ISSN : 2393-8188 (print)
2393-8196 (online)
www.milliyasrcollege.org.journal.php

CLASSIFICATION OF BRAIN TUMOR BY FEATURE VECTOR USING CART AND C5.0 ALGORITHMS

Kailash D. Kharat*, Anil V. Turukmane*

Assistant Professor, Department of Computer Science, Dr. BAMU University, Aurangabad, Maharashtra, India

*Corresponding author email: kailashdkharat@yahoo.co.in

* *Both the authors contributed equally*.

ABSTRACT: This paper describes a Naive Bayesian classifier (NBC) that simply uses only those features that C5.0 would use in its decision tree when learning a small example of a training set, a combination of the two different natures of classifiers. The brain tumor image Classification is a difficult task due to the variance and complexity of tumors. This paper uses data mining classification algorithms-- C5.0, CART and Naïve Bayesian classifier algorithms to get useful information to decision-making of brain tumor classification and tumor grade and tumor behaviors. These Decision Tree techniques can both be applied in the tumor classification with grade model and can obtain a quite accurate result for the classification of the magnetic resonance human brain images. The Decision Tree technique consists of three stages, namely feature extraction, dimensionality reduction, and classification. In the first stage, we have obtained the features related with MRI images using discrete wavelet transformation (DWT). In the second stage, the features of magnetic resonance images (MRI) have been reduced using principles component analysis (PCA) to the more essential features. In the classification stage, two classifiers based on supervised machine learning have been developed. The first classifier C5.0 algorithm and the second Classification and Regression Trees (CART) algorithm. The classifiers have been used to classify subjects as normal or abnormal MRI brain images. The classification analysis is by analyzing the data in the demonstration database, to make the accurate description or establish the accurate model or mine the classifying rule for each category, and then use the classifying rule to classify records in other databases.

KEYWORDS: MRI; Feature Extraction; Feature Selection; Tumor Classification; C5.0decision tree, CART decision tree algorithm

1. INTRODUCTION

Early detection and classification of brain tumors is very important in clinical practice. Many researchers have proposed different techniques for the classification of brain tumors based on different sources of information. In this paper we propose a process for brain tumor classification, focusing on the analysis of Magnetic Resonance (MR) images and Magnetic Resonance Spectroscopy (MRS) data collected for patients with benign and malignant tumors. Our aim is to achieve a high accuracy in discriminating the two types of tumors through a combination of several techniques for image segmentation, feature extraction and classification. The proposed technique has the potential of assisting clinical diagnosis.

Necessary preprocessing steps prior to characterization and analysis of regions of interest (ROIs) are segmentation and registration. Image registration is used to determine whether two subjects have ROIs in the same location. However, in this work we do not take into account the location of the tumor in the classification model so we do not employ registration. Image segmentation is required to delineate the boundaries of the ROIs ensuring, in our case, that tumors are outlined and labeled consistently across subjects. Segmentation can be performed manually, automatically, or semi-automatically. The manual method is time consuming and its accuracy highly depends on the domain knowledge of the operator. Specifically, various approaches have been proposed to deal with the task of segmenting brain tumors in MR

images. The performance of these approaches usually depends on the accuracy of the spatial probabilistic information collected by domain experts. In previous work, we proposed an automatic segmentation algorithm that is based on the fuzzy connectedness concept. The main idea is to assign to every pair of voxels, *x*, *y*, in the image, a real number between 0 and 1 indicating their connectedness. Starting with several seed points, all the voxels are automatically assigned to the structure to which they have the highest connectedness value. Utilizing the statistical information cumulated during the segmentation process, this method can provide satisfying results even in cases where the boundaries of the ROIs cannot be easily identified. Two of the most widely used and successful methods of classification are C4.5 decision trees and Naïve Bayesian learning (NB). While C4.5 constructs decision trees by using features to try and split the training set into positive and negative examples until it achieves high accuracy on the training set, NB represents each class with a probabilistic summary, and finds the most likely class for each example it is asked to classify.

Several researchers have emphasized on the issue of redundant attributes, as well as advantages of feature selection for the Naïve Bayesian Classifier, not only for induction learning. Pazzani explores the methods of joining two (or more) related attributes into a new compound attribute where the attribute dependencies are present. Another method, Boosting on Naïve Bayesian classifier [10] has been experimented by applying series of classifiers to the problem and paying more attention to the examples misclassified by its predecessor. However, it was shown that it fails on average in a set of natural domain. Langley and Sage use a wrapper approach for the subset selection to only select relevant features for NB. Cardie [5] uses the attributes from decision trees in combination with nearest neighbor methods. This leads to improved classification accuracy.

