SUPPORT VECTOR MACHINES FOR CLASSIFICATION OF MALIGNANT AND BENIGN LESIONS

Anatoli Nachev, Mairead Hogan

Abstract: *This paper presents an exploratory study of the effectiveness of support vector machines used as a tool for computer-aided breast cancer diagnosis. We explore the discriminatory power of heterogeneous mammographic and sonographic descriptors in solving the classification task. Various feature selection techniques were tested to find a set of descriptors that outperforms those from similar studies. We also explored how choice of the SVM kernel function and model parameters affect its predictive abilities. The kernels explored were linear, radial basis function, polynomial, and sigmoid. The model performance was estimated by ROC analysis and metrics, such as true and false positive rates, maximum accuracy, area under the ROC curve, partial area under the ROC curve with sensitivity above 90%, and specificity at 98% sensitivity. Particular attention was paid to the latter two as lack of specificity causes unnecessary surgical biopsies. Experiments registered that an appropriate reduction of variables can greatly improve the predictive power of the model, as long as the choice of the kernel affects the model performance marginally. We also found that the SVM is superior to the common classification technique used in the field - MLP neural networks.*

Keywords: *data mining, support vector machines, heterogeneous data; breast cancer diagnosis, computer aided diagnosis.*

ACM Classification Keywords: I.5.2- Computing Methodologies - Pattern Recognition – Design Methodology - Classifier design and evaluation.

Introduction

A proper treatment of breast cancer disease requires timely, reliable, and accurate diagnosis, which allows radiologists and physicians to differentiate between benign and malignant lesions. Many computer-aided detection/diagnosis (CAD) tools currently support medical practices by capturing knowledge from previous cases and applying that knowledge to the new cases. CAD is a typical machine-learning problem, which has been dealt with by various data mining techniques and tools such as linear discriminant analysis (LDA), logistic regression analysis (LRA), multilayer perceptions (MLP), etc. [Chen et al., 2009].

Most of the current implementations tend to use only one information source, usually mammographic data in the form of data descriptors defined by the Breast Imaging Reporting and Data System (BI-RADS) lexicon, developed by the American College of Radiology (ACR) in order to standardize the mammographic language and interpretations, and to facilitate communication between clinicians [BI-RADS, 2003], [Kopans, 1992]. Jesneck et al. [2007] have used a novel combination of BI-RADS mammographic and sonographic descriptors and some proposed by Stavros et al. [1995] in order to build a predictive model based on MLP, which shows superior characteristics to those that use one data source. Our study takes that approach, but investigate another predictive technique - support vector machines (SVM). We also address the problem of high false positive rate of indication for biopsy (specificity rate), which causes unnecessary surgical biopsies, lowers the efficiency of the diagnosis, exposes patients to discomfort, and creates financial burden as procedures cost thousands of euros each [Lacey et al., 2002].

Further to the study of Jesneck et al. [2007] who used a set of fourteen descriptors to train and test a MLP neural network, we explore the discriminatory power of all descriptors in order to seek alternative sets that when applied to SVM can ensure even higher sensitivity and specificity.

The paper is organized as follows: Section 2 provides a brief overview of the support vector machines used as data mining tools; Section 3 introduces the dataset used in this study and discusses variable selection as part of the data preprocessing; Section 4 presents and discuses results obtained from experiments; and Section 5 gives the conclusions.

Support Vector Machines

Support vector machines are common machine learning techniques. They belong to the family of generalized linear models, which achieve a classification or regression decision based on the value of the linear combination of input features. Using historical data along with supervised learning algorithms, SVM generate mathematical functions to map input variables to desired outputs for classification or regression prediction problems.

