

Magnetic Resonance Images Classification through Relevance Vector Machine

Zia Ur Rahman^{*1}, Izaz Ahmad Khan¹, Safyan Mukhtar², Shah Suhail², Muhammad Safdar¹,
Izaz ur Rahman³

¹Department of Computer Science, Bacha Khan University, Charsadda, KPK, Pakistan

²Department of Mathematics & Statistics, Bacha Khan University, Charsadda, KPK, Pakistan

³Department of Computer Science, Abdul Wali Khan University, Mardan, KPK, Pakistan

Received: October 3, 2016

Accepted: December 11, 2016

ABSTRACT

Magnetic Resource Images (MRI) are crucial for the identification of human brain diseases. The correct segmentation of the MR images into various tissues group e.g. gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) is helpful in diagnosis of many brain diseases which includes neurodegenerative disorders such as Alzheimer disease, movements disorders such as Parkinson or Parkinson related syndrome, white matter metabolic or inflammatory disease, congenital brain malformations or perinatal brain damage and post traumatic syndrome. The automatic segmentation of MR images is a prolonged dilemma mainly because of the inhomogeneity bias field in the image. There are many classification techniques used but most suffers from the intensity-based classification. Support Vector Machines (SVM) are used for the classification of brain tissues because of the efficient results shown in many pattern recognition system and use of Kernel functions which results in high generalization ability. The motivation of using Relevance Vector Machines (RVM) for the MR images comes from the fact that it has similar functional form as that of SVM while using less Kernel functions by adopting Bayesian approach which gives probabilistic outcomes and giving sparser results. In this research SVM has been used for the classification of MR images through combination of LIBSVM and SVMLITE. The results obtained through the experiments is an improvement over past.

KEYWORDS: Magnetic Resonance Image, Relevance Vector Machine (RVM), Support Vector Machines (SVM), Neural Network, LIBSVM, SVMLITE.

1. INTRODUCTION

Magnetic Resonance imaging is an important procedure for the identification of the human brain diseases. MR image segmentation has been used for many diagnosing purposes but automatic segmentation of MR scans is of significant value in research and clinical study of neurological pathology. The correct segmentation of the MR images into various tissues group e.g. gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) is helpful in diagnosis and prognosis of many brain diseases which includes neurodegenerative disorders such as Alzheimer disease, movements disorders such as Parkinson or Parkinson related syndrome, white matter metabolic or inflammatory disease, congenital brain malformations or perinatal brain damage and post traumatic syndrome. The automatic segmentation of MR images is a prolonged dilemma mainly because of the inhomogeneity bias field in the image. There are many classification techniques used but most suffers from the intensity-based classification [1].

Image segmentation is extremely important part of image processing and pattern recognition. In image segmentation an image is separated into dissimilar regions in such a way that there is nothing common in them. There are some important techniques for image segmentation such as histogram thresholding, characteristic feature clustering, edge detection, region-based methods, fuzzy techniques, neural networks etc [2], support vector machines [3] and relevance vector machines [4].

Classification is the mechanism in which separation of many objects is performed which are then allotted to one of a number of mutually exhaustive and exclusive groups called classes e.g. customers who are more probable or not probable to buy an item from a shop [5]. Support Vector Machines (SVM) are used for the classification of brain tissues because of the efficient results shown in many pattern recognition system and use of Kernel functions which results in high generalization ability. Back Progration Multi-Layer Perceptron (BP-MLP), a supervised learning methodology and Fuzzy C-Mean, an unsupervised learning methodology are the two commonly used techniques for the classification of brain tissues in Magnetic Resonance Images (MRI) but the SVM shows considerable improvements over them [6]. Similarly, many techniques are used for the image segmentation e.g. Edge Detection, Histogram Thresholding, Feature space, Clustering, Region Based methods and Neural Networks etc. The SVM is used for image segmentation of scenery images which outperforms other methods because of the Structural Risk Minimization [7].

The motivation of using Relevance Vector Machines (RVM) for the MR images comes from the fact that it has similar functional form as that of SVM while using less Kernel functions by adopting Bayesian approach which gives probabilistic outcomes and giving sparser results [4].

