MICROCALCIFICATION CLASSIFICATION ASSISTED BY CONTENT-BASED IMAGE RETRIEVAL FOR BREAST CANCER DIAGNOSIS

Yongyi Yang[a], Liyang Wei[b], and Roberts M. Nishikawa[c]

[a]Dept. of Electrical and Computer Engineering, [b]Dept. of Biomedical Engineering,
Illinois Institute of Technology, 3301 S. Dearborn Street, Chicago, IL 60616
[c]Dept. of Radiology, University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637

ABSTRACT

In this paper we propose a microcalcification classification scheme, assisted by content-based mammogram retrieval, for breast cancer diagnosis. We recently developed a machine learning approach for mammogram retrieval where the similarity measure between two lesion mammograms is modeled after expert observers. In this work we investigate how to use retrieved similar cases as references to improve the performance of a numerical classifier. Our rationale is that by adaptively incorporating local proximity information into a classifier, it can help improve its classification accuracy, thereby leading to an improved "second opinion" to radiologists. Our experimental results on a mammogram database demonstrate that the proposed retrieval-driven approach with an adaptive support vector machine (SVM) could improve the classification performance from 0.78 to 0.82 in terms of the area under the ROC curve.

Index Terms— microcalcification classification, adaptive support vector machine, image retrieval

1. INTRODUCTION

Breast cancer remains to be a leading cause of death among women in the developed countries. Currently mammography is the dominant method for detection of breast cancer. Clustered microcalcifications (MC) can be an important early sign of breast cancer. As an example, Fig. 1 shows a mammogram with a cluster of microcalcifications. Due to the subtlety in the appearance of individual MCs, there is a significant risk that a radiologist may misclassify some cases in breast cancer diagnosis [1].

Recently we developed a content-based mammogram retrieval system as a diagnostic aid to radiologists in their interpretation of mammograms [2]. We conjecture that by presenting perceptually similar mammograms with known pathology to the one being evaluated, the radiologists could reach a better informed decision in their diagnosis. Our proposed mammogram retrieval system involves two major components: 1) retrieving similar mammogram images from a database by using learning based similarity measure, and 2) classifying the query mammogram image based on retrieved results (retrieval-driven classification). This retrieval framework is illustrated with a functional diagram in Fig. 2.

This work was supported by NIH/NCI grant CA89668.

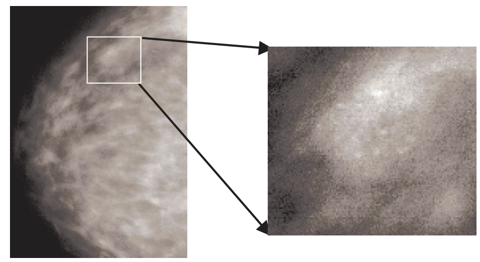


Fig. 1. Left: A mammogram in craniocaudal view; Right: expanded view showing clustered microcalcifications (MCs)

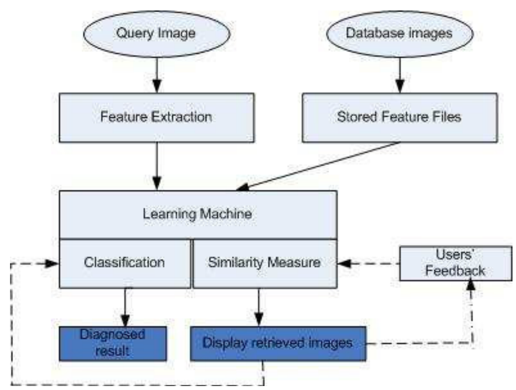


Fig. 2. The proposed content-based mammogram retrieval and classification framework

In [2], we explored a similarity measure for mammogram retrieval based on supervised learning from expert readers. We evaluated this approach using data collected from an observer study with a set of clinical mammograms. It was demonstrated that the proposed machine learning approach can be used to model the notion of similarity as judged by expert readers in their interpretation of mammogram images and that it can outperform alternative similarity measures derived from unsupervised learning.

In this work, we focus on the second component of our proposed mammogram retrieval and classification system: microcalcification classification assisted by retrieval. The traditional approach in this field is to present the human observer with examples in the database that are similar to the one being examined. In this study we consider how to use the retrieved similar cases as references to improve a numerical classifier's performance. We conjecture that by adaptively incorporating proximity information to the cost function of a classifier, it can help to improve its classification accuracy, thereby leading to an improved "second opinion" to radiologists. Toward this goal, we propose a retrieval-driven approach with an adaptive support vector machine (SVM) for improving the classfication performance. We choose SVM since it has been demonstrated to outperform many of the competing methods in microcalcification classification [1].