SVM, originally introduced by Vapnik [1995], provide a new approach to the problem of pattern recognition with clear connections to the underlying statistical learning theory. They differ radically from comparable approaches such as neural networks because SVM training always finds a global minimum in contrast to the neural networks. SVM can be formalized as follows. Training data is a set of points of the form

$$D = \{(x_i, c_i) \mid x_i \in \Re^p,\ c_i \in \{-1, 1\}\}_{i=1}^{n}, \quad (1)$$

where the c_i is either 1 or -1, indicating the class to which the point x_i belongs. Each data point x_i is a p-dimensional real vector. During training a linear SVM constructs a p-1-dimensional hyperplane that separates the points into two classes (see Figure 1). Any hyperplane can be represented by: $w \cdot x - b = 0$ where w is a normal vector and $\cdot$ denotes dot product. Among all possible hyperplanes that might classify the data, SVM selects one with maximal distance (margin) to the nearest data points (support vectors).

When the classes are not linearly separable (there is no hyperplane that can split the two classes), a variant of SVM, called soft-margin SVM, chooses a hyperplane that splits the points as cleanly as possible, while still maximizing the distance to the nearest cleanly split examples. The method introduces slack variables, ξ_i, which measure the degree of misclassification of the datum x_i. Soft-margin SVM penalizes misclassification errors and employs a parameter (the soft-margin constant C) to control the cost of misclassification. Training a linear SVM classier solves the constrained optimization problem (2).

$$\min_{w,b,\xi_k} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \qquad s.t. \quad w \cdot x_i + b \geq 1 - \xi_i \tag{2}$$

In dual form the optimization problem can be represented by (3)

$$\min_{\alpha_i} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i \cdot x_j - \sum_{i=1}^{n}\alpha_i \qquad s.t. \quad 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^{n}\alpha_i c_i = 0 \tag{3}$$

The resulting decision function $f(x) = w \cdot x + b$ has a weight vector $w = \sum_{k=1}^{n}\alpha_k y_k x_k$. Data points x_i for which $\alpha_i > 0$ are called support vectors, since they uniquely define the maximum margin hyperplane. Maximizing the margin allows one to minimize bounds on generalization error.

If every dot product is replaced by a non-linear kernel function, it transforms the feature space into a higher-dimensional one, thus though the classifier is a hyperplane in the high-dimensional feature space (see Figure 2). The resulting classifier fits the maximum-margin hyperplane in the transformed feature space. The kernel function can be defined as

$$k(x_i, x_j) = \Phi(x_i)\Phi(x_j) \tag{4}$$

where $\Phi(x)$ maps the vector *x* to some other Euclidean space. The dot product $x_i \times x_j$ in the formulae above is replaced by $k(x_i, x_j)$ so that the SVM optimization problem in its dual form can be redefined as: maximize (in α_i)

$$\tilde{L}(\alpha) = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j k(x_i, x_j) \text{, s. t. } \sum_i \alpha_i y_i = 0; \quad \alpha_i \geq 0 \text{ for all } 1 \leq i \leq N \qquad (5)$$

A non-linear SVM is largely characterized by the choice of its kernel, and SVMs thus link the problems they are designed for with a large body of existing work on kernel-based methods. Some common kernels functions include:

- Linear kernel: k(x,x') = (x·x')
- Polynomial kernel: $k(x, x') = (sx \cdot x' + c)^d$
- RBF kernel: $k(x, x') = \exp(-\gamma(x - x')^2)$
- Sigmoid kernel: $k(x, x') = \tanh(s(x \cdot x') + c)$

Once the kernel is fixed, SVM classifiers have few user-chosen parameters. The best choice of kernel for a given problem is still a research issue. Because the size of the margin does not depend on the data dimension, SVM are robust with respect to data with high input dimension. However, SVM are sensitive to the presence of outliers, due to the regularization term for penalizing misclassification (which depends on the choice of C). The SVM algorithm requires $O(n^2)$ storage and $O(n^3)$ to learn.

The SVM method can also be applied to the case of regression. A version of SVM for regression, called support vector regression (SVR), was proposed by Drucker et al. [1997]. The basic idea of SVR is that a non-linear function learns by a linear learning method in a kernel-induced higher dimensional space. Similarly to how SVM classification ignores data points that are not support vectors, the SVR depend on a small subset of training data points.

