

Breast Density Classification: a comparison of Different Classification Schemes

Hooshiar Zolfagharnasab
University of Porto, Porto, Portugal
dee12012@fe.up.pt

Inês Domingues
University of Porto, Porto, Portugal
inesdomingues@gmail.com

Jaime S. Cardoso
University of Porto, Porto, Portugal
jaime.cardoso@inescporto.pt

Abstract—Although the density of breasts is not a disease, it can lead to indiscrimination of some diseases, since it is hard for specialists to discern tumours in dense tissues. In this paper, based on tissue density, a mammogram is classified in four ordinal classes. To do this, an ordinal classification method is implemented with two different machine learning methods, SVM and Neural Network. Experimental results indicate that by using preprocessing followed by Support vector regression one attain better results than the other implemented methods.

Keywords-breast density; classification; Support vector regression

I. INTRODUCTION

Breast density is a radiologic phenomenon dissimilar to the common concept of density which is weight per unit. Breast density is not discernible by palpation but rather relates to the fact that x-rays permeate different types of breast tissue differently. Although high density breasts are not perilous, they will raise the chance of cancer, since tumours would not be clearly visible through dense tissues.

Previously, interpreting mammogram via unprocessed images, specialists used to classify breasts subjectively which increased both diagnosing time and miss-prediction rate. Also such subjective diagnosing prevents comparison between different decisions.

Decision support systems are worthy choice to solve such ill-posed problem objectively. These systems can automatically classify breasts into 4 common classes [1, 2]. Also machine learning algorithms can impress decision support system to have less miss-classification. Additionally, in this paper, we aim to increase the accuracy more than previous by applying ordinal classification.

II. DATA AND METHODS

More than 200 images which are used to extract vectors are from INBreast [3]. Each image is processed due to topographical properties as explained in [2]. The output of each image is a vector with 21 features representing the density of the breast depicted in the image.

Two learning methods are used in this paper, Support Vector Machine Regression (SVR) and Neural Network (NN) Regression. Both are tested on the same data and results are compared.

(1)- Calculating z

$$z_i = \begin{cases} c_1 + (1 - \omega_{x_i^{(1)}}) \cdot \frac{1}{Q} & , \text{if } q = 1 \\ c_q - (1 - \omega_{x_i^{(q)}}) \cdot \frac{1}{2Q} & , \text{if } q \in \{2, \dots, Q-1\} \text{ and } \partial(x_i^{(q)}, q-1) \leq \partial(x_i^{(q)}, q+1) \\ c_q + (1 - \omega_{x_i^{(q)}}) \cdot \frac{1}{2Q} & , \text{if } q \in \{2, \dots, Q-1\} \text{ and } \partial(x_i^{(q)}, q-1) > \partial(x_i^{(q)}, q+1) \\ c_Q - (1 - \omega_{x_i^{(Q)}}) \cdot \frac{1}{Q} & , \text{if } q = Q \end{cases}$$

In each method, two approaches are used: direct learning from feature vectors, and learning via preprocessed value calculated by feature vectors. Preprocessing projects each feature vector into a single value. Then, by using a regressor which could be mapped to classes of training data, learning takes place. Same regressor will be used to predict unknown data which are mapped to value via preprocessing. Evaluation is obtained by Average Mean Absolute Error.

In the other approach, same regressor is used to map full vector (instead of a preprocessed value) to classes. We aim to compare the impact of preprocessing on learning. Such preprocessing seems to be worthy in ordinal classification.

Finally common classification methods are carried out in order to compare the different classification schemes.

Besides, to reduce the correlation effects between features, a technique is used to select best features in order to raise the performance of learning.

A. Feature Selection

After processing the images from the database, each image is described by a 21-feature vector. Using all features from an image in order to learn a classifier, may not lead to expected result, since some features have correlation with the others. Thus, by applying a feature selection method, a set of features are selected in order to decrease feature space.

In feature selection method, all features are checked one by one to find the best set of features. Starting with a set consisting of just the best feature (one feature with least error) two features are added to current set in each further step. Then, calculating error with both previous and added feature, the one which has higher error is omitted. So in each step, two features are added and one is omitted. This procedure continues until all features are checked. At last, the set which has the lowest error is selected to be presented as the result of feature selection [4].

It should be noted that in the feature selection method, a 5-fold cross validation methodology was used.

B. Exploitation of pairwise class distances for ordinal classification

The aim of the paper [5] which is used as preprocessing in this work is: “Striving to project a vector into a value due to class number and distance from adjacent classes”, This value is then used to learn a classifier in order to decide for unknown data.

(2)- Calculating w

$$\omega_{x_i^{(q)}} = \begin{cases} \frac{\partial(x_i^{(q)}, q+1)}{\text{MAX}_{x_n^{(q)}}\{\partial(x_n^{(q)}, q+1)\}} & , \text{if } q = 1 \\ \frac{\partial(x_i^{(q)}, q-1) + \partial(x_i^{(q)}, q+1)}{\text{MAX}_{x_n^{(q)}}\{\partial(x_n^{(q)}, q-1) + \partial(x_n^{(q)}, q+1)\}} & , \text{if } q \in \{2, 3, \dots, Q-1\} \\ \frac{\partial(x_i^{(q)}, q-1)}{\text{MAX}_{x_n^{(q)}}\{\partial(x_n^{(q)}, q-1)\}} & , \text{if } q = Q \end{cases}$$

In this method, distance (Euclidean) of each pattern (or vector) within other instances of adjacent classes (two classes) is calculated and normalized. The difference between each class is magnified by adding the information about the center of each class to the current value. Finally Equation 1 is used to calculate the mentioned ‘z’ value.