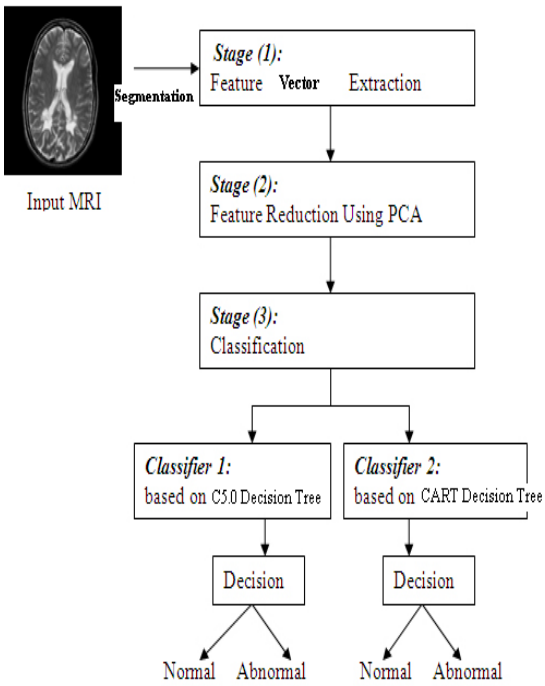

Figure 1. Proposed Work Model

2. METHODOLOGY

There are four major steps in the proposed approach for brain tumor classification: (a) ROI segmentation: delineating the boundary of the tumor (ROI) in an MR image; (b) feature extraction: getting meaningful features of the ROI identified in the previous step; (c) feature selection: removing the redundant features; (d) classification: learning a classification model using the features.

A. Segmentation

Within the segmentation process, each image region confined by a rectangular window is represented by a feature vector of length **R**. These vectors computed for **Q** selected regions are organized in the pattern matrix $\mathbf{P}_{R,Q}$ and form clusters in the R-dimensional space. The **Q** pattern vectors in **P** are fed into the input NN layer, while the number C of the output layer elements represents the desired number of segmentation classes. In each epoch of the network training process, the network weights WC,R are recalculated by minimizing the distances between each input pattern vector and the corresponding weights of the winning neuron characterized by its coefficients closest to the current pattern. In case that the process is successfully completed, the network weights belonging to separate output elements represent typical class individuals. In this paper, the region segmentation process comprises of training the NN on all image regions extracted by a rectangular sliding window with half overlap, and subsequent exploitation of the trained network for region classification. The algorithm comprises of the following successive steps:

1. Feature vectors computation to create the feature matrix P using the sliding window
2. Initialization of the learning process coefficients and the network weights matrix W
3. Iterative application of the competitive process and the Kohonen learning rule [10] for all feature vectors during the learning stage
4. NN simulation to assign class numbers to individual feature vectors
5. Evaluation of the regions classification results

B. Feature Extraction

The proposed system uses the Discrete Wavelet Transform (DWT) coefficients as feature vector. The wavelet is a powerful mathematical tool for feature extraction, and has been used to extract the wavelet coefficient from MR images. Wavelets are localized basis functions, which are scaled and shifted versions of some fixed mother wavelets. The main advantage of wavelets is that they provide localized frequency information about a function of a signal, which is particularly beneficial for classification. A review of basic fundamental of Wavelet Decomposition is introduced as follows:

The continuous wavelet transform of a signal x(t), square-integrable function, relative to a real-valued wavelet, (t) is defined as:

(1)

$$W\psi(a,b) = \int_{-\infty}^{\infty} f(x) * \psi_{a,b}(t)dx$$

Where $\psi_{a,b}(t) = 1/\sqrt{|a|}$

and the wavelet $\Psi_{a,b}$ is computed from the mother Ψ wavelet by translation and dilation, wavelet, a the dilation factor and b the translation parameter (both being real positive numbers). Under some mild assumptions, the mother wavelet Ψ satisfies the constraint of having zero mean.