The SVM's major advantage lies with their ability to map variables onto an extremely high feature space. This, in essence facilitates a means for the exploration of nonlinear kernel-based classifiers [Oladunni and Singhal, 2009; Burges, 1998], however, it has been discovered they do not favour large datasets, due to the demands imposed on virtual memory, and the training complexity resultant from the use of such a scaled collection of data [Horng et al., 2010]. Work from Fei et al. [2008] highlighted three "crucial problems" in the use of support vector machines. These are attaining the optimal input subset, correct kernel function, and the optimal parameters of the selected kernel, all of which are prime considerations within this study.

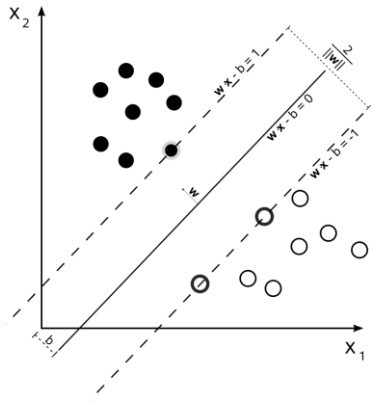

Fig. 1. Maximum-margin hyperplane for a SVM trained with samples from two classes. Samples on the margin are support vectors

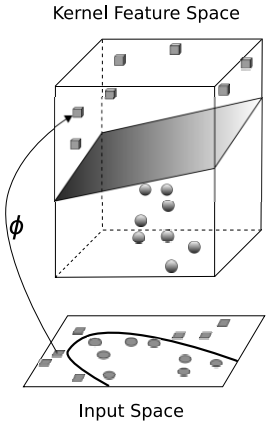

Fig. 2. Kernel function: a linearly inseparable input space can be mapped to a linearly separable higher-dimmensional space

Dataset and Preprocessing

This study uses a dataset that contains data from physical examination of patients, including mammographic and sonographic examinations, family history of breast cancer, and personal history of breast malignancy, all collected at Duke University Medical Centre [Jesneck et al., 2007]. Samples included in the dataset are those selected for biopsy only if the lesions corresponded to solid masses on sonograms and if both mammographic and sonographic images taken before the biopsy were available for review. Data contain 803 samples, 296 of which are malignant and 507 benign. Out of 39 descriptors, 13 are mammographic BI-RADS, 13 sonographic BI-RADS, 6 sonographic suggested by Stavros et al. [1995], 4 sonographic mass descriptors, and 3 patient history features. There is also a class label, -1 and 1, that indicates if a sample is malignant or benign.

The data features are as follows: mass size, parenchyma density, mass margin, mass shape, mass density, calcification number of particles, calcification distribution, calcification description, architectural distortion, associated findings, special cases (as defined by the BI-RADS lexicon [BI-RADS, 2003]: asymmetric tubular structure, intramammary lymph node, global asymmetry, and focal asymmetry), comparison with findings at prior examination, and change in mass size. The sonographic features are radial diameter, antiradial diameter, anteroposterior diameter, background tissue echo texture, mass shape, mass orientation, mass margin, lesion boundary, echo pattern, posterior acoustic features, calcifications within mass, special cases (as defined by the BI-RADS lexicon: clustered microcysts, complicated cysts, mass in or on skin, foreign body,

intramammary lymph node, and axillary lymph node), and vascularity. The six features suggested by Stavros [Stavros et al., 1995] are mass shape, mass margin, acoustic transmission, thin echo pseudocapsule, mass echogenicity, and calcifications. The four other sonographic mass descriptors are edge shadow, cystic component, and two mammographic BI-RADS descriptors applied to sonography—mass shape (oval and lobulated are separate descriptors) and mass margin (replaces sonographic descriptor angular with obscured). The three patient history features were family history, patient age, and indication for sonography [Jesneck et al., 2007].

