): 201 − 221.

[7] Lewis D, Gale W. A sequential algorithm for training text classifiers [A]. In: *Proc of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval* [C]. Dublin, Ireland, 1994. 3 − 12.

[8] Tong S, Chang E. Support vector machine active learning for image retrieval [A]. In: *Proc of ACM Mutimedia* [C]. Ottawa, Canada, 2001. 107 − 118.

[9] Campbell C, Cristianini N, Smola A. Query learning with large margin classifiers [A]. In: *Proc of the Seventeenth International Conference on Machine Learning* [C]. Stanford University, 2000. 111 − 118.

[10] Schohn G, Cohn D. Less is more: active learning with support vector machines [A]. In: *Proc of the Seventeenth International Conference on Machine Learning* [C]. Stanford University, 2000. 839 − 846.

[11] Vapnik V. *Statistical learning theory* [M]. New York: Addison Wiley, 1998.

[12] Carson C, Belongie S, Greenspan H. et al. Blobworld: image segmentation using expectation-maximization and its application to image querying [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, **2002, 24**(8): 1026 − 1038.

图像检索中基于最大信息获取量的主动学习算法

徐 杰 施鹏飞

(上海交通大学电子信息与电气工程学院, 上海 200030)

摘要: 本文提出一种基于内容的图像中的主动学习算法. 首先用支撑向量机学习得到初始查询概念, 然后用相似性测度对其进行检验, 选取信息量最大的样本来请求用户标记, 最后在相关反馈的迭代优化过程中获取用户的图像查询概念. 算法通过支撑向量机二值分类器与相似性测度2种不同学习模型的融合, 来减轻它们各自所存在的模型偏置. 实验结果显示, 所提算法能够显著提高图像检索的精确度, 在少量的反馈迭代之后即能准确地获取目标概念.

关键词: 主动学习; 基于内容的图像检索; 相关反馈; 支撑向量机; 相似性测度

中图分类号: TP391