For a very long time, humans wanted to develop a system that learns from experience with two aspects, the technical aspect benefits from the development in electronic computers to help in the ability. In real world we have problems which cannot be solved by classical programming as no mathematical model exist for the solution of such problem e.g. how to write a program for computer to recognize hand written “A” just like humans learn to read [3]. This is known as learning methodology. Other such examples are finding a gene in DNA sequence, filtering email, detecting or recognizing objects in machine vision. The key to all these examples is machine learning algorithms.

2. MATERIALS AND METHODS

Neural Networks: The main advantage of computer over human is that computers are much faster than humans but there are certain areas in which human perform well especially in speech and image recognition. So, a human like performance was desired from computers which led to the development of Neural Networks (NN) which are based on the neurons in human brain. In 1940s first neural model was developed by McCulloch and Pits then a solution to simple classification problems was given by Rosenblatt through Perceptron Model in 1962. Now there are powerful NN with better training algorithms and whose performance is boosted by improved hardware. NN is made of highly interconnected nodes which similar to human neurons structure.

Each NN is made of a nodes where a unit is given as U while X_i is the real valued input to U where $x_i = (x_1, x_2, \dots, x_n)$. A weight is connected to each input which shows the strength of the connection between the connected nodes where $w_i = (w_1, w_2, \dots, w_n)$ is a real valued number which shows weight vector corresponding to X_i . When these weighted inputs are applied to U then it produces a net sum given as,

$$s = \text{sum}(w_i * x_i)$$

$$s = w.v \quad (1)$$

This allows weight to be modified dynamically. The activation value of U is represented by a numeric value A which shows its state while the new activation value of the node is calculated from the activation function f which is the net sum and current activation of the node. A is given as

$$A = f(s) \quad (2)$$

The weights are connected indirectly. The field of NN is further subdivided into two types which are Single-layer Linear Networks and Multilayer Networks. In Single-layer Linear a number of nodes are arranged in single layer where each node receives a weighted input. Multilayer Networks has more than one layer where output from one layer serves input for other.

Long training time used to be the major disadvantage which is reduced by improved hardware and now NN is used in many applications for example speech and image recognition because it offers high level of parallelism.

- Ford motor company is using it for the identification of problems in the engine which has shown better result than the manual.
- Banking sector is using it for assessing the value of credit cards.
- Used for the classification of a given input into different classes and has shown better result even with the noisy data.
- Boeing Aircraft Company has used it for time cutting and saving money in the production process.
- Used in the biomedical research for analysis and classification of outcome of experiments

Relevance Vector Machines (RVM): RVM is a Bayesian form of the SVM. The over fitting is a major challenge for the modelling of classification and regression problems especially when regression is performed on noisy data and class overlap in classification [8]. SVM gives a solution for the over fitting and gives good generalisation by minimising number of errors on training set and maximising the margin between the two classes in a feature space defined by the Kernel functions [4]. SVM gives sparse model dependent only on a subset of Kernel functions which are associated with those who lies on margin or wrong side of it but the results given by SVM are not probabilistic, make a very liberal use of the Kernel functions, estimation of the error/margin trade-off parameter ‘C’ (insensitivity parameter ‘ ϵ ’ in regression) is required which in turns requires data and computation procedure of cross validation and need Mercer’s condition to be satisfied by a Kernel function. RVM solves these problems by using Bayesian approach which introduces a prior by the set of hyper parameters on the weights attached with a one to one correspondence whose probability is estimated iteratively from data [4]. RVM gives sparser results because the posterior distribution is sharply peaked around zero, also the nonzero weights shows prototypical example of classes which is a difference from the SVM where they are associated with decision boundary, these examples are known as relevance vectors. Another advantage of RVM is the use of less Kernel functions while producing same level of efficiency. The RVM is further subdivided into Relevance Vector Regression (RVR) and Relevance Vector Classification (RVC). The research is based on image segmentation so only RVC is explained.