Using the dataset in its original format for classification with SVM would be problematic due to the large amplitude of feature values caused by the different nature of the data variables and different units of measurements used. For example, the mass size values range from 0 to 75, as long as calcification ranges from 0 to 3. Such an inconsistency could affect the predictive abilities of a SVM classifier as some variables can be viewed as more 'influential' than others. The approach we used to address that problem was to process each data variable (data column) separately by transformation (6). It scales down the variables within the unit hypercube.

$$x_i^{new} = \frac{x_i^{old} - \min_i}{\max_i - \min_i} \tag{6}$$

We also explored how presence or absence of variables presented to the model for training and testing affects the classifier performance. Removing most irrelevant and redundant features from the data helps to alleviate the effect of the curse of dimensionality and to enhance the generalization capability of the model, and to speed up the learning process and to improve the model interpretability. The feature selection also helps to acquire better understanding about data and how they are related with each other. The exhaustive search approach that considers all possible subsets of variables is best for datasets with small cardinality, but impractical for large number of features as in our case. Jesneck et al. [2007] proposed a feature subset of 14 descriptors (s14) for their experiments with neural networks. They were derived by the stepwise feature selection technique. There is no guarantee, however, that an optimal variable selection for one classification technique will be optimal for another. In order to find the alternative selections for the SVM model we considered several feature selection algorithms, which generally fall into two categories: feature ranking and subset selection. The latter is more advanced and widely used in practice, which made us focus on it. We considered best first, subset size forward selection, race search, scatter search and genetic search combined with a set evaluation technique that considers individual predictive ability of each feature along with the degree of redundancy between them [Goldberg, 1989;

Hall, 1998]. We propose a set of 17 variables (s17) derived by the linear forward selection technique, proposed by Guetlien et al. [2009]. The feature set we obtained consists of the following variables: patient age, indication for sonography, mass margin, calcification number of particles, architectural distortion, anteroposterior diameter, mass shape, mass orientation, lesion boundary, special cases, mass shape, mass margin, thin echo pseudocapsule, mass echogenicity, edge shadow, cystic component, and mass margin. Two of these are general descriptors; three - mammographic BI-RADS; five - sonographic BI-RADS; four - Stavros'; and three - sonographic mass descriptors. The feature set is relatively balanced in representing different categories of data. In our experiments we also used the set of 14 variables (s14) mentioned above and the original full set of 39 variables (s39).

Empirical Results and Discussion

Using the training and testing datasets described above, we built a classification model based on SVM. For the purposes of the ROC we used the support vector regression technique, which outputs predictions as real numbers between -1 and 1, which mapped to the class labels (either -1 or 1). In order to minimize the bias in results associated with the random sampling of the training and testing data samples, we applied five-fold cross-validation, a.k.a. rotation estimation. The dataset was randomly spit into five mutually exclusive subsets (folds) of equal size. The model was trained and tested five times so that each time it was trained on one combination of four folds and tested on the remaining one. The cross-validation estimate of the overall model accuracy was calculated by (7).

$$CVA = \frac{1}{k}\sum_{i=1}^{k} A_i \quad , \tag{7}$$

where the number of folds k=5, CVA is the cross-validation accuracy, and A_i is the accuracy measure of the i-th fold (e.g. hit-rate, sensitivity, specificity).

The primary source for estimating the model accuracy is the confusion matrix (a.k.a. contingency table), illustrated in Figure 3. Results from experiments were summarized in four categories: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The numbers along the primary diagonal in the matrix represent correct predictions, as long as those outside the diagonal represent the errors.

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive Count (TP)	False Positive Count (FP)
	Negative	False Negative Count (FN)	True Negative Count (TN)

Fig. 3. Confusion matrix for tabulation of classification results

In order to estimate the model performance we used the following derivations from the confusion matrix:

- True Positive Rate (TPR), a.k.a. sensitivity, hit rate, or recall is the ratio of correctly classified positives divided by the total positive count.

$$TPR = \frac{TP}{TP + FN} \tag{8}$$

- False positive rate (FPR), a.k.a. fall-out, or (1-specificity) is the ratio of incorrectly classified positives divided by the total negative count.

$$FPR = \frac{FP}{FP + TN} \tag{9}$$

- Accuracy (ACC) is the ratio of correctly classified instances (both positives and negatives) divided by the total number of instances.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

Estimating the accuracy of the built SVM model is important for the following two reasons: first, it can be used to estimate the future prediction accuracy, which could imply the level of confidence, the potential users may have; secondly, it can be used for choosing a particular instance of the SVM model among available options, e.g. selection of kernel function and parameter settings.