The eq. (1) can be discretized by restraining a and b to a discrete lattice ($a = 2^b$; a € R_+; b € R) to give the discrete wavelet transform (DWT). The discrete wavelet transform (DWT) is a linear transformation that operates on a data vector whose length is an integer power of two, transforming it into a numerically different vector of the same length. It is a tool that separates data into different frequency components, and then studies each component with resolution matched to its scale. DWT can be expressed as.

(2)

$$DWT_{x(n)} = \begin{cases} dj,k = \sum (x(n)h^* j(n-2\,jk)) \\ dj,k = \sum (x(n)g^* j(n-2\,jk)) \end{cases}$$

The coefficients $d_{j,k}$, refer to the detail components in signal x(n) and correspond to the wavelet function, whereas $a_{j,k,}$ refer to the approximation components in the signal. The functions h(n) and g(n) in the equation represent the coefficients of the high-pass and low-pass filters, respectively, whilst parameters j and k refer to wavelet scale and translation factors. The main feature of DWT is multiscale representation of function. By using the wavelets, given function can be analyzed at various levels of resolution. Fig. 2 illustrates DWT schematically. The original image is process along the x and y direction by h(n) and g(n) filters which, is the row representation of the original image. As a result of this transform there are 4 sub band (LL, LH, HH, HL) images at each scale. (Fig.2). Sub band image LL is used only for DWT calculation at the next scale. To compute the wavelet features in the first stage, the wavelet coefficients are calculated for the LL sub band using Harr wavelet function.

C. Feature Selection and Reduction

One of the most common forms of dimensionality reduction is principal components analysis. Given a set of data, PCA finds the linear lower-dimensional representation of the data such that the variance of the reconstructed data is preserved. Using a system of feature reduction based on a combined principle component analysis on the feature vectors that calculated from the wavelets limiting the feature vectors to the component selected by the PCA should lead to a n efficient classification algorithm utilizing supervised approach. So, the main idea behind using PCA in our approach is to reduce the dimensionality of the wavelet coefficients. This leads to more efficient and accurate classifier.

The feature extraction process was carried out through two steps: firstly the wavelet coefficients were extracted by the DWT and then the essential coefficients have been selected by the PCA.

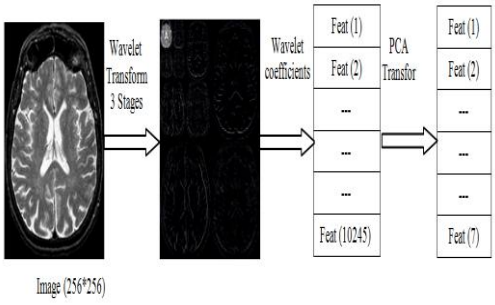

Figure3: Schematic diagram for the used feature extraction and reduction scheme

3. MODEL LEARNING

A. *Naïve bayesian classifier decision tree algorithm*

The Naïve Bayesian classifier is a straightforward and frequently used method for supervised learning. It provides a flexible way for dealing with any number of attributes or classes, and is based on probability theory. It is the asymptotically fastest learning algorithm that examines all its training input. It has been demonstrated to perform surprisingly well in a very wide variety of problems in spite of the simplistic nature of the model. Furthermore, small amounts of bad data, or "noise," do not perturb the results by much.

The Naïve Bayesian classification system is based on Bayes' rule and works as follows. There are classes, say C_k for the data to be classified into. Each class has a probability $P(C_k)$ that represents the prior probability of classifying an attribute into C_k; the values of $P(C_k)$ can be estimated from the training dataset. For n attribute values, v_j, the goal of classification is clearly to find the conditional probability $P(C_k | v_1 \wedge v_2 \wedge \dots \wedge v_n)$.

For classification, the denominator is irrelevant, since, for given values of the v_j, it is the same regardless of the value of C_k. The central assumption of Naïve Bayesian classification is that, within each class, the values v_j are all independent of each other. Then by the laws of independent probability,

$P(v_i | \{\text{all the other values of } v_j\}, C_k) = P(v_i | C_k)$ and therefore

$P(v_1 \wedge v_2 \wedge \dots \wedge v_n | Ck) = P(v_1 | C_k)P(v_2 | C_k)\dots P(v_n | C_k)$.