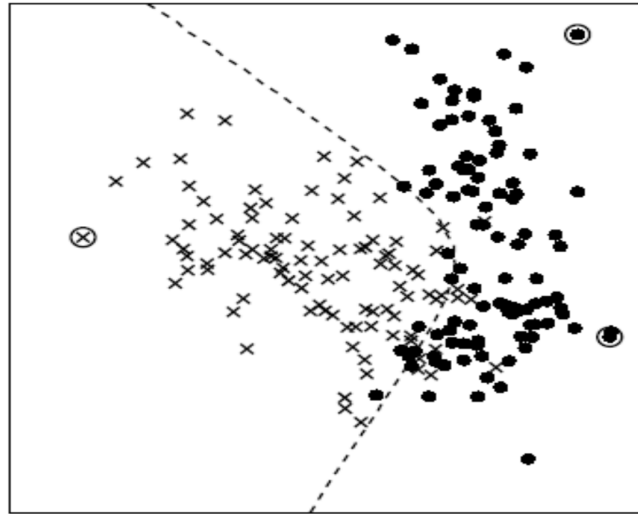


Fig 1: Relevance Vector Classification function [8]

Classification: In Relevance Vector Classification (RVC) the class probability of class membership is predicted given the input X [8]. The linear model is generalized by applying logistic sigmoid function $\sigma(y) = 1 / (1 + e^{-y})$ to $y(x)$ and the probability is given as 10

$$P(t | w) = \prod_{n=1}^N \sigma\{y(x_n)\}^{t_n} [1 - \sigma\{y(x_n)\}]^{1-t_n} \quad (3)$$

An interactive method based on Mackay is used for finding probability because weights cannot be integrated through analytical method. The current most likely weights WMP for current fixed value of α is find out which is same as the standard optimisation of a regularised logistic model and efficient iteratively reweighed least-square algorithm to find maximum.

Hessian at WMP is calculated as:

$$\nabla \nabla \log p(t, w | \alpha) |_{w_{MP}} = -(\Phi^T B \Phi + A)_{As} \quad (4)$$

$$B_{nn} = \sigma\{y(x_n)\} [1 - \sigma\{y(x_n)\}] \quad (5)$$

This is negated and inverted to give covariance \sum for a Gaussian approximation to posterior over weights and from hyper parameter is α are updated using:

$$\alpha_i^{new} = \frac{\gamma_i}{\mu_i^2} \quad (6)$$

There is no noise variance α^2 . This procedure continues until a convergence criterion is satisfied. Another major advantage of RVC is its standard form as a probabilistic generalised linear model due to which it can be extended directly to the multi-class case while in SVM it is not that easy because training is needed. The estimation of posterior probabilities of class membership is also an advantage because it gives a principle way of measuring the uncertainty of prediction and necessary for adaptation of varying class priors along with incorporation of asymmetric misclassification costs. The major disadvantage of the RVM is the complexity of the training phase because it continuously compute and reverse Hessian matrix due to which storage time is $O((N^2))$ and computing time is $O((N^3))$ [8].

Algorithm: The algorithm is divided into three major parts. In the first part the suitable Kernel function is selected, associated parameters are supplied to that Kernel function and the value of the regularization parameter is

supplied. In the second part of the algorithm training is performed in which both input and output pairs are supplied to algorithm which is then mapped to output. After training of the algorithm testing is performed. In testing phase only unseen inputs are provided which are mapped into output on the basis of the learning performed during the testing phase.

The data provided to the algorithm as input are MR images taken by MR machines. The output is one of the three types of tissues or background. So, the output can be any one of the four types while input is in the form of pixels. The data will be divided into two equal parts. One part will be used for training purposes and the second part will be used for the testing purpose.

First part: The algorithm uses types of kernel functions. Kernel functions help in performing operations in the input space rather than potentially high dimensional feature space. First the type of the kernel function is selected then the values of the associated parameters are selected after which the values of the regularization parameter “C” are supplied as the result depends upon its value.

Second Part: This information is passed to the training phase where input and its associated output pair is supplied for the training of the algorithm. The input and its associated output are supplied as the algorithm is based on the principal of supervised learning methodology. So, in this part of algorithm both the data as pixels and output as type of tissues and background is supplied.