Accuracy is a common performance estimator in machine learning and data mining, but in many cases and problem domains it is not sufficient metric. Sometimes, accuracy can be

misleading, for example where important classes are underrepresented in the datasets and class distribution is skewed. In that case accuracy is helpless in counting different costs and consequences from misclassifications. This is the case of the domain we consider, as misdiagnosed malignant and benign samples have different consequences and even may cost life. Another drawback of the accuracy is that it depends on the classifier's operating threshold. When SVM runs as a regression function that outputs real numbers between -1 and 1, mapping outputs to class labels requires defining a threshold between -1 and 1, so that the output can fall below or above it, i.e. mapped to one or another class label. Applying different thresholds produces different instances of the classification model, each of which features a specific accuracy.

In order to address those accuracy deficiencies, we did Receiver Operating Characteristics (ROC) analysis [Fawcett, 2006]. This is a graphical assessment technique where the true positive rate is plotted on the Y-axis and false positive rate is plotted on the X-axis (Figure 4). In the ROC space, a classification model is a step curve plotted by connecting all model instances made by varying the threshold value.

The line that links (0,0) and (1,1) is the no-discrimination line. It represents the worst possible model, which predicts by a completely random guess. Any other classifier should appear above that line. If it pops below the line, a negation of its predictions would move it above the no-discrimination line.

On the other hand, the 'ideal' classifier would be represented by the point (0,1), the top-left corner, which shows that all true positives are found and no false positives are found. Any model performance can be measured by its proximity to the 'ideal' classifier. The closer the ROC curve is as a whole to the north-west corner, the better. That is also the most distant from the no-discrimination line, the better. Given a curve, the most 'north-west' point of the curve represents the model instance with maximal accuracy.

The ROC analysis also provides means for quantification of a model performance, these are Area Under the ROC Curve (AUC) and partial Area Under the ROC curve (pAUC) where sensitivity is above a certain value (p). The AUC / pAUC are scalars that measure the overall model performance, regardless of the operational threshold. The bigger the values, the better the model is. As long as AUC provides an overall estimation of the model, the pAUC is more relevant to the application area, as the potential users of CAD tools are particularly interested in working with high levels of sensitivity. It is believed that sensitivity above 90% ($_{0.90}$AUC) is relevant to the application field. Another clinically relevant metric that we estimated is specificity at given sensitivity. In order to be consistent with previous studies [Jesneck et al., 2007], we considered specificity at 98% sensitivity.

Table 1 summarizes results from numerous experiments where the SVM model was trained and tested using four kernels: linear, polynomial, RBF, and sigmoid. For each of those kernels we experimented with three different sets of variables: s39 that contains all

Table 1. Performance of SVM with linear, polynomial, RBF, and sigmoid kernels. Metrics for comparison include: area under the ROC curve (AUC), partial AUC at sensitivity above 90% ($_{0.90}$AUC), specificity at 98% sensitivity, and maximal accuracy (ACC_{max}). Models have been tested with three variable selections: s39, s17, and s14. Typical radiologist assessment values are also included. Figures in bold show best values.