By calculating ‘z’ as proposed in [5], each vector has one value as identification of both class number and feature value. This calculation is called preprocessing in the current paper.

C. Performance Measurement

Mean Absolute Error (*MAE*) is a metric which is the average of absolute deviation of the predicted classes [2, 5, 6]. Current metric ranges from 0 to Q-1, indicating the deviation among predicted class and reality. It can be obtained by Equation 3.

$$MAE = \frac{1}{N} \sum_{i=1}^N e(x_i) = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (3)$$

Since each class contains different number of vectors, it is necessary to use a metric which would be robust to quantity. Such robust metric can be Average *MAE* or *AMAE* as it is presented in Equation 4.

$$AMAE = \frac{1}{Q} \sum_{j=1}^Q MAE_j = \frac{1}{Q} \sum_{j=1}^Q \frac{1}{n_j} \sum_{i=1}^{n_j} |\hat{y}_i - y_i| \quad (4)$$

Both metrics range from 0 to Q-1 while *AMAE* indicates how far all predicted classes deviate from actual classes. Results near to zero indicate that predicted classes have less deviation from reality [5, 6]. In this paper, confronting with an ordinal classification *AMAE* is considered as performance metric.

III. IMPLEMENTATION

Using both SVM and Neural Network (NN), a system is trained to classify breast via two approaches as explained before. Algorithms are implemented in a machine running MATLAB R2012b.

A. Support Vector Machine

Support Vector Machine (SVM) is used in this paper with three approaches. In the first approach, by using nu-SV Regression (SVR) and epsilon-SV Regression (SVR), the explained preprocessing is used to find a suitable regressor.

The second approach is carried out without using preprocessing. In this approach, a regressor is used to map vectors into corresponding classes directly. Using this regressor, unknown data are mapped directly to the correct classes from feature vectors.

Finally, a third approach is performed by learning directly from vector space into classes. First, a model is trained with a

TABLE I RESULTS COMPUTED BY SVM-BASED METHOD

		AMAE Before Feature Selection *	AMAE After Feature Selection *
With z	epsilon-SVR	1.00	0.57
	nu-SVR	1.00	0.59
Without z	epsilon-SVR	1.00	0.64
	nu-SVR	0.99	0.70
	SVM	1.00	0.76

*. Remember that metric ranges between 0 and 3, smaller values means less deviation from actual classes.

training set and then using test set different from training, the learning process is qualified. Reported results in Table I represent both *AMAE*, and the impact of feature selection in different SVM approaches.

B. Neural Network

Neural networks (NN) are also used to predict labels for each class. As explained before, three approaches are used, namely, NN with and without preprocessing and NN classification. Unlike SVM, regression without preprocessing has less error. Results are reported in Table II.

IV. CONCLUSION

Since mammograms with high density tissues have low visibility, tumours may not be detected correctly. It seems to be worth to train a system which can classify mammographic images into four classes, so specialists can decide to check some images more accurately.

This paper contains comparison between two learning methods to classify breasts tissue density; SVM and NN. For each method, the same preprocessing described in [5] is performed and results are compared. It is shown that using combination of preprocessing and *epsilon-SVR* will result the least average mean absolute error. *nu-SVR* took second place in ranking among implemented methods. Third rank among regression methods belongs to NN regression method.

To have a full comparison, a traditional classification is performed without preprocessing. Same as before, the best set of features are selected for both SVM and NN. As it is reported in Table II, since different model which is trained in each running, an approximate error is obtained after 100 iterations. But using average error, it is concluded that NN performs better when there is no preprocessing.

All in all, it is concluded that using preprocessing in this paper will lead to better results. Also, it was shown that feature selection has a positive impact on the accuracy.

REFERENCES

- [1] American College of Radiology, American College of Radiology Breast Imaging Reporting and Data System (BIRADS), American College of Radiology, 4th Edition, 2003.
- [2] J. Chen et al., Automated mammographic risk classification based on breast density estimation, *submitted*, 2012.
- [3] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, ‘INbreast: Toward a Full-field Digital Mammographic Database’, *Academic Radiology*, 2012: 19(2), pp. 236–248,.
- [4] M. Sadeghi, S. Khoshrou, J. Kittler: SVM-Based Selection of Colour Space Experts for Face Authentication. ICB 2007: 907-916
- [5] J. Sanchez-Monedero, P. A. Gutiérrez, P. Tino, and C. Hervás-Martínez, ‘Exploitation of pairwise class distances for ordinal classification’, *Neural Computation*, accepted, 2013.
- [6] J. F. Costa, H. Alonso, J. S Cardoso, The unimodal model for the classification of ordinal data, *Neural Netw.* 2008: 21(1): 78-91.

TABLE II RESULTS COMPUTED BY NEURAL NETWORK-BASED METHOD

		AMAE Before Feature Selection *	AMAE After Feature Selection *
With z	Neural Network Regression	1.40	0.84
	Neural Network Classification	0.96 – 0.69 **	0.54 – 0.76 **

*. Remember that metric ranges between 0 and 3, smaller values mean less deviation from actual classes.
 **. Different running resulted in different values.