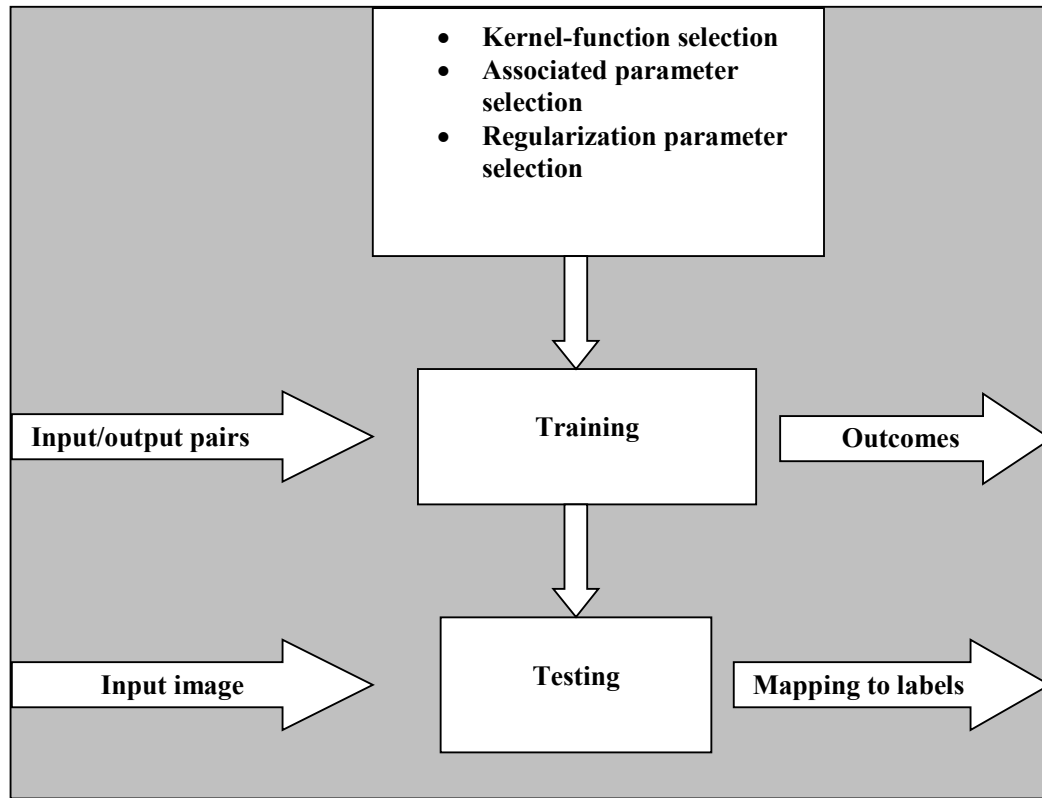


Fig 2: Classification Algorithm

Third part :When algorithm is trained then only input image data is provided and the algorithm map the value of each input pixel to the most optimum label or output which will be one out of the four types of labels. The testing is performed on the basis of the prior knowledge already learnt from training part through examples.

3. RESULTS AND DISCUSSIONS

The data is a simulated MR image. The simulated MRI volumes for normal brain are based on the anatomical model of normal brain which is the base for any analysis procedure. This technique uses four parameters. These are three modalities that are T1, T2 and PD, second is five slice thicknesses with 1, 3, 5, 7, 9 mm, third parameter is the six level of noise that includes 0%, 1%, 3%, 5%, 7%, 9%, while last parameter is the three levels of intensity non-uniformity that is 0%, 20% and 40%.

The size of the image volume is 181 x 217 x 60 pixels whose thickness is 3mm which is simulated at high resolution T1-weighted 2D scans with 3% noise level and 40 % intensity non-uniformity. The classifiers will determine which one of the brain tissues (WM, GM, and CSF) and the background the pixel belongs to. The training of the datasets is

performed by assigning each class a label (0: background; 1: WM; 2: GM; 3: CSF). The training is performed by randomly sub-sampling a 2dimensional 181 x217 image. The testing is performed by taking data from the same 2dimensional MR image and from other 2dimensional MR images. There are large numbers of background pixels as compared to the three brain tissues in the labeled image due to which the training set for the background class can be impractically large. This problem is solved by randomly selecting four classes of training samples of 2dimensional image at different random proportion to keep the number of training examples of each class approximate.