SVM linear	*s39*	*s17*	*s14*	*Radiologist*
AUC	**0.91**	**0.91**	0.89	0.92
$_{0.90}$AUC	0.71	0.74	0.62	0.52
Spec /98% sens	**0.36**	0.30	0.22	0.52
ACC_{max}	0.84	**0.85**	**0.85**	n/a

SVM polynomial	*s39*	*s17*	*s14*	*Radiologist*
AUC	**0.91**	**0.91**	0.89	0.92
$_{0.90}$AUC	0.72	0.74	0.62	0.52
Spec /98% sens	**0.36**	0.29	0.23	0.52
ACC_{max}	0.84	**0.85**	0.84	n/a

SVM FBF	*s39*	*s17*	*s14*	*Radiologist*
AUC	0.90	**0.91**	0.88	0.92
$_{0.90}$AUC	0.64	**0.75**	0.58	0.52
Spec /98% sens	0.29	0.32	0.20	0.52
ACC_{max}	0.83	**0.85**	0.83	n/a

SVM sigmoid	*s39*	*s17*	*s14*	*Radiologist*
AUC	**0.91**	**0.91**	0.88	0.92
$_{0.90}$AUC	0.67	**0.75**	0.62	0.52
Spec /98% sens	0.27	**0.36**	0.20	0.52
ACC_{max}	**0.85**	0.84	0.83	n/a

The table figures show that the selection of descriptors for training and testing plays a significant role in the SVM performance. According to all metrics and no matter which kernel is selected, it is evident that the variable set s14, proposed by Jesneck et al. [2007] for classification with MLP is outperformed by both s39 and s17. As mentioned before, that is not surprising as an optimal variable selection for one classification model would not be optimal for another. We also show that the alternative selection of variables, s17, can outperform both s14 and s39. That selection significantly improves the $_{0.90}$AUC of s14 from 12% to 17%, depending on which kernel is used, and also outperforms the radiologist value by 23%. The other clinically relevant metric, specificity at 89% sensitivity, is also improved by s17 in comparison with s14 - from 6% to 16%. In some cases s39 performs as well as s17, but it never gets better. The variable set s17 shows itself as the best performer regarding AUC and ACC_{max} with only few exceptions.