Each factor on the right-hand side of this equation can be determined from the training data, because (for an arbitrary v_i),

$$P(v_i | C_k) \approx [\#(v_i \wedge C_k)] / [\#(C_k)]$$

Where "#" represents the number of such occurrences in the training set data. Therefore, the classification of the test set can now be estimated by

$P(C_k | v_1 \wedge v_2 \wedge \dots \wedge v_n)$ which is proportional to $P(C_k)\, P(v_1 | C_k)\, P(v_2 | C_k)\, P(v_3 | C_k) \dots P(v_n | C_k)$.

B. *C5.0 Decision Tree Algorithm.*

C5.0 algorithm is the algorithm in the Clementine decision tree model. C5.0 is the classification algorithm which applies in big data set. C5.0 is better than C4.5 on the efficiency and the memory.

The C5.0 model can split samples on basis of the biggest information gain field. The sample subset that is get from the former split will be split afterward. The process will continue until the sample subset cannot be split and is usually according to another field. Finally, examine the lowest level split, those sample subsets that don't have remarkable contribution to the model will be rejected or the trimmed.

Winnowing attributes

The decision trees and rule sets constructed by C5.0 do not generally use all of the attributes. The hypothyroid application has 22 predictive attributes (plus a class and a label attribute) but only six of them appear in the tree and the rule set. This ability to pick and choose among the predictors is an important advantage of tree-based modeling techniques.

Some applications, however, have an abundance of attributes! For instance, one approach to text classification describes each passage by the words that appear in it, so there is a separate attribute for each different word in a restricted dictionary.

When there are numerous alternatives for each test in the tree or rule set, it is likely that at least one of them will appear to provide valuable predictive information. In applications like these it can be useful to pre-select a subset of the attributes that will be used to construct the decision tree or rule set. The C5.0 mechanism to do this is called "winnowing" by analogy with the process for separating wheat from chaff (or, here, useful attributes from unhelpful ones).

Specifying the classes

The first entry in the names file specifies the classes in one of three formats:

- A list of class names separated by commas, e.g.
  Primary, compensated, secondary, negative.

- The name of a discrete attribute (the *target* attribute) that contains the class value, e.g.:
  Diagnosis.

- The name of a continuous target attribute followed by a colon and one or more thresholds in increasing order and separated by commas. If there are *t* thresholds *X1*, *X2*, ..., *Xt* then the values of the attribute are divided into *t+1* ranges:

  - less than or equal to *X1*
  - greater than *X1* and less than or equal to *X2*
  - Greater than *Xt*.

- Each range defines a class, so there is *t+1* class. For example, a hypothetical entry age: 12, 19. Would define three classes: age <= 12, 12 < age <= 19, and age > 19.

This first entry defining the classes is followed by definitions of the attributes in the order that they will be given for each case.

Improvements in C5.0 algorithm

- **Speed** - C5.0 is significantly faster than C4.5 (several orders of magnitude)
- **Memory usage** - C5.0 is more memory efficient than C4.5
- **Smaller decision trees** - C5.0 gets similar results to C4.5 with considerably smaller decision trees.
- **Support for boosting** - Boosting improves the trees and gives them more accuracy.
- **Weighting** - C5.0 allows you to weight different cases and misclassification types.
- **Winnowing** - a C5.0 option automatically winnows the attributes to remove those that may be unhelpful.

C. Classification and Regression Trees (CART) Decision Tree Algorithm.

Classification and Regression Trees (CART) is one of the classification algorithms. It is a flexible method to describe how the variable Y distributes after assigning the Forecast vector X.

This model uses the binary tree to divide the forecast space into certain subsets on which Y distribution is continuously even. Tree's leaf nodes correspond to different division areas which are determined by Splitting Rules relating to each internal node. By moving from the tree root to the leaf node, a forecast sample will be given an only leaf node, and Y distribution on this node also be determined.

Data descriptions

A data file also requires a description file which names and classifies the features in a data files. Features must haves names so they can be referred to in the decision tree (or other model output) and also be classified into their type. The basic types available for features are

- **continuous** for features that range over reals (e.g. duration of phones)
- **categorial** for features with a pre-defined list of possible values (e.g. phone names)
- **string** for features with an open class of discrete values (e.g. words)

The data description consists of a parenthesized list of feature descriptions. Each feature description consists of the feature name and its type (and/or possible values). Feature names, by convention, should be features names in the sense for features (and pathnames) used throughout the utterance structures in the Edinburgh Speech Tools.

The expected method to use models generated from features sets in the Edinburgh Speech Tools is to apply them to items. In that sense having a feature name be a feature of an item (or relative) pathname will avoid having the extra step of extracting features into a separated table before applying the model. However it should also be stated that to wagon these names are arbitrary tokens and their semantic irrelevant at training time.

4. CONCLUSIONS

In this paper, we propose two approaches for Brain Tumor Detection based on decision tree algorithm. The trees were categorized into C5.0 Decision Tree Algorithm and Classification and Regression Trees (CART) Decision Tree Algorithm.

A simple method that uses C5.0 decision trees to select features has been described. This is to be used to improve Naïve Bayesian learning. The empirical evidence shows that this method is very fast and surprisingly successful, given the very different natures of the two classification methods. This Selective Bayesian classifier is asymptotically at least as accurate as the best of C5.0, Naïve Bayes, and Augmented Bayes on each of the domains on which the experiments were

performed. Further, it learns faster than both C5.0 and NB on each of these domains.

This work suggests that C5.0 decision trees systematically select good features for Naïve Bayesian classifier to use. We believe the reasons are that C5.0 does not use redundant attributes in constructing decision trees, since they cannot generate different splits of training data. When few training examples are available, C5.0 uses the most relevant features it can find. The high accuracy NBC achieves with few training examples is indicative of the fact that using these features for probabilistic induction leads to higher accuracy both in Bayesian classifier and C4.5 itself in each of the domains we have examined.

REFERENCES

[1] Chotirat "Ann" Ratanamahatana Dimitrios Gunopulos

"Scaling up the Naive Bayesian Classifier: Using Decision Trees for Feature Selection" Computer Science Department University of California Riverside, CA 92521 1-909-787-5190 {ratana}.

[2]K.M. Iftekharuddin, On techniques in fractal analysis and their applications in brian MRI, in: T.L. Cornelius (Ed.), Medical imaging
systems: technology and applications, Analysis and Computational Methods, vol. 1, World Scientific Publications, 2005, ISBN 981-
256-993-6.

[3] L.P. Clarke, R.P. Velthuizen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, R.W. Thatcher, M.L. Silbiger, MRI segmentation: methods and applications, Magn. Reson. Imaging 13 (3) (1995) 343–368.

[4] K.D.Kharat,P.P.Kulkarni,M.B.Nagori, Brain Tumor Classification Using Neural Network Based Methods 62 (2012) 304-309.

[5] J.C. Bezdek, L.O. Hall, L.P. Clarke, Review of MR image segmentation techniques using pattern recognition, Med. Phys. 20 (4)
(1993) 1033–1048.

[6] Almuallim, H. and Dietterich, T.G. (1991). Learning with many irrelevant features. In Proceedings AAAI-91, volume 2, pp. 547-552, Anaheim, CA.

[7] Cardie, C. Using Decision Trees to Improve Case-based Learning. ICML 1993, pp. 25-32.

[8] Domingos, P. and Pazzani, M. (1997). Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier.

[9] Ming, K. and Zheng, Z. Improving the Performance of Boosting for Naïve Bayesian Classification. In Proceedings of the PAKDD-99, pp.296-305, Beijing, China.

[10] Khan M. Iftekharuddin et al., Fractal-based brain tumor detection in multimodal MRI, Appl. Math. Comput. (2008), doi:10.1016/j.amc.2007.10.063

[11] N. Pal and S. Pal, "A review on image segmentation techniques," Pattern Recognition, vol. 26, pp. 1277-1294, 1993.

[12] Josin G. M and Liddle P.F. Neural Network Analysis of the Pattern of Functional Connectivity between Cerebral Areas in Schizophrenia.
Biological Cybernetics, 2001; 84, 117-122.

[13] M. Sasikala and N. Kumaravel, "Comparison of Feature Selection Techniques for Detection of Malignai Tumor in Brain Images", *IEEE Indicon 20C Conference India*, Dec. 2005.

[14] Yan Zhu and Hong Yan, "Computerized Tumor Boundary Detection Using a Hopfield Neural Network", *IEEE Transactions on Medical Imaging*, **l6**, No. l, February l997.

[15] Alan Wee-Chung Liewand Hong Yan, "An Adaptive Spatial Fuzzy Clustering Algorithm for 3-D MR Image Segmentation", *IEEE Transactions C Medical Imaging*, **22**, No. 9, September 2003.