2. METHODOLOGY

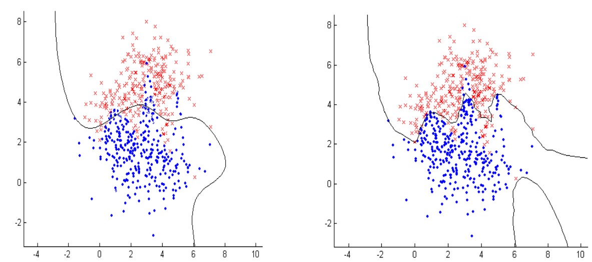
2.1. Adaptive SVM

SVM is a constructive learning procedure rooted in statistical learning theory [3]. It is based on the principle of structural risk minimization, which aims at minimizing the bound on the generalization error. The SVM decision function is predetermined through a training process using a set of examples before it can be applied to data outside the training set.

Despite its success, the performance of an SVM classifier can be hampered by several factors in practice. First, the nature of the problem is often complicated and not well understood, as is the case of breast cancer diagnosis[1], and it is not even clear that the classification task could be well described by a single decision function. Secondly, even when such a decision function indeed exists, the "true" decision boundary is rarely obtainable because of the limited number of available training samples. In such a case, a challenging problem is how to strike a balance between over-fitting and under-fitting in the classifier model. This is especially the case when it is too expensive or simply impossible to obtain enough training samples in many practical problems. Consequently, it becomes impossible to determine the "optimal" classifier function. For example, in [1] the best classification performance achieved by the SVM was still far from being perfect, which we believe is largely due to the presence of many hard-to-classify cases in the database we used [1].

In this work we propose a locally adaptive SVM classification scheme. In the proposed scheme, we attempt to adapt the decision function of the SVM classifier according to how it performs on samples that are close to the one being examined (called query). Specifically, before the SVM decision function is applied to the query, it is first tested and adapted based on the knowledge of the samples that are in its neighborhood. Our motivation is as follows: if the SVM function is found to perform poorly on known samples close to the query, it implies that the decision function is not well trained for samples in the neighborhood of the query, and thus, it will also likely not perform well on the query. In such a case, we will adjust the SVM classifier using these similar samples accordingly, which in turn can lead to improved classification accuracy on the query.

In our retrieval-driven classification scheme the SVM classifier is adaptive in that its decision boundary is adjusted according to the "local" information of the case to be classified (i.e., retrieved similar cases). To demonstrate the concept, we show a classification example in Figure 3, where Figure 3(a) shows the decision boundary between the two classes of the SVM obtained from a set of training samples; Figure 3(b) shows the decision boundary obtained using the proposed adaptive SVM classifier (with the same parametric setting as the SVM in (a)). In this example, the five nearest neighbors according to the Euclidean distance were used as the similar cases to the query sample. As can be seen, the proposed adaptive SVM could achieve better classification than the SVM in this example.



(a) (b)

Fig. 3. (a) The classification boundary learned by SVM; (b) the classification boundary learned by adaptive SVM

We note that there exist several algorithms related to adaptive SVM classification in the literature which aim to improve a classifier's performance by treating each test sample differently, e.g., [4]. In our own previous work [5], we used the conept of adaptive SVM in a content-based image retrieval system, where the SVM regression function was adjusted according to the relevance feedback samples provided by the user. To our best knowledge, these methods are quite different from our proposed approach here.

2.2. Algorithm

Consider a general two-class classification problem of assigning a class label $y \in \{-1, +1\}$ to an input feature vector $\mathbf{x} \in R^N$. We are given input-output training data pairs $\{(\mathbf{x}_1, y_1)$, ..., $(\mathbf{x}_n, y_n)\}$. The SVM classification function can be written in the following form: $f_{SVM}(\mathbf{x}) = \mathbf{w}^T\Phi(\mathbf{x}) + b$, where $\Phi(\mathbf{x})$

is a (nonlinear) mapping function, and $\mathbf{w}$ and b are parameters determined through training.

In the proposed adaptive SVM, we modify the SVM cost function as follows:

$$\begin{aligned} \tilde{J}(\mathbf{w}, \xi) &= \frac{1}{2}\|\mathbf{w}\|^2 + C^{(s)} \sum_{\mathbf{x}_i \in N(\mathbf{x})} \xi_i + C \sum_{\mathbf{x}_i \notin N(\mathbf{x})} \xi_i \quad (1) \\ s.t. \quad & y_i f_{SVM}(\mathbf{x}_i) \geq 1 - \xi_i \\ & \xi_i \geq 0; i = 1, 2, ..., N \end{aligned}$$

where $N(\mathbf{x})$ denotes the set of training samples that are in a defined neighborhood of a query sample $\mathbf{x}$, and $C^{(s)}$ is a penalty parameter introduced for the training samples in $N(\mathbf{x})$.