- The data is collected from the McConnell Brain Image Centre, Montreal Neurological Institute, McGill University which is available at <http://www.bic.mni.mcgill.ca/brainweb/> [6].
- MR images of brain: A data set of the MR images of the brain which is a test collection used as a resource for research in classification of the brain tissues.

The solution to the validation problem is provided by simulated brain database, which contains a set of the realistic MRI data volumes generated by an MRI simulator. Neuroimaging community uses this data to evaluate the performance of various image analysis methods. The simulated brain MRI data is based on two anatomical models i.e. normal and multiple sclerosis (MS).

Sample Data Format

Modality=T1, Protocol=ICBM, Phantom_name=normal, Slice_thickness=3mm, Noise=3%, INU=40%

MINC volume info:

Image: signed__ short 0 to 4095

Image dimensions: zspace yspace xspace

Table 1: Data Dimensions

Dimension name	Length	Step	start
Zspace	60	3	-72
Yspace	217	1	-126
Xspace	181	1	-90

The data is represented in three formats i.e. MINC, rawbyte (unsigned) and raw short (12 bit) but MINC are used as it is compatible with the MATLAB while in all the three data formats data is binary. This format is developed at the McConnell Brain Imaging Center (McBIC) at the Montreal Neurological Institute (MNI). Following are some of the packages for working with the MINC files:

The files in the MINC format have both data and header information. A sample header is shown below.

For example, the displayed header info:

Image dimensions: zspace yspace xspace

Table 2: Data Dimensions

Dimension name	Length	Step	start
Zspace	36	5	-72
Yspace	217	1	-126
Xspace	181	1	-90

Should be interpreted as follows:

- The file scans the 3D image volume such that the 'X' coordinates changes fastest and the 'Z' changes slowest.
- The image sizes along the X-Y-Z axes are 181x217x36 voxel (pixels).
- The voxel sizes along the X-Y-Z axes are 1x1x5 mm.

Using this header info, you should be able to correctly read the raw volumes on any kind of computer platform out there (using C, Matlab, etc). There is a function known as `getimageinfo` (handle, whatinfo), which is used for getting information related to the image. Its handle parameter is used for an already open image or can be created with a new image while the whatinfo parameter is used which describe the type of information which the user wants from a particular handle with many possible values. It can take one of the standards MINC image dimensions i.e. time, zspace, yspace or xspace and give the length of the dimension or zero in case the dimension does not exist. The time is equivalent to NumFrames while the entire three spatial dimensions have complex equivalences. Zspace is NumSlices, yspace is Image Height and xspace is ImageWidth for transverse images.