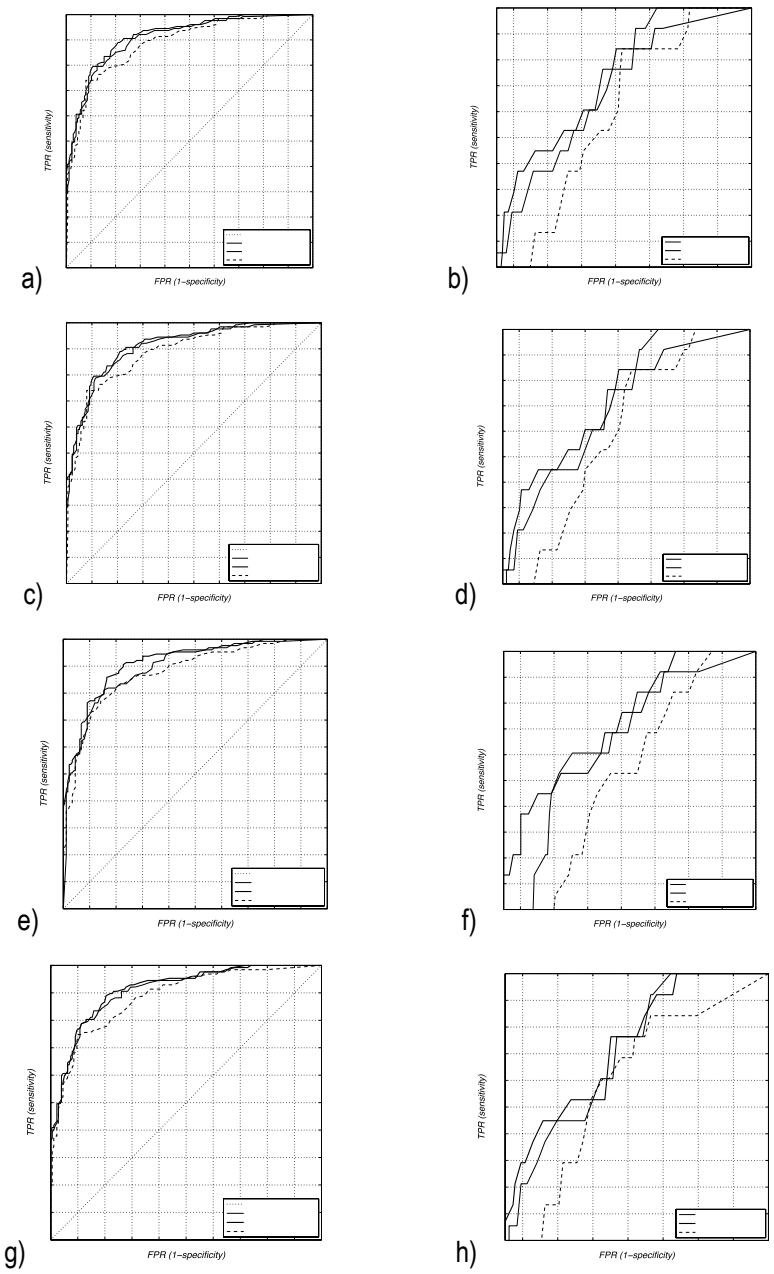

Fig. 4. Performance of SVM with linear, polynimial, RBF, and sigmoid kernels with three variable sets: all attributes (s39); selection of 17 attributes based on the subset size forward selection method (s17) Guetlin et al. [2009]; and selection of 14 attributes proposed by Jesneck et al. [2007]

We also explored how choice of the kernel function influences the SVM predictive abilities. This is particularly important when data belong to classes, which are not linearly separable. In those cases we can expect that the SVM model with linear kernel wouldn't perform well in contrast to the non-linear ones. Table 1 shows the results for each kernel function. The SVM works well with all of them. Considering s17 only, AUC for all four kernels is 91% and $_{0.90}$AUC is from 74% to 75%. Specificity at high sensitivity, however, varies with different kernels. Regarding this metric, the sigmoid kernel outperforms the others, followed by the RBF, linear, and polynomial.

Our findings also could be compared with those from studies that use models based on the most common neural networks - MLPs, given that all methods use the same dataset and variable sets (Nachev & Stoyanov, 2010). SVM shows the same AUC as MLP, but improves $_{0.90}$AUC by 7% (68% vs. 75%) and max accuracy by 2% (83% vs. 85%).

Fig. 4 gives further details on the SVM ROC analysis, The left-hand figures illustrate the ROC curves and AUC for each kernel and variable set; the right-hand figures illustrate the area of sensitivity above 90% and list $_{0.90}$AUC.

Conclusion

This study explores support vector machines utilized as predictors of malignant breast masses, trained and tested with data from mammographic and sonographic examinations. We used data collected from Duke University Medical Centre, which contains 39 descriptors. Our study was focused on two issues: how reduction of dimensionality of the training and testing data affect the discriminatory power of the model; and how choice of the SVM kernel function and model parameters affect its predictive abilities.

In order to quantify the model performance we did ROC analysis and utilized metrics, such as true positive rate, false positive rate, area under the ROC curve, partial area under the ROC curve, and specificity at high sensitivity.

Our results show that the reduction of dimensionality plays a significant role in the model performance. We propose a set of 17 variables, which outperforms the 14 variables set of Jesneck et al. [2007]. The choice kernel function among linear, polynomial, RBF, and sigmoid, however, does not influences the model performance, with exception of one metric - specificity at high sensitivity. In that case the sigmoid kernel is the best performer. The fact that the linear kernel shows similar performance to that of the non-linear kernels is an indication that the feature space is linearly separable and data points are distributed in a way that makes the classification task linear in terms of complexity.

We also found experimentally that the SVM outperform a common classification technique used in the field - MLP neural networks. SVM shows the same AUC, but improves $_{0.90}$AUC by 7% (68% vs. 75%) and max accuracy by 2% (83% vs. 85%).

In conclusion, we believe that SVM is a promising technique for breast cancer diagnosis, but when used, it requires a careful reduction of dimensionality and well-selected model parameters.

Bibliography

[BI-RADS, 2003] American College of Radiology. BI-RADS: ultrasound, 1st ed. In: Breast imaging reporting and data system: BI-RADS atlas, 4th ed. Reston, VA: American College of Radiology, 2003

[Burges, 1998] Burges, C. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2, 121-167, 1998.