In the modified cost function $\tilde{J}(\mathbf{w}, \xi)$ above, the training samples in $N(\mathbf{x})$ are closer (hence more similar) to the query $\mathbf{x}$ than the others. We write $C^{(s)} = tC$, where $1 < t < \infty$ is a *penalty factor*. This will have the effect to impose a greater emphasis ($C^{(s)}$) on those samples similar to the query $\mathbf{x}_i$ over other samples. The rationale is that those similar samples should have a greater impact on the classification of the query. Thus, a larger penalty is assessed in the cost function $\tilde{J}(\mathbf{w}, \xi)$ when a similar sample is misclassified. Indeed, when $t \to 1$, the adaptive SVM simply becomes a regular SVM where the same factor C is used for all training samples; on the other hand, when $t \to \infty$, the cost function $\tilde{J}(\mathbf{w}, \xi)$ will be dominated by the samples similar to the query. In this latter case, the adaptive SVM decision function will depend on only the similar samples. Interestingly, this would be similar in spirit to a class of classification algorithms, such as the powerful K nearest neighbor classifier (KNN) [6], which makes use of only local neighborhood information in the decision function. In this sense, the adaptive SVM functions can be viewed as playing a role of joining a global SVM classifier with a local classifier.

In the SVM cost function, the purpose of using model complexity to constrain the optimization of empirical risk is to avoid over-fitting, a situation in which the decision boundary too precisely corresponds to the training data, and thereby fails to perform well on data outside the training set. As in the choice of the parameter C in regular SVM, the newly introduced penalty factor t in the adaptive SVM will have to be determined during the training phase.

2.3. Implementation issues

Compared to the regular SVM, it may seem that the adaptive SVM will be much more demanding computationally, because the modified cost function $\tilde{J}(\mathbf{w}, \xi)$ in (1) would vary with the query sample $\mathbf{x}$, which would need to be re-optimized for every $\mathbf{x}$. Fortunately, this is not the case. Instead, we can greatly reduce the extra computation burden by employing a regular SVM. Specifically, we adopt the following procedure for training the adaptive SVM: for each query, we first apply a regular SVM classifier on its similar cases. If it can correctly classify all of them, then we apply this SVM classifier to the query as well; otherwise, we invoke the adaptive SVM procedure. The rationale behind this is that if the SVM classifier performs well on the similar cases, it will likely perform well on the query as well. Our experiments show that this can result in marked saving in computation time, especially for those easy-to-classify cases.

As in SVM, the optimization of the cost function in (1) can be carried out by solving its dual problem using quadratic programming. To speed up the numerical algorithm, the so-called incremental learning technique can be applied [5].

As in regular SVM, the parameters of the adaptive SVM will have to be determined during the training phase. In particular, the newly introduced penalty factor t can be determined using a leave-one-out cross-validation procedure on the set of similar samples to the query. Too large a value for t can lead to over-emphasis on the training samples near the query, which may cause over-fitting. On the other hand, too small a value for t may not have enough impact on the cost function. For each query, we pick the t value that corresponds to the lowest error rate resulting from this procedure. In our experiments the penalty factor was found typically to be in the range of $2 \leq t \leq 10$.

Another issue for the adaptive SVM is how to determine the similar samples to use for the query. In this work we use the mammogram retrieval framework reported previously in [2]. For each query mammogram image, we invoke the retrieval system to obtain a set of similar mammograms from the database, which is then used for the adaptive SVM. For comparison purposes, we also experimented with using other distance based similarity measures, including K-nearest neighbors based on the Euclidean distance [6], and discriminant adaptive nearest neighbors (DANN) [7].

3. EVALUATION STUDY

3.1. Data set

In our study we used a database of mammogram images collected by the Department of Radiology at the University of Chicago. The database consists of a total of 200 different mammogram images from 104 cases (46 malignant, 58 benign), digitized with a spatial resolution of 0.1 mm/pixel and 10-bit grayscale. All these images contain clustered microcalcifications, as shown in the example earlier in Fig. 1.

3.2. Experiment setup

For the retrieval system, we used the same learning based similarity measure reported in [2], where 600 image pairs had been scored in a human observer study for training the similarity function. To test the proposed retrieval-driven adaptive SVM classifier, the 200 mammogram images were used in a leave-one-out procedure. During each round, one mammogram image was used as the test sample (i.e., query); similar mammogram images were then retrieved for this query from the database based on the learned similarity function; subsequently, the test mammogram was classified by the adaptive SVM.

It is important to note that, to avoid any potential bias, during the leave-one-out procedure the left-out image for testing was also removed from training the retrieval stage. This achieved complete isolation of the test sample from any of the training sets. For this study we used the same set of 12 image features for characterizing the clustered MCs as described in [2].