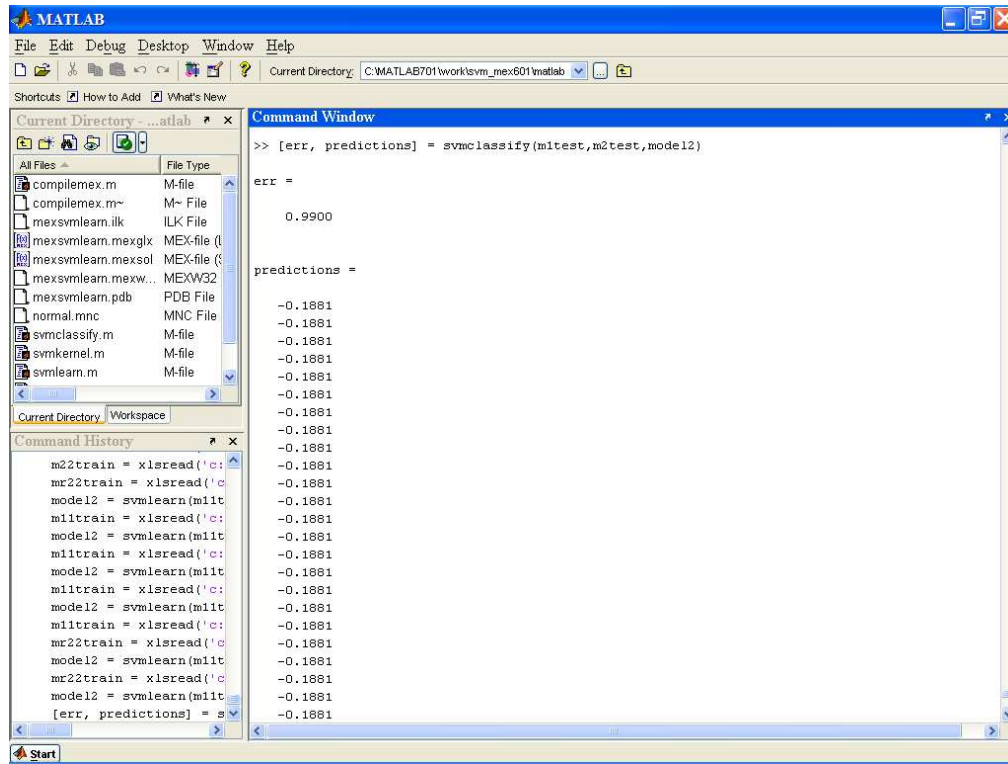


Fig 3. MATLAB snapshot

The fig 3 shows the err as 0.9900 which is the error, so if we take error rate it will be 9.9 % which is an much better as compared to the previous classification error rate of SVM, BP-MLP and FCM techniques with 10.1%, 18.1% and 13.2% [6]. It has also given the predictions.

Table 3: Different classification techniques and classification error rate

Different techniques	SVM (LIBSVM + SVMLITE)	SVM	BP-MLP	FCM
Classification error rate	9.9%	10.1%	18.1%	13.2%

4. FUTURE WORK

Then the SVM and RVM are compared with each other. The comparison has given special attention to the superior features of the RVM over SVM. The RVM is a probabilistic approach which uses less kernel functions due to which it can give better classification of the brain tissues in the MR images. An algorithm has been proposed for image segmentation through RVM, but due to the unavailability of RVM software and lack of time it is not implemented. So, for future research the same technique should be implemented through RVM based software that can give better classification results. Currently, LIBSVM has been used which is non-probabilistic software along with SVMLITE.

5. CONCLUSIONS

There are many harmful brain diseases in human i.e. neurodegenerative disorders such as Alzheimer disease, movements disorders such as Parkinson or Parkinson related syndrome, white matter metabolic or inflammatory disease, congenital brain malformations or perinatal brain damage and post traumatic syndrome. MR imaging is an important procedure for the identification of the human brain diseases. MR image segmentation has been used for many diagnosing purposes but automatic segmentation of MR scans is of significant value in research and clinical study of neurological pathology. The correct segmentation of the MR images into various tissues group e.g. gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) is helpful in diagnosis of many brain diseases. The work described has implemented the classification of the brain tissues in the MR image into four classes i.e. GM, WM, CSF and background with each one has been assigned a different class label

REFERENCES

- [1] Gu, Jian-Wen; Lin, Pan; Yang, Yong; Zheng, Chong-Xun (2005) Full automatic framework for segmentation of MR brain image: *Journal of Computer Science & Technology*
- [2] Cheng, H. D., Jiang, X. H., Sun, Y., Wang, J.: Color image segmentation: advances and prospects, *Pattern Recognition*, 34(12), 2001, 2259-2281.
- [3] Cristianini, N. & Taylor, J.S. (2006) Support vector machines and other kernel -based learning methods. Cambridge: Cambridge University press
- [4] Tipping, M.P. & Smola, A. (2001): Sparse Bayesian learning and the Relevance Vector Machine: *Journal of Machine Learning Research*. 1 211-244
- [5] Bramer, M (2007) what is classification? *Principles of Data Mining* springer-verlag London pp. 23
- [6] Zhang, X., Xiao, X.L., Tian, J.W., Liu, J., & Xu, G.Y.: Application of Support Vector Machines in classification of magnetic resonance images, *International Journal of Computers and Applications*, 28(2), 2006 122-128
- [7] Yu, Y., Chang, C: Scenery Image Segmentation Using Support Vector Machines, *Fundamental Informaticae* 61 (2004) 379–388
- [8] Tipping, M.P., (2000), The Relevance Vector Machine: Advances in Neural Information Processing Systems, San Mateo, CA, Available at: <http://citeseer.ist.psu.edu/tipping00relevance.html>