[Chen et al., 2009] Chen, S., Hsiao, Y., Huang, Y., Kuo, S., Tseng, H., Wu, H., Chen, D.: Comparative Analysis of Logistic Regression, Support Vector Machine and Artificial Neural Network for the Differential Diagnosis of Benign and Malignant Solid Breast Tumors by the Use of Three-Dimensional Power Doppler Imaging. Korean J Radiol vol. 10, 464-471 2009.

[Drucker et al., 1997] Drucker, H. Burges, C., Kaufman, L., Smola, A., and Vapnik, V., Support vector regression machines, Advances in Neural Information Processing Systems 9, pages 155-161, Cambridge, MA, MIT Press, 1997.

[Fawcett, 2006] Fawcett, T. “An introduction to ROC analysis”; Pattern Recognition Letters, Vol. 27 Issue 8, pp. 861-874, 2006.

[Fei et al., 2008] Fei, L., Li, W. & Yong, H. Application of least squares support vector machines for discrimination of red wine using visible and near infrared spectroscopy. Intelligent System and Knowledge Engineering, ISKE' 08, 2008.

[Goldberg, 1989] Goldberg, D.: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley: Reading, MA, 1989.

[Guetlein et al., 2009] Guetlein, M., Frank, E., Hall, M., Karwath, A. “Large Scale Attribute Selection Using Wrappers”; In Proc. IEEE Symposium on CIDM, pp.332-339, 2009.

[Hall et al., 2009] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten. I., 2009. The WEKA data mining software: an update. SIGKDD Explor. Newsl. 11, 1, 10-18, 2009.

[Horng et al., 2010] Horng, S., Su, M., Chen, Y., Kao, T., Chen, R., Lai, J. and Perkasa, C. A novel intrusion detection system based on hierarchical clustering and support vector machines. Expert Systems with Applications, 38, 306-313, 2010.

[Jesneck et al., 2007] Jesneck, J., Lo, J., Baker, J. “Breast Mass Lesions: Computer-Aided Diagnosis Models with Mamographic and Sonographic Descriptors”; Radiology, vol.244, Issue 2, pp 390-398, 2007.

[Kopans, 1992] Kopans D. “Standardized mammographic reporting”; Radiol Clin North Am, Vol. 30, pp. 257–261, 1992

[Lacey et al., 2002] Lacey, J., Devesa, S., Brinton, L. “Recent Trends in Breast Cancer Incidence and Mortality.”; Environmental and Molecular Mutagenesis, Vol. 39, pp. 82–88, 2002.

[Nachev & Stoyanov, 2010] Nachev, A. and Stoyanov, B., “An Approach to Computer Aided Diagnosis by Multi-Layer Preceptrons”, In Proceedings of International Conference Artificial Intelligence (IC-AI’10), Las Vegas, 2010.

[Oladunni and Singhal, 2009] Oladunni, O. O. & Singhal, G. 2009. Piecewise multi-classification support vector machines. International Joint Conference on Neural Networks, IJCNN'09, 2009.

[Stavros et al., 1995] Stavros, A., Thickman, D., Rapp, C., Dennis, M., Parker, S., Sisney, G. "Solid Breast Modules: Use of Sonography to Destinguish between Benign and Malignant Lesions"; Radiology, Vol. 196, pp. 123-134, 1995.

[Vapnik, 1995] Vapnik, V., The Nature of Statistical Learning Theory. Springer, New York (1995)

Authors' Information

Anatoli Nachev *– Business Information Systems, Cairnes Business School, National University of Ireland, Galway, Ireland; e-mail: anatoli.nachev@nuigalway.ie*

Major Fields of Scientific Research: data mining, neural networks, support vector machines, adaptive resonance theory.

Mairead Hogan *– Business Information Systems, Cairnes Business School, National University of Ireland, Galway, Ireland; e-mail: mairead.hogan@nuigalway.ie*

Major Fields of Scientific Research: HCI, usability and accessibility in information systems, data mining.