3.3. Performance Evaluation for Classification

To evaluate the performance of a classifier, we use the so-called ROC analysis [8], which is now used routinely for many classification tasks. As a summary measure of overall diagnostic performance, the area under the ROC curve (denoted by A_z) is used here. A larger A_z means better classification.

4. EXPERIMENTAL RESULTS

Figure 4 summarizes the classification results achieved by the proposed retrieval-driven approach (Ada-SVM), where the obtained A_z value is plotted against N, the number of most similar cases used for the adaptive SVM classifier. For comparison, the best A_z value obtained by a regular SVM is also shown in Fig. 4.

As can be seen from Fig. 4, the classification accuracy has been improved from $A_z = 0.7752$ for SVM to $A_z = 0.8223$ for Ada-SVM with N=5. A statistical comparison between SVM and Ada-SVM using the ROCKIT program yielded a two-tailed p-value 0.0285 (one-tailed p-value 0.0142) for rejecting the null hypothesis that their corresponding ROC curves have the same area under them. These results show that the proposed retrieval-driven approach can lead to meaningful improvement in classification accuracy over the SVM, which was demonstrated to outperform many other state-of-the-art methods in our previous study [1].

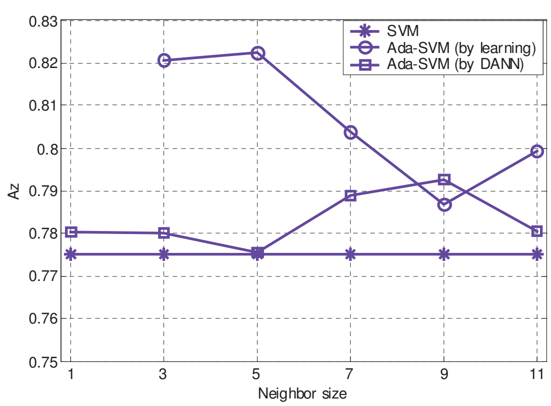


Fig. 4. Classification results achieved by Ada-SVM with different similarity measures for CBIR

We also note from Fig. 4 that as the size N of retrieved images is further increased the classification performance A_z value starts to decrease. We believe that this is due to the fact that the database is limited in size, which in turn limits the number of truly "similar" cases to the query. Thus, further increase of the number of retrieved images will no longer be beneficial. For both SVM and Ada-SVM, the same model setting (RBF kernel, $\sigma = 2.5$, $C = 100$) was determined in the leave-one-out procedure.

Finally, for comparison, in Fig. 4 we also show the classification results obtained by the adaptive SVM using a different similarity measure, the discriminant adaptive nearest neighbors [7] (DANN), for retrieving similar images. Note that the classification result could still be improved from 0.7752 (SVM) to 0.7925 (Ada-SVM with N=9).

5. CONCLUSION

In this paper we proposed a classification approach assisted by content-based image retrieval to improve the classification accuracy in computer aided diagnosis for breast cancer. Our results using a clinical database show that the proposed adaptive SVM classifier can lead to reduced generalization error. Encouraged by this initial success, we plan to further develop and validate the proposed approach using clinical evaluations.

6. REFERENCES

[1] L. Wei, Y. Yang, and R. M. Nishikawa, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Trans. on Medical Imaging*, vol. 24, no. 3, pp. 371–380, 2005.

[2] L. Wei, Y. Yang, R. M. Nishikawa, and M. N. Wernick, "Learning of perceptual similarity from expert readers for mammogram retrieval," *IEEE Int'l Symposium on Biomedical Imaging*, 2006.

[3] V. Vapnik, *Statistical Learning Theory*, John Wiley, 1998.

[4] R. Herbrich and J. Weston, "Adaptive margin support vector machines for classification," *Ninth International Conference on Artificial Neural Networks*, vol. 2, pp. 880–885, 1999.

[5] I. El-Naqa, Y. Yang, N. P. Galatsanos, and M. N. Wernick, "Relevance feedback based on incremental learning for mammogram retrieval," *Intel. Conf. on Image Processing*, vol. 1, pp. 729–732, 2003.

[6] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, New York: Wiley, 1973.

[7] Trevor Hastie and Robert Tibshirani, "Discriminant adaptive nearest neighbor classification," *IEEE Trans. on Pattern Anal. Mach. Intell*, vol. 18, no. 6, pp. 607–616, 1996.

[8] C. E. Metz, B. A. Herman, and J. Shen, "Maximum-likelihood estimation of receiver operating (roc) curves from continuously-distributed data," *Stat. Med.*, vol. 17, pp. 1033–1053